

2013

**University of North Carolina Wilmington
Master of Science in
Computer Science and Information Systems
Proceedings**

<https://csbapp.uncw.edu/mscsis>

A MULTI-UNIVERSITY DATA ANALYTICS SYSTEM
FOR UNDERSTANDING CAMPUS CRIME

Uche Iheadindu

A Capstone Project Submitted to the

University of North Carolina Wilmington in Partial Fulfillment of
the Requirements for the Degree of

Master of Science

Department of Computer Science

Department of Information Systems and Operations Management

University of North Carolina Wilmington

2013

Advisory Committee

Dr. Douglas Kline

Dr. Ulku Yaylalicegi Clark

Dr. Ron Vetter

Chief David Donaldson

Abstract

The budget challenges the University of North Carolina at Wilmington (UNCW) faced over the last few years and the projected increase in student population impose serious challenges on the UNCW Police Department (UPD). The UPD is expected to improve its activity and presence with limited funds on a growing campus. In order to make the right decisions, it is essential for the UPD to understand crime and crime prevention resources across similar campuses. This paper describes the process of developing and implementing a multi-university data analytics solution for the UPD. This solution provides interactive visualization, univariate statistics, and comparative analytics among thirty-eight university police departments. The paper provides a detailed explanation of the data collection process, technologies utilized, system architecture, and challenges encountered during the development process. Additionally, the paper explores statistical methods for analyzing data.

Table of Contents

- Abstract..... i
- 1 Introduction..... 2
- 2 Background..... 3
 - 2.1 University of North Carolina Wilmington 3
 - 2.2 University of North Carolina Wilmington Police Department 4
 - 2.3 The Data Warehouse..... 8
 - 2.3.1 Microsoft SQL Server..... 8
 - 2.3.2 SQL Server Reporting Services 9
 - 2.3.3 SQL Server Analysis Services 11
- 3 Problem Statement..... 12
 - 3.1 The UPD Challenges..... 12
 - 3.2 Motivation..... 13
 - 3.3 Organizational Objectives..... 13
- 4 System Analysis..... 14
 - 4.1 Process 14
 - 4.1.1 User Requirements Gathering Process..... 14
 - 4.1.2 Data Collection 15
 - 4.1.3 Prototyping..... 17
 - 4.2 Stakeholders..... 18
 - 4.3 Architecture..... 19
 - 4.4 Data Model..... 20
 - 4.4.1 Entities 21
 - 4.5 Reporting..... 27
 - 4.5.1 Views Structure..... 27
 - 4.5.2 Datasets 28
 - 4.5.3 Reports 29
 - 4.7 Benefits of the System 30
- 5 Data Analysis..... 31
 - 5.1 Benchmarking Analysis..... 31
 - 5.2 Correlation Analysis/Regression Analysis..... 32

5.3 Data Envelopment Analysis (DEA).....	33
6 Future Work.....	35
6.1 System Migration.....	35
6.2 Data Entry.....	35
6.3 System Expansion.....	36
6.4 Improvement of Statistical Models.....	36
7 Lessons Learned.....	37
7.1 Importance of Soft and Hard Skills.....	37
7.2 Requirements Validation.....	38
7.3 Naming Conventions.....	38
8 Conclusion.....	39
Bibliography.....	41
Appendix 1. UNCW Police Department Organizational Chart.....	45
Appendix 2. Research on Data Warehouse.....	46
Appendix 3. Log of meetings UNCW Police Department Project.....	56
Appendix 4. IPEDS UNCW Feedback Reports 2012.....	64
Appendix 5. List of Universities Included in the Database.....	72
Appendix 6. Description of Clery Act and the Crime Metrics.....	73
Appendix 7. Stakeholder Register.....	75
Appendix 8. System Architecture.....	76
Appendix 9. Data Access.....	77
Appendix 10. Data Model Progression.....	78
Appendix 11. The Carnegie Foundation for the Advancement of Educational Research.....	84
Appendix 12. Views in the Database.....	86
Appendix 13. Dependency of objects in the Database.....	88
Appendix 14. Examples SSRS Reports.....	89
Appendix 15. Correlation Analysis SPSS Output.....	94
Appendix 16. Regression Analysis Output.....	95

1 Introduction

Performance management allows an organization to improve its efficiency while ensuring that its activities are in line with its strategic goals and objectives. Performance management involves the monitoring and evaluation of organizational processes to ensure that all business parts are in sync towards achieving organization-wide goals. Critical to this process is the identification of benchmarks and metrics against which performance can be measured.

The recent economic crisis had a significant impact on the University of North Carolina at Wilmington (UNCW). Between 2008 and 2013, the university has registered significant budget cuts, and these cuts had a negative impact on the UNCW Police Department (UPD). It is essential for the UPD to engage in performance management activities to assess its activity on a regular basis and evaluate its performance. The UPD needs to identify areas that require improvement and direct its attention towards those areas. Furthermore, the UPD must justify the allocation of resources towards the department. The UPD can evaluate its performance by analyzing and comparing data reflecting the number of incidents registered on the UNCW campus in recent years to the incidents registered by a group of benchmark universities. In order to facilitate these actions, the UPD established several organizational objectives that stress the importance of gathering data over several years and improving the access and quality of information used by decision makers.

Over the last decade, technology enhanced performance management processes through solutions such as business intelligence applications and dashboards. These solutions synchronize and consolidate information by transforming data and enabling decision makers to take the necessary actions to ensure the overall success of the organization. The UPD is in need of a system that would provide support to activities related to performance measurement. The system should allow the UPD to translate consolidated data, trends, and metrics into information that ultimately leads to actionable strategies. The system should empower the UPD with the information it needs to identify causes that hinder performance and resolve them by taking corrective actions, while aligning its activities with the overall objectives of UNCW.

This paper describes the process of developing and implementing a system that stands as a decision support structure and allows the UPD to engage in performance management activities. The system was useful in establishing a set of benchmark schools to use for performance comparisons. This project implements the following:

- Data repository for multi-campus crime data using SQL Server
- Interactive reporting using SQL Server Reporting Services
- Statistical techniques for establishing and evaluating benchmarking universities

The paper provides detailed explanations regarding the motivation behind this project, the data collection process, the data warehouse implementation, and reporting tools utilized. Moreover, the paper explores the statistical methods utilized to determine benchmarks, while proposing a set of additional statistical techniques that may be utilized to analyze data.

Section 2 covers the background of the project, including a description of UNCW and its police department. Section 3 provides information regarding the challenges faced by the UPD, the motivation behind the project, and UPD's organizational objectives. Section 4 describes the system analysis process. This section provides details regarding the user requirements gathering process, data collection, prototyping, project stakeholders, system architecture, data modeling, and reporting structure. Section 5 covers the data analysis stage of the project by describing the steps taken in determining the benchmarks while exploring several statistical techniques that may be utilized to analyze the data. Section 6 explores future work opportunities, while Section 7 discusses the lessons learned during the lifetime of the project.

2 Background

2.1 University of North Carolina Wilmington

The University of North Carolina Wilmington (UNCW) is located in Wilmington, North Carolina. The university is one of the seventeen public institutions composing the University of North Carolina (UNC) multi-campus system (The University of North Carolina, 2013). The institution opened its doors in 1947 under the name of Wilmington College, as a county institution under the control of the New Hanover County Board of Education. At its inception,

UNCW had only 238 students (University of North Carolina Wilmington, 2013). UNCW has quickly become one of the top universities in North Carolina. As of fall 2012, a total of 13,075 undergraduate and graduate students were enrolled at UNCW, with circa 40% of students living on campus. (University of North Carolina Wilmington, 2012). UNCW sits on 661 acres, containing campus housing complexes, various administrative and department buildings, wetlands, and The Bluenthal Wildflower Preserve (University of North Carolina Wilmington, 2012). UNCW is organized into six colleges: College of Arts and Sciences, College of Health and Human Sciences, Cameron School of Business, School of Nursing, Watson College of Education, and the Graduate School (University of North Carolina Wilmington, 2012). The university offers fifty-two Bachelor's majors, thirty-eight Master's Degrees, and two Doctoral programs (University of North Carolina Wilmington, 2012). As of June 2012, UNCW had an endowment of \$67.7 million, and a total annual budget of \$286 million, with 33% of the operating revenue being state-funded (University of North Carolina Wilmington, 2012).

The recent economic crisis has had a significant impact on UNCW. From 2008 to 2011, on average, cuts to the school's annual allotment have amounted to around \$8.2 million per year (The Seahawk, 2011). In July 2011, the North Carolina General Assembly approved a budget cut of \$414 million to the UNC System, and, as a result, the UNC System Board of Governors administered budget cuts of an average of 16% across all institutions composing the UNC System (Starnewsonline, 2011). In July 2011, UNCW experienced a cut of 15.8%, which led to negative consequences, such as reduction in staff, fewer available classes, tuition increases, and increases in average class size (Starnewsonline, 2011). Additionally, the North Carolina General Assembly has proposed funding cuts of over \$100 million for fiscal year 2013-2014 (Dailytarheel, 2013). With enrollment projected to increase and significant budget cuts, UNCW is presented with a series of severe challenges.

2.2 University of North Carolina Wilmington Police Department

The UNCW Police Department (UPD) plays a vital role in helping the university accomplish its mission and vision, while also adhering to the university's core values. The sixth goal included in the UNCW strategic plan is to

“enhance the quality of UNCW’s environment and provide a sustainable campus that is attractive, functional and, above all, safe” (University of North Carolina Wilmington, 2009).

This goal highlights the importance of the UPD to the university. The primary responsibilities of the UPD are to provide police, security, and law enforcement services to the campus of UNCW and its properties, in order to ensure that the students, faculty, and staff live, learn, and work in a safe and secure environment (University of North Carolina Wilmington, 2011). Additionally, the UPD engages in crime prevention activities by providing a variety of educational programs ranging from self-defense to property security and community response protocols (University of North Carolina Wilmington, 2011). The following mission statement has been adopted by the UNCW Police Department:

“The University of North Carolina Wilmington Police will provide quality police service to our community on a twenty four hour basis. We will continually develop and deliver the best law enforcement services in North Carolina. We will enforce all laws of the State of North Carolina and all rules and regulations set forth by the University in a manner that will be of maximum benefit to all, and will offer whatever assistance we can lawfully render without regard to the parameters of traditional policing responsibilities. We will encourage substantial interaction with the students, faculty and staff to coadunate the community” (University of North Carolina Wilmington, 2013).

The UPD adheres to three important core values:

- 1) *Excellence in problem solving,*
- 2) *Excellence in Community Policing,*
- 3) *Active support for University programs* (University of North Carolina Wilmington, 2013).

The UPD has 50 fulltime positions consisting of 34 sworn personnel, seven telecommunicators, 6 guards, and 3 administrative support personnel (Donaldson, 2013). Appendix 1 provides a visual representation of the organizational chart. Sworn personnel, also

called university police officers, are certified law enforcement officers who are armed, have full powers of arrest, and they are empowered with the same authority as other law enforcement officers in the state of North Carolina (Donaldson, 2013). Most of the university's police officers have extensive military and civil law enforcement experience, and they collaborate closely with other law enforcement agencies in the region, specifically the Wilmington Police Department and New Hanover County Sheriff's Office (Donaldson, 2013).

The UPD is organized under four primary functional divisions that include the Office of the Chief, Support Services, Patrol, and Investigations (University of North Carolina Wilmington, 2013). The Office of the Chief is comprised of the Chief of the Police, the ASSAstant Chief, the Professional Standard Unit, and administrative support personnel. The division provides executive oversight, direction, and support through activities associated with planning, personnel recruitment, budget, finance, resource allocation, and policy (Donaldson, 2013).

The Support Services Division has among its primary responsibilities statistical data processing, records administration and Clery reporting compliance, research and development, crime mapping and analysis, and logistics and maintenance support of departmental equipment and infrastructure needs (University of North Carolina Wilmington, 2013). The Division consists of sworn officers and telecommunicators.

The Patrol Division is the largest division, and also the most visible division of the UPD (University of North Carolina Wilmington, 2013). The division operates twenty-four-seven, patrolling university property in patrol vehicles, on bicycles, and on foot (University of North Carolina Wilmington, 2013). The main responsibility of the Patrol Division is to deter criminal activity and remain accessible to members of the community for service functions such as jump starting/unlocking cars, security checks, and escorts (University of North Carolina Wilmington, 2013).

The Investigation Division performs follow-up investigations of reported crimes while also providing expert insight with regard to matters of crime prevention, substance abuse, sexual assault, and other areas of concern to the UNCW community (University of North Carolina

Wilmington, 2013). The division cooperates with local, state, and federal agencies in the conduct of investigations and has access to the resources of these agencies (University of North Carolina Wilmington, 2013).

The UPD strives to perform its activity in a manner that supports UNCW's strategic plan. The sixth goal within UNCW's strategic plan, stressing the importance of providing a campus that is attractive, functional, and safe, has among its objectives the following:

- Objective 5:
“Improvement of processes and procedures that ensure safety and security in all facets of campus life”
- Objective 6:
“Establishment of education and training initiatives for campus violence prevention (University of North Carolina Wilmington, 2009).”

The UPD has been active in supporting these objectives throughout the years by constantly implementing programs that increase the level of safety and security on campus, while also providing crime prevention education. Programs such as Police and Community Networking (PAC'N), and Rape Aggression Defense (RAD) have proven to be beneficial over time (University of North Carolina Wilmington, 2013). However, for the last several years, as a result of the recent economic crisis, UNCW has faced a series of budget cuts that have had a negative impact on the UPD. Additionally, the fact that the campus population is projected to increase over the next few years imposes an additional challenge on the UPD. Therefore, it is vital for the UPD to assess its activity on a regular basis and to evaluate its performance. By assessing its performance, the UPD can identify areas requiring more attention and channel resources or personnel activity towards those areas, while also justifying the allocation of resources towards the department.

One of the best ways of measuring the performance of UPD is to analyze and compare data reflecting the number of incidents registered on the UNCW campus in recent years to the incidents registered by a group of universities with similar characteristics. Thus, UPD is in need of a data repository to compile and allow easy access to data, and mechanisms that would facilitate the standardization, filtering, aggregation, sorting, and formatting of the data collected.

2.3 The Data Warehouse

A data warehouse, containing the data relevant to assessing the performance of the UPD, provides a response to the needs of the UPD. The data warehouse concept was developed in the late '80s and revolutionized in the '90s (Johnston & Weis, *Managing Time in Relational Databases*, 2010). A data warehouse is developed in an effort to provide support to organizational leaders in their decision making process. Applications such as business intelligence systems are among the available tools for gathering, storing, and analyzing the data required for supporting the decision making process. In most cases, at the core of such systems sits a data warehouse (Hwang, *Success Factors for Business Intelligence: Perceptions of Business Professionals*, 2009). Data from various sources are collected and compiled in a central database, and, based on this data, managers are able to perform data analysis and base their decisions on any trends identified. A data warehouse processes high volume data from various sources and transforms it into a format that is easy to understand. It also provides the storage space for the data, and basically represents the backend or the infrastructure of most analytical systems (Negash, *Business Intelligence*, 2004). Appendix 2 provides more information about the data warehouse and the different ways of implementing it.

After an extensive review of UPD's organizational resources and capabilities, and the products available to satisfy the needs of the UPD, Microsoft SQL Server was chosen as the database engine that will support business intelligence tools such as SQL Server Reporting Services, and SQL Server Analysis Services.

2.3.1 Microsoft SQL Server

A variety of factors have contributed to choosing Microsoft SQL Server (MSS) as the data repository. The first step was to determine whether to use an open source or a proprietary solution. At first glance, open source database solutions such as MySQL seemed very attractive, since the software is free of cost. However, open source software requires a certain level of technical expertise to manage content, and the costs incurred once the software is obtained may be substantial if there are no resources in place to manage it (Lindsey & Rogers, 2012). Hence, the cost of ownership of open source software can outweigh the cost of some proprietary option. Moreover, since open source database solutions lack support for basic database features that

allow the manipulation of data, in most cases, the development of an interface to the database is required in order to simplify user interaction (Lindsey & Rogers, 2012). This process can be extremely costly in terms of time and monetary resources.

Depending on the complexity of the system, the cost of proprietary solutions can be very high when first purchased. However, the user is paying for a customized product from a reliable brand that guarantees functionality, scalability, and, perhaps most important, a lower requirement for technical skills (CoreDna, 2009). Out of the available proprietary database solution, MSS proves to be the most viable solution for addressing the needs of the UPD. MSS is rich in out-of-the-box features that simplify the process of manipulating, securing, and managing data, without the need of feature packs, options, or separate products (Anderson, Fox, & Bolton, 2010). MSS comes with a set of tools that ease up the process of data movement, management, tuning, monitoring, development, and configuration (Anderson, Fox, & Bolton, 2010). The MSS brand also includes business intelligence features such as SQL Server Reporting Services (SSRS) and SQL Server Analysis (SSA) Services, which are essential in creating value from the available information (Anderson, Fox, & Bolton, 2010). These features can be easily integrated with the MSS database engine (Anderson, Fox, & Bolton, 2010). Furthermore, the Information Technology Systems (ITS) and the Business Applications divisions of UNCW, serving the information technology needs of the university, already have production servers that run MSS with features such as SSRS. Therefore, UPD can host the database on a campus server and access the reports generated through SSRS or SSA through the intranet. By hosting the database on a server managed by ITS or Business Applications, concerns regarding database management, security, back-up, and restore and recovery would be eliminated.

2.3.2 SQL Server Reporting Services

SSRS is a server based platform that provides an environment for creating reports from a number of different data sources and supports report delivery via the internet or intranet in a secure environment (Larson, 2012). SSRS delivers graphical, interactive, or tabular reports from relational, multidimensional, or XML-based data sources (Larson, 2012). Moreover, SSRS has features that enable developers to integrate data and report processing in custom applications (Larson, 2012). SSRS tools can be accessed through Microsoft Visual Studio or the Report

Builder application that can be accessed through the Report Manager. The Report Manger is a web-based interface that interacts with the report server (Larson, 2012). SSRS was chosen as the report generation software for the following reasons:

- UNCW has production servers that run MSS
- SSRS is free
- SSRS tools are fully integrated with SQL Server tools and components
- SSRS is used as a reporting tool at UNCW
- SSRS requires a very low level of programming
- Simplicity of report development
- Reports can be accessed over a browser, or as part of a SharePoint site
- SSRS incorporates security features
- Reports can be exported to other formats such as Microsoft Excel, Word, or PDF
- SSRS is printer friendly

The fact that UNCW already has production servers running MSS constitutes perhaps the main reason for choosing SSRS. SSRS is a feature incorporated in MSS that is fully integrated with SQL Server tools and components, and, with a licensed copy, it is free of charge (Larson, 2012). Also, SSRS is already used on the UNCW campus, which simplifies the process of setting up a report server that will host the UPD reports. Additionally, in case of any technical issues that may be encountered in the future, the UPD can receive support and guidance from the ITS division of UNCW.

SSRS contains a drag-and-drop approach to creating reports using database information (Larson, 2012). Any user with basic knowledge of Microsoft products such as Excel, Access, or Word can easily generate reports through the Report Server Project Wizard (Larson, 2012). In order to generate highly customized reports, basic programming knowledge is required, and Visual Basic expressions can be used to manipulate the reports (Larson, 2012). However, these expressions are not always necessary to create useful reports.

SSRS can be delivered to users several different ways. By using the Report Manager feature of SSRS, reports can be accessed via a browser over the internet, or, in the case of the

UPD, over the intranet. Furthermore, SSRS can be integrated with SharePoint, and users can take advantage of the benefits of the collaborative environment provided by SharePoint sites.

SSRS also includes security features that guarantee access only to the appropriate users (Larson, 2012). Moreover, SSRS uses Windows Integrated authentication, and, based on a specific network topology, it assumes trusted relationships where client and network resources are in the same domain or in a trusted domain (Microsoft, 2013).

SSRS includes a feature that allows the reports to be exported to several formats such as Excel, Word, TIFF, or PDF (Larson, 2012). Through this feature, users can manipulate the report data in an application of their choice. Furthermore, the reports can be easily printed from these applications.

2.3.3 SQL Server Analysis Services

SQL Server Analysis Services (SSAS) is an on-line analytical processing database that is highly optimized for queries and calculations common in business intelligence environments (Russo, Ferrari, & Webb, 2012). SSAS is used in conjunction with a relational database such as MSS. SSAS can be perceived as an extra layer of metadata sitting on top of a data warehouse in a relational database that contains information on how fact and dimensional tables should be joined, how users should explore data through hierarchies, and how measures should aggregate up (Russo, Ferrari, & Webb, 2012). Also, SSAS simplifies queries through the use of models containing the business logic of a data warehouse. Users can run queries on these models rather than the actual relational database, and, as a result, achieve better query performance and speed up reporting.

During the initial stages of the project SSAS was taken in consideration in an effort to expand reporting and analytics capabilities for the UPD. However, after performing research on the product and consulting with experts, it was determined that for the project underhand SSAS would not expand any of the reporting and analytics capabilities that SSRS already offers. SSAS is usually recommended under the following scenarios (Microsoft, 2013):

- Data is merged from Heterogeneous Data Stores.

- Populating Data Warehouses and Data Marts occurs frequently
- Data requires Cleaning and Standardizing
- Building Business Intelligence into a Data Transformation Process
- Automating Administrative Functions and Data Loading

Building Business Intelligence into a Data Transformation Process refers to the fact that data transformation process requires built-in logic to respond dynamically to the data it accesses and processes, while Automating Administrative Functions and Data Loading refers to functions such as backing up and restoring databases, copying objects, and loading data (Microsoft, 2013).

The UPD data warehouse does not fall under the scenarios described above for the following reasons:

- All data sits on a single storage system,
- The data warehouse will be populated on a yearly basis
- Every piece of data is collected manually and it does not require any cleaning
- Data transformation can be accomplished by using SSRS
- Since updates are made on a yearly basis, administrative functions will performed on the same basis

Under the premises described above, SSAS was not used for the UPD data analytics system.

3 Problem Statement

3.1 The UPD Challenges

The budget challenges that UNCW faced over the last few years and the projected increase in student population make it difficult for the UPD to continue delivering the best services. Additionally, the UPD is confronted with the arduous task of improving its activity and presence with limited resources on a growing campus.

3.2 Motivation

The UPD is committed to ensuring that the UNCW community receives the best safety and security services possible. Data plays an important role in the decision making process, since it may provide an in depth understanding of internal and external processes while also allowing decision makers to identify trends and potential opportunities. Data reflecting the characteristics of various universities across the country, including the number of incidents occurring on-campus, is readily available from a number of reliable, online sources. By using this information to compare itself against a pool of similar universities, the UPD would be able to identify the factors that impact performance, areas of improvement, or assess current capabilities. The UPD is in need of tools that would enable accessibility to data and allow its decision makers to assess the performance of the department, while also developing an understanding of areas that require more attention. The ultimate business purpose is to develop a tool that would improve the performance of the UPD.

3.3 Organizational Objectives

The UPD has a series of strategic objectives:

- Improve the accessibility and quality of information used for decision making
- Generate consistent on-demand reporting across years
- Establish benchmarking capabilities
- Identify areas of improvement
- Provide information to justify resource allocation

The UPD seeks to facilitate access to data from various sources, and collect this data on a yearly basis. Also, in order to create valuable information, the UPD seeks mechanisms that would format, standardize, filter, sort, and aggregate the available data. The mechanism in place should allow the data to be manipulated, exported, and analyzed. The ultimate goal is to be able to generate consistent on-demand reporting across years that would allow the establishment of benchmarking capabilities. Benchmarking would allow the UPD to assess

its performance and identify areas that require a higher level of attention. Moreover, the UPD can use the information to justify the allocation and distribution of resources.

4 System Analysis

4.1 Process

4.1.1 User Requirements Gathering Process

The requirements of the project were established over numerous meetings with representatives of the UPD. See Appendix 3 for a log of the meetings from the initial to the final phases of the project. During the initial meeting, Chief David Donaldson, the Chief of the UPD, expressed interest in collecting data that would allow the department to perform benchmarking and performance management. Chief Donaldson was looking for a solution that would enable the department to record and analyze data from previous years, and also update the data on a yearly basis once it became available. Even though there was no clear direction as to what type of data was to be recorded, the data sources, or how the data should be stored, accessed, and utilized, all of the details provided by Chief Donaldson indicated the need for a database solution, and a business intelligence tool that would allow the generation of reports. Carol Strickland, Director of Business Affairs Planning & Budgets at UNCW, also attended the meeting, and stated the fact that there is an interest in undertaking a similar initiative throughout various divisions under the Business Affairs unit at UNCW. However, it was determined to first focus on the UPD in order to develop a better understanding of the process, tools, and capabilities of such a project before extending it to additional business divisions.

During several subsequent meetings with Chief Donaldson in which Dr. Douglas Kline, Professor of Information Systems and Database Expert, also participated, it was agreed that MSS, with its database solution and business intelligence feature, is the best option for the project. All parties agreed that during the initial development stages, both the database and SSRS would sit on a development machine (personal laptop) in order to allow flexibility in the requirements-gathering process, while also determining the exact capabilities of the tools. However, the database will ultimately be deployed on a UNCW production server, and the

reports will be deployed and accessed from a UNCW SSRS server through the intranet. Authorized personnel from the UPD will be able to access specific reports through a browser based on their domain privileges.

4.1.2 Data Collection

From the offset, it was decided to focus on a total of thirty-eight universities including UNCW. The universities are classified as peer (17), sister (15), and universities identified by the Education Advisory Board as similar to UNCW (5). Peer universities were chosen based on the UNCW Integrated Postsecondary Education Data System (IPED) 2012 Feedback Report generated by the National Center for Education Statistics (NCES). IPED is a system operated by NCES that gathers information from colleges, universities, and other institution participating in the federal student financial aid programs. Annually, IPED releases a feedback report to each university (see Appendix 4 for more information on the IPED Feedback Report). Sister universities are all the institutions that are part of the University of North Carolina System. Appendix 5 contains a complete list of the universities chosen. The ultimate goal was to identify six universities (benchmarks) that are most similar to UNCW from this group. For each university, data associated with the following metrics was recorded:

- Total Acreage
- Student Headcount
- Total Number of Employees
- Employees FTE
- Dormitory Capacity
- Students Living on Campus
- Total Operating Expenses
- Operation and Maintenance of Plant Expenses

The UPD representatives chose the metrics for which to collect data since they were considered important characteristics of a university. The data associated with these metrics was collected manually from sources such as the National Center for Education Statistics (NCES),

IPEDS, universities' websites, and universities' common data sets, starting with the most recent academic year, 2012/2013, and ending with academic year 2008/2009. Later, in the lifetime of the project, data reflecting Operation and Maintenance Plant Expenses was collected. After contacting the UNCW Budget office and performing research on the metric, it was determined that expenses related to campus police departments are listed under the Operation and Maintenance of Plant category of expenses. Since data reflecting the exact budget of each campus police department was not readily available, this metric was used in the study in order to refer to a common category of expenses that includes an aggregate total of what each university spends on public safety services.

Additionally, the UPD had an interest in developing an understanding on the characteristics of the areas where the universities are located, and how these characteristics impact campuses. Therefore, data on the metropolitan statistical areas (MSA) of each university was collected and integrated in the data model. Data associated with the following metrics were added to the study:

- MSA Population
- MSA Murder and non-negligent manslaughter
- MSA Forcible Rape
- MSA Robbery
- MSA Aggravated Assault
- MSA Burglary
- MSA Larceny-Theft
- MSA Motor-Vehicle Theft

The data on MSA Population was collected from the US Census Bureau website. The rest of MSA metrics are crime metrics, and they were collected in case the UPD decides to perform an analysis that determines whether the crimes taking place in a specific community may have some kind of impact on campus crimes. The MSA crime metrics were collected from the Federal Bureau of Investigation website.

All the metrics described so far, are categorized as Control Metrics that are held constant and remain unchanged in the experiment.

Once the available data was analyzed and the benchmarks were identified, data associated with the following metrics was collected (see section 4.6 for a description of the analysis process):

- Sworn Police Officers
- Communicators/Dispatchers
- Non-Sworn/Security Personnel
- Administrative staff
- Police Department Budget

The data was collected by contacting representatives of each benchmark campus police department. These metrics are used to develop an understanding of the characteristics of campus police departments. These metrics were selected by the UPD representatives. All these metrics were categorized as Discretionary Metrics that may be changed or vary over the course of the investigation. The data shows numbers registered in calendar year 2013.

4.1.3 Prototyping

Once the MSS database was populated with the first group of data collected, the requirement gathering process continued over several meetings. Queries were generated from the available data in order to make the client aware of the types of reports that could be generated. Over these meetings, it was determined to add a set of new metrics representing the incidents that occurred on campuses over the last two available years (2010, 2011). These metrics fall under the Clery Act and more information regarding this act can be found on Appendix 6. The metrics were grouped under three categories of offenses: referrals, arrests, and criminal offenses. Criminal offenses can be described as violent, or property, while arrests and referrals can be described as enforcement. Additionally, the criminal offenses category has a hate crime sub-category. Appendix 6 provides a detailed explanation of each category and the metrics recorded under each category. The source for the data was the NCES, and the data was collected manually.

The next step in the prototyping phase involved the generation of reports using SSRS. A series of customized, parameterized, drilldown, and drill-through reports were generated in an effort to allow the client to develop an understanding of the capabilities of SSRS.

4.2 Stakeholders

In order to ensure the success of the project, it is essential that the stakeholders are identified in the early stages, and that their roles, expectations, level of involvement, and importance to the project are well-known. The stakeholders involved in this project are the following:

- Chief David Donaldson
- Uche Iheadindu
- Dr. Douglas Kline
- Dr. Ulku Yaylacicegi Clark
- Dr. Ron Vetter
- Stanley Edwards
- Lori Speakman

The Stakeholder Register is included in Appendix 7, and it contains information regarding all stakeholders.

The UPD is the client and user of the system. Chief David Donaldson is the representative of the UPD, and the main champion of the project. The UPD expects to have a central data warehouse in place that can be updated on a yearly basis, a set of customized standard reports, general data-sets for ad-hoc reporting, and the ability to access reports with ease. Chief Donaldson has been highly involved since the early stages of the project, and his importance to the project is vital. His input is extremely valuable in terms of understanding the data, the type of reports needed, and determining if data presentation and accessibility meet the needs of the UPD.

Uche Iheadindu is the Consultant/Developer of the system. He is expected to choose the appropriate tools for storing and accessing the data, perform the initial phase of data collection,

develop interactive analytical tools, use statistical methods for determining universities similar to UNCW, and provide the documentation necessary for maintaining the system on a long term.

Dr. Douglas Kline has the role of Consultant/Database Expert/Committee Chair. Dr. Kline is a Professor of Information Systems with extensive knowledge of database principles. His involvement in the project is mainly based on providing database expertise, and determining if the development process is headed in the right direction. Dr. Kline has been involved from the early stages of the project, and his expertise is essential to the project.

Dr. Ron Vetter, Professor of Computer Science at UNCW, and Dr. Ulku Yaylacicegi, Professor of Information Systems at UNCW, are project committee members, and they are monitoring the project progress, while also providing input with regard to any technical issues encountered.

Stanley Edwards is the Director of Applications & Access Management for the Business Affairs-Business Services unit at UNCW. His role in the project is that of a Technical Liaison, and his expectation is to enable the deployment of the database on a production server and ensure that the appropriate data connections can be established. Mr. Edwards's importance to the project is vital since failure to meet his expectations may lead to the inability to utilize the system in a production environment.

Lori Speakman is an ITS System Administrator at UNCW, and she has the role of Technical Liaison. Ms. Speakman's role is very important to the project since she is expected to enable the deployment of reports and data sets on an SSRS Report Server, and allow access to appropriate users based on the existing Windows/Domain authentication policies. Ms. Speakman's role to the project is vital since failure to meet her expectations may lead to the inability of utilizing the system.

4.3 Architecture

Appendix 8 provides a visual representation of the system architecture. The data warehouse is hosted on a database server, and a set of views representing specific information are created in order to facilitate access to data. These views are the foundation of the datasets and the reports, which are located on a separate SSRS Report Server. Appendix 9 provides a visual representation of the manner in which data will be accessed, and Section 4.5 of this document

provides more information regarding the views, datasets, and reports. The users will access the datasets and the reports through the domain by using a browser, and access will be granted based on windows domain authentication.

4.4 Data Model

The design and implementation of the data model is one of the aspects of the project that required perhaps the most attention and time. The data model is a representation of the data collected, and it ensures that all data objects are adequately represented in the database. Lack of careful planning in the design of the data model may lead to a database implementation that omits the data necessary for creating critical reports and produces results that are incorrect or inconsistent. Additionally, since the requirements of the project were not clearly defined, the data model had to be designed in a manner that allowed any changes in the user requirements to be easily accommodated.

In most cases, business intelligence solutions generate their queries from a database that is designed under the Online Analytical Processing (OLAP) principles (Inmon, 2002). OLAP data is usually historical data that is aggregated into structures that permit sophisticated analysis (Inmon, 2002). Under the OLAP principles, the data model is typically de-normalized with fewer tables in order to speed up the reading process (Inmon, 2002). However, for the purposes of this project, the Online Transactional Processing (OLTP) principles were followed, and the data model is normalized with more tables included. Under the OLTP principles, the data model is flexible, and, as a result, new information can be added to the database overtime without having to make any drastic changes to the database (Ozsu & Valduriez, 2011). Appendix 2 provides more information about OLAP and OLTP. Since the UPD had no clear definition of the data to be stored in the database, the implementation of a data model under the OLTP principles was the most suitable choice for this project. The data model implemented allowed multiple changes to be made, while also simplifying the process of adding new metrics to the database. An Entity-Relationship (ER) model was used to design the data model. ER model is a conceptual data model that presents objects in the real world in the form of entities and relationships (Chen, 1976). The ER diagram is a basic component of the model, and it visually represents data objects.

In order to accommodate the changing user requirements and the addition of new metrics, over the lifetime of the project, the data model experienced a few changes. Appendix 10 provides a visual representation of the progression of the data model over the lifetime of the project.

4.4.1 Entities

The data model started with the following entities:

- *University*
- *Reason*
- *UniversityReason*
- *Source*
- *AcademicType*
- *Metric*
- *MetricType*
- *AcademicYear*
- *MetricValue*

The *University* entity contains the following eight attributes:

- *ID* (primary key)
- *FullName* (the complete name of the university)
- *Acronym* (the acronym under which the university goes by)
- *City* (city the university is located in)
- *State* (state the university is located in)
- *Zip* (zip code of the area)
- *Website* (the url of the university)
- *AcademicType* (foreign key associating a specific university with its academic classification)

The *Reason* entity refers to the reason a specific university has been taken in consideration, and it contains the following attributes:

- *ID* (primary key)

- *Reason* (peer vs. sister vs. recommended by the Education Advisory Board)

UniversityReason is an associative entity used to associate a specific university with the reason it was included in the project. An associative entity is a relationship that has been turned into an entity as a result of meeting the following conditions: it is a many-to-many binary relationship, or it is a ternary relationship or a relationship of an even higher degree (Abhijit & Ravindra , 2007). *UniversityReason* has the following attributes:

- *ID* (primary key)
- *ReasonID* (foreign key associating a specific university to a reason)
- *UnivID* (foreign key associating a specific reason to a university)

The *Source* entity refers to the source from where a specific piece of data has been retrieved, and it includes the following attributes:

- *ID* (primary key)
- *Name* (the name of the source)
- *URL* (the unified resource allocator from where the data was retrieved)

The *AcademicType* entity refers to the academic type and classification of a specific university according to the Carnegie Foundation for the Advancement of Educational Research. Appendix 11 provides detailed information regarding the Carnegie Foundation for the Advancement of Educational Research, and the different levels of classifications used for universities. The attributes of the *AcademicType* entity are:

- *ID* (primary key)
- *Type* (the academic type of an university)
- *Acronym* (acronym used to describe the academic type of a university)
- *PreviousType* (the prior academic type used for a university)
- *Description* (a detailed description of a specific academic type)
- *Classification* (the classification of a specific academic type)

The *Metric* entity refers to the variables/metrics stored with regard to a specific university. The attributes of the *Metric* entity are:

- *ID* (primary key)
- *Name* (the name of a specific metric)
- *Description* (a detailed description of a specific metric)

Initially, the *MetricType* entity contained information with regard to the different types of metrics. The metrics have been classified as control, discretionary, and crime metrics, and a specific metric was associated with a metric type through a *MetricTypeID* foreign key in the *Metric* table (see Appendix 10). However, once the user requirements changed and the client required recording the location where the crime took place (residence halls vs. campus), and the number of hate crimes taking place on campus, some adjustments were made to the data model. The following entities were included in the data model:

- *ControlMetric*
- *DiscretionaryMetric*
- *CrimeMetric*
- *HateCrime*
- *HateBias*
- *Location*

Generalization hierarchy was implemented in the data model in order to accommodate the changes in an appropriate manner. Under generalization hierarchy a structured group of entities share common attributes, while also preserving their differences (Buxton et al., 2009). Generalization hierarchies improve the stability of a data model by allowing changes to be made only to those entities that are pertinent to the change (Buxton et al., 2009). A generalization hierarchy includes a supertype entity, containing the common attributes, and subtype entities containing unique attributes, which are associated through a one-to-one relationship (Buxton et al., 2009). In the case of the data model at hand, the *Metric* entity is the supertype, and three additional subtype entities (e.g. *ControlMetric*, *DiscretionaryMetric*, and *CrimeMetric*) were added to the data model to accommodate user requirements. Furthermore, the *CrimeMetric* entity acts like a supertype to the subtype entity *HateCrime* in order to distinguish the regular incidents or offenses from those labeled as hate crimes. The *HateBias* entity was added in an effort to identify the bias under which a crime was committed (see Appendix 10).

The *ControlMetric* entity contains those metrics that are held constant and unchanged in an experiment, while the *DiscretionaryMetric* contains those metrics that may change or vary over the course of the investigation. Both the *ControlMetric* and the *DiscretionaryMetric* contain the following attributes:

- *ID* (primary key)
- *Notes* (providing unique details with regard to a specific metric)

The *CrimeMetric* entity contains information with regard to all the incidents or offenses taking place on a campus, and it contains the following attributes:

- *ID* (primary key)
- *LocationID* (foreign key associating a location to an incident or offense)
- *Classification* (describing the type of incident or offense, e.g. arrests vs. criminal offenses vs. referrals)
- *Notes* (providing details with regard to a specific incident or offense)

The *HateCrime* entity contains those metrics that are labeled as hate crimes, and it includes the following attributes:

- *ID* (primary key)
- *BiasID* (foreign key associating a specific bias to a crime)
- *Location* (key associating a specific location to a crime)

The *HateBias* entity contains data reflecting the bias under which a specific crime was committed. Appendix 6 provides an explanation of the biases recorded. The *HateBias* entity has the following attributes:

- *ID* (primary key)
- *BiasType* (identifies a specific type of bias)

The *Location* entity contains information about a specific location, and it includes the following attributes:

- *ID*

- *Location* (the location where a crime was recorded, e.g. campus vs. residence halls)

Initially, the *AcademicYear* entity was added to the data model in order to associate a piece of data with a specific academic year. However, over the lifetime of the project, it was determined that not all metrics were associated with an academic year. As an example, metrics such as Total Operating Budget were associated with a fiscal year, while metrics such as total Student Headcount were associated with an academic year. Moreover, the crime metrics record data based on a calendar year. Therefore, in order to adequately accommodate these differences, the *AcademicYear* entity was eliminated, and the following entities were added to the data model:

- *YearType*
- *YearValue*
- *YearKind*

The *YearType* entity was included in an effort to emulate the reports that the University of North Carolina System generates. These reports revealed the fact that a specific academic year and fiscal year is associated with a calendar year. As an example, the number of crimes recorded in calendar year 2011 is compared against the student headcount numbers registered for academic year 2011/2012 and the faculty and staff numbers registered for fiscal year 2010/2011. The *YearType* entity contains the following attributes:

- *ID* (primary key)
- *AcademicYear*
- *FiscalYear*
- *CalendarYear*

The *YearValue* entity contains information regarding the years related to a specific piece of data. The following attributes are under this entity:

- *ID* (primary key)
- *YearValue* (identifying a specific year)
- *YearKindID* (foreign key associating a specific year value with a type)

The *YearKind* entity identifies whether a specific year value stored in the *YearValue* entity is an academic year, a fiscal year, or a calendar year. *YearKind* has the following attributes:

- *ID* (primary key)
- *Type* (identifying the type of year)
- *Acronym* (the acronym used for a specific type of year)

The addition to the data model of the *YearValue* and the *YearKind* entities allows flexibility in the manner in which reports are generated. Users have the possibility to compare data from separate years.

The *MetricValue* entity is perhaps the most important entity in the data model. The table stores every single piece of data recorded. Several entities in the data model are associated through this entity (see Appendix 10). The attributes of the *MetricValue* entity are:

- *ID* (primary key)
- *UnivID* (foreign key associating a university with a specific piece of data)
- *Value* (the data piece)
- *SourceID* (foreign key associating a source with a specific piece of data)
- *YearID* (foreign key associating a piece of data with a specific academic year)
- *MetricID* (foreign key associating a value with a specific metric)
- *YID* (foreign key associating a piece of data to a specific year value and its type)

At the request of the client, who had an interest in developing an understanding as to how the characteristics of a specific area might have an impact on the crimes recorded at a specific university, data on the metropolitan areas of each university integrated in the data model. Since these metrics were not taken in consideration during the initial stages of the project, and the project was at a phase where a significant number of views and reports were created, the new metrics had to be integrated without making any major changes that could potentially affect the structure of the system. At a first glance, there was an inclination to include these metrics in the current *MetricValue* table. However, since several universities may be in the same metropolitan

area, there was a risk of creating data redundancy. As a result, two new entities were added to the data model (see Appendix 10):

- *MetropolitanArea*
- *MSAValues*

The *MetropolitanArea* entity contains specific information on metropolitan areas. This entity contains the following attributes:

- *ID* (primary key)
- *MsaArea* (the name of the metropolitan area)
- *MetropolitanVsMicropolitan* (identifying whether an area is metropolitan or micropolitan)

The *MSAValues* entity contains the values of the metropolitan area metrics. The entity has the following attributes:

- *ID* (primary key)
- *MetricValue* (the data piece)
- *MetricID* (foreign key associating a value with a specific metric)
- *MsaID* (foreign key associating a value with a specific metropolitan area)
- *YearID* (foreign key associating a piece of data to a specific year value and its type)
- *SourceID* (foreign key associating a piece of data with a specific source)
- *YearTypeID* (foreign key associating a piece of data with a specific academic year)

4.5 Reporting

4.5.1 Views Structure

Views are preplanned mapping and representation of the data stored in tables (Vieira, Being Objective: Re-examining Objects in SQL Server, 2009). They can be perceived as virtual tables that are based on the result of a specific SQL statement. Views can be tailored based on

what data the users desire to see, without having to search through unnecessary information (Vieira, *Being Objective: Re-examining Objects in SQL Server*, 2009).

In the case of the project underhand, in order to facilitate the access and partition of the information in the data warehouse in a useful manner, views have been utilized as data marts. A data mart is a smaller collection of summary or dimensional data that focuses on a subset of the data warehouse as a whole (Vieira, *Data Warehousing*, 2009). A total number of 112 views have been created throughout the lifetime of the project. Appendix 12 provides a detailed description of the views created. In most part, these views provide fast access to specific information such as the types of crimes, the control variables, or the discretionary metrics stored in the data warehouse. Additionally, views were used to simplify the complexity of queries. As an example queries retrieving information where control metrics were compared against crime metrics were written using views that were identifying these pieces of information. Sequentially, these queries were stored in the database as views. This process allowed the database to be decoupled and allow the separation of implementation from manipulation.

Since the requirements of the project were not entirely defined in the early stages of the project, a variety of changes were made to the database over time. The use of views to decouple the database weakened object dependency, and changes to the tables did not affect the overall structure of the system. Only basic views were updated, triggering the update of all other dependent views in the database. If each view was created directly from the tables, a very strong dependency between the table structure in the database and the views would have been created. Any changes to the tables would have required a manual update of each view and its other depending parts of the system. Appendix 13 illustrates the UDP database dependency diagram.

Views were also created in order to perform the data analysis process. Section 4.6 provides an explanation of how the views were utilized in the data analysis process. Some views were designed for experimental purposes in order to determine whether certain data analysis methods may be implemented dynamically.

4.5.2 Datasets

A SSRS a dataset is composed of a data source definition, a query of the data source and its resulting fields, and possibly parameters and calculated fields (Microsoft, 2012). There are

two types of datasets: shared datasets and embedded datasets (Microsoft, 2012). Once a shared dataset is created, a variety of SSRS reports can be generated from it. A SSRS report can contain multiple datasets that can be used by different data regions on the report, or used to provide dynamic lists of parameters (Microsoft, 2012).

The database views described in the previous section were used in order to simplify the process of generating the datasets, and provide an abstraction layer.

4.5.3 Reports

A series of 73 reports were designed. The following types of reports were created:

- Tabular
- Charts
- Parameterized
- Drill-down
- Drill-through

Appendix 14 shows examples of each SSRS report.

Parameterized reports use input values to generate a report. The report changes its output based on the values chosen. The input value can be passed by either typing, or selecting it from a dynamic dropdown list (see Appendix 15).

Drill-down reports are interactive reports that initially present only the high-level information, such as the sum or the average, and, once the user clicks on a specific area of the report, detailed lower-level information is presented. Drill-through reports are interactive reports that are accessed through a hyperlink in a designated initial report. Drill-through reports can stand alone, but in most cases they utilize a parameter retrieved from the initial report.

Most reports are customized through the use of expressions. In SSRS, expressions are used to specify or calculate values for parameters, queries, filters, report item properties, group and sort definitions, text box properties, bookmarks, document maps, dynamic page header and footer content, images, and dynamic data source definitions (Microsoft, 2012). Expressions are written in Microsoft Visual Basic, and in some cases JavaScript code can be inserted in order to accomplish certain capabilities such as displaying a new report in a separate tab.

4.7 Benefits of the System

The data analytics system provides a series of benefits to the UPD:

- Dynamic reporting capabilities
- Faster delivery of information
- Intuitive display options
- Configure what data is returned by the report
- Integration with MS Office
- Granular analytic capability

Dynamic reports are created at runtime and they display the most current information stored in the data warehouse. Prior to the data analytic system, the UPD relied mainly on static reports, created upon requested using MS Excel. This process was time consuming, and, since there was no capability to automatically update reports, specific information had to be extracted manually from disparate sources each time a report was built. The data analytic system eliminates any limitations associated with static reporting.

Since the data warehouse contains a plethora of data on campus crimes and the characteristics of universities, the UPD has now the capabilities to faster deliver specific information. As an example, the UPD can easily extract information comparing a certain type of crimes against metrics such as student headcount, employee headcount, or sworn officers at all the universities included in the repository. Prior to the data analytics system, this information could have been produced only by manually researching and manipulating the required data.

Some information can be better perceived through the use of the appropriate report formats. The data analytics system allows the UPD to find the most intuitive representation of data. The UPD has different possibilities of viewing and interacting with reports that display information in the most appropriate layout such as charts, tables, or matrix formats.

The UPD has the capabilities to configure what data is returned through the use of parameterized reports. As an example certain reports have parameters that allow users to choose the year for which they wish to see specific information.

The granular analytic capability of SSRS allows users to drill through summarized information in the reports in order to identify the details composing this information.

Perhaps one of the most beneficial features of the system is the fact that reports can be easily exported to formats such as MS Excel, MS Word, or PDF. As an example, users can export a report to an MS Excel format and manipulate the data as they wish.

5 Data Analysis

5.1 Benchmarking Analysis

The benchmarking analysis component of this project involved a few stages. The initial stage involved the identification of a set of six universities that are most similar to UNCW. The UPD had an interest in narrowing down the number of benchmarks in an effort to eventually engage law enforcement representatives from these campuses in discussions regarding crime prevention and reduction. The following metrics were label by the UPD representatives as important in determining the group of benchmarks:

- *Student Headcount*
- *Total dormitory capacity*
- *Employee Headcount*
- *Msa Population*
- *Total Acreage*
- *Total Operating Budget Expenses*
- *Operation and Maintenance Plant Expenses*

The Euclidean Distance matrix analysis was utilized to determine the six campuses. The Euclidean Distance is a measure that is commonly used to determine the distance between two objects, and it essentially measures the length of a straight line between objects (Hair, Anderson, & Black, Cluster Analysis, 1998). The Euclidean Distance can be calculated in n-dimensional space by using the formula $\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$, where $x = (x_1, x_2, x_n)$, and $y = (y_1, y_2, y_n)$ are the coordinates of objects in the space (Hair, Anderson, & Black, Cluster Analysis, 1998). The formula was implemented using SQL by taking the following steps:

1. Create a view that identifies the minimum and maximum value for each metric chosen
2. Create a view that normalizes the metrics for each university to scale between 0 and 1 by using the formula $(x_i - x_{min}) / (x_{max} - x_{min})$, where x_i = the metric value of a specific university, x_{min} = the minimum value for metric x, and x_{max} = the maximum value for metric x
3. Create a view that normalizes the metrics for UNCW, by using the formula described above
4. Create a view that joins the view normalizing the values for each university and the view that normalizes the values of UNCW, calculates the Euclidian Distance, and ranks universities in ascending order based on the calculated score

The results were also filtered by placing a specific emphasis on universities that were identified as coastal communities. The following is the list of universities that were identified as very similar to UNCW, and used as the benchmark set in subsequent analyses:

- *University of Maine*
- *Western Washington University*
- *University of Northern Iowa*
- *Southern Connecticut State University*
- *North Dakota State University*
- *College of Charleston*

5.2 Correlation Analysis/Regression Analysis

One of the goals of the project was to determine whether any of the discretionary metrics collected may have some type of relationship with campus crimes. One limitation to this phase of the analysis process was the fact that there was not enough discretionary data. Unlike the data for the control metrics, which was collected from a number of reliable online sources, the data for the discretionary metrics was not readily available (section 4.1.2 of this paper describes the data collection process). In order to build a reliable statistical model, future work should focus on

collecting data on discretionary metrics for a bigger sample size. Furthermore, data over several years for these metrics will allow the development of a solid statistical model.

The correlation analysis was performed using a statistical analysis software package, SPSS. Appendix 15 shows the output of the analysis. The analysis revealed that none of the correlations are significant since all levels of significance are higher than 0.05. Additionally, regression analysis does not seem to be a suitable method with a sample size of six universities. By using a sample size calculator for regression analysis, it was determined that in order to have a significance level/alpha level of 0.05, with four predictors, and a statistical power level of 0.80, a minimum sample size of eighty-four universities is required (www.danielsoper.com/statcalc). Furthermore, the SPSS outputs of the regression analysis shows that the regression model is not significant (see Appendix 16). In this model, alcohol violation on campus was used as the dependent variable since it registered the lowest significance levels during the correlation analysis (see appendix 15). The ANOVA in the regression output shows there were no statistically significant differences between group means ($p = 0.076$).

Since the administrative staff variable had the highest significance level (0.136), and the correlation analysis revealed a high correlation with the rest of the variables included in the model, the variable was eliminated in an effort to test whether the model can improve. However, the results showed that the model did not improve since the significance level registered a value higher than 0.05 ($p = 0.085$).

The final phase of testing involved the removal of the variable showing the number of dispatchers, since it registered the highest significance level during the previous test (0.194), and testing the model without it. The results showed that the model remained insignificant ($p = 0.066$). Similar analysis were performed with other crimes categorizes as dependent variables and similar results were registered.

5.3 Data Envelopment Analysis (DEA)

DEA is an application of linear programming used to measure the relative efficiency of operating units with comparable goals and objectives. The technique is commonly used to determine the efficiency of hospitals, banks, or schools (Anderson, Sweeny, & Willimas, 1994). DEA measures the performance of a unit/producer in a group relative to the performance of all

other units in the same group, based on a hypothetical unit constructed from all units in the group. Usually, units have different level of inputs and produce outputs with various levels. The inputs and outputs of each unit in the group are used to form a virtual unit with composite inputs and composite outputs (Anderson, Sweeny, & Willimas, 1994). The virtual unit is basically a weighted composite of the inputs and output of all units in the group.

For the purposes of this project, DEA may be used to determine the efficiency of the UPD in comparison with other police departments at the benchmarking universities identified. The inputs of each university are the discretionary metrics and the outputs are the crimes recorded on each campus. The measure of efficiency for UNCW can be determined through the following linear programming model:

$$\begin{aligned}
 & \text{Min } E \\
 \text{Inputs: } & \sum_i w_i * X_i \leq X_0 * E \\
 \text{Outputs: } & \sum_i w_i * Y_i \geq Y_0 \\
 \text{Weights: } & \sum_i w_i = 1 \\
 & E, w_i \geq 0
 \end{aligned}$$

Where E = the efficiency index; w_i = weights for universities; X_i = inputs for universities; X_0 = inputs for UNCW; Y_i = outputs for universities; Y_0 = outputs for UNCW.

This model was implemented using the MS Excel Solver add-in. Efficiency measures for each university were developed. The results were not very helpful because:

- There were not enough universities relative to the number of dimensions, which cause almost all universities to appear 100% efficient
- The relationships between inputs and outputs were not strictly positively correlated, as is typical in DEA, e.g., more officers do not increase crime
- The relationships between inputs and outputs are not linear, as is assumed by DEA. This is clear from the scatterplots.
- The input and output measures available were not from matching years (discretionary metrics were only available for 2013, but crime metrics were only available for years prior to 2013.)

For these reasons, the DEA results were not included in the paper. In future work, these issues could be addressed, and DEA is likely to be a useful technique for examining the data.

6 Future Work

6.1 System Migration

The system is currently set up on a research server for experimental purposes. This decision was made as a result of time constraints, compatibility problems between software utilized on the development environment and software on the production server, and delay in action of some of the stakeholders involved in the project. Additionally, having the system on a research server allows the appropriate testing of all the components of the systems. Once it is determined that the system meets all the expectations, the process of migrating the system to a production environment should be initiated.

6.2 Data Entry

Another important aspect of the project which experienced essentially the same issues as the process of deploying the system to a production environment was the implementation of a data entry solution. The initial goal of the project was to develop a web interface that would allow UPD employees to upload new data to the system. This initiative was delayed since representatives of the Business Application division at UNCW had various security concerns.

Currently, new data is uploaded to the system through inserts performed directly on the database. As a temporary solution, the UPD established a relationship with the MSCSIS program at UNCW, and there is an agreement that the MSCSIS program will provide the technical expertise required for performing data entry against the database. Since it is anticipated that new data would be added to the system on a yearly basis, and the system is for the moment on a research server, this solution does not seem to create any inconveniences. However, this solution does not seem feasible once the decision to migrate the system to a production server is made. Therefore, it is vital that discussions with the Business Application division continue in order to identify solutions to the security concerns raised, and allow the implementation of a web interface for uploading data to the system. The interface would simplify the data entry process while providing the UPD with flexibility in the process. The UPD would not have to rely on any external parties to upload new data to the system.

6.3 System Expansion

Expanding the system to allow other business units at UNCW to engage in performance management activities would be an interesting continuation of the project. During the first meeting, where the initial details of the project were discussed, Carol Strickland, Director of Business Affairs Planning & Budgets at UNCW, already expressed interest in having a system that would allow multiple divisions under business affairs to benchmark against other similar divisions at other universities (see Appendix 3). The feasibility of this initiative is high since the current data model is flexible enough to support the addition of multiple divisions and any relevant metrics.

Another interesting continuation of the project would be the development of a web portal that would allow all the police departments in the UNC System and the benchmarks to upload data and also access reports. However, the development process of such a project would be very challenging. The challenges would not be technical, but a matter of policy and procedure. In order to make this initiative possible, representatives of all the targeted departments would have to be willing to participate and provide all the resources necessary for developing the portal.

6.4 Improvement of Statistical Models

The data analysis component of this project had several limitations. The data collected so far was not sufficient for building an appropriate regression model, or to allow a proper implementation of the DEA. The data warehouse does not contain sufficient discretionary metrics, and those included have data only for calendar year 2013. Additionally, data on discretionary metrics were collected only for the universities that were identified as benchmarks.

In order to allow a better implementation of the statistical methods proposed in this project, future work should focus on the identification and collection of data on discretionary metrics over several years. Furthermore, data on discretionary metrics were collected only for the universities that were identified as benchmarks. The analysis would provide a deeper and more meaningful understanding of campus crime deterrents if it included data on discretionary metrics for all the universities in the data warehouse. Also, increasing the number of universities in the data warehouse would allow the implementation of strong statistical models.

7 Lessons Learned

7.1 Importance of Soft and Hard Skills

Gathering user requirements and facilitating communication between project stakeholders are often referred to as important steps in the system development process. In order to successfully accomplish these tasks, an individual is required to possess a good combination of technical, business, and people skills (Satzinger, Jackson, & Burd, Chapter 1-The World of the Information Systems Analyst, 2010). If the requirements are misunderstood or incorrectly communicated, the whole project is in danger of failing. The primary sources of information for the requirements are the various stakeholders of the system, and facilitating communication with and between stakeholder groups is vital to the success of the project (Satzinger, Jackson, & Burd, Chapter 1-The World of the Information Systems Analyst, 2010). Interpersonal skills are perhaps the most important to the entire development process. Taking a system from initial idea to final implementation does not rely strictly on developers, and the active participation of parties such as users, technical specialists, customers, and vendors is also vital to the process (Satzinger, Jackson, & Burd, Chapter 1-The World of the Information Systems Analyst, 2010).

While working on my graduate degree, I was involved in a variety of school projects where only hard skills were sufficient for success. As a result, I was minimizing the importance of having soft skills. However, while working on the UPD data analytics system, I had the privilege to be the analyst and the developer, and I realized how important it is to have good business and people skills. I found that the most challenging part of the development process was developing an understanding of how campus police operate and how campus crimes are recorded, translating technical jargon and details into terms that nontechnical personnel can easily understand, and ensuring that all project stakeholders are in common agreement with the system specifications. There were cases where some stakeholder groups were resistant to negotiating some of the system needs, and the success of the project was in jeopardy.

At the start of the project, I anticipated that the most difficult challenges would be related to the technical components. However, in order to ensure that the project was heading in the right direction, I spent a significant amount of time engaging in communication with stakeholder groups, and also facilitating interactions between these groups.

7.2 Requirements Validation

After ensuring all the system requirements were collected, in order to allow my client to understand the capabilities of the system, I engaged in prototyping. However, when looking back at the prototyping phase of the project, I realize that I could have anticipated some of the challenges related to changing requirements if I paid greater attention to this process.

I focused on prototyping during the development of the database and during the initial phase of generating reports. Once I received the list of reports required by the client, I rushed into the development process without allowing my client to provide feedback and ensure the requirements were validated. After presenting the reports developed, a variety of changes to the style and content of the reports were requested by the client. Even though these changes seemed minor at a first glance, there were cases where I had to rebuild reports from scratch in an effort to incorporate the client's requirements.

Overall, the whole process of incorporating the changes requested by the client was extremely costly. This situation allowed me to better understand the importance of a principle that is often stressed in the system analysis field—structured walkthrough. A structured walkthrough is a powerful technique of validating the requirements against users' needs, and it involves the review of the requirements collected and the models built based on those requirements (Satzinger, Jackson, & Burd, Chapter 4-Investigating System Requirements, 2010). Structured walkthroughs coincide with progress made in each project lifecycle phase, and it allows the identification of errors, problems, omissions, or inconsistencies in the requirements.

7.3 Naming Conventions

Naming conventions are extremely important in the application development field, and, throughout my courses at UNCW, professors stressed the importance of it. During the early stages of the development process, I did not place as much attention to the names I was assigning the database tables, columns, datasets, data sources, reports, and especially views. Initially, since my project was small in nature, I did not feel the effects of not using an appropriate naming convention, but once the project started growing and the requirements started to change, I struggled to identify certain needed elements and made a variety of errors. I realized that I

needed a consistent naming convention in order to avoid any other issues, and I decided to go back and review and, in some cases, rename some of the objects in the system. This process was time-consuming, and it caused me to waste time that could have been dedicated towards making progress in other areas of the project. There were cases where I could not change some of the original names since any amendments would have had a domino effect and ruined some of the features of the system that were already in place.

A consistent naming convention makes a complex system easier to understand, it makes the search for relevant elements easier, and it is more intuitive for someone trying to understand the system. Additionally, I found that the use of names that accurately describe the functionality of objects is extremely useful. Therefore, being somewhat long to allow for more accurate descriptors is better than being too short when naming objects.

8 Conclusion

To help the UPD address their challenging decision-making environment, the following were designed and implemented:

- SQL Server Database on a shared server
- SSRS Report Server on a shared server
- Customized SSRS reports
- Euclidean Distance model to establish benchmarks
- A set of statistical techniques that may be utilized to analyze data

The reports can be easily and securely accessed by decision makers. The UPD now has the capability to compare the crimes taking place on multiple campuses and the characteristics of these campuses through interactive data visualization. Furthermore, the information generated from the data repository allowed the identifications of a benchmark set, and, as a result, the UPD can engage in a more detailed process of evaluating its performance in comparison with similar campus police departments. This comparison enables the UPD to perform an accurate assessment of its current activity and identify whether changes can be made to current practices, or learn whether, and how, things may be done better. On the other hand, the UPD also gained

the capability to identify areas of strength and take the necessary steps to maintain those strengths.

The overall value of the data analytics system is that it provides a better decision support system for the UPD. The system improves the accessibility and quality of the information utilized in the decision making process. The system contributes to the UPD's efforts to ensure that UNCW receives the best possible safety and security services. The analytic capabilities of the system help the UPD to provide a response to the challenges raised by the budget cuts and the projected increase in student population. By analyzing and comparing data, the UPD can identify areas requiring more attention and channel resources or personnel activity towards those areas, while also justifying the allocation of resources towards the department.

Bibliography

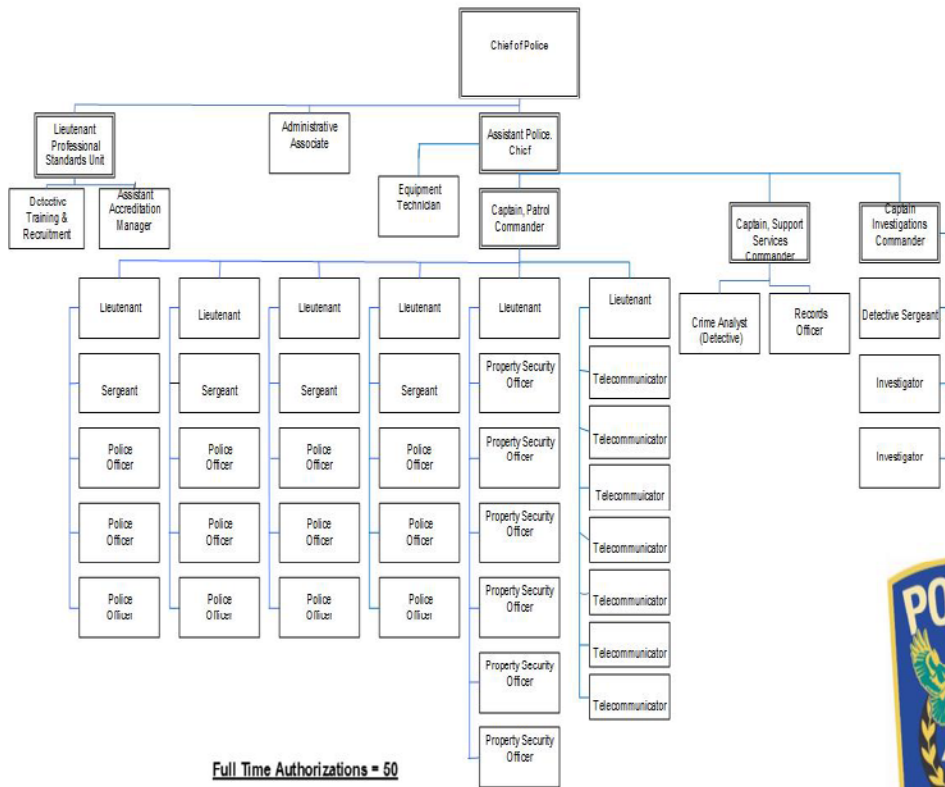
- Abhijit , P. A., & Ravindra , A. K. (2007). *Developing Web-Enabled Decision Support Systems*. Belmont: Dynamic Ideas.
- Akinde, M. O., Bohlen, , M., Johnson, T., Lakshmanan, L. V., & Srivastava, D. (2002). *University at Buffalo*. Retrieved November 4, 2012, from Efficient OLAP Query Processing in Distributed Data Warehouses:
<http://ftp.cse.buffalo.edu/users/azhang/disc/springer/0558/papers/2287/22870336.pdf>
- Anderson, M., Fox, J., & Bolton, C. (2010). *Microsoft SQL Server 2008 Administration for Oracle DBAs*. New York: McGraw-Hill Osborne Media.
- Anderson, R. D., Sweeny, D. J., & Willimas, T. A. (1994). Chapter 4 Linear Programming Applications. In R. D. Anderson, D. J. Sweeny, & T. A. Willimas, *An Introduction to Management Science - Quantitative Approaches to Decision Making* (pp. 130-187). New York: West Publishing Company.
- Breslin, M. (2004). Data Warehousing Battle of the Giants: Kimball and Inmon Models. *BUSINESS INTELLIGENCE JOURNAL*, 18-20.
- Buxton et al., S. (2009). *Database Design: Know It All*. New York: Morgan Kaufmann.
- Buxton, S. (2009). *Database Design: Know It All*. New York: Morgan Kaufmann.
- Chen, P. P.-s. (1976). *The Entity-Relationship Model: Toward a Unified View of Data*. Retrieved August 14, 2013, from CiteSeer: <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.123.1085>
- CoreDna. (2009). *Open Source vs. Closed Source (Proprietary) Software* . Retrieved July 29, 2013, from CoreDna: <http://www.coredna.com/files/openvsclosed.coredna.pdf>
- Dailytarheel. (2013, June 13). *N.C. House proposes more funding cuts, new program for UNC system*. Retrieved July 2013, 2013, from Dailytarheel: <http://www.dailytarheel.com/article/2013/06/n-c-house-proposal-cuts-149-million-from-unc-system-adds-new-program>
- Donaldson, D. M. (2013, July 26). Chief UNCW Police Department. (U. Iheadindu, Interviewer)
- Hair, J. F., Anderson, R. E., & Black, W. C. (1998). Cluster Analysis. In J. F. Hair, R. E. Anderson, & W. C. Black, *Multivariate Data Analysis* (pp. 471, 476, 478). New Jersey: Prentice Hall Inc.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Multiple Regression Analysis. In J. F. Hair, R. E. Anderson, R. L. Tatham, & W. C. Black, *Multivariate Data Analysis* (pp. 156-157). Upper Saddle River, NJ: Prentice Hall.

- Hwang, M. I. (2009). *Success Factors for Business Intelligence: Perceptions of Business Professionals*. Retrieved July 29, 2013, from THE ASSOCIATION FOR CHINESE MANAGEMENT EDUCATORS: <http://www.myacme.org/ACMEProceedings09/p44.pdf>
- Hwang, M. I. (2009). *Success Factors for Business Intelligence: Perceptions of Business Professionals*. Retrieved November 2, 2012, from THE ASSOCIATION FOR CHINESE MANAGEMENT EDUCATORS: <http://www.myacme.org/ACMEProceedings09/p44.pdf>
- Inmon, W. H. (2002). *Building the Data Warehouse*. New York: Wiley Computer Publishing.
- Johnston, T., & Weis, R. (2010). *Managing Time in Relational Databases*. San Francisco: Morgan Kaufmann Publishers.
- Johnston, T., & Weis, R. (2010). *Managing Time in Relational Databases*. San Francisco: Morgan Kaufmann Publishers.
- Larson, B. (2012). *Microsoft SQL Server 2012: Reporting Services*. New York: McGraw-Hill.
- Lindsey, T. D., & Rogers, P. (2012, July). *Open-Source Software: Value, Cost, and Supporting Open Government*. Retrieved July 29, 2013, from California State Library: <http://www.library.ca.gov/crb/12/S-12-02.pdf>
- Microsoft. (2012, October 10). *Embedded and Shared Datasets*. Retrieved October 10, 2012, from Microsoft SQL Server 2012: <http://technet.microsoft.com/en-us/library/ee633675.aspx>
- Microsoft. (2012, October 3). *Understanding Simple and Complex Expressions*. Retrieved October 3, 2012, from Microsoft SQL Server: [http://technet.microsoft.com/en-us/library/ms345242\(v=sql.100\).aspx](http://technet.microsoft.com/en-us/library/ms345242(v=sql.100).aspx)
- Microsoft. (2012, October 10). *Working with Data in Reporting Services*. Retrieved October 10, 2012, from Microsoft SQL Server 2012: [http://technet.microsoft.com/en-us/library/ms160324\(v=sql.90\).aspx](http://technet.microsoft.com/en-us/library/ms160324(v=sql.90).aspx)
- Microsoft. (2013, July 31). *Authentication with the Report Server*. Retrieved July 31, 2013, from Microsoft: <http://msdn.microsoft.com/en-us/library/bb283249.aspx>
- Microsoft. (2013, September 23). *Microsoft Sql Server 2012*. Retrieved September 23, 2013, from Typical Uses of Integration Services: [http://technet.microsoft.com/en-us/library/ms137795\(v=sql.105\).aspx](http://technet.microsoft.com/en-us/library/ms137795(v=sql.105).aspx)
- Moller, R. (2001). *Distributed Data Warehousing Using Web Technology: How to Build a More Cost-Effective and Flexible Warehouse*. AMACOM.
- Negash, S. (2004). Business Intelligence. *Communications of the Association for Information Systems*, 178.

- Negash, S. (2004). Business Intelligence. *Communications of the Association for Information Systems*, 178.
- Noaman, A. Y. (2000). *Library and Archives Canada*. Retrieved November 4, 2012, from Distributed data Warehouse Architecture and Design:
http://www.collectionscanada.gc.ca/obj/s4/f2/dsk1/tape3/PQDD_0027/NQ51662.pdf
- Ozsu, T. M., & Valduriez, P. (2011). *Principles of distributed Database Systems*. New York: Springer Science Business Media.
- Russo, M., Ferrari, A., & Webb, C. (2012). *Sql Server 2012 Analysis Services - The BISM Tabular Model*. Sebastopol: Microsoft Corporation.
- Satzinger, J. W., Jackson, R. B., & Burd, S. D. (2010). Chapter 1-The World of the Information Systems Analyst. In J. W. Satzinger, R. B. Jackson, & S. D. Burd, *System Analysis and Design in a Changing World* (pp. 12-15). Boston: Course Technology.
- Satzinger, J. W., Jackson, R. B., & Burd, S. D. (2010). Chapter 4-Investigating System Requirements. In J. W. Satzinger, R. B. Jackson, & S. D. Burd, *System Analysis and Design In a Changing World* (pp. 116-158). Boston: Course Technology.
- Starnewsonline. (2011, September 2). *UNCW students feeling pinch of budget cuts*. Retrieved July 24, 2013, from Starnewsonline:
<http://www.starnewsonline.com/article/20110902/ARTICLES/110909944?tc=ar>
- The Seahawk. (2011, August 22). *Budget cuts and their effects on UNCW students*. Retrieved July 24, 2013, from The Seahawk: <http://www.theseahawk.org/news/budget-cuts-and-their-effects-on-uncw-students-1.2210294#.UfAGtY2siSq>
- The University of North Carolina. (2013, July 24). *About the University*. Retrieved July 23, 2013, from The University of North Carolina: <http://www.northcarolina.edu/about/index.htm>
- University of North Carolina Wilmington. (2009, April 22). *Strategic Plan*. Retrieved July 27, 2013, from University of North Carolina Wilmington: <http://uncw.edu/planning/docs.html>
- University of North Carolina Wilmington. (2011). *UNCW Crime and Security Report*. Retrieved July 27, 2013, from University of North Carolina Wilmington:
<http://uncw.edu/police/documents/crime.pdf>
- University of North Carolina Wilmington. (2012, August). *Academic Catalogue*. Retrieved July 27, 2013, from University of North Carolina Wilmington:
<http://catalogue.uncw.edu/content.php?catoid=3&navoid=126>

- University of North Carolina Wilmington. (2012, August 15). *UNCW Fact Sheet*. Retrieved July 24, 2013, from University of North Carolina Wilmington:
<http://uncw.edu/aboutuncw/documents/UNCWfactsheet4.2013.pdf>
- University of North Carolina Wilmington. (2013, July 28). *Police Department/Programs*. Retrieved July 28, 2013, from University of North Carolina Wilmington:
<http://uncw.edu/police/program.html#pacn>
- University of North Carolina Wilmington. (2013, July 27). *Police Department-Divisions*. Retrieved July 27, 2013, from University of North Carolina Wilmington: <http://uncw.edu/police/divisions.html>
- University of North Carolina Wilmington. (2013, July 24). *UNC Wilmington History and Traditions*. Retrieved July 24, 2013, from University of North Carolina Wilmington:
<http://uncw.edu/aboutuncw/aboutHistory.html>
- University of North Carolina Wilmington. (2013, July 24). *UNCW Police Mission Statement*. Retrieved July 24, 2013, from University of North Carolina Wilmington: <http://uncw.edu/police/mission.html>
- Vieira, R. (2009). Being Objective: Re-examining Objects in SQL Server. In R. Vieira, *Sql Server 2008 Programming* (pp. 8-10). Indiana: Wiley Publishing.
- Vieira, R. (2009). Data Warehousing. In R. Vieira, *SQL Server 2008 Programming* (p. 734). Indiana: Wiley Publisher.
- Yeruva, S., Kumar, P., & Padmanabham, P. (2012). Distributed Warehouses: A Review on Design Methods and Recent Trends. *International Journal of Computers and Distributed Systems*, 60-61.

Appendix 1. UNCW Police Department Organizational Chart



Full Time Authorizations = 50

SWORN PERSONNEL TELECOMMUNICATORS GUARDS ADMIN / SUPPORT

34 7 6 3



July 25, 2013

The Data Warehouse

Research Paper

Executive Summary

This paper explores the data warehouse concept. The concept was developed in the late 80's and revolutionized in the 90's. The data warehouse was developed in an effort to provide support to organizational leaders in their decision making process. Data from various departmental operational systems is collected and compiled in a central repository. Based on this data, managers are able to perform data analysis and base their decisions on any trends identified. Many online analytical processing systems (OLAP) are built on a data warehouse. The data warehouse is a relational database and there are cases where the queries necessary for an OLAP system are performed directly on a database designed for supporting online transactional processing (OLTP). However, the performance of both systems is affected as a result of the massive amount of data present and the operations taking place. Therefore, the literature strongly recommends the separation of data required to run OLAP systems and OLTP systems. Various studies argue several methodologies of implementing a data warehouse. Bill Inmon proposes a top down approach where data is collected from disparate systems and stored in a central repository. On the other hand, Richard Kimball is an advocate of a bottom down approach where data marts specific to individual departmental units are developed first and gradually combined into a central data warehouse. Both Inmon's and Kimball's approach present advantages and disadvantages and a successful implementation of each is strictly related to organizational characteristics. Furthermore, studies discuss two of the most widely used data warehouse systems: the centralized data warehouse and the distributed data warehouse. A centralized data warehouse involves the creation of a central data repository that is accessed by all departments within an organization. Additionally, under this approach, individual departments may create data marts containing information applying strictly to their business processes. This approach is widely recommended for small organizations that do not have units distributed geographically. A distributed data warehouse involves a collection of database systems located at the local level of each individual department, and a global database located at the headquarters. The systems are connected across the nodes of a network and they allow a user to access data as if it were located in a central system. The local database contains information specific to each unit, while the global database contains information that applies to the organization as a whole. This approach was developed in an effort to improve speed and performance, especially in environments where business is conducted over locations distributed geographically.

Introduction

In order to be successful in a highly competitive environment, businesses have to stay proactive. Managers are required to thoroughly understand the operating environment and make decisions that may create an advantage over their competition. Data plays an important role in the decision making process since it may provide an in depth understanding of internal and external processes while also allowing managers to identify trends or potential opportunities (Noaman, 2000). Most times, the success of an organization is strictly related to the manner in which data is recorded and utilized. Applications such as business intelligence (BI) systems are among the available tools for gathering, storing, and analyzing the data required for supporting the decision making process. In most cases, at the core of such systems sits a data warehouse (Hwang, Success Factors for Business Intelligence: Perceptions of Business Professionals, 2009). A data warehouse processes high volume data from various operational systems and transforms it into a format that is easy to understand. It also provides the storage space for the data and basically represents the backend or the infrastructure of most analytical systems (Negash, Business Intelligence, 2004). Data warehousing involves the extraction and transformation of data from an online transaction processing (OLTP) database and loading it separately to an online analytical processing (OLAP) system (Inmon, 2002). Worldwide, the largest and most successful organizations have been utilizing data warehousing for quite some time (Hwang, Success Factors for Business Intelligence: Perceptions of Business Professionals, 2009).

The Data Warehouse

The concept of the data warehouse was first developed in the late 1980's by IBM researchers Paul Murphy and Barry Devlin (Johnston & Weis, Managing Time in Relational Databases, 2010). In the 90's, Bill Inmon and Ralph Kimball revolutionized the concept. Inmon defines a data warehouse as a *"subject-oriented, integrated, time-variant, non-volatile collection of data in support of management decisions"* (Inmon, 2002). The development of a data warehouse involves the extraction of historical data from an operational database, the transformation of data in a format that is easy to understand, and storage of data in a separate database (Inmon, 2002). Therefore, a data warehouse is basically a database containing strictly historical information (time-variant). The idea behind this concept arose out of a need to mitigate issues related to the performance of the OLTP system. An OLTP system is design to handle real time transactions by capturing data about daily business processes. It requires extensive data control and accessibility, high multiuser throughput, and predictable fast response times (Ozsu & Valduriez, 2011). Inserts, updates, and deletes are the operations usually encountered in an OLTP system. OLAP involves the analysis of historical information and

its operations may also be performed directly on the OLTP system (Ozsu & Valduriez, 2011). However, OLAP requires complex queries over large tables, and, as a consequence, there is a negative impact on the performance of OLTP systems (Ozsu & Valduriez, 2011). The data warehouse concept was introduced as a response to these issues.

Inmon, considered by most the father of the data warehouse, is the advocate of a top-down approach where data from disparate units within a specific organization is stored and integrated in a centralized data repository (Breslin, 2004). The database is designed using the traditional database modeling techniques (Inmon, 2002). Under Inmon's principles, subject-oriented refers to the fact that data is organized around business processes or functional areas about which the organization needs information. In some cases data marts, containing information needed by specific departments, are created from the central repository in an effort to provide better access to personalized data. Non-volatile refers to the fact that once the data is stored in the database, it cannot be over-written, and users only have permission to read the data (Inmon, 2002). Scholars recommend the use of Inmon's approach in situations where an organization has a large team of data warehouse specialists, plans a large project that requires access throughout the whole enterprise, stores data that is not primarily business metrics, and can wait to see results (Breslin, 2004). The disadvantages are that the up-front cost for implementing a data warehouse is extremely high, and the timeline between the start of the project and the benefits expected can be very long (four to nine months). Additionally, the approach can be very inflexible and cannot adapt to changing departmental needs during the implementation phases (Breslin, 2004). It seems that the top down approach provides the guidelines necessary for implementing a centralized data warehouse, a concept discussed further in the paper.

Unlike Inmon, Kimball proposes a bottom-down approach where first data marts containing information specific to certain departments are created and eventually combined under a data warehouse. Kimball defines the central repository as the "data warehouse bus architecture" (Johnston & Weis, *Managing Time in Relational Databases*, 2010). This approach is widely perceived as a faster approach where business value can be returned quickly as a result of the initial focus on individual departmental data marts (Breslin, 2004). Kimball's approach is recommend for organizations that are better able to deploy smaller teams of data warehouse specialists, and also expect to store mostly business metrics (Breslin, 2004).

Various studies claim that there is not a right or wrong approach between Inmon's and Kimball's data warehouse implementation methodologies (Breslin, 2004). The adoption of one is strictly dependent on organizational characteristics such as staffing, skills, time to delivery, cost to deploy, and the organization's decision support requirements (Breslin, 2004). However,

both approaches have fundamental similarities. The use of time-stamped data and the extract, transform, and load (ETL) process is found in both methodologies, despite the different manner in which they are executed (Breslin, 2004).

Centralized versus Distributed Data Warehouse

Widely, when considering the implementation of a data warehouse, organizations choose between a centralized or distributed data warehouse. In most cases, the decision is strictly dependent on the needs, size, and characteristics of an organization.

A centralized data warehouse is a single database containing data integrated from a variety of operational systems, designed under a single data model which serves the needs of multiple disparate business units within an organization (Inmon, 2002). After the central database is developed, each individual department can build their own data marts containing information addressing their specific needs (Inmon, 2002). Appendix 1 provides a visual representation of a centralized data warehouse. According to some studies, most organizations tend to adopt a centralized data warehouse structure, despite having units geographically scattered (Inmon, 2002). Inmon states that a centralized data warehouse has a variety of advantages:

- Saves costs related to constructing, staffing, securing and managing the warehouse. A single data warehouse would require investment only in a central IT team.
- Provides common data of interest to the whole organization in one central location.
- The development of new business applications is simplified.
- Redundant data is eliminated since data is maintained in a central point.

However, a variety of studies have highlighted several disadvantages that make a centralized data warehouse unappealing:

- Developing a single data model can be extremely cumbersome and time consuming (Moller, 2001).
- Some departments within an organization may refuse to provide full access to their data due to resistance to change or any other political issues (Moller, 2001).
- A high number of users attempting to access data at the same time may cause delays and performance issues (Noaman, 2000).
- The inherently distributed nature of data collection and the huge amount of data extracted at each collection point make it impractical to gather all data at a centralized site (Akinde, Bohlen, Johnson, Lakshmanan³, & Srivastava, 2002).
- A centralized data warehouse is a single point of failure (Moller, 2001).

Moreover, industry surveys have shown that half of the centralized data warehouse projects have failed within their first year (Moller, 2001). Even Bill Inmon, the main advocate of a centralized data warehouse, has recognized the fact that in some cases a centralized data warehouse may not be so efficient. The implementation of a centralized data warehouse may not be suitable in environments where the business is distributed geographically, the data warehouse will hold a lot of data and the volume of data will be distributed over multiple processors, or the data warehouse grows up in an uncoordinated manner (Inmon, 2002). With the current business environment where organizations conduct their business activities from and among multiple location throughout the world, it seems that the principles advanced under a centralized data warehouse approach are not so appealing.

A distributed data warehouse is widely accepted and presented as an optimal solution to the challenges imposed by a centralized data warehouse. In organizations geographically distributed, two groups of decision makers emerge. The first group is composed of managers who make decision for the entire organization, while the second group is composed of managers that make decisions at the local level (Noaman, 2000). The first group of managers requires highly summarized information that is strategic in nature and focuses on long term trends. On the other hand, the second group of managers focuses on local information with an emphasis on short term trends (Noaman, 2000). According to the literature, distributed data warehouses address the needs of both types of decision makers in a more efficient manner than a centralized data warehouse (Noaman, 2000). A distributed data warehouse is defined as “a logically integrated collection of shared data that is physically distributed across the nodes of a computer network” (Moller, 2001). Behind this concept the data gathered at each individual unit is stored in a local data warehouse and only data pertinent to the organization as a whole is stored at a center point. The advancement of network technologies enables the implementation of a distributed warehouse where parts of the data may reside in different locations and the system functions to the outside world as a single global access-transparent repository (Yeruva, Kumar, & Padmanabham, 2012).

Inmon proposes an approach that includes a global and local data warehouse. Appendix 2 provides a visual representation of the approach. Unlike the concept of data marts advanced during the centralized data warehouse approach, in a distributed environment, the local data warehouse is actually an individual database modeled under the requirements of a specific business unit, containing only data specific to that location. The global data warehouse, usually located at the headquarters of the organization, has as a scope the organization as a whole and it contains only data that is of common interest to the whole organization, such as financial information, customer information, and vendor information (Inmon, 2002). The sources of data for the global data warehouse are the local data warehouses. At the global level, the data must

go through a transformation process since information generating from various places may be under metrics characteristic to a specific location (Inmon, 2002). According to Inmon, mapping the data from local systems to the structure of the global data warehouse is the most important and also the most difficult aspect of building the data warehouse. Mapping determines which data goes into the global warehouse and the conversion process it must go through (Inmon, 2002). Some advantages of a distributed data warehouse include:

- Multiple table queries will not cause major technical problems and, as a result, performance is not an issue. (Inmon, 2002)
- The system is reliable since there is not a single point of failure (Moller, 2001).
- Each unit has control over the design and resources required to implement the data warehouse (Inmon, 2002).
- Data can be added without concerns regarding space and capabilities (Moller, 2001).

Perhaps the main concern with regard to Inmon's approach is data redundancy. Inmon claims that there is mutual exclusivity between the data in the local and global warehouse (Inmon, 2002). However, some studies show that the partitioning of data in a multi-database system is impossible to accomplish since data tends to migrate to either the local or the global level (Noaman, 2000). Other disadvantages include:

- A network is required to sustain the distributed data warehouse and issues related to security or network traffic may raise some concerns (Yeruva, Kumar, & Padmanabham, 2012).
- The overall management and maintenance of the system can be very expansive (Moller, 2001).
- Development efforts may not move collectively and the implementation process may take some time (Moller, 2001).

Conclusion

Data management is an important process that allows organizations to remain proactive and create an advantage over their competitors. A data warehouse plays an important role in the data management process. Various studies are proposing different methodologies of data warehouse implementation. Overall, there seems to be a common agreement that a data warehouse is essential to the success of an organization. The implementation of a data warehouse takes the burden imposed by the OLAP queries on the OLTP system and maximizes the performance of both. Furthermore, a data warehouse sits at the core of analytical systems such as the BI. However, there are disagreements in the manner in which the data warehouse is implemented and the type of data warehouse that may eventually guarantee success. A

successful implementation of a data warehouse is strictly related to characteristics of an organization. The available literature implies that smaller organizations that do not have units scattered geographically should lean towards the implementation of a centralized data warehouse. On the other hand, the impact of globalization on businesses seems to favor the implementation of a distributed data warehouse. Both implementation approaches have a series of advantages and disadvantages, but overall, the suggestion is that architects have to develop a very good understanding of the organization and its processes in an effort to choose the most suitable implementation approach.

Bibliography

Akinde, M. O., Bohlen, J. M., Johnson, T., Lakshmanan, L. V., & Srivastava, D. (2002). University at Buffalo. Retrieved November 4, 2012, from Efficient OLAP Query Processing in Distributed Data Warehouses: <http://ftp.cse.buffalo.edu/users/azhang/disc/springer/0558/papers/2287/22870336.pdf>

Breslin, M. (2004). Data Warehousing Battle of the Giants: Kimball and Inmon Models. BUSINESS INTELLIGENCE JOURNAL , 18-20.

Hwang, M. I. (2009). Success Factors for Business Intelligence: Perceptions of Business Professionals. Retrieved November 2, 2012, from THE ASSOCIATION FOR CHINESE MANAGEMENT EDUCATORS: <http://www.myacme.org/ACMEProceedings09/p44.pdf>

Inmon, W. H. (2002). Building the Data Warehouse. New York: Wiley Computer Publishing.

Johnston, T., & Weis, R. (2010). Managing Time in Relational Databases. San Francisco: Morgan Kaufmann Publishers.

Moller, R. (2001). Distributed Data Warehousing Using Web Technology: How to Build a More Cost-Effective and Flexible Warehouse. AMACOM.

Negash, S. (2004). Business Intelligence. Communications of the Association for Information Systems , 178.

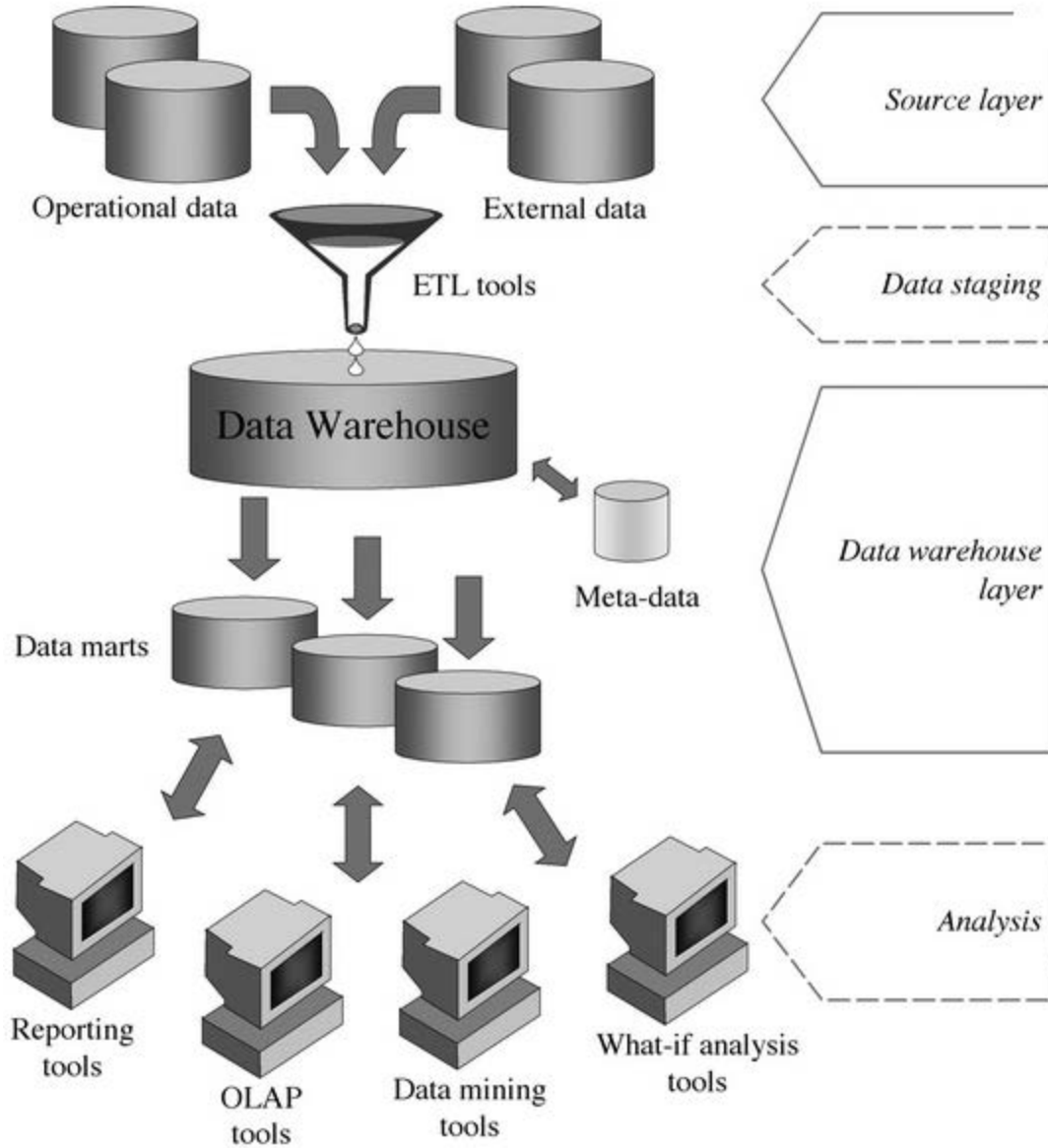
Noaman, A. Y. (2000). Library and Archives Canada. Retrieved November 4, 2012, from Distributed data Warehouse Architecture and Design: http://www.collectionscanada.gc.ca/obj/s4/f2/dsk1/tape3/PQDD_0027/NQ51662.pdf

Ozsu, T. M., & Valduriez, P. (2011). Principles of distributed Database Systems. New York: Springer Science Business Media.

Yeruva, S., Kumar, P., & Padmanabham, P. (2012). Distributed Warehouses: A Review on Design Methods and Recent Trends. International Journal of Computers and Distributed Systems , 60-61.

Appendix 1.

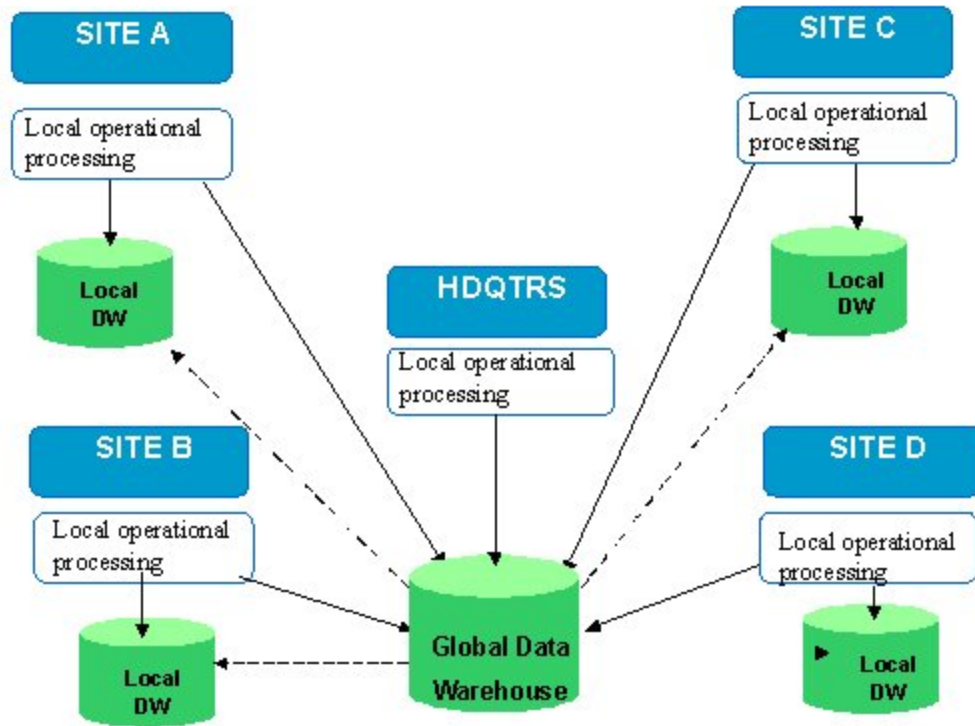
Centralized Data Warehouse



Source: <http://datasolutions.searchdatamanagement.com>

Appendix2.

Distributed Data Warehouse



Appendix 3. Log of meetings UNCW Police Department Project Meeting 1 - 04/1/2013

Duration: 11 am – 12 pm

Attendees: Dr. Ron Vetter (Interim Associate Provost for Research and Dean)

Chief David Donaldson (Chief of UNCW Police Department)

Chris Bertram (Major-UNCW Police Department)

Strickland Carol (Director of Business Affairs Planning & Budgets)

Uche Iheadindu (Graduate Student)

The representatives of the UNCW Police Department expressed interest in collecting and analyzing data to conduct performance management/benchmarking.

The Chief had a desire to keep the data and collect new data yearly.

The Chief of police, David Donaldson, wanted tools that would allow the collection of data and the generation of reports from the data collected.

From the interest expressed by the Chief, it was determined that there is a need of a database and a business intelligence solution.

Carol Strickland expressed interest in having several divisions under business affairs perform activities similar to the ones described by Chief Donaldson.

Since the scope and the capabilities of the project are not yet clearly defined, it was determined to focus first on the UNCW Police Department.

Further research will have to be performed in order to determine the type of data repository and the business intelligence solution.

The project was scheduled to start May 20, 2013.

Meeting 2 – 04/29/2013

Duration: 1 pm – 3pm

Attendees: Dr. Ron Vetter (Interim Associate Provost for Research and Dean)

Dr. Douglas Kline (Professor of Information Systems, Database Expert)

Uche Iheadindu (Graduate Student)

Possible solutions for the needs of the UNCW Police Department were discussed.

After analyzing the existing solutions, it was determined that MS SQL Server with its Business Intelligence features would meet the needs of the UNCW Police Department.

Meeting 3 – 05/06/2013

Duration: 11 am – 1pm

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Dr. Douglas Kline (Professor of Information Systems, Database Expert)

Uche Iheadindu (Graduate Student)

The database and business intelligence solutions were presented to Chief Donaldson

The first metrics to be recorded were determined.

It was determined that the data should be recorded in an excel spreadsheet first, in order to develop an understanding of the information available.

Meeting 4 – 05/21/2013

Duration: 1:30 pm – 3:30 pm

Attendees: Dr. Douglas Kline (Professor of Information Systems, Database Expert)

Uche Iheadindu (Graduate Student)

A data model was developed to accommodate the variables collected.

Since all the metrics/variables are not known yet, the data model was designed in a manner that would allow the addition of new metrics over time.

Meeting 5 – 05/29/2013

Duration: 1:00 pm – 2:00 pm

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented the data collected and the data model to the Chief of Police.

Chief required the addition of metrics reflecting the incidents recorded on campus (arrests and criminal offenses).

Meeting 6 – 06/07/2013

Duration: 9:00 am– 10:00 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented some of the queries that can be generated from the database.

Since there seems to be a discrepancy between the years associated with specific data (Fiscal Year, Academic Year, and Calendar Year), it was determined that changes should be made to the model in order to reflect such differences.

Chief required the addition of data reflecting the referrals on campus, and hate crimes.

Meeting 7 – 06/19/2013

Duration: 9:00 am– 10:00 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented some of the queries that can be generated from the database.

It was determined to add data reflecting the area under Total Operating expenses that relates directly to the Police Department.

It was determined that hate crimes should be separated into two categories (violent vs. other property)

Meeting 8 – 07/10/2013

Duration: 2:00 pm – 3:00 pm

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented some of the reports that can be generated with SSRS.

Chief Donaldson was interested in showing the numbers associated with UNCW when looking at the characteristics or incidents associated with a specific university.

Meeting 9 – 07/13/2013

Duration: 9:00 am– 9:30 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented some of the reports that compare UNCW with a specific university selected.

Chief Donaldson expressed interest in finding a group of 6 universities that are most similar to UNCW.

Meeting 10 – 07/19/2013

Duration: 9:00 am– 10:00 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented the method for determining the group of universities that are most similar to UNCW (Euclidian Distances).

It was determined that benchmarking universities should be determined on specific variables chosen by the Chief.

Meeting 11 – 07/25/2013

Duration: 12:00 pm– 1:00 pm

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Dr. Douglas Kline (Professor of Information Systems, Database Expert)

Uche Iheadindu (Graduate Student)

Discussed and evaluated the progress of the project.

Discussed the steps necessary for implementing the database on a production server.

Meeting 12 – 08/02/2013

Duration: 9:00 am– 10:00 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Dr. Douglas Kline (Professor of Information Systems, Database Expert)

Stanley Edwards (Director of Applications & Access Management, UNCW)

Lori Speakman (Systems Programmer, UNCW)

Philip Martin (Operations & Systems Analyst)

Discussed with the Business affairs team the implications of hosting the database and the SSRS reports on UNCW production servers.

Meeting 13 – 08/14/2013

Duration: 11:00 am– 11:30 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented the results registered during the first phase of identifying benchmarking universities.

Chief Donaldson expressed interest in collecting data that identifies the population of the metropolitan areas where each university is included.

Meeting 14 – 09/02/2013

Duration: 10:00 am– 10:30 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented the results registered during the second phase of identifying benchmarking universities (including metropolitan population in the analysis).

Chief Donaldson expressed interest in placing a specific emphasis on universities that are located in a coastal community. Furthermore, it was determined that it is important to identify the crime level in each metropolitan area by recording the crime stats in each area in the database.

Meeting 15 – 09/11/2013

Duration: 10:00 am– 10:30 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented the results registered during the second phase of identifying benchmarking universities (including metropolitan population in the analysis).

Chief Donaldson expressed interest in placing a specific emphasis on universities that are located in a coastal community. Furthermore, it was determined that it is important to identify the crime level in each metropolitan area by recording the crime stats in each area in the database.

Meeting 16 – 09/18/2013

Duration: 12:00 pm– 12:30 pm

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented the final results of the benchmarking analysis.

It was determined that each of the 6 universities identified should be contacted in order to retrieve specific data such as the number of sworn officers, dispatchers, security guards, administrative staff, and departmental budget.

Meeting 17 – 10/07/2013

Duration: 10:00 am– 10:30 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Uche Iheadindu (Graduate Student)

Chief requested the design of a series of customized reports to be presented during a meeting with representatives of the Governor's office.

Meeting 18 – 10/14/2013

Duration: 10:00 am– 10:30 am

Attendees: Dr. Douglas Kline (Professor of Information Systems, Database Expert)

Uche Iheadindu (Graduate Student)

Discussed the implications of setting up MS SQL Server and SSRS on a research server.

Started the installation process.

Meeting 19 – 10/21/2013

Duration: 10:00 am– 10:30 am

Attendees: Chief David Donaldson (Chief of UNCW Police Department)

Chris Bertram (Major-UNCW Police Department)

Cynthia Deacon (Captain-UNCW Police Department)

Charles Bost (Investigator-UNCW Police Department)

Uche Iheadindu (Graduate Student)

Presented to the UNCW Police Department the process of accessing the reports, and the capabilities of the reports generated.

Determined aggregate values and sorting capabilities to be included in the reports, and additional reports to be designed.

Appendix 4. Integrated Postsecondary Education Data System (IPEDS)

NATIONAL CENTER FOR EDUCATION STATISTICS

IPEDS DATA FEEDBACK REPORT 2012

What Is IPEDS?

The Integrated Postsecondary Education Data System (IPEDS) is a system of survey components that collects data from about 7,500 institutions that provide postsecondary education across the United States. IPEDS collects institution-level data on students (enrollment and graduation rates), student charges, program completions, faculty, staff, and finances.

These data are used at the federal and state level for policy analysis and development; at the institutional level for benchmarking and peer analysis; and by students and parents, through the College Navigator (<http://collegenavigator.ed.gov>), to aid in the college search process. For more information about IPEDS, see <http://nces.ed.gov/ipeds>.

What Is the Purpose of This Report?

The Data Feedback Report is intended to provide institutions a context for examining the data they submitted to IPEDS. Our goal is to produce a report that is useful to institutional executives and that may help improve the quality and comparability of IPEDS data.

What Is in This Report?

The figures provided in this report are those suggested by the IPEDS Technical Review Panel. They were developed to provide selected indicators and data elements for your institution and a comparison group of institutions. The figures are based on data collected during the 2011-12 IPEDS collection cycle and are the most recent data available. Additional information about these indicators is provided in the Methodological Notes at the end of the report. On the next page is a list of the institutions in your comparison group and the criteria used for their selection. Please refer to "Comparison Group" in the Methodological Notes for more information.

Where Can I Do More with IPEDS Data?

The Executive Peer Tool (ExPT) is designed to provide campus executives easy access to institutional and comparison group data. Using the ExPT, you can produce reports using different comparison groups and access a wider range of IPEDS variables. The ExPT is available through the IPEDS Data Center (<http://nces.ed.gov/ipeds/datacenter>).



University of North Carolina Wilmington
Wilmington, NC



IPEDS DATA FEEDBACK REPORT

COMPARISON GROUP

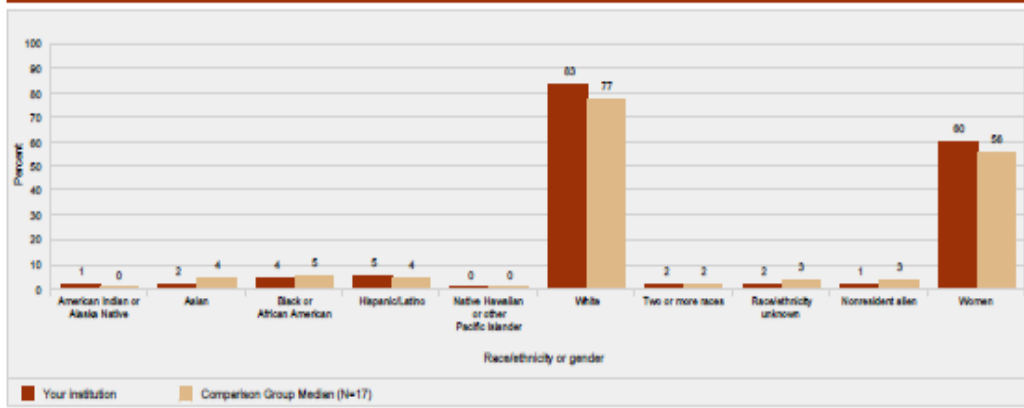
Comparison group data are included to provide a context for interpreting your institution's statistics. If your institution did not define a Custom Comparison Group for this report by July 15, NCES selected a comparison group for you. (In this case, the characteristics used to define the comparison group appears below.) The Executive Peer Tool (ExPT)(<http://nces.ed.gov/ipeds/datacenter/>) can be used to reproduce the figures in this report using different peer groups.

The custom comparison group chosen by University of North Carolina Wilmington includes the following 17 institutions:

- ▶ California Polytechnic State University-San Luis Obispo (San Luis Obispo, CA)
- ▶ California State University-Chico (Chico, CA)
- ▶ College of Charleston (Charleston, SC)
- ▶ College of William and Mary (Williamsburg, VA)
- ▶ James Madison University (Harrisonburg, VA)
- ▶ Murray State University (Murray, KY)
- ▶ Rowan University (Glassboro, NJ)
- ▶ SUNY at Binghamton (Vestal, NY)
- ▶ The College of New Jersey (Ewing, NJ)
- ▶ The University of Texas at Dallas (Richardson, TX)
- ▶ Towson University (Towson, MD)
- ▶ Truman State University (Kirksville, MO)
- ▶ University of Maine (Orono, ME)
- ▶ University of Maryland-Baltimore County (Baltimore, MD)
- ▶ University of Northern Iowa (Cedar Falls, IA)
- ▶ University of Wisconsin-Eau Claire (Eau Claire, WI)
- ▶ Western Washington University (Bellingham, WA)

IPEDS DATA FEEDBACK REPORT

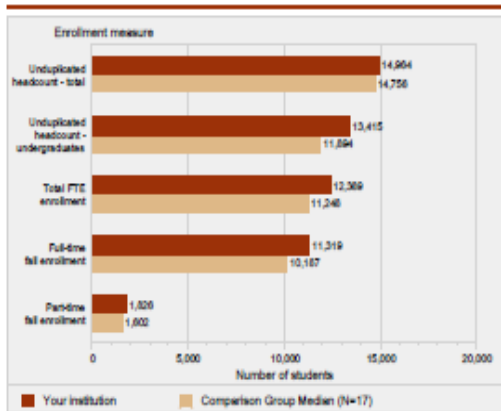
Figure 1. Percent of all students enrolled, by race/ethnicity and percent of students who are women: Fall 2011



NOTE: For more information about disaggregation of data by race and ethnicity, please see the Methodological Notes at the end of this report. Median values for the comparison group will not add to 100 percent. See "Use of Median Values for Comparison Group" in the Methodological Notes at the end of this report for how median values are determined. N is the number of institutions in the comparison group.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Spring 2012, Fall Enrollment component.

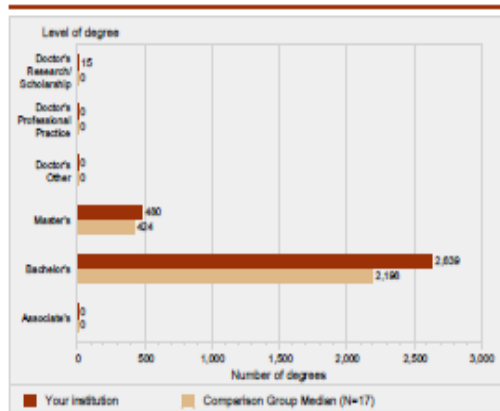
Figure 2. Unduplicated 12-month headcount of all students and of undergraduate students (2010-11), total FTE enrollment (2010-11), and full- and part-time fall enrollment (Fall 2011)



NOTE: For details on calculating full-time equivalent (FTE) enrollment, see Calculating FTE in the Methodological Notes at the end of this report. Total headcount, FTE, and full- and part-time fall enrollment include both undergraduate and postbaccalaureate students, when applicable. N is the number of institutions in the comparison group.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Fall 2011, 12-month Enrollment component and Spring 2012, Fall Enrollment component.

Figure 3. Number of degrees awarded, by level: 2010-11

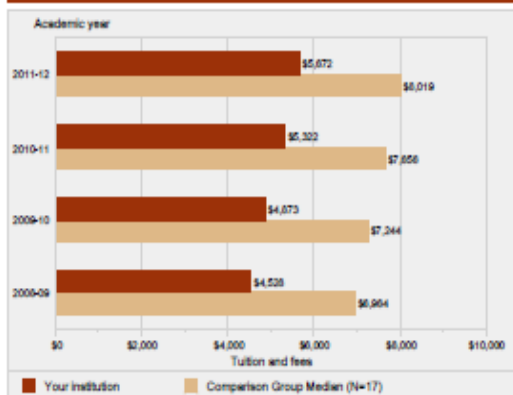


NOTE: For additional information about postbaccalaureate degree levels, see the Methodology Notes. N is the number of institutions in the comparison group.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Fall 2011, Completions component.

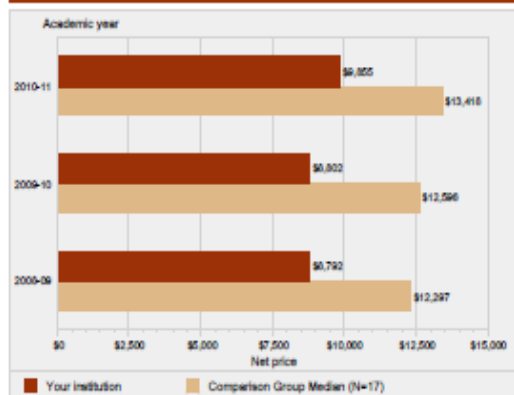
IPEDS DATA FEEDBACK REPORT

Figure 4. Academic year tuition and required fees for full-time, first-time, degree/certificate-seeking undergraduates: 2008-09–2011-12



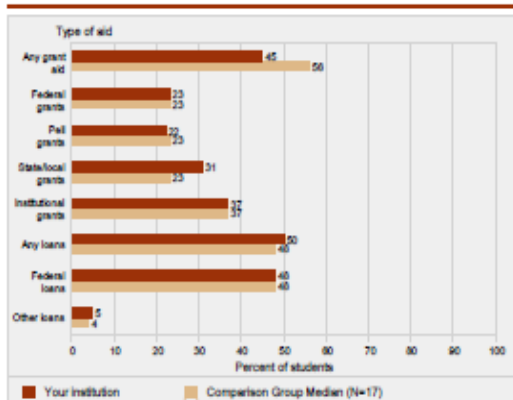
NOTE: The tuition and required fees shown here are the lowest reported from the categories of in-district, in-state, and out-of-state. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Fall 2011, Institutional Characteristics component.

Figure 5. Average net price of attendance for full-time, first-time, degree/certificate-seeking undergraduate students receiving grant or scholarship aid: 2008-09–2010-11



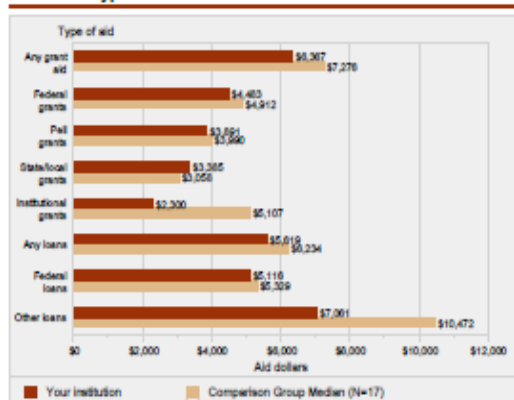
NOTE: Average net price is for full-time, first-time, degree/certificate-seeking undergraduate students and is generated by subtracting the average amount of federal, state/local government, and institutional grant and scholarship aid from the total cost of attendance. For public institutions, this includes only students who paid the in-state or in-district tuition rate. Total cost of attendance is the sum of published tuition and required fees, books and supplies, and the average room and board and other expenses. For more information, see the Methodological Notes at the end of this report. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Fall 2011, Institutional Characteristics component; Winter 2011-12, Student Financial Aid component.

Figure 6. Percent of full-time, first-time degree/certificate-seeking undergraduate students who received grant or scholarship aid from the federal government, state/local government, or the institution, or loans, by type of aid: 2010-11



NOTE: Any grant aid above includes grant or scholarship aid from the federal government, state/local government, or the institution. Federal grants includes Pell grants and other federal grants. Any loans includes federal loans and other loans to students. For details on how students are counted for financial aid reporting, see Cohort Determination in the Methodological Notes at the end of this report. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Winter 2011-12, Student Financial Aid component.

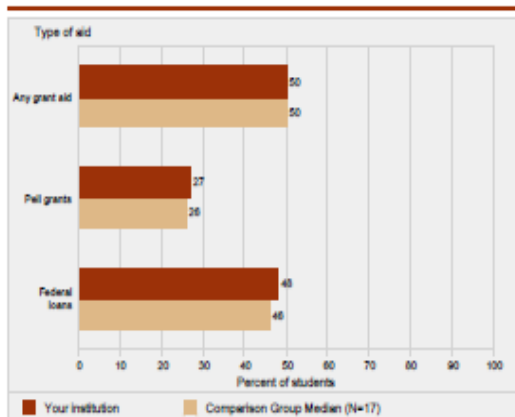
Figure 7. Average amounts of grant or scholarship aid from the federal government, state/local government, or the institution, or loans received, by full-time, first-time degree/certificate-seeking undergraduate students, by type of aid: 2010-11



NOTE: Any grant aid above includes grant or scholarship aid from the federal government, state/local government, or the institution. Federal grants includes Pell grants and other federal grants. Any loans includes federal loans and other loans to students. Average amounts of aid were calculated by dividing the total aid awarded by the total number of recipients in each institution. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Winter 2011-12, Student Financial Aid component.

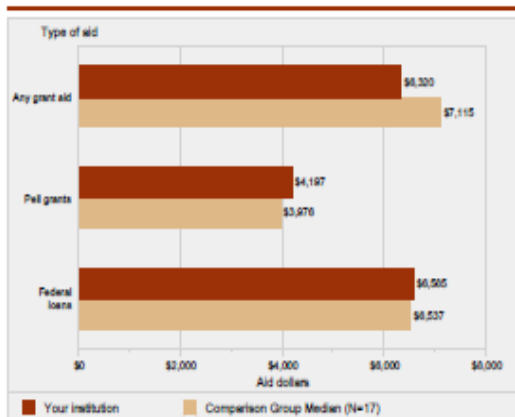
IPEDS DATA FEEDBACK REPORT

Figure 8. Percent of all undergraduates receiving aid by type of aid: 2010-11



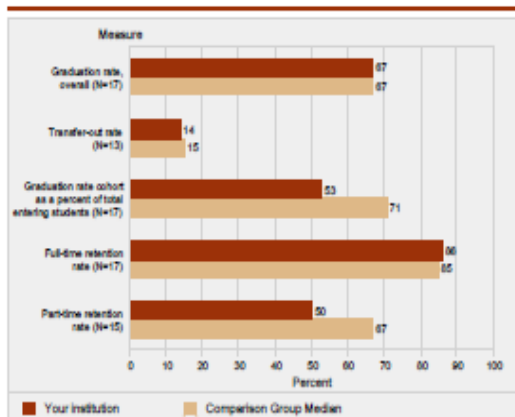
NOTE: Any grant aid above includes grant or scholarship aid from the federal government, state/local government, the institution, or other sources. Federal loans includes only federal loans to students. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Winter 2011-12, Student Financial Aid component.

Figure 9. Average amount of aid received by all undergraduates, by type of aid: 2010-11



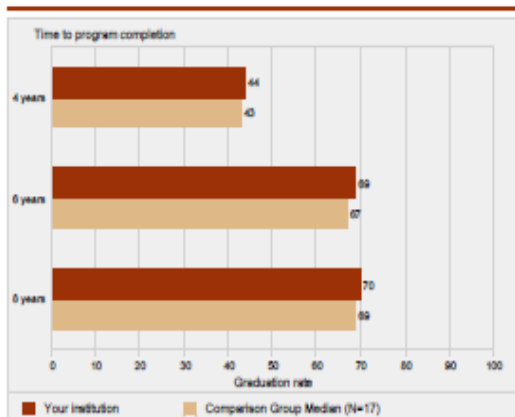
NOTE: Any grant aid above includes grant or scholarship aid from the federal government, state/local government, the institution, or other sources. Federal loans includes federal loans to students. Average amounts of aid were calculated by dividing the total aid awarded by the total number of recipients in each institution. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Winter 2011-12, Student Financial Aid component.

Figure 10. Graduation rate and transfer-out rate (2005 cohort); graduation rate cohort as a percent of total entering students and retention rates of first-time students (Fall 2011)



NOTE: Graduation rate cohort includes all full-time, first-time, degree/certificate-seeking undergraduate students. Entering class includes all students coming to the institution for the first time. Only institutions with a mission to prepare students to transfer are required to report transfers out. Graduation and transfer-out rates are the Student Right-to-Know rates. Retention rates are measured from the fall of first enrollment to the following fall. 4-yr institutions report retention rates for students seeking a bachelor's degree. Median values for the comparison group will not add to 100 percent. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Spring 2012, Graduation Rates component and Fall Enrollment component.

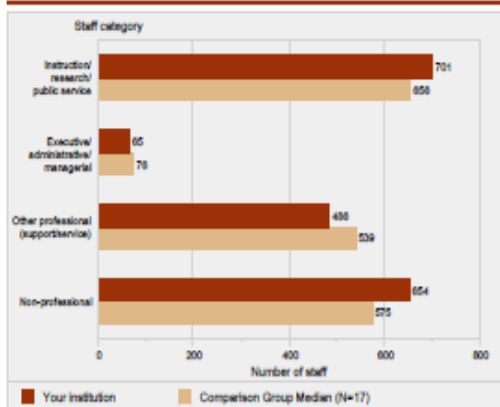
Figure 11. Bachelor's degree graduation rates of full-time, first-time, degree/certificate-seeking undergraduates within 4 years, 6 years, and 8 years: 2003 cohort



NOTE: The 6-year graduation rate is the Student Right-to-Know (SRK) rate; the 4- and 8-year rates are calculated using the same methodology. For more information see the Methodological Notes at the end of the report. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS): Spring 2012, 200% Graduation Rates component.

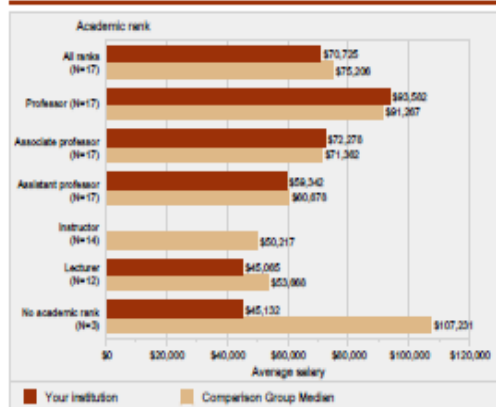
IPEDS DATA FEEDBACK REPORT

Figure 12. Full-time equivalent staff, by assigned position: Fall 2011



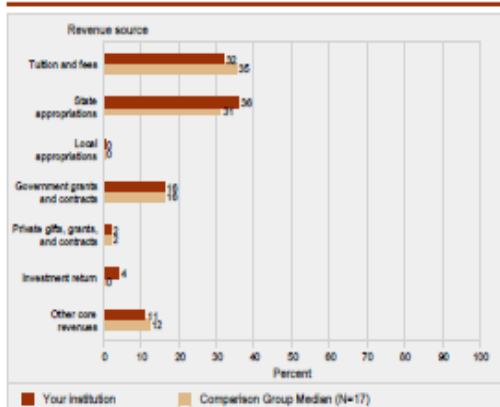
NOTE: Graduate assistants are not included in this figure. For information on the calculation of FTE of staff, see the Methodological Notes. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS); Winter 2011-12, Human Resources component.

Figure 13. Average salaries of full-time instructional staff equated to 9-month contracts, by academic rank: Academic year 2011-12



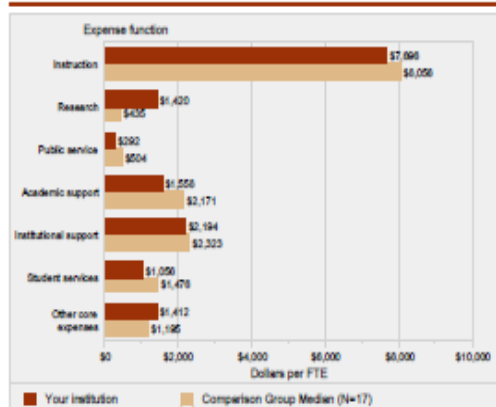
NOTE: Average full-time instructional staff salaries for 11/12-month contracts were equated to 9-month average salaries by multiplying the 11/12-month salary by .8182. Salaries based on less than 9-month contracts are not included. Medical school salaries are not included. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS); Winter 2011-12, Human Resources component.

Figure 14. Percent distribution of core revenues, by source: Fiscal year 2011



NOTE: The comparison group median is based on those members of the comparison group that report finance data using the same accounting standards as the comparison institution. For a detailed definition of core revenues, see the Methodological Notes. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS); Spring 2012, Finance component.

Figure 15. Core expenses per FTE enrollment, by function: Fiscal year 2011



NOTE: Expenses per full-time equivalent (FTE) enrollment, particularly instruction, may be inflated because finance data includes all core expenses while FTE reflects credit activity only. For details on calculating FTE enrollment and a detailed definition of core expenses, see the Methodological Notes. N is the number of institutions in the comparison group.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS); Fall 2011, 12-month Enrollment component and Spring 2012, Finance component.

IPEDS DATA FEEDBACK REPORT

METHODOLOGICAL NOTES

Overview

This report is based on data supplied by institutions to IPEDS during the 2011-12 survey year. Response rates exceeded 99 percent for most surveys. Detailed response tables are included in IPEDS First Look reports, which can be found at

<http://nces.ed.gov/pubsearch/qetpubcats.asp?sid=010>.

Use of Median Values for Comparison Group

The value for the comparison institution is compared to the median value for the comparison group for each statistic included in the figure. If more than one statistic is presented in a figure, the median values are determined separately for each indicator or statistic. Medians are not reported for comparison groups with less than three values. Where percentage distributions are presented, median values may not add to 100 percent. Through the ExPT, users have access to all of the data used to create the figures included in this report.

Missing Statistics

If a statistic is not reported for your institution, the omission indicates that the statistic is not relevant to your institution and the data were not collected. As such, not all notes listed below may be applicable to your report.

Use of Imputed Data

All IPEDS data are subject to imputation for total (institutional) and partial (item) nonresponse. If necessary, imputed values were used to prepare your report.

Data Confidentiality

IPEDS data are not collected under a pledge of confidentiality.

Disaggregation of Data by Race/Ethnicity

When applicable, some statistics are disaggregated by race/ethnicity. Data disaggregated by race/ethnicity have been reported using the 1997 (new) Office of Management and Budget categories. Detailed information about the recent race/ethnicity changes can be found at <http://nces.ed.gov/ipeds/reloresource.asp>.

Postbaccalaureate Degree Categories

The use of new postbaccalaureate degree categories was mandatory in the 2011-12 collection year. These categories are: doctor's degree-research/scholarship, doctor's degree-professional practice, and doctor's degree-other. (The first-professional degree and certificate categories and the single doctor's degree category have been eliminated.)

Cohort Determination for Reporting Student Financial Aid and Graduation Rates

Student cohorts for reporting Student Financial Aid and Graduation Rates data are based on the reporting type of the institution. For institutions that report based on an academic year (those operating on standard academic terms), student counts and cohorts are based on fall term data. Student counts and cohorts for program reporters (those that do not operate on standard academic terms) are based on unduplicated counts of students enrolled during a full 12-month period.

Description of Statistics Used in the Figures

Average Institutional Net Price

Average net price is calculated for full-time, first-time degree/certificate-seeking undergraduates who were awarded grant or scholarship aid from the federal government, state/local government, or the institution anytime during the full aid year. For public institutions, this includes only students who paid the in-state or in-district tuition rate. Other sources of grant aid are excluded. Average net price is generated by subtracting the average amount of federal, state/local government, and institutional grant and scholarship aid from the total cost of attendance. Total cost of attendance is the sum of published tuition and required fees, books and supplies, and the average room and board and other expenses.

For the purpose of the IPEDS reporting, aid received refers to financial aid that was awarded to, and accepted by, a student. This amount may differ from the aid amount that is disbursed to a student.

Core Revenues

Core revenues for public institutions reporting under GASB standards include tuition and fees; state and local appropriations; government grants and contracts; private gifts, grants, and contracts; sales and services of educational activities; investment income; other operating and non-operating sources; and other revenues and additions (federal and capital appropriations and grants and additions to permanent endowments). Core revenues for private, not-for-profit institutions (and a small number of public institutions) reporting under FASB standards include tuition and fees; government appropriations (federal, state, and local); government grants and contracts; private gifts, grants, and contracts (including contributions from affiliated entities); investment return; sales and services of educational activities; and other sources. Core revenues for private, for-profit institutions reporting under FASB standards include tuition and fees; government appropriations, grants, and contracts (federal, state, and local); private grants and contracts; investment income; sales and services of educational activities; and other sources. At degree-granting institutions, core revenues exclude revenues from auxiliary enterprises (e.g., bookstores, dormitories), hospitals, and independent operations. Nondegree-granting institutions do not report revenue from auxiliary enterprises in a separate category. These amounts may be included in the core revenues from other sources.

Core Expenses

Core expenses include expenses for instruction, research, public service, academic support, institutional support, student services, scholarships and fellowships (net of discounts and allowances), and other expenses. Expenses for operation and maintenance of plant, depreciation, and interest are allocated to each of the other functions. Core expenses at degree-granting institutions exclude expenses for auxiliary enterprises (e.g., bookstores, dormitories), hospitals, and independent operations. Nondegree-granting institutions do not report expenses for auxiliary enterprises in a separate category. These amounts may be included in the core expenses as other expenses.

IPEDS DATA FEEDBACK REPORT

Equated Instructional Staff Salaries

Total salary outlays for full-time instructional staff on 11/12-month contracts were equated to 9-month outlays by multiplying the outlay for 11/12-month contracted instructional staff by 0.8182. The equated outlays were then added to the outlays for 9/10-month instructional staff to determine an average salary for each rank. Salaries are not included for medical school staff or staff on less-than-9-month contracts.

FTE for Enrollment

The full-time equivalent (FTE) enrollment used in this report is the sum of the institution's FTE undergraduate enrollment and FTE graduate enrollment (as calculated from or reported on the 12-month Enrollment component). Undergraduate and graduate FTE are estimated using 12-month instructional activity (credit and/or contact hours). See "Calculation of FTE Students (using instructional activity)" in the IPEDS Glossary at <http://nces.ed.gov/ipeds/glossary/>.

FTE for Staff

The full-time equivalent (FTE) of staff is calculated by summing the total number of full-time staff from the Employees by Assigned Position (EAP) section of the Human Resources component and adding one-third of the total number of part-time staff.

Graduation Rates and Transfer-out Rate

Graduation rates are those developed to satisfy the requirements of the Student Right-to-Know and Higher Education Opportunity Acts and are defined as the total number of individuals from a given cohort of full-time, first-time, degree/certificate-seeking undergraduates who completed a degree or certificate within a given percent of normal time (for the degree or certificate) before the ending status date of August 31, 2011, divided by the entire cohort of full-time, first-time, degree/certificate-seeking undergraduates minus any allowable exclusions. Institutions are permitted to exclude from the initial cohort students who died or were totally and permanently disabled; those who left school to serve in the armed forces or were called to active duty; those who left to serve with a foreign aid service of the federal government, such as the Peace Corps; and those who left to serve on an official church mission. Transfer-out rate is the total number of students from the cohort who are known to have transferred out of the reporting institution within the same time period, divided by the same adjusted cohort. Only institutions with a mission that includes preparing students to transfer are required to report transfers out.

Retention Rates

Full-time retention rates are defined as the number of full-time, first-time, degree/certificate-seeking undergraduate students who enter the institution for the first time in the fall and who return to the same institution the following fall (as either full- or part-time), divided by the total number of full-time, first-time, degree/certificate-seeking undergraduates in the fall of first entrance. Part-time retention rates are similarly defined. For 4-year institutions offering a bachelor's degree, this rate is reported only for those first-time students seeking a bachelor's degree. For less than 4-year institutions, the rate is calculated for all first-time degree/certificate-seeking students.

Salaries, Wages, and Benefits

Salaries, wages, and benefits, for public institutions under GASB standards, and private, not-for-profit institutions under FASB standards. Include amounts paid as compensation for services to all employees

regardless of the duration of service, and amounts made to or on behalf of an individual over and above that received in the form of a salary or wage. Frequently, benefits are associated with an insurance payment. Private, for-profit institutions under FASB standards do not report salaries.

Total Entering Undergraduate Students

Total entering students are students at the undergraduate level, both full- and part-time, new to the institution in the fall term (or the prior summer term who returned in the fall). This includes all first-time undergraduate students, students transferring into the institution at the undergraduate level, and nondegree/certificate-seeking undergraduates entering in the fall. Only degree-granting, academic year reporting institutions provide total entering student data.

Tuition and Required Fees

Tuition is defined as the amount of money charged to students for instructional services; required fees are those fixed sum charges to students for items not covered by tuition that are required of such a large proportion of all students that the student who does not pay the charge is an exception. The amounts used in this report are for full-time, first-time, degree/certificate-seeking undergraduates and are those used by the financial aid office to determine need. For institutions that have differential tuition rates for in-district or in-state students, the lowest tuition rate is used in the figure. Only institutions that operate on standard academic terms will have tuition figures included in their report.

Additional Methodological Information

Additional methodological information on the IPEDS components can be found in the publications available at <http://nces.ed.gov/pubsearch/getpubcats.asp?sjid=010>. Additional definitions of variables used in this report can be found in the IPEDS online glossary available at <http://nces.ed.gov/ipeds/glossary/>.

Gary L. Miller, Chancellor
University of North Carolina Wilmington (ID:
199218)
601 South College Road
Wilmington, NC 28403-5963

Appendix 5. List of Universities Included in the Database

University	Acronym	Reason
California Polytechnic State University	CalPoly	Peer
California State University-Chico	CSUChico	Peer
College of Charleston	COFC	Peer
College of New Jersey	TCNJ	Peer
College of William & Mary	W&M	Peer
James Madison University	JMU	Peer
Murray State University	MSU	Peer
Rowan University	RU	Peer
SUNY Binghamton	BU	Peer
Towson University	TU	Peer
Truman State University	TSU	Peer
University of Maine	U Maine	Peer
University of MD-Baltimore County	UMBC	Peer
University of Northern Iowa	UNI	Peer
University of Texas at Dallas	UT Dallas	Peer
University of Wisconsin-Eau Claire	UWEC	Peer
Western Washington University	WWU	Peer
Appalachian State University	APP State	Sister
East Carolina University	ECU	Sister
Elizabeth City State University	ECSU	Sister
Fayetteville State University	FSU	Sister
NC A & T State University	NCAT	Sister
NC Central University	NCCU	Sister
NC State University	NCSU	Sister
UNC Asheville	UNCA	Sister
UNC Chapel Hill	UNC	Sister
UNC Charlotte	UNCC	Sister
UNC Greensboro	UNCG	Sister
UNC Pembroke	UNCP	Sister
UNC School of Arts	UNCSA	Sister
Western Carolina University	WCU	Sister
Winston-Salem State University	WSSU	Sister
North Dakota State University	NDSU	Proposed By Education AB
University of Tennessee-Chattanooga	UTC	Proposed By Education AB
Austin Peay State University	APSU	Proposed By Education AB
Southern Connecticut State University	SOUTHERNCT	Proposed By Education AB
California State University-San Bernardino	CSU Bernardino	Proposed By Education AB

Appendix 6. Description of Clery Act and the Crime Metrics

Clery Act:

According to the US Department of Education, the *Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics Act (Clery Act)*, act requires all postsecondary institutions participating in HEA's Title IV student financial assistance programs to disclose campus crime statistics and security information. The act was amended in 1992, 1998 and 2000. The 1998 amendments renamed the law the *Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics Act* in memory of a student who was slain in her dorm room in 1986¹.

Crime Metrics:

Referrals – According to the US Department of Education, *referred for disciplinary action* is defined as the referral of any person to any official who initiates a disciplinary action of which a record is kept and which may result in the imposition of a sanction².

Under Referrals, the following metrics are recorded:

- Illegal weapons possession
- Drug law violations
- Liquor law violations

A description of each offense can be found in The Handbook for Campus Safety and Security Reporting³.

Arrests – Refers to any individuals that were detained as result of committing a specific violation.

Under Arrests, the following metrics are recorded:

- Illegal weapons possession
- Drug law violations
- Liquor law violations

A description of each offense can be found in The Handbook for Campus Safety and Security Reporting⁴.

¹ <http://www2.ed.gov/admins/lead/safety/handbook.pdf>

² [ibid](#)

³ [ibid](#)

⁴ [ibid](#)

Criminal Offenses – Arrests taking place as a result of a specific set of offenses are labeled as criminal offenses. Criminal offenses can be described as violent and property crimes. The following criminal offenses metrics are recorded, and they represent each incident labeled as criminal offense:

- Murder/Non-negligent manslaughter (violent)
- Negligent manslaughter (violent)
- Sex offenses Forcible (violent)
- Sex offenses Non-forcible (violent)
- Robbery (violent)
- Aggravated assault (violent)
- Burglary (property)
- Motor vehicle theft (property)
- Arson (property)

Any of the above-mentioned offenses can be part of a sub-category named Hate Crimes if they are committed under a specific bias. The following is the list of biases:

- Race
- Religion
- Sexual Orientation
- Gender
- Disability
- Ethnicity/National origin

Additionally, any of following offenses can be labeled as hate crimes:

- Larceny-Theft (property)
- Simple Assault (violent)
- Intimidation (violent)
- Destruction/Damage/Vandalism of Property (property)

All the above metrics are recorded in the database.

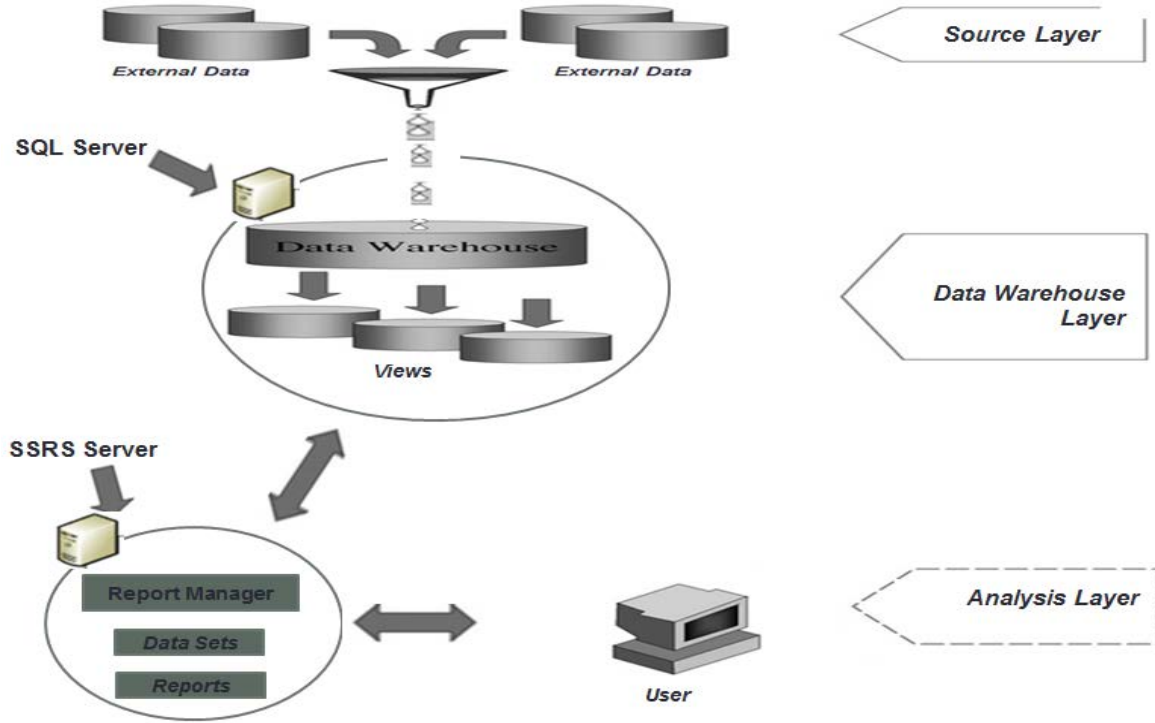
A description of each criminal offense and hate crime can be found in The Handbook for Campus Safety and Security Reporting⁵.

⁵ [ibid](#)

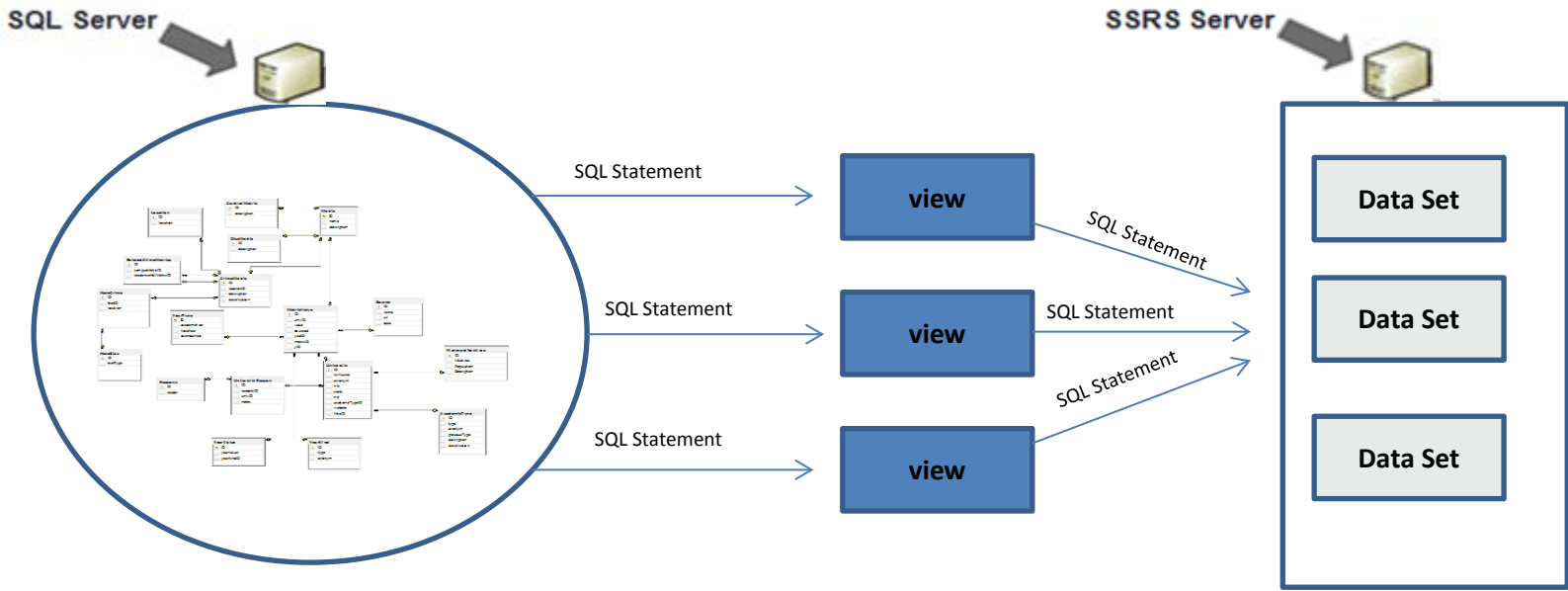
Appendix 7. Stakeholder Register

Project Name: Campus Crime Data Analytics for the UNCW Police Department					
Prepared by: Uche Iheadindu					
Date:					
Project Stakeholder Name	Department/Title	Role	Expectations	Level of Involvement	Importance/Impact on Project
Chief David Donaldson	Chief UNCW Police Department	Client/User	-Central data warehouse -Set of standard reports -General data-sets for ad-hoc reporting -Ability to access reports with ease	High	High
Uche Iheadindu	Graduate Student-UNCW-MSCSIS	Consultant/Developer	-Choose the appropriate tools -Data collection -Develop interactive analytic tools -System Documentation -Use statistical method for determining universities similar to UNCW	High	High
Dr. Douglas Kline	Professor of Information Systems-UNCW-MSCSIS	Consultant/Database Expert	-Provide database expertise -Provide feedback -Monitor project progress	High	High
Dr. Ron Vetter	Professor of Computer Science - UNCW	Committee Member	-Monitor project progress	Low	Low
Dr. Ulku Yaylacicegi	Professor of Information Systems-UNCW	Committee Member	-Monitor project progress	Low	Low
Stanley Edwards	Director Application and Access Management-UNCW Business Affairs	Technical Liaison	-Enable the deployment of the database on a production server -Allow appropriate data connections to be established	Low	High
Lori Speakman	Systems Administrator ITS UNCW	Technical Liaison	-Provide access to reports based on existent policies	Low	High

Appendix 8. System Architecture

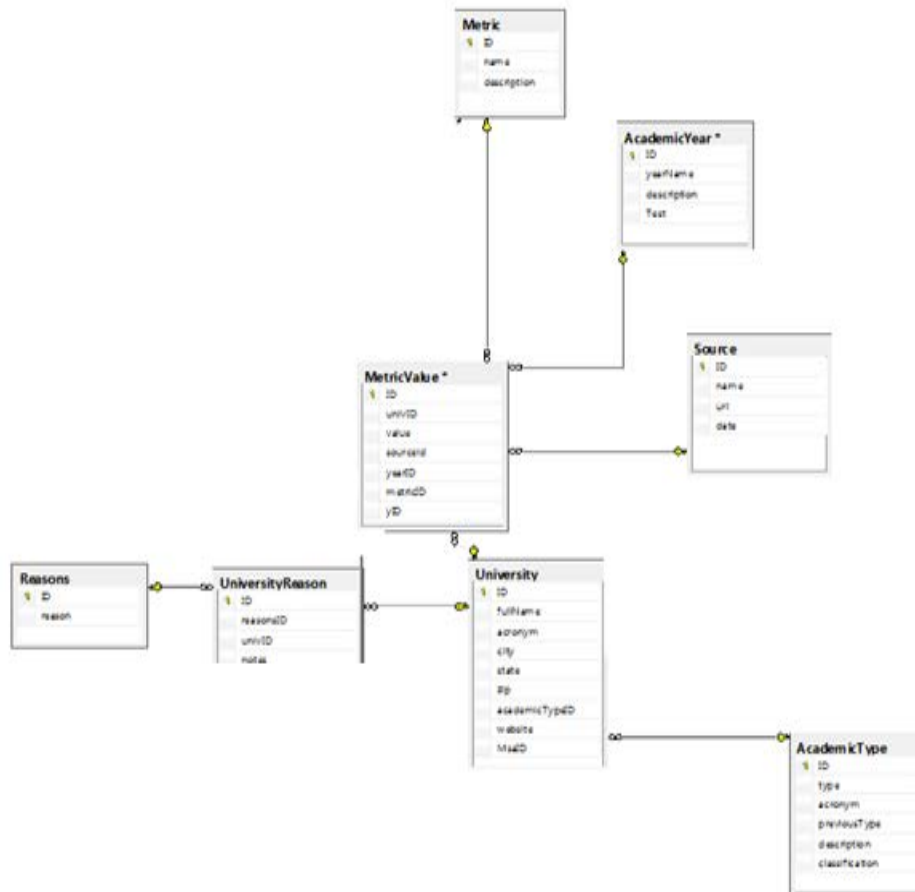


Appendix 9. Data Access

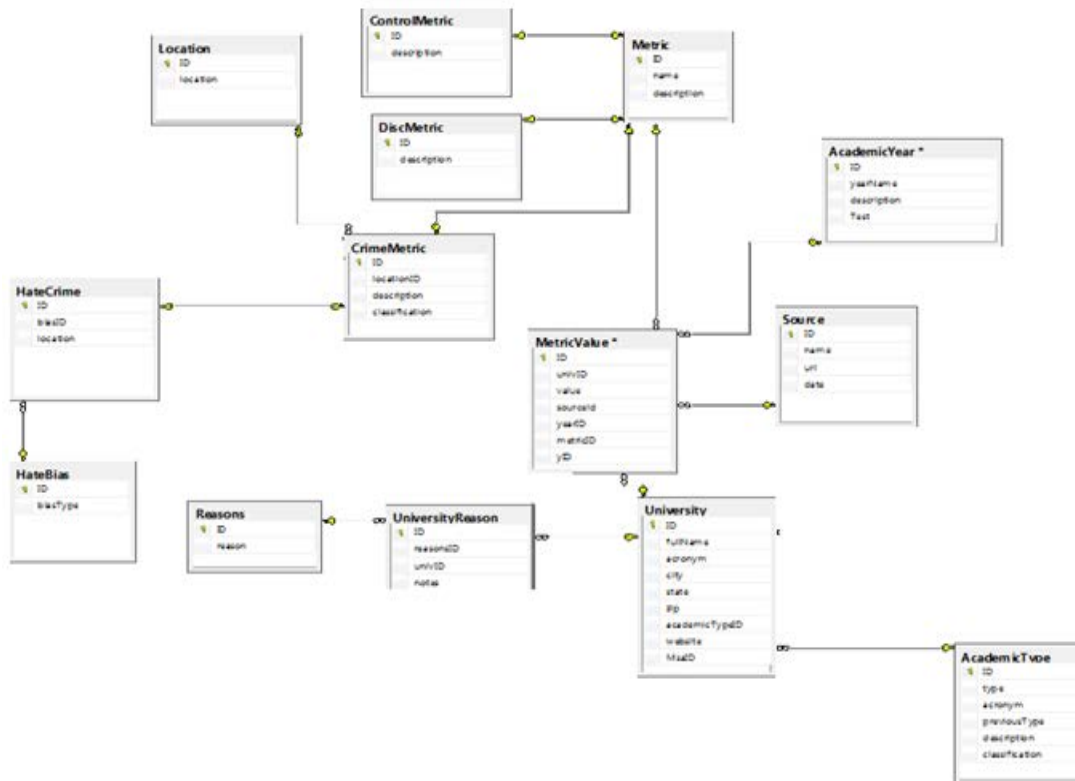


Appendix 10. Data Model Progression

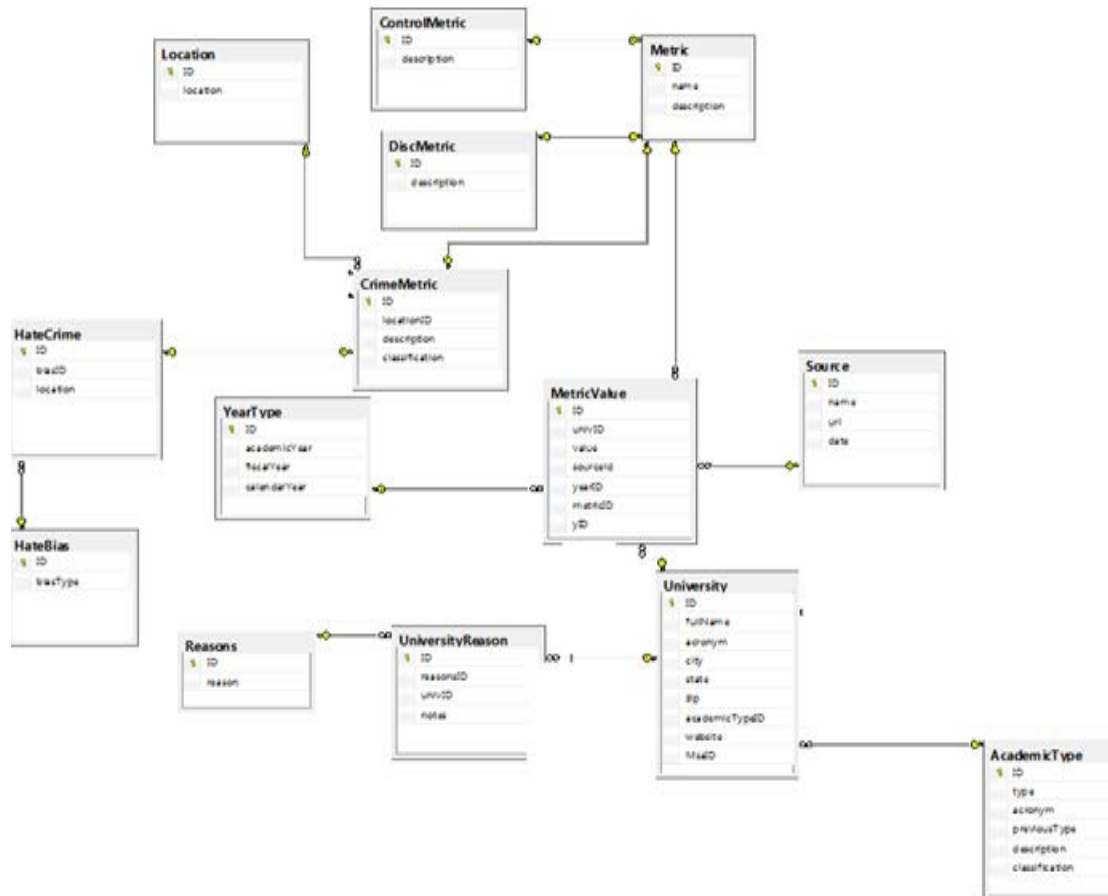
Data Model 1



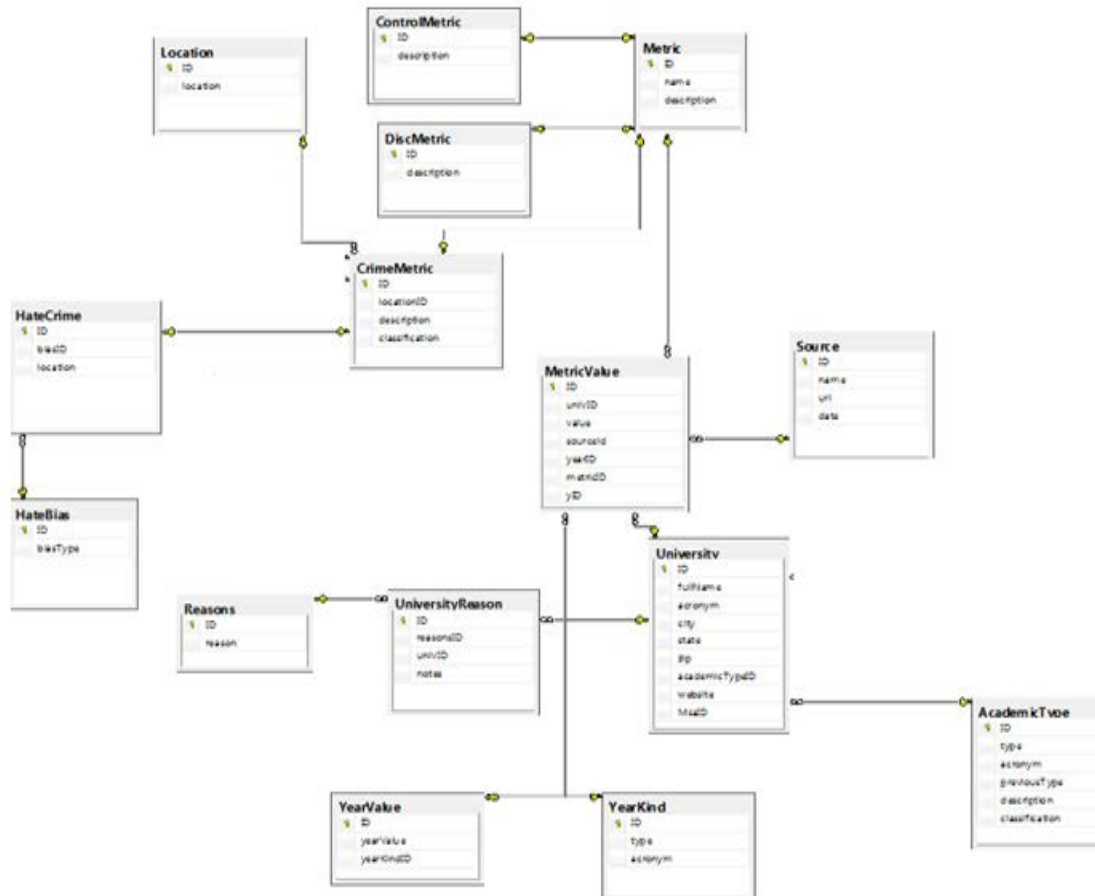
Data Model 2



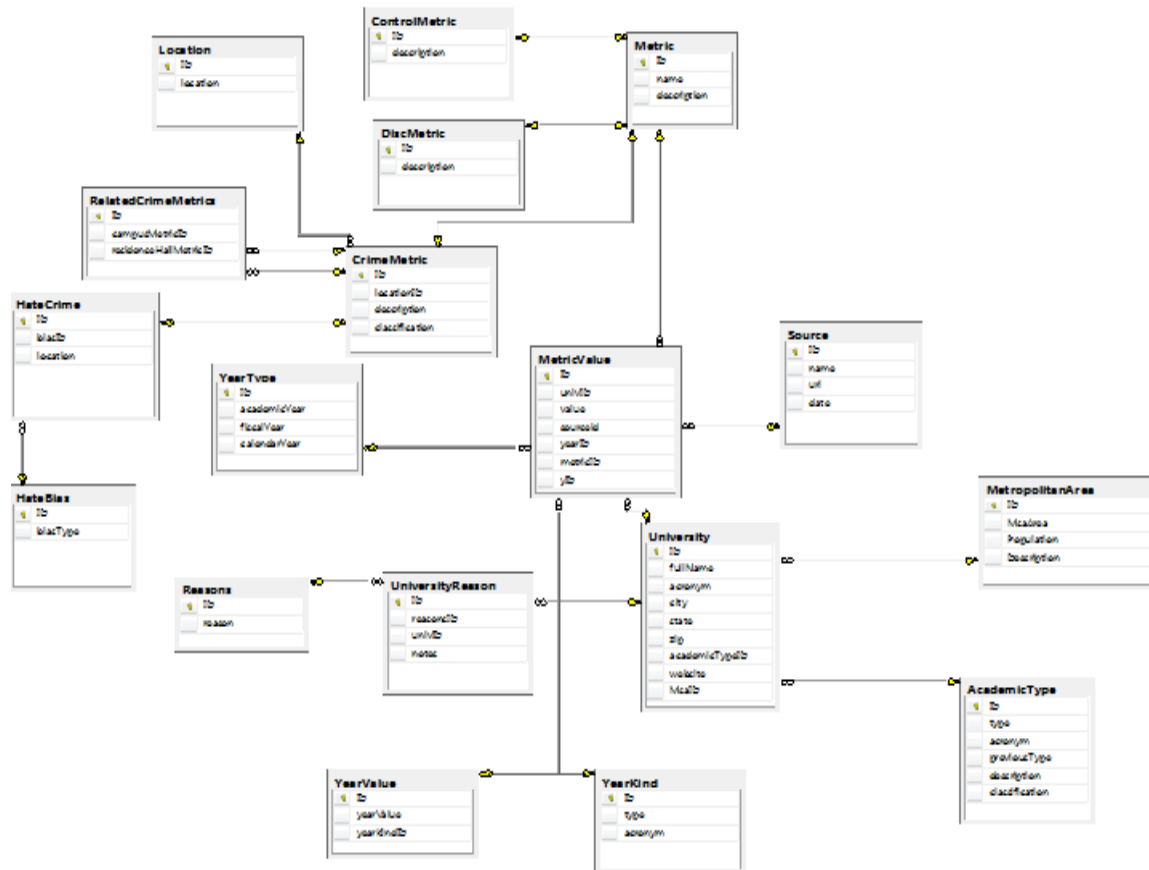
Data Model 3



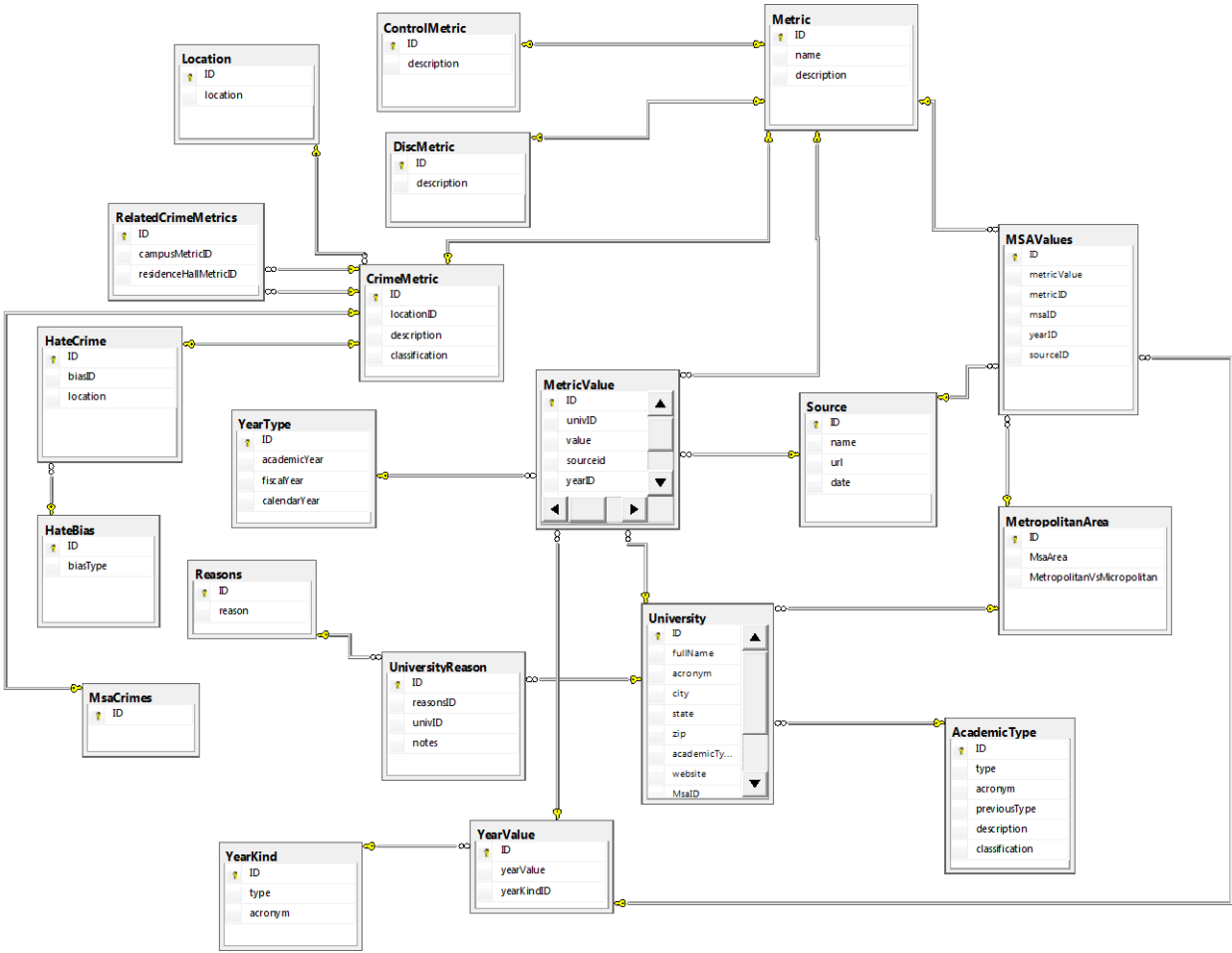
Data Model 4



Data Model 5



Final Data Model



Appendix 11. The Carnegie Foundation for the Advancement of Educational Research, and the different levels of classifications for academic institutions

The Carnegie Foundation for the Advancement of Educational Research is an independent policy and research center founded in 1905 and chartered in 1906 by an act of Congress⁶. The Carnegie Foundation for the Advancement of Educational Research center provides a framework for classifying colleges and universities in the United States⁷. The classification process started in 1970 and it derives from empirical data on colleges and universities⁸. The Carnegie Classification was originally published in 1973 and subsequently updated in 1976, 1987, 1994, 2000, 2005, and 2010 to reflect changes among colleges and universities⁹. The Carnegie Classification framework has been widely used in the study of higher education¹⁰.

The Carnegie Classification groups institution in the U.S. under the following classifications:

- Undergraduate Instructional Program
- Graduate Instructional Program
- Enrollment Profile
- Undergraduate Profile
- Size and Setting
- Basic Classification

The universities included in the UPD database are categorized under the Basic Classification category. The Basic Classification category is an update of the traditional classification framework developed by the Carnegie Commission on Higher Education in 1970 to support its research program¹¹. The Basic classification was last updated in 2005¹². The table below shows the Basic Classification categories included in the UPD database. The Previous Classification column contains the classification names that were used prior to the 2005 update¹³.

⁶ <http://www.carnegiefoundation.org/about-us/about-carnegie>

⁷ [ibid](#)

⁸ <http://classifications.carnegiefoundation.org/>

⁹ [ibid](#)

¹⁰ [ibid](#)

¹¹ <http://classifications.carnegiefoundation.org/descriptions/basic.php>

¹² [ibid](#)

¹³ <http://classifications.carnegiefoundation.org/resources/faqs.php>

Classification	Acronym	Previous Classification	Type
Research University-Very High Research Activity	RU/VH	Research 1	Doctorate-granting Universities
Research University-High Research Activity	RU/H	Research 2	Doctorate-granting Universities
Master's Colleges and Universities-Larger Programs	Master's/L	Master 1	Master's Colleges and Universities
Master's Colleges and Universities-Medium Programs	Master's/M	Master 2	Master's Colleges and Universities
Master's Colleges and Universities-Smaller Programs	Master's/S	Masters 3	Master's Colleges and Universities
Baccalaureate Colleges—Arts & Sciences	Bac/A&S	Bac 1	Baccalaureate Colleges
Baccalaureate Colleges—Diverse Fields	Bac/Diverse	Bac 2	Baccalaureate Colleges
Baccalaureate/Associate's Colleges	Bac/Assoc	Bac 3	Baccalaureate Colleges
Doctoral/Research Universities	DRU	Research 3	Doctorate-granting Universities
Theological seminaries, Bible colleges, and other faith-related institutions	Spec/Faith		Special Focus Institutions
Medical schools and medical centers	Spec/Medical		Special Focus Institutions
Other health profession schools	Spec/Health		Special Focus Institutions
Schools of engineering	Spec/Engg		Special Focus Institutions
Other technology-related schools	Spec/Tech		Special Focus Institutions
Schools of business and management	Spec/Bus		Special Focus Institutions
Schools of art, music, and design	Spec/Arts		Special Focus Institutions
Schools of law	Spec/Law		Special Focus Institutions
Other special-focus institutions	Spec/Other		Special Focus Institutions

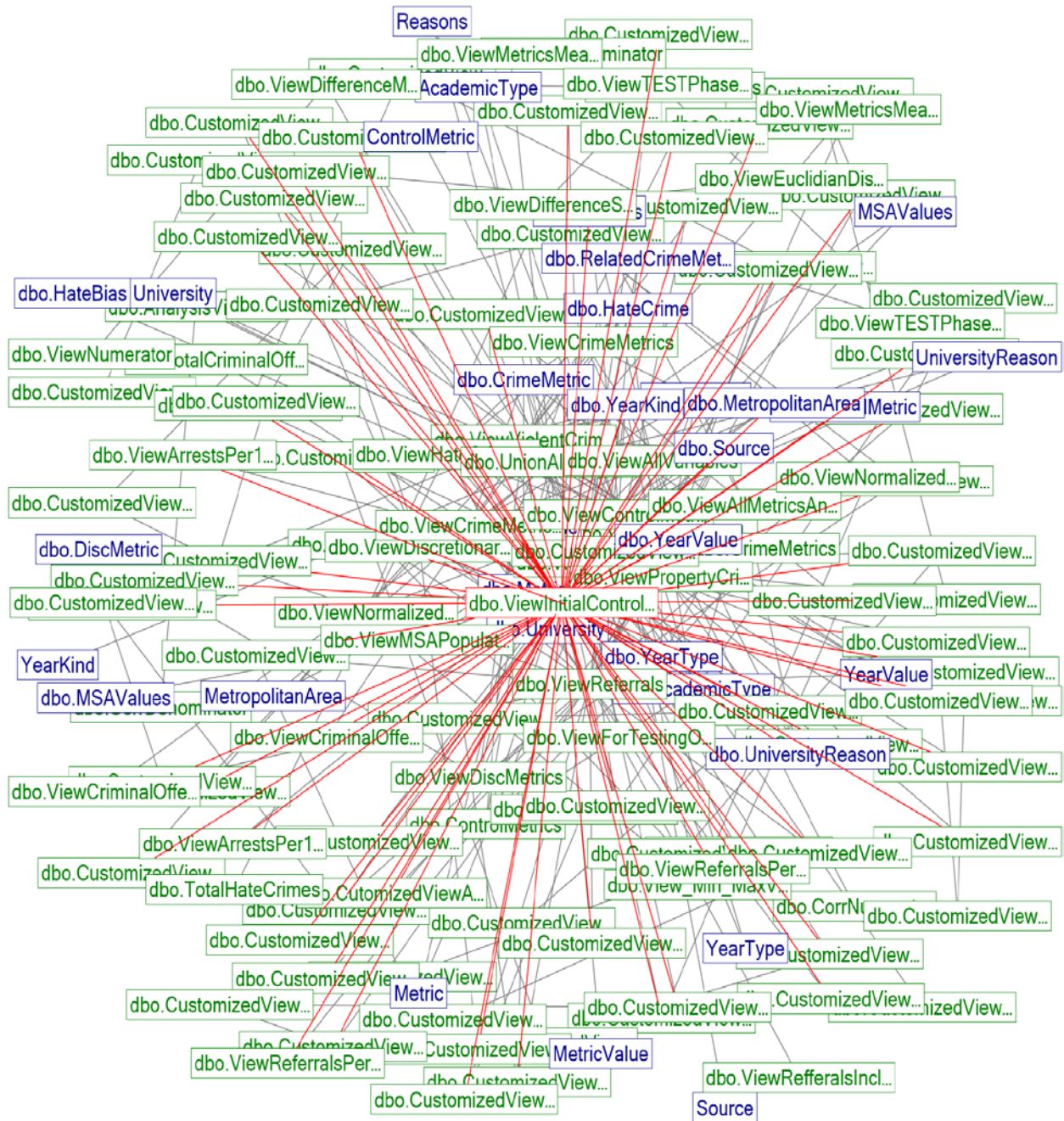
Appendix 12. Views in the Database

	<u>View Name</u>	<u>Description</u>	<u>Included In the Study</u>
1	ViewForTestingOffResidenceHallsCrimes	View breaking down crimes on campus (residence halls vs off residence halls)	Yes
2	ViewReferralsIncludingOFFResidenceHalls	View breaking down referrals on campus (residence halls vs off residence halls)	No
3	ViewCriminalOffensesIncludingOffResidenceHalls	View calculating criminal offenses taking place outside residence halls	Yes
4	CorrDenominator	View calculating the denominator in the Pearson Correlation Coefficient	No
5	ViewEuclidianDistances	View calculating the Euclidean Distance	Yes
6	ViewMetricsMeansAndStandDev	View calculating the means and standard deviation (grouped by year)	Yes
7	ViewMetricsMeansAndStandDevNotIncludingYears	View calculating the means and standard deviation (no year grouping)	Yes
8	CorrNumerator	View calculating the numerator in the Pearson Correlation Coefficient	No
9	ViewReferralsPer100TotPop	View calculating the referrals per total population (students and employees)	No
10	ViewReferralsPer100Students	View calculating the referrals per total students	No
11	ViewControlMetrics	View identifying the control metrics	Yes
12	ViewHateCrimesMetrics	View identifying the hate crimes	
13	ViewNormalizedMetricsByYear	View normaling metrics between 0 and 1 (grouped by year)	Yes
14	ViewNormalizedMetricsByYearUNCW	View normaling metrics between 0 and 1 for UNCW(grouped by year)	Yes
15	ViewArrests	View showing all the arrests on campus.	Yes
16	ViewCrimeMetrics	View showing all the crimes on campus	
17	UnionAllMetricsWithMsaPop	View showing all the data in the database	Yes
18	ViewArrestsPer100TotalPopulation	View showing arrests per 100 population on campus (students and employees)	
19	ViewCriminalOffensesPer100TotPop	View showing criminal offenses per 100 population on campus (students and employees)	No
20	ViewAllVariables	View showing only a specific metrics included in the data analysis process.	No
21	ViewInitialControlMetrics	View showing the control metrics that were initially taken in consideration	Yes
22	ViewCrimeMetricsCriminalOffenses	View showing the criminal offenses taking place on campus	No
23	ViewDiscretionaryMetrics	View showing the discretionary metrics collected	Yes
24	ViewMSAPopulation	View showing the Population of Metropolitan Areas	Yes
25	ViewPropertyCrimes	View showing the property crimes on campus	Yes
26	ViewReferrals	View showing the referrals on campus	Yes
27	ViewViolentCrimes	View showing the violent crimes on campus	Yes
28	ViewAllMetricsAndTheirValue	View used prior to adding MSA data. It show all the metrics and their value.	
29	ViewCorrelations	View used to calculate Pearson Correlation Coefficient	No
30	ViewDifferenceMeanActualValues	View used to calculate the difference between the mean and the value of a specific university metric (grouped by year)	Yes
31	ViewDifferenceMeanActualValuesNoYearsIncluded	View used to calculate the difference between the mean and the value of a specific university metric (no year grouping)	
32	ViewDifferenceSquared	View used to calculate the Euclidian Distances.	Yes
33	View_Min_MaxValue_ByMetric_Year	View used to calculate the minimum and maximum value for each metric in the database per year.	Yes
34	AnalysisViewSumSquaredVal	View used to calculate the sum of the squared values of each metric. The view was designed as a step in calculating regression analysis.	No
35	CustomizedViewAlcoholViolationsResidenceHallPer100TotalDormitoryCapacity		No
36	CustomizedViewAlcoholArrestsAndReferralsAsPercentOfTotalStudents		Yes
37	CustomizedViewAlcoholArrestsAndReferralsPer100TotalDormitoryCapacity		No
38	CustomizedViewAlcoholArrestsAndReferralsRHasPercentOfTotalDormitoryCapacity		Yes

39	CustomizedViewAlcoholArrestsAsPercentOfTotalStudents	Yes
40	CustomizedViewAlcoholArrestsAsPercentTotalDormitoryCapacity	Yes
41	CustomizedViewAlcoholArrestsRHasPercentTotalDormitoryCapacity	Yes
42	CustomizedViewAlcoholReferralsAsPercentOfTotalDormitoryCapacity	Yes
43	CustomizedViewAlcoholReferralsAsPercentOfTotalStudents	Yes
44	CustomizedViewAlcoholReferralsPer100Students	No
45	CustomizedViewAlcoholReferralsPer100TotalDormitoryCapacity	No
46	CustomizedViewAlcoholReferralsResidenceHallPer100TotalDormitoryCapacity	No
47	CustomizedViewAlcoholReferralsRHasPercentTotalDormitoryCapacity	Yes
48	CustomizedViewAlcoholViolationPer100Students	No
49	CustomizedViewAlcoholViolationsAndReferralsPer100Students	No
50	CustomizedViewAlcoholViolationsAndReferralsPer100TotalDormitoryCapacity	No
51	CustomizedViewAlcoholViolationsAndReferralsResidenceHallsPer100TotalDormCap	No
52	CustomizedViewAlcoholViolationsPer100TotalDormitoryCapacity	No
53	CustomizedViewDrugArrestsAndReferralsAsPercentTotalDormitoryCapacity	Yes
54	CustomizedViewDrugArrestsAndReferralsAsPercentTotalStudents	Yes
55	CustomizedViewDrugArrestsAndReferralsPer100TotalDormitoryCapacity	Yes
56	CustomizedViewDrugArrestsAndReferralsRHasPercentTotalDormitoryCapacity	Yes
57	CustomizedViewDrugArrestsAsPercentOfTotalStudents	Yes
58	CustomizedViewDrugArrestsAsPercentTotalDormitoryCapacity	Yes
59	CustomizedViewDrugArrestsRHasPercentTotalDormitoryCapacity	Yes
60	CustomizedViewDrugReferralper100Students	No
61	CustomizedViewDrugReferralPer100TotalDormitoryCapacity	No
62	CustomizedViewDrugReferralsAsPercentOfTotalStudents	Yes
63	CustomizedViewDrugReferralsAsPercentTotalDormitoryCapacity	Yes
64	CustomizedViewDrugReferralsResidenceHallPer100TotalDormitoryCapacity	No
65	CustomizedViewDrugReferralsRHasPercentTotalDormitoryCapacity	Yes
66	CustomizedViewDrugViolationsAndReferralsPer100Students	No
67	CustomizedViewDrugViolationsAndReferralsPer100TotalDormitoryCapacity	No
68	CustomizedViewDrugViolationsAndReferralsResidenceHallPer100TotalDormitoryCa	No
69	CustomizedViewDrugViolationsPer100Students	No
70	CustomizedViewDrugViolationsPer100TotalDormitoryCapacity	No
71	CustomizedViewDrugViolationsResidenceHallPer100TotalDormitoryCapacity	No
72	CustomizedViewPoliceStaffPer100Students	No
73	CustomizedViewPropertyCrimesAsPercentEmployeeHeadcount	Yes
74	CustomizedViewPropertyCrimesAsPercentStudents	Yes
75	CustomizedViewPropertyCrimesAsPercentTotalDormitoryCapacity	Yes
76	CustomizedViewPropertyCrimesPer100Acreage	No
77	CustomizedViewPropertyCrimesRHasPercentTotalDormitoryCapacity	Yes
78	CustomizedViewPropertyCrimesRHPer100TotalDormitoryCapacity	No
79	CustomizedViewSwornOfficersPer100Acreage	Yes
80	CustomizedViewSwornOfficersPer100Students	No
81	CustomizedViewSwornOfficersPer100TotalDormitoryCapacity	No
82	CustomizedViewTotalPoliceEmployeesPer100Acres	No
83	CustomizedViewTotalPoliceEmployeesPer100Students	No
84	CustomizedViewViolentCrimesAsPercentEmployeeHeadcount	Yes
85	CustomizedViewViolentCrimesAsPercentStudents	Yes
86	CustomizedViewViolentCrimesAsPercentTotalDormitoryCapacity	Yes
87	CustomizedViewViolentCrimesPer100Acreage	No
88	CustomizedViewViolentCrimesPer100Students	No
89	CustomizedViewViolentCrimesRHPer100TotalDormitoryCapacity	No
90	CustomizedViewViolentCrimesRHPercentTotalDormitoryCapacity	No
91	CustomizedViewWeaponArrestsPer100TotalDormitoryCapacity	No
92	CustomizedViewWeaponReferralsPer100TotalDormitoryCapacity	No
93	CustomizedViewWeaponsArrestAndReferralsRHper100TotalDormitoryCapacity	No
94	CustomizedViewWeaponsArrestsAndReferralsAsPercentTotalDormitoryCapacity	Yes
95	CustomizedViewWeaponsArrestsAndReferralsAsPercentTotalStudents	Yes
96	CustomizedViewWeaponsArrestsAndReferralsPer100Students	No
97	CustomizedViewWeaponsArrestsAndReferralsPer100TotalDormitoryCapacity	No
98	CustomizedViewWeaponsArrestsAndReferralsRHasPercentTotalDormitoryCapacity	Yes
99	CustomizedViewWeaponsArrestsAsPercentOfTotalDormitoryCapacity	Yes
100	CustomizedViewWeaponsArrestsAsPercentTotalStudents	Yes
101	CustomizedViewWeaponsArrestsPer100Students	No
102	CustomizedViewWeaponsArrestsPer100TotalDormitoryCapacity	No
103	CustomizedViewWeaponsArrestsRHasPercentTotalDormitoryCapacity	Yes
104	CustomizedViewWeaponsArrestsRHper100TotalDormitoryCapacity	No
105	CustomizedViewWeaponsReferralAsPercentOfTotalStudents	Yes
106	CustomizedViewWeaponsReferralsAsPercentTotalDormitoryCapacity	Yes
107	CustomizedViewWeaponsReferralsPer100Students	Yes
108	CustomizedViewWeaponsReferralsRHasPercentTotalDormitoryCapacity	No
109	CustomizedViewWeaponsReferralsRHper100TotalDormitoryCapacity	Yes
110	CustomizedViewAlcoholReferralsAsPercentOfTotalDormitoryCapacity	Yes
111	TotalCriminalOffenses	No
112	TotalHateCrimes	No
113	ViewTESTPhase1EuclidDistance	No
114	ViewTESTPhase2EuclidDistance	No

Appendix 13. Dependency of objects in the Database

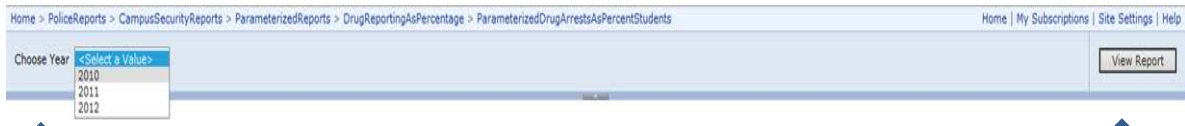
The blue squares represent the tables in the database, and the green squares represent the views. The red line highlights the dependencies associated with a specific view, *ViewInitialControlMetrics*.



Created with NodeXL (<http://nodexl.codeplex.com>)

Appendix 14. Examples SSRS Reports

Parameterized Tabular SSRS Reporting:



1) Select a value from the dropdown



2) Click on the button to display the report

3) The size of the report can be adjusted by choosing a size from the dropdown list

4) To search for an item in the report, enter a key word in the search box, such as as the name of an university, and press Find



5) To save the report in a different format, click the dropdown list and select a format from the list (excel, pdf, word, pdf)

Police Department Reporting Services

Drug Arrests as % Total Students on Campus

*Source of data: US Department of Education - Office of Postsecondary Education
-Integrated Postsecondary Education Data System

University	Relation to UNCW	Total Student Headcount	Drug Arrests	Drug Arrests as % Total Students	Year
College of Charleston	Benchmark	11,532	5	0.043	
University of Maine	Benchmark	11,501	2	0.017	2010
University of Northern Iowa	Benchmark	13,201	31	0.235	2010
Western Washington University	Benchmark	14,979	12	0.080	2010
Appalachian State University	Sister	17,222	123	0.714	2010
East Carolina University	Sister	27,783	82	0.295	2010
Elizabeth City State University	Sister	3,307	3	0.091	2010
Fayetteville State University	Sister	5,781	8	0.138	2010
NC A & T State University	Sister	10,795	38	0.352	2010
NC Central University	Sister	8,645	6	0.069	2010
NC State University	Sister	34,376	32	0.093	2010
UNC Asheville	Sister	3,967	10	0.252	2010

6) The report can be sorted by a specific column, in ascending and descending order, by pressing the column's sort button

Drill-Down, Drill-Through, Chart Reporting:

Home > PoliceReports > CampusSecurityReports > UniversityCrimes Home | My Subscriptions | Site Settings

1) Click the plus sign to move from summary information to detailed information (drill-down)

University	Location	Incident Classification	Total Occurrences	Year
Appalachian State University	Campus		1,748	Includes Years: 2010 to 2011
	Residence Hall		1,340	Includes Years: 2010 to 2011
Austin Peay State University	Campus		196	Includes Years: 2010 to 2011
	Residence Hall		131	Includes Years: 2010 to 2011
California Polytechnic State University	Campus		115	Includes Years: 2010 to 2011
	Residence Hall		74	Includes Years: 2010 to 2011
California State University - Chico	Campus		312	Includes Years: 2010 to 2011
	Residence Hall		250	Includes Years: 2010 to 2011

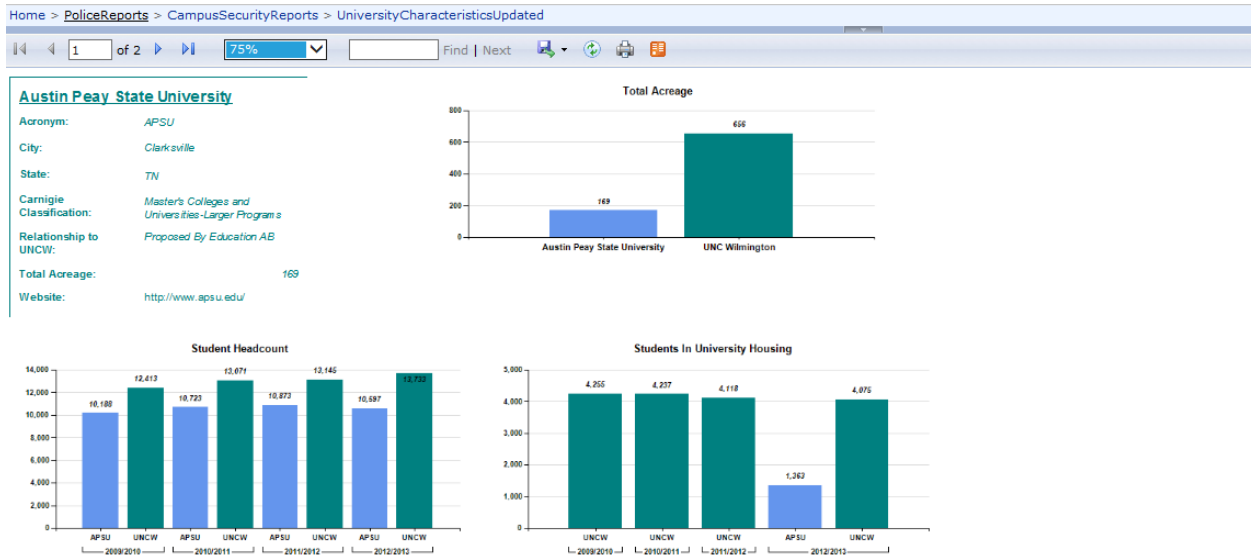
The detailed information is revealed

Home > PoliceReports > CampusSecurityReports > UniversityCrimes

Click on the university name to access institutional specifics (drill-through)

University	Location	Incident Classification	Total Occurrences	Year
Appalachian State University	Campus		1,748	Includes Years: 2010 to 2011
	Residence Hall		1,340	Includes Years: 2010 to 2011
Austin Peay State University	Campus		196	Includes Years: 2010 to 2011
		assault	7	2011
		criminal_offense	13	2011
		referral	46	2011
		arrest	11	2010
		criminal_offense	55	2010
	referral	64	2010	
Residence Hall		131	Includes Years: 2010 to 2011	
California Polytechnic State University	Campus		115	Includes Years: 2010 to 2011
	Residence Hall		74	Includes Years: 2010 to 2011



The institutional information is revealed



Click on the browser's back button to return to the previous page

Home > PoliceReports > CampusSecurityReports > UniversityCrimes

1 of 2 ? 100% Find | Next

Police Department Reporting Services

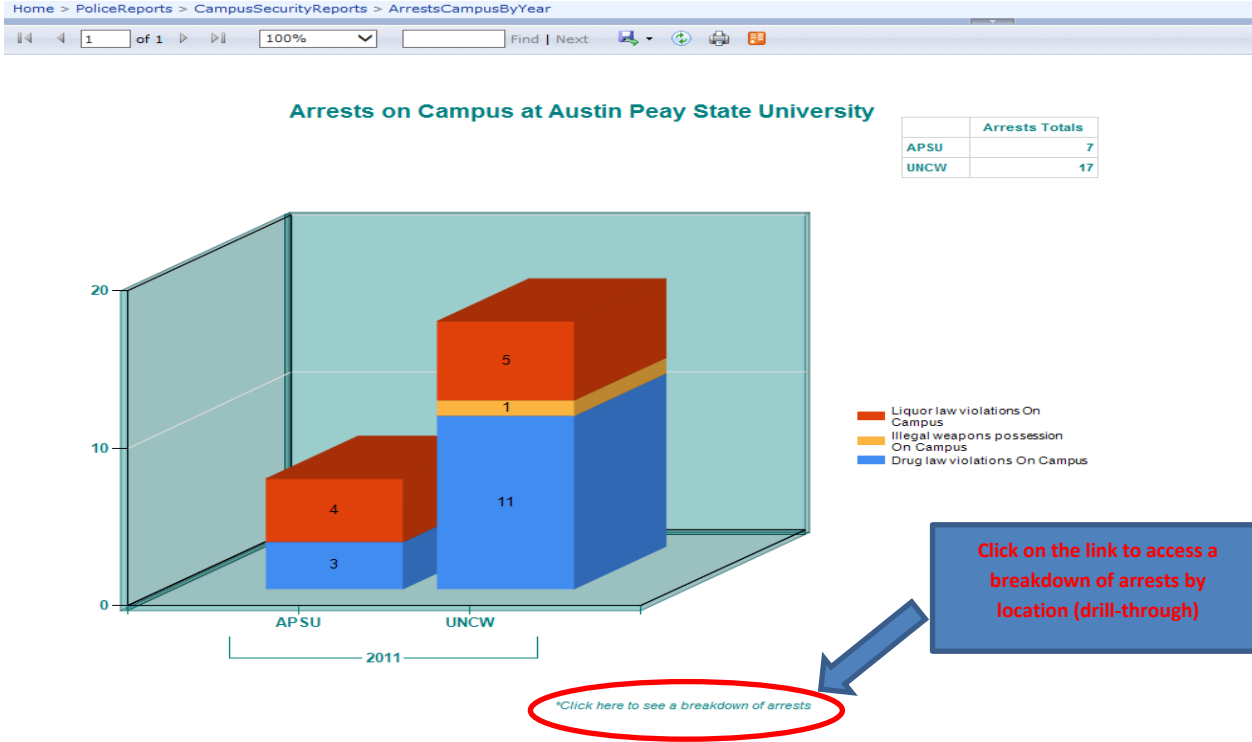
Breakdown of Crimes by University

*Source of data: US Department of Education - Office of Postsecondary Education

University	Location	Incident Classification	Total Occurrences	Year
Appalachian State University	Campus		1,748	Includes Years: 2010 to 2011
	Residence Hall		1,340	Includes Years: 2010 to 2011
Austin Peay State University	Campus	arrest	7	2011
		criminal offense	13	2011
		referral	46	2011
		arrest	11	2010
		criminal offense	55	2010
		referral	64	2010
	Residence Hall		131	Includes Years: 2010 to 2011

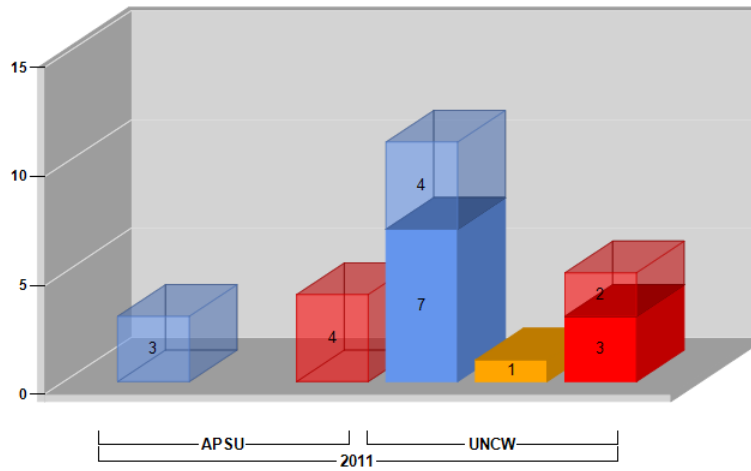
Click on a specific crime category to access a breakdown of the category and the year (drill-through)

A breakdown of the category is presented in a stacked bar chart (SSRS Chart Report). The values are compared against UNCW's values.



The arrests and their location are displayed (SSRS Chart Reporting)

Arrest at Austin Peay State University and UNCW



Metric & CampusTotals	APSU	UNCW
Drug Law Violations	3	11
Illegal Weapons Possession	0	1
Liquor Law Violations	4	5

*The transparent part of the bar represents incidents in Residence Halls and the nontransparent part incidents outside Residence Halls

Appendix 15. Correlation Analysis SPSS Output

Correlations

		Admin Staff	Dispatchers	Security Personnel	Sworn Officers	Alcohol Violations	Drug Violations	Weapons Violations	Property Crimes	Violent Crime
AdminStaff	corr	1	0.716	0.518	0.788	0.347	-0.06	0.198	-0.204	0.255
	p value		0.11	0.293	0.063	0.501	0.91	0.707	0.698	0.625
	N	6	6	6	6	6	6	6	6	6
Dispatchers	corr	0.716	1	0.245	0.795	0.771	-0.138	0.584	0.302	-0.051
	p value	0.11		0.64	0.059	0.073	0.794	0.223	0.561	0.924
	N	6	6	6	6	6	6	6	6	6
Security Personnel	corr	0.518	0.245	1	0.768	-0.097	-0.231	-0.603	-0.495	0.524
	p value	0.293	0.64		0.074	0.855	0.66	0.205	0.318	0.286
	N	6	6	6	6	6	6	6	6	6
Sworn Officers	corr	0.788	0.795	0.768	1	0.507	-0.309	0.043	-0.149	0.213
	p value	0.063	0.059	0.074		0.304	0.551	0.936	0.778	0.686
	N	6	6	6	6	6	6	6	6	6
Alcohol Violations	corr	0.347	0.771	-0.097	0.507	1	-0.37	0.79	0.352	-0.509
	p value	0.501	0.073	0.855	0.304		0.47	0.062	0.494	0.302
	N	6	6	6	6	6	6	6	6	6
Drug Violations	corr	-0.06	-0.138	-0.231	-0.309	-0.37	1	-0.085	0.618	0.698
	p value	0.91	0.794	0.66	0.551	0.47		0.872	0.191	0.123
	N	6	6	6	6	6	6	6	6	6
Weapons Violations	corr	0.198	0.584	-0.603	0.043	0.79	-0.085	1	0.509	-0.604
	p value	0.707	0.223	0.205	0.936	0.062	0.872		0.302	0.205
	N	6	6	6	6	6	6	6	6	6
Property Crimes	corr	-0.204	0.302	-0.495	-0.149	0.352	0.618	0.509	1	0.108
	p value	0.698	0.561	0.318	0.778	0.494	0.191	0.302		0.839
	N	6	6	6	6	6	6	6	6	6
Violent Crimes	corr	0.255	-0.051	0.524	0.213	-0.509	0.698	-0.604	0.108	1
	p value	0.625	0.924	0.286	0.686	0.302	0.123	0.205	0.839	
	N	6	6	6	6	6	6	6	6	6

Appendix 16. Regression Analysis Output Phase 1

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	SwornOfficers, SecurityPersnnel, AdminStaff, Dispatchers	.	Enter

a. All requested variables entered.

b. Dependent Variable: AlcoholViolations

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999 ^a	.997	.987	12.45632

a. Predictors: (Constant), SwornOfficers, SecurityPersnnel, AdminStaff, Dispatchers

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	60402.840	4	15100.710	97.323	.076 ^a
	Residual	155.160	1	155.160		
	Total	60558.000	5			

a. Predictors: (Constant), SwornOfficers, SecurityPersnnel, AdminStaff, Dispatchers

b. Dependent Variable: AlcoholViolations

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	805.687	107.200		7.516	.084
	AdminStaff	-36.861	7.970	-.392	-4.625	.136
	Dispatchers	-195.841	32.628	-1.489	-6.002	.105
	SecurityPersnnel	-41.924	3.776	-2.602	-11.104	.057
	SwornOfficers	45.717	4.313	3.999	10.599	.060

a. Dependent Variable: AlcoholViolations

Phase 2

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	SwornOfficers, SecurityPersnnel, Dispatchers		Enter

a. All requested variables entered.

b. Dependent Variable: AlcoholViolations

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.971 ^a	.943	.857	41.67594

a. Predictors: (Constant), SwornOfficers, SecurityPersnnel, Dispatchers

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	57084.233	3	19028.078	10.955	.085 ^a
	Residual	3473.767	2	1736.884		
	Total	60558.000	5			

a. Predictors: (Constant), SwornOfficers, SecurityPersnnel, Dispatchers

b. Dependent Variable: AlcoholViolations

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	863.839	356.189		2.425	.136
	Dispatchers	-209.463	108.720	-1.592	-1.927	.194
	SecurityPersnnel	-42.060	12.632	-2.611	-3.330	.080
	SwornOfficers	43.204	14.317	3.779	3.018	.094

a. Dependent Variable: AlcoholViolations

Phase 3

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	SwornOfficers, SecurityPersnnel	.	Enter

a. All requested variables entered.

b. Dependent Variable: AlcoholViolations

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.914 ^a	.836	.727	57.50626

a. Predictors: (Constant), SwornOfficers, SecurityPersnnel

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	50637.090	2	25318.545	7.656	.066 ^a
	Residual	9920.910	3	3306.970		
	Total	60558.000	5			

a. Predictors: (Constant), SwornOfficers, SecurityPersnnel

b. Dependent Variable: AlcoholViolations

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	185.884	76.160		2.441	.092
	SecurityPersnnel	-19.151	5.883	-1.189	-3.255	.047
	SwornOfficers	16.245	4.175	1.421	3.891	.030

a. Dependent Variable: AlcoholViolations