May 1, 2017

Data Warehouse for University

PROJECT REPORT

ANAND IYER, ANSHUL PRAKASH, BRIAN FERNANDES, MALLIKA MALHOTRA, SUMIT SINGH

ISYS 637-603 | GROUP 5

Contents

1. Executive Summary 3

2. Business Case 4

2.1 Project Objective *4*

2.2 Business Drivers *4*

2.3 Business Requirements *4*

3. Data Warehouse Architecture 7

3.1 Components of the Architecture *9*

3.2 Data Warehouse Bus Matrix *9*

4. Dimensional Model Design 10

4.1 Dimensions *11*

4.2 Facts *12*

5. Data Extraction, Transformation and Loading 14

6. Reporting 20

6.1 Reports related to Students and Applicants *20*

6.2 Reports related to students' GPA and academic performance *25*

6.3 Reports related to Placements of the students *27*

7. Meta data 32

8. Appendices 40

8.1 Appendix A: OLAP definitions *40*

8.2 Appendix B: SSIS Terminology *40*

9. Glossary 41

10. Bibliography 42

11. Project Work Breakdown 43

1. Executive Summary

As of late, Wales University has experienced major problems in taking decisions and reporting. The management is struggling to carry out complex reporting tasks and gathering data to generate such reports. The data required for reporting / decision-making is often dispersed / duplicated across the systems. As the education sector is increasingly becoming more complex and competitive, Wales University needs to ensure that it is using its significant data resources to obtain the information that it needs to make informed, timely evidence-based decisions. Wales University also needs to ensure that it has the data resources to optimally respond to the demands placed on American universities to report statistical information to external / government agencies.

Our project aims to create a data warehouse for the Wales University to help them in taking informed strategic decisions and report timely to maintain compliance. Our comprehensive dataset incorporates applicant data along with their entrance exam scores, student enrollment data, departments of the university, all the majors and the courses offered by Wales University, student graduation data, GPAs and other scores of students of the university for each term and student's placement / job-related data over the years.

This report encompasses the different components we have considered ultimately resulting into a data warehouse. The major components of this report are:

1. Data Warehouse Architecture
2. Dimensional Model
3. ETL process depiction and integration through SSIS
4. Reporting and trend analysis through Tableau

The recommendation to our client, Wales University, is to make efficient use of our data warehouse solution to increase efficiency of operations, improve reporting efficiency and enhance decision-making analysis in various fields across the university. Impact of the data warehouse can be maximized by using the data available as a shared asset. The stakeholders of the University can, thus, use our data warehouse which has gathered data from all the separate information systems maintained by different departments in the University to view University-level, Student-level, Faculty-Wise, Student-wise and Course-wise reports that supports decision making.

In conclusion, the purpose of this project is to help Wales University improve its rankings and functions by helping the management take data-driven decisions.

2. Business Case

Wales University has a range of transactional or operational systems. These support high volumes of repetitive functions and store the associated very large volumes of data (such as faculty and student records). These systems are used in the day-to-day operations of the university and the systems and their data sets are optimally designed for this purpose. Therefore, transactional systems are not well designed for the higher-level aggregation, extraction and analysis that is required in the reporting, monitoring, planning and decision making associated with management and strategic planning of the University. The University wants a way to generate all the reports / statistics required for compliance and aid the stakeholders in making better strategic decisions for the University.

2.1 Project Objective

The aim of this project is to help Wales University improve its global rankings and functions and to store all the data generated by various departments and administration of the University in a cohesive way allowing the management to take data-driven decisions and maintaining compliance. This will be achieved through an extensive analysis on the University's data for the past few years. Our exhaustive dataset includes applicant data along with their entrance exam scores, student enrollment data, student's placement data over the years, departments of the university, all the majors and the courses offered by Wales University, student graduation data, student grade point averages for each term and cumulative grade point averages as well as other scores of students of the university for each term.

2.2 Business Drivers

- Enrollment / Admission of students
- Course Registration
- Current Academic Performance of students
- Academic Performance of the students over the years
- Placement of the students

2.3 Business Requirements

We started with the identification of important business questions the answers to which would help the university in making effective decisions. Identification of the questions was followed by data preparation and analysis to come up with qualitative and quantitative comparisons which will form the foundation for all the decisions that Wales University would take.

The business questions important for the University are listed below along with their business justification:

1. **Does the exams score used in admissions act as an indicator of academic performance at University?**

   University uses scores from various exams like SAT, GRE, GMAT etc. as a criterion while handing out admits. There are other criteria's as well such as essays, recommendations, work experience etc. This analysis can help make sure if these exam scores as a criterion are useful or not. If they are not a good indicator of academic performance then admissions committee can rethink on the exams being considered or decrease the weightage of them among other criteria's. They can also provide this as a feedback to testing agencies that conduct the exams.

2. **Which is the most applied for major, whose capacity can be increased?**

   Trend analysis for applications coming for each major can help determine which majors are in demand and which are not. Based on this university can decide to increase intake for those majors with corresponding plan for resources required to maintain same standards with increased intake. If there is a decreasing trend for a major then university can find out the factors causing the decrease and make appropriate action plan.

3. **Which state has the most number of employing companies?**

   This information can be used by the university to identify which companies hire our students the most and in which location. This will give the university a list of potential companies who can be invited to the career fair at the university or to host networking events at the university. Very often, companies which hire candidates from a particular university also invest in the university by making donations, offering scholarships or sponsoring events. Hence, this information will be great to increase the donations from companies to the university. The state information will help us promote our university among companies in states where our students are hired.

4. **Is the average salary affected by the VISA status of the students?**

   This question will help us identify if there is a trend in the average annual salary that students receive based on their visa status. For example, we can see if F1 students receive a higher average annual salary than local students. Trends can also be identified based on different courses/ departments. This information can be used to promote the courses to the respective students which will help improve student enrollment in the university.

5. **What is the effect of scholarship on student acceptance rate?**

   This information can be used by the university to decide whether the scholarships they offer each year to the students play a role in increasing the admission count. Based on this information, the university can take steps to increase / decrease the funding for incoming students. If the trends over the years show that applicants who are offered scholarships, do not join the university, then the management can decrease the funding and use it for other purposes.

6. **Which 10 companies offer the highest compensation to students of Wales University?**

This question will help us identify the top 10 companies that provide the students across different majors with the highest compensation. These companies can then be used as leads that can become sponsors for the departments at Wales Universities. The management could also contact the universities identified as a part of this question to the career fairs or recruiting events and can work with them to understand the skills these companies are looking for in students and accordingly enhance their programs/courses.

7. **Is the highest education level correlated to a candidate's GPA?**

The university admit the students hailing from diverse backgrounds. The open education system provides the opportunity to students from diverse backgrounds and education level to take admission in completely different area of education. This report will provide insights about student performance based on their highest education level. We examine the prediction of individual performance of each candidate based on his previous education level. This report also provides the insight about likelihood of a student choosing majors and minors based upon the previous educational qualification.

8. **Were there any instances of grade inflation for any major over the years in Wales University?**

Grade inflation is also an important aspect of grading system. In grading system sometimes defining grades and corresponding scaling measure become a nightmare. So, we need to provide a quantitative analysis of grades awarded in a course. This report will provide a bull's eye view about a certain course and its corresponding division of grades over the years. This can also be utilized while deciding grading scheme during the curve fitting of a class.

9. **Does the academic standing affect the employment status of a candidate?**

Companies which generally come for on-campus hiring, place restrictions on the minimum GPA a candidate must have in order to be eligible for employment with them. This is generally a status statement for companies to showcase its clients that they employ only academically bright students. But are academics and employment really correlated? It has become a common trend now following the likes of Bill Gates, Steve Jobs and Mark Zuckerberg to drop out of college and be successful.

10. **Which factors can be attributed to length of job search?**

By understanding the factors that are directly and indirectly proportional to the length of job search for a student, the career center can make informed decisions while assisting students in their pursuit. Though there can be factors that affect a student's job search, one of the most important parameters is the major enrolled. For instance, a major like Computer Science would be very popular in a state like California, with bright job prospects in Bay Area, A degree in Petroleum Engineering would not be so useful at the same place.

3. Data Warehouse Architecture

When it comes to designing data warehouse for any business, the two most commonly discussed methods are Inmon approach and Kimball approach. Both philosophies have their own advantages and disadvantages, and enterprises continue to use either of these as per their business / sector.

Bill Inmon's approach (the Top-Down Design): A normalized data model is designed first. Then the dimensional data marts, which contain data required for specific business processes or specific departments are created from the data warehouse.

Ralph Kimball's approach (the Bottom-Up design): The data marts facilitating reports and analysis are created first; these are then combined together to create a broad data warehouse.

For our data warehouse architecture, we decided to go with the Inmon approach. Thus, for our data warehouse, a university wide data warehouse is organized first and then separate data marts are created for specific business processes so that all the data elements relating to the same real-world event or object are linked together. The changes to data in the database will be tracked and recorded so that reports can be generated showing changes over time. This architecture makes sure that data warehouse contains data from all of university's operational applications, and that this data is made consistent.

Inmon defines data warehouse as "*A centralized repository for the entire enterprise*". Data warehouse stores the atomic data at the lowest level of detail. Dimensional data marts are created only after the complete data warehouse has been created. Thus, data warehouse is at the center of the information factory, which provides a logical framework for delivering business intelligence.

The diagram on the next page gives the look and feel of our data warehouse model.

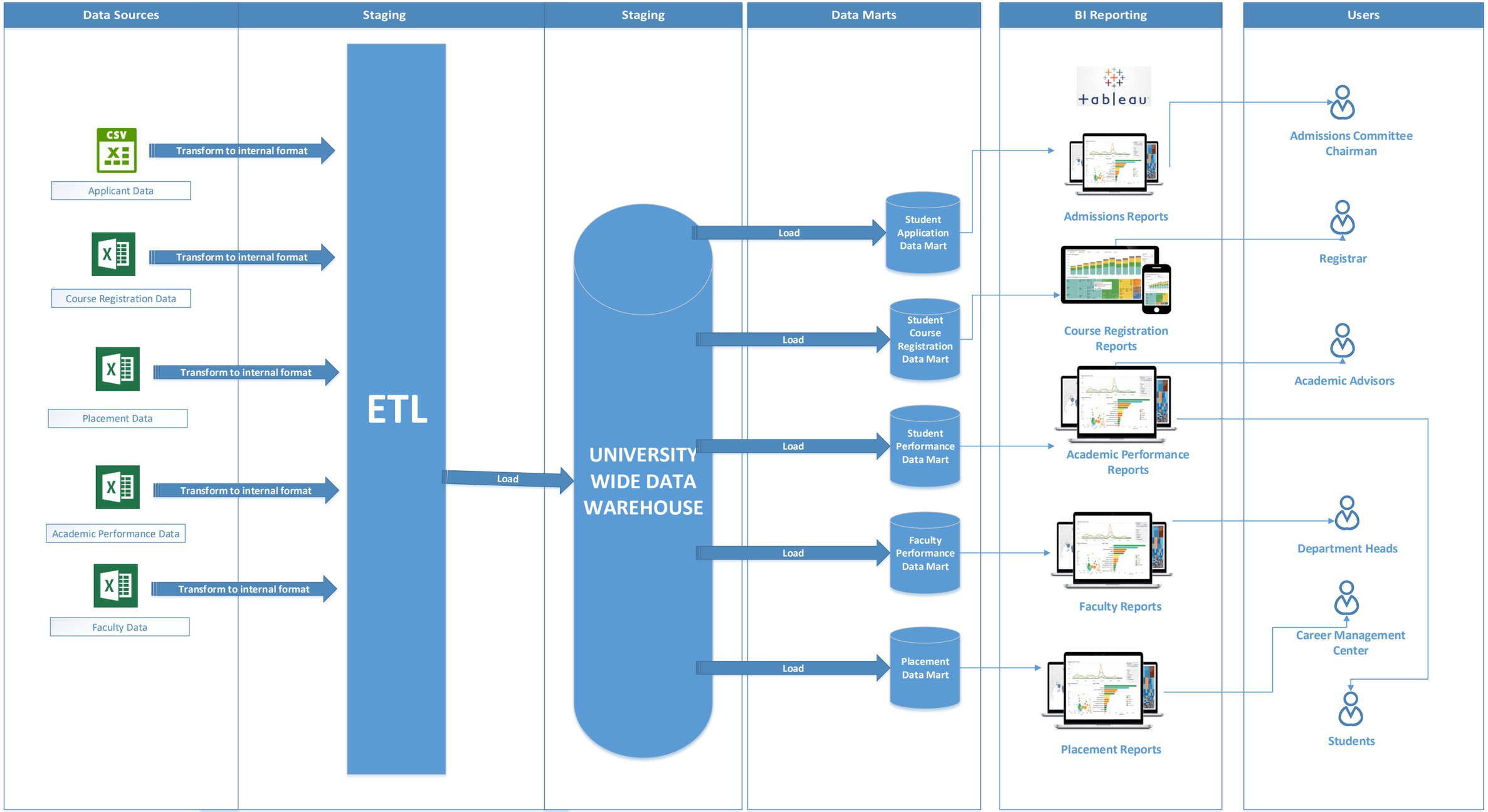

Figure: Data Warehouse Architecture

3.1 Components of the Architecture

Components / Processes	Description
Source Data	We used CSV and Excel files to get the raw data for our data warehouse model
Data Store	The data had to be first extracted and transformed into a format suitable for querying analysis before it is uploaded. We used SQL Server and SQL Server Integration Services (SSIS)
Data Extraction, Transformation and Loading	SSIS was used to extract data from the CSV and Excel files and load into a database server. The data was transformed (cleaned) in SSIS as it was raw and had junk values. The clean and standardized data was finally loaded to the data warehouse
Data Presentation Area	Tableau was used to organize and store data. Analysis could be performed on this data. For the business users, it is all that they see and they interact with the data presentation area via the data access tools.
Data Access Tools	Data Access Tools help in running reports, creating organized listings and to perform cross-table querying. We used Tableau to create multiple complex reports for statistical and strategic purposes
Reports	The business questions are answered by various reports generated

3.2 Data Warehouse Bus Matrix

The bus architecture decomposes the data warehousing or business intelligence planning process into manageable pieces by focusing on the organization's core business processes, along with the associated conformed dimensions. It allows to present the relations between measure group and dimensions in a tabular form that is easy to comprehend.

Business Processes	Common Dimensions									
	Dim_ Applicants	Dim_ Exam	Dim_ Major	Dim_ Date	Dim_ Student	Dim_ Course	Dim_ Section	Dim_ Instructor	Dim_ Company	Dim_Company _Location
Fact_Applications	X	X	X	X						
Fact_StudentCourseMapping				X	X	X	X	X		
Fact_SectionWiseGPA				X		X	X	X		
Fact_SWCumulativeGPA				X	X			X		
Fact_Placement					X				X	X

4. Dimensional Model Design

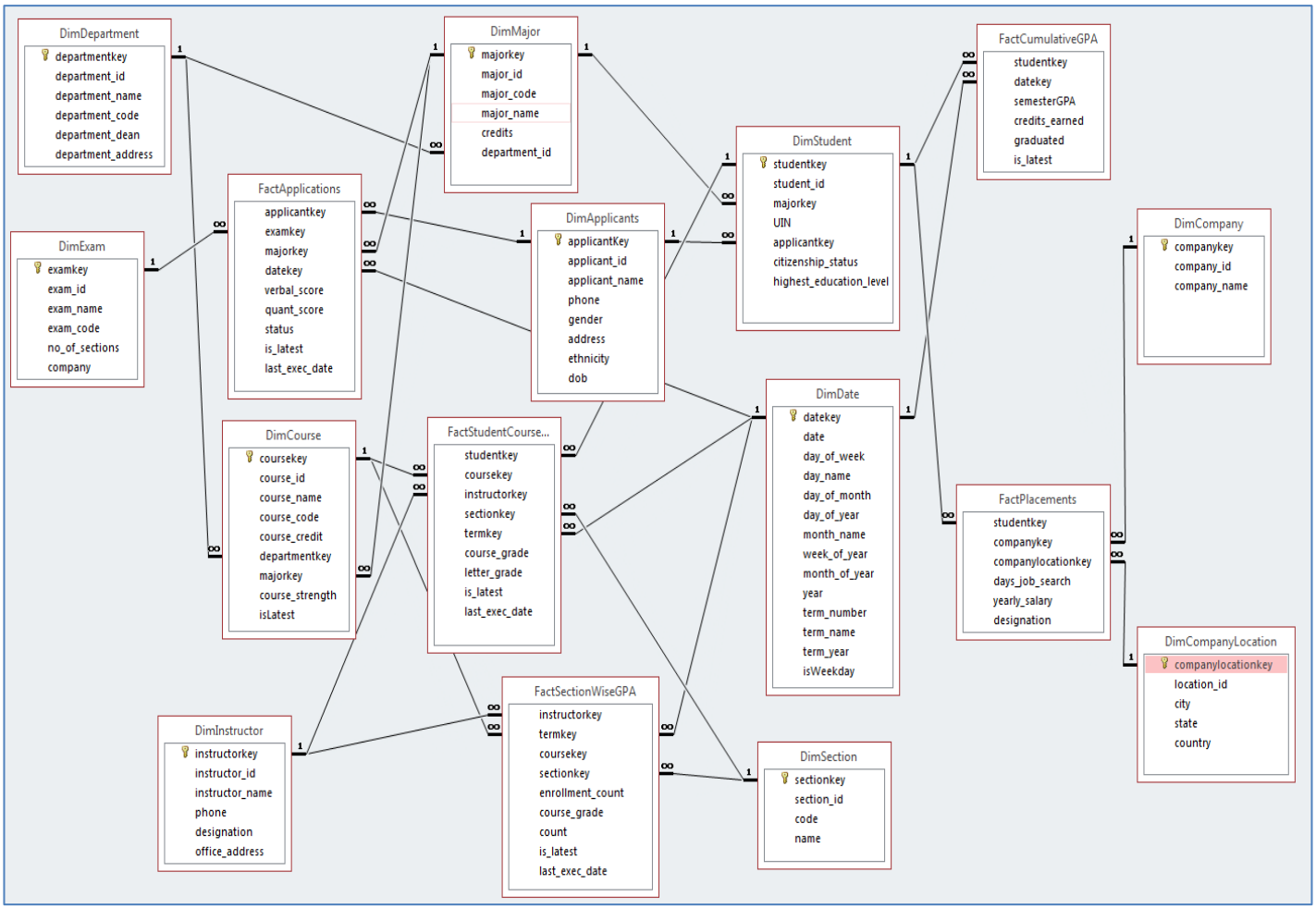

Figure: Dimensional Model

While designing the dimensional model, we tried to follow the best possible approach and decided to use snowflake model.

The snowflake schema is represented by centralized fact tables which are connected to multiple dimensions. In the snowflake schema, dimensions are normalized into multiple related tables. When it is completely normalized along all the dimension tables, the resultant structure resembles a 'snowflake' with the fact table in the middle. Snowflake schemas in general use less space to store dimension tables as any normalized database produces far fewer redundant records.

We have created **eleven dimensions** and **five fact tables** in our dimensional model to store data for applicants who apply to the various courses offered by the Wales University along with their entrance exam results, students who get admitted to the University and their CGPAs, scores etc., the various departments of the University and the majors and courses they offer, the instructors for these courses and the various sections they teach, the placement data for the students along with the details of their companies and location after graduation.

4.1 Dimensions

4.1.1 Dim_Applicants

This dimension contains information about applicants such as name, phone, address, gender and ethnicity that apply to university for admissions. Applicant_id is the **primary key** used within the source system. Applicant_key is a **surrogate key** generated while uploading data using SSIS.

4.1.2 Dim_Exam

This dimension contains information about the exams, which are used by university to make decision for admitting applicants. A new column is added if there is any change in attributes of exam to preserve attribute's current and prior values. Exam_key is a **surrogate key** generated during ETL.

4.1.3 Dim_Major

This dimension contains information about various majors offered by a department. Department_key is foreign key, which connects this dimension with department dimension. A new column is added if there is any change in attributes such as name of major to preserve attribute's current and prior values. Major_key is a **surrogate key** generated during ETL.

4.1.4 Dim_Department

Dim_Department contains information such as name, dean and address for various departments of university. Department_Key is a **surrogate key** generated while uploading data using SSIS.

4.1.5 Dim_Student

Dim_student contains information about student such as UIN, name, age, contact details, exam scores and other demographic information. Student_Key is a **surrogate key** generated while uploading data using SSIS. Major_key and applicant_key act as foreign keys to link this dimension to major and applicant dimensions.

4.1.6 Dim_Course

This dimension contains information about courses, which are offered by departments of university. Course_key is a **surrogate key** generated during ETL.

4.1.7 Dim_Section

Dim_Section contains information about sections that a course may have. Section_Key is a **surrogate key** generated while uploading data using SSIS.

4.1.8 Dim_Instructor

Dim_Instructor contains information about instructors such as name, phone, designation and

address. Instructor_Key is a **surrogate key** generated while uploading data using SSIS. A new column is added if there is any change in attributes of instructor such as designation and address to preserve attribute's current and prior values. It is a **conformed** dimension that is used in fact tables related to student course mapping and cumulative GPA for a course.

4.1.9 Dim_Company

This dimension contains information about companies where students are placed. The attributes in this dimension do not change with time.

4.1.10 Dim_Company_Location

This dimension contains information about locations where students are placed. The attributes in this dimension do not change with time.

4.1.11 Dim_Date

Dim_Date contains information about a date like day, month year and term. Day, month and year are derived using ETL tools during upload. The attributes in this dimension do not change with time. Date_Key is a surrogate key generated during ETL.

4.2 Facts

4.2.1 Fact_Applications

Applications Fact table contains four foreign keys from the following dimensions:

- Applicant
- Exam
- Major
- Date

This table contains information about the applicant including his/her scores for exams taken to apply to the Wales University and his/her registration information. The information in this fact table can be used to evaluate the student's academic potential and to ensure if he/she is a good fit for the University.

4.2.2 Fact_StudentCourseMapping

StudentCourseMapping Fact table contains five foreign keys from the following dimensions:

- Student
- Course
- Instructor
- Section
- Date

This table contains information about the student and the course details including the instructor and section for the course the student is interested in taking. The fact table also has information on

the student's grade on the course and his/her letter grade.

4.2.3 Fact_SectionWiseGPA

SectionWiseGPA Fact table contains four foreign keys from the following dimensions:

- Course
- Instructor
- Section
- Date

This table contains information about the student and the various scores obtained by the student in a course taught by a particular instructor in a particular section. The fact table also has information on the grade point averages (GPA) for the students including the average GPA and maximum GPA section wise.

4.2.4 Fact_SWCumulativeGPA

StudentCourseMapping Fact table contains three foreign keys from the following dimensions:

- Student
- Instructor
- Date

This table contains information about the student including the grade point average of the student for a particular term / semester and the cumulative grade point average for the students. The fact table also has information on the credits earned by the student and his/her graduation status.

4.2.5 Fact_Placement

Placement Fact table contains three foreign keys from the following dimensions:

- Company
- Student
- Company_Location

This table contains information about the student and his/her placement details including the company the student started working with after his/her graduation and the location of the company. The fact table also has information on the student's position and salary in the company.

5. Data Extraction, Transformation and Loading

The data in a Data Warehouse system is loaded with an ETL (Extract, Transform, Load) tool. The ETL tool performs the following three operations:

- Extracts the data from your transactional system which can be an Oracle, Microsoft, or any other relational database
- Transforms the data by performing data cleansing operations
- Loads the data into the data warehouse

Our primary data source for this project has been various CSV and Excel files that we either found on the internet or generated ourselves. These files had junk and null values and therefore, had to be cleaned before loading them onto SQL server. We used SSIS to clean the data before uploading it to the data warehouse.

Below screenshots depict transformations that we performed on our data to clean it before loading it into the data warehouse:

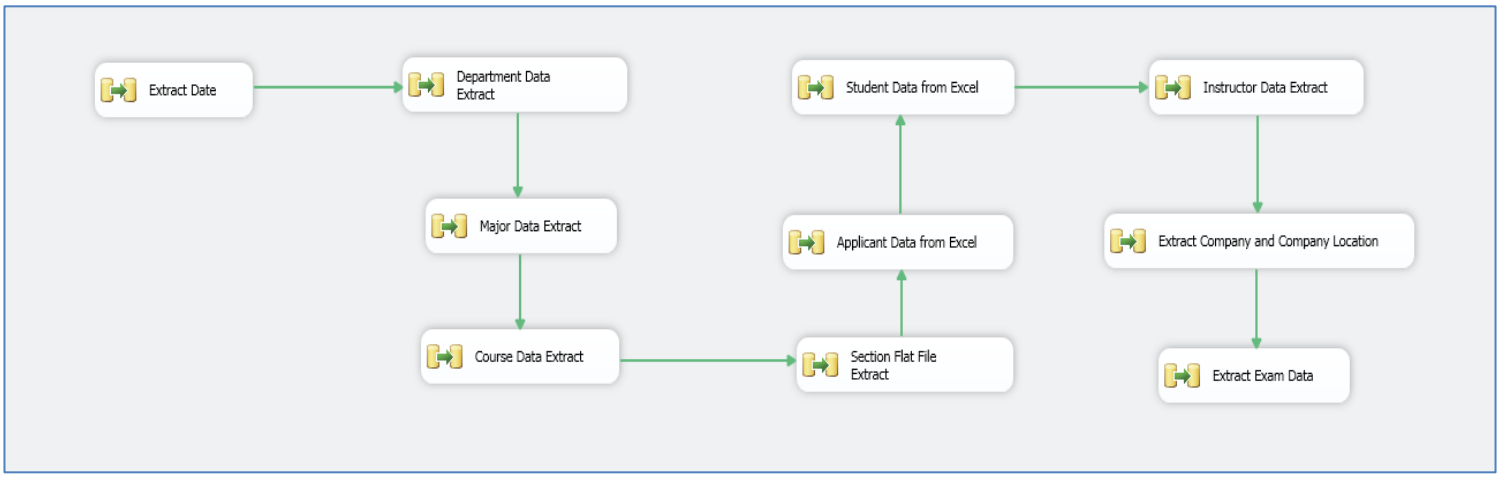

Figure: SSIS – ETL data flow

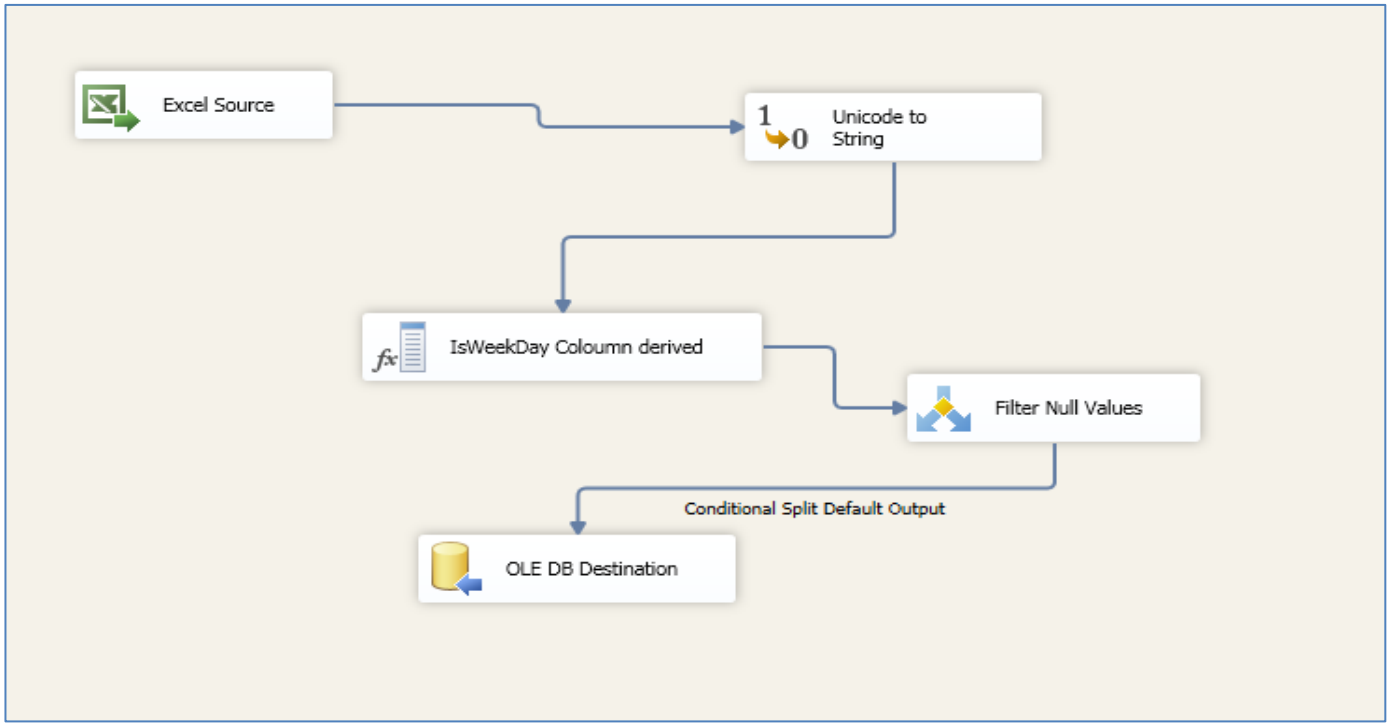

Figure: Dim_Date Creation and Extract

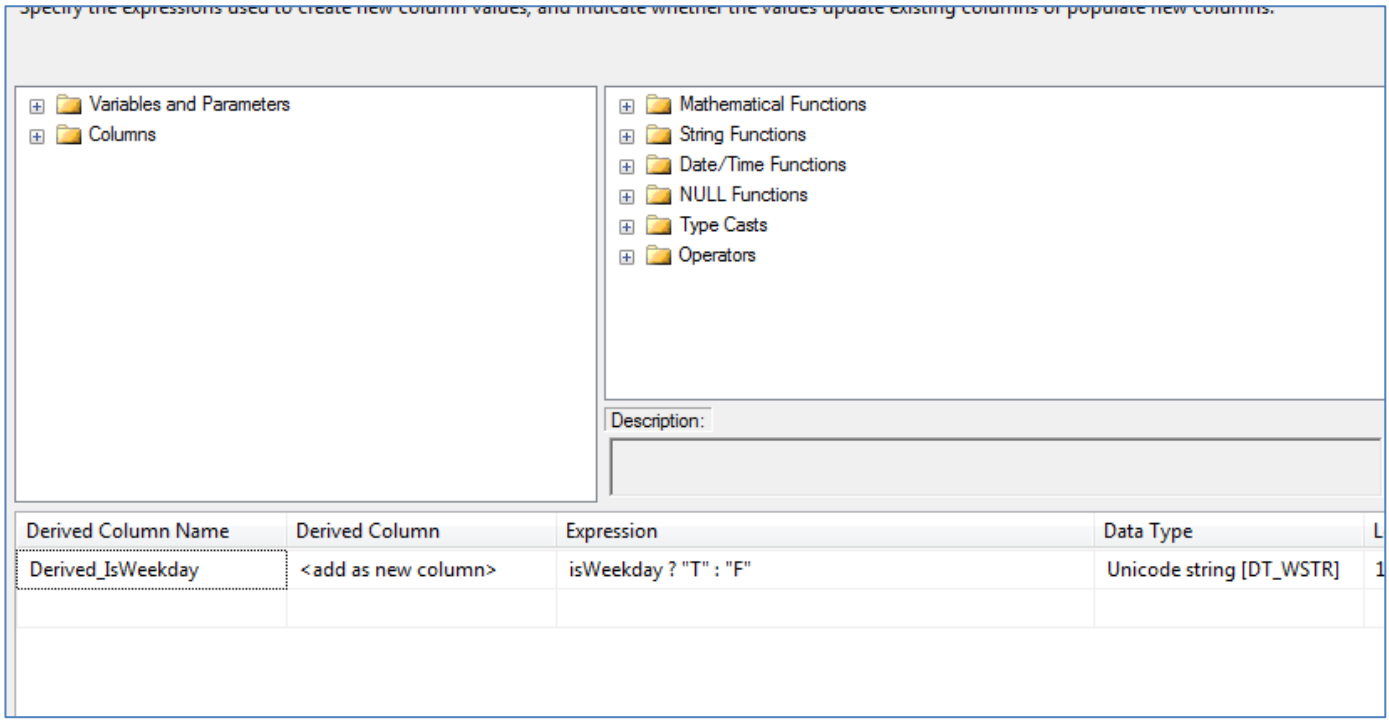

Figure: Deriving day, month and year from date to create Dim_Date

We used the derived column functionality to derive full names of applicants and students

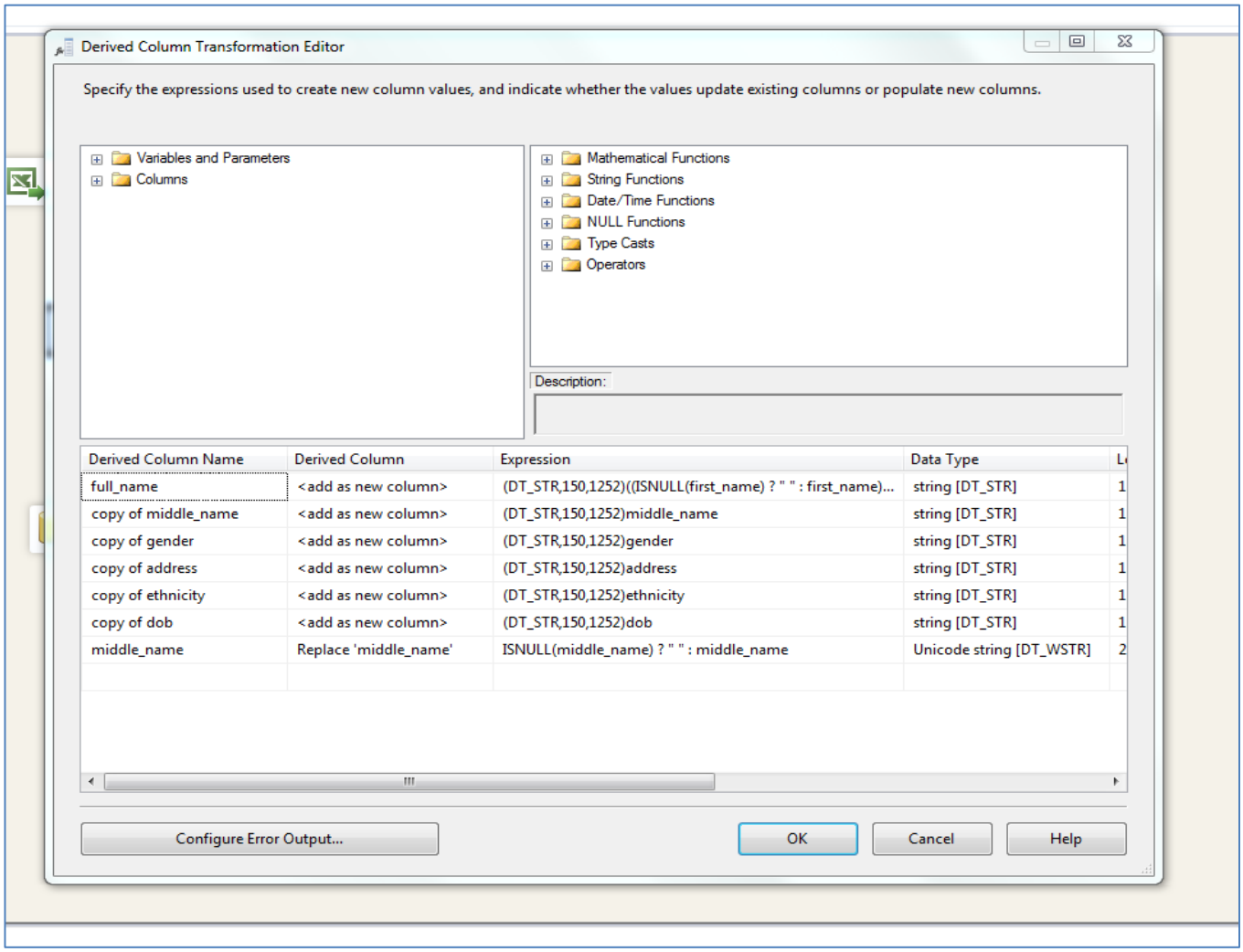

Figure: Deriving full names of students and applicants

We then performed null checks to filter out the null values from our datasets before loading them into the data warehouse

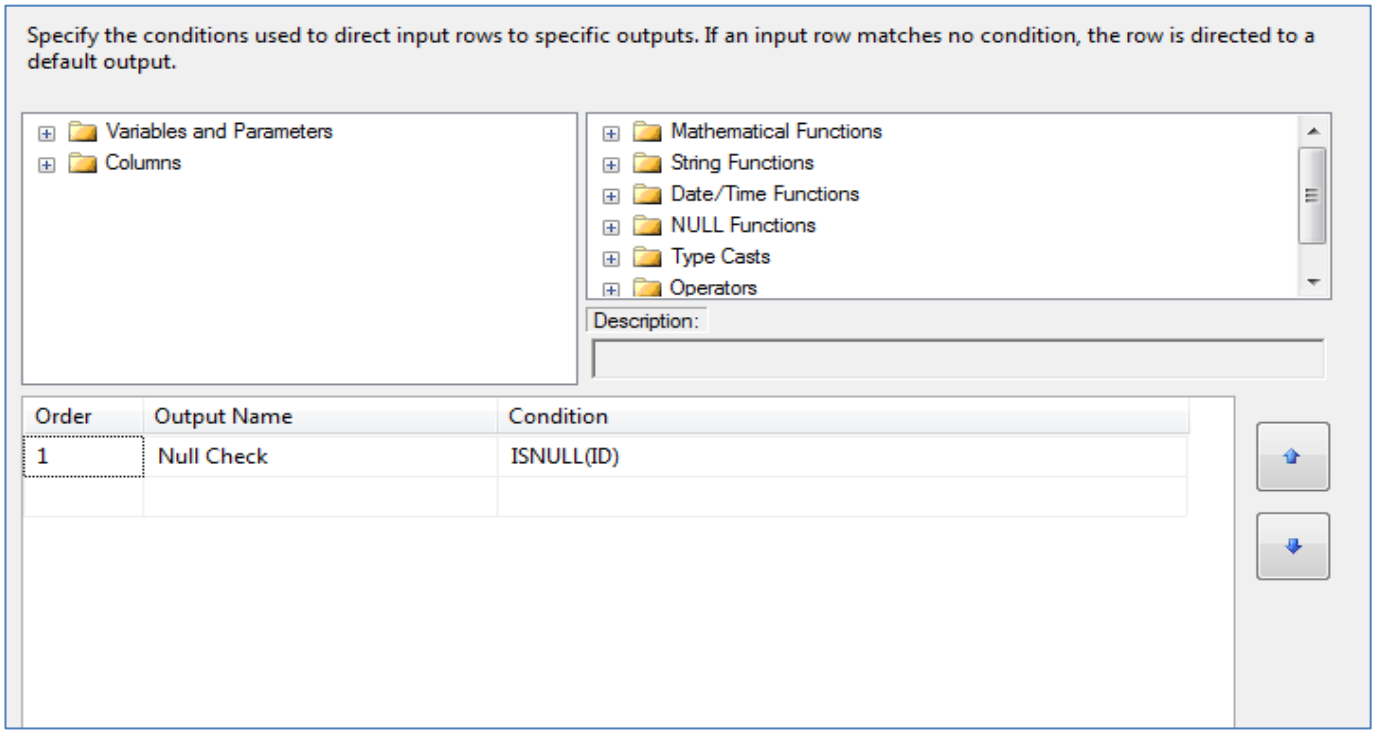

Figure: Performing null check on data before loading

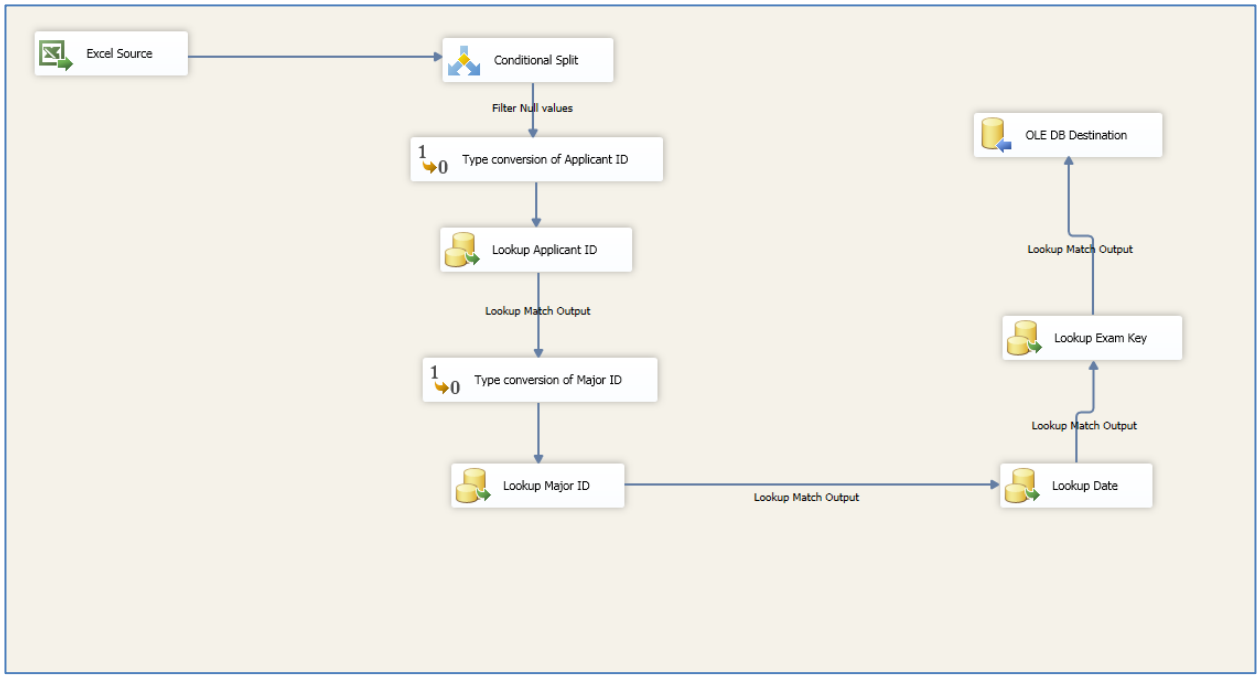

Figure: Data Flow for the Applicant Dimension

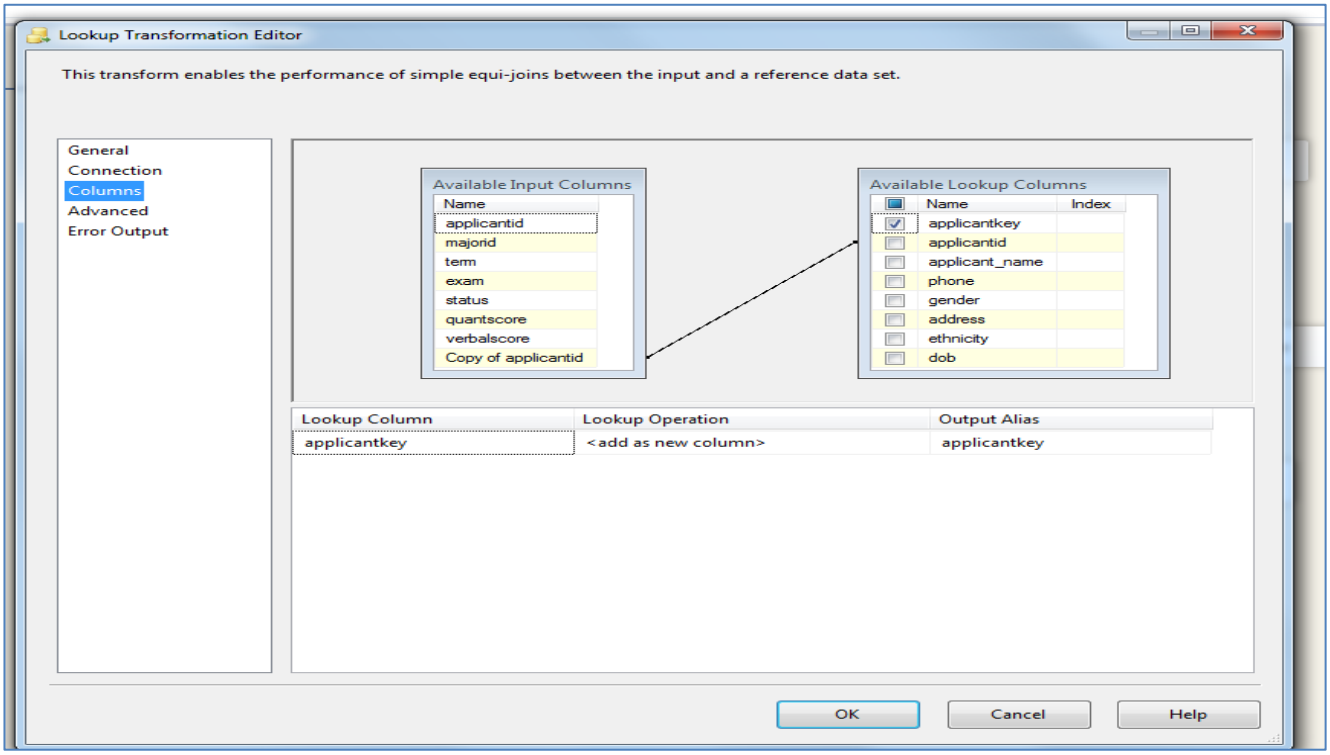

Figure: Applicantkey Lookup

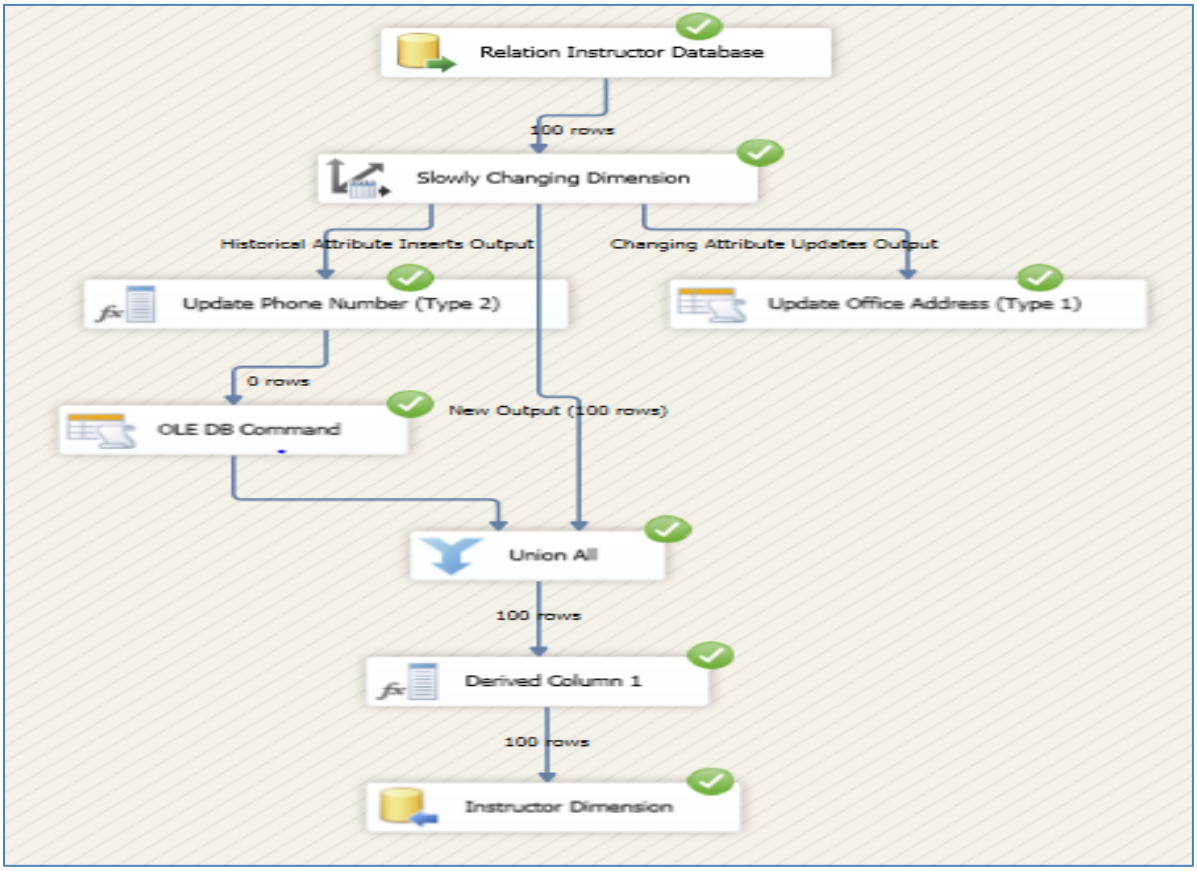

Figure: Handling Slowly Changing Dimensions (SCD) in Instructor Dimension

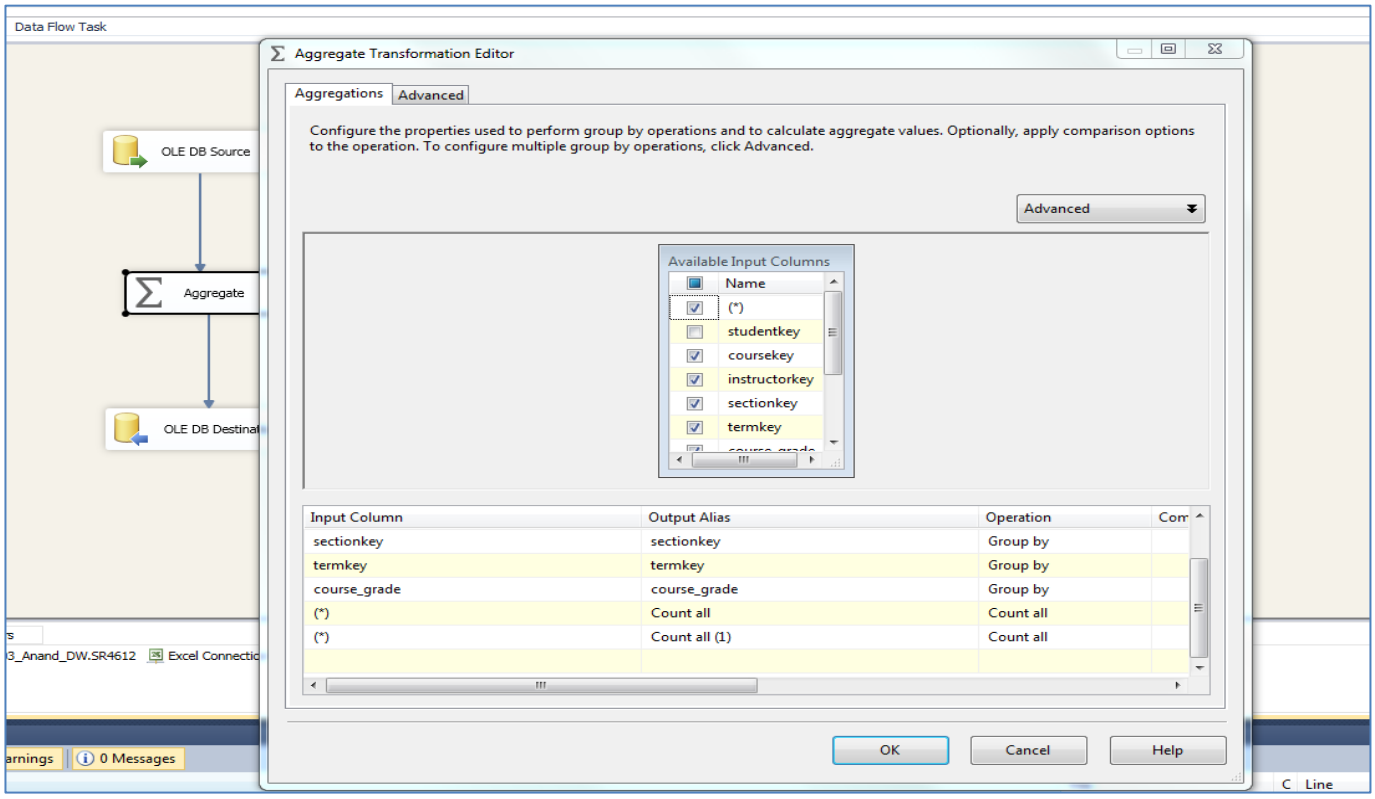

Figure: Performing Section-wise aggregation of results

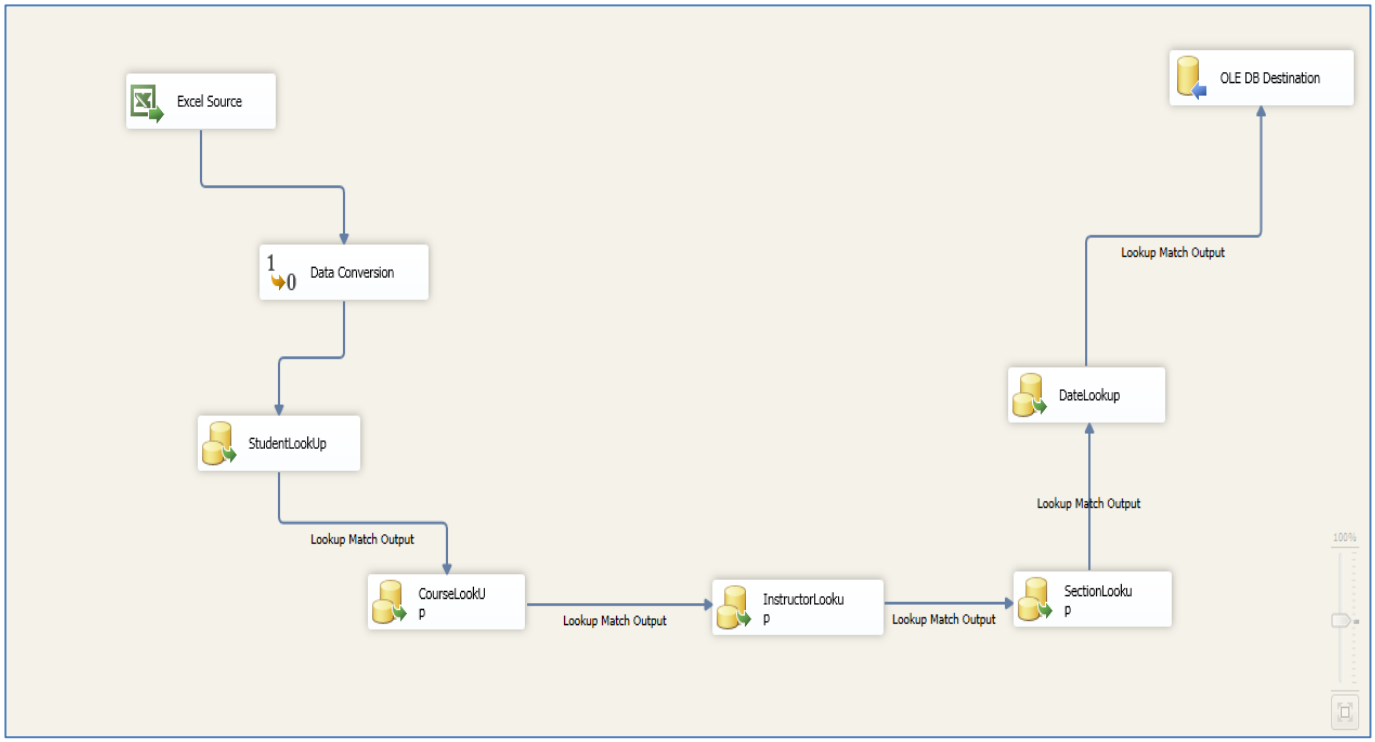

Figure: Student Course Mapping Fact Table

After performing the transformations, cleaning and loading of data, further analysis and reporting was performed using Tableau. The details are discussed in the following sections.

6. Reporting

6.1 **Reports related to Students and Applicants**

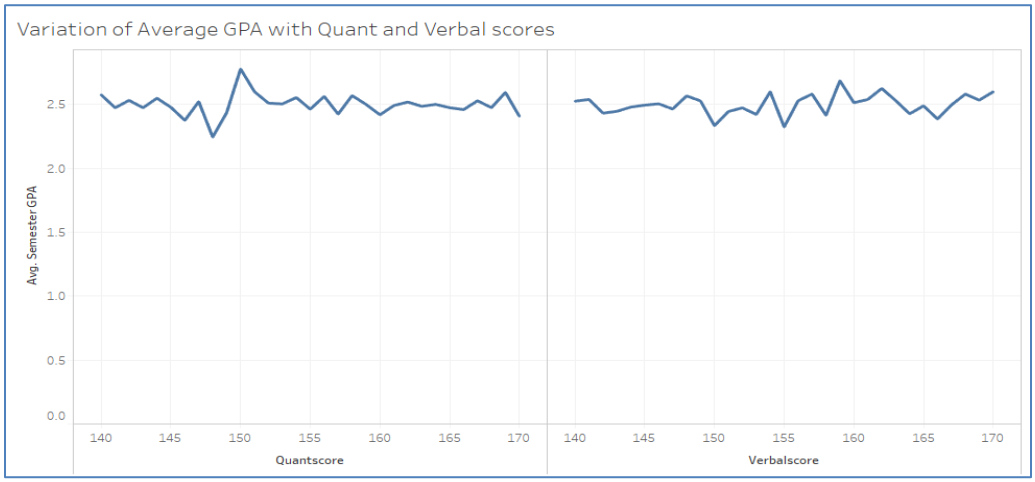

Figure: Variation of Average GPA with Quant & Verbal Scores

This report provides a comparison between the average GPA of students of the University with their Quantitative and Verbal section scores on the entrance exams they took to apply to the Wales University. The report can be generated for any year as we have made sure that all the years are added in the filter. It is important to understand how exam scores which were used a criterion for admissions are indicative of academic performance of students at university. For example, GMAC added integrated reasoning as a new section to GMAT. So, university can use this analysis to find out if scores of students in integrated reasoning are a good indicator of academic performance in future. If there is a correlation then that score can we given weightage while handing out admits.

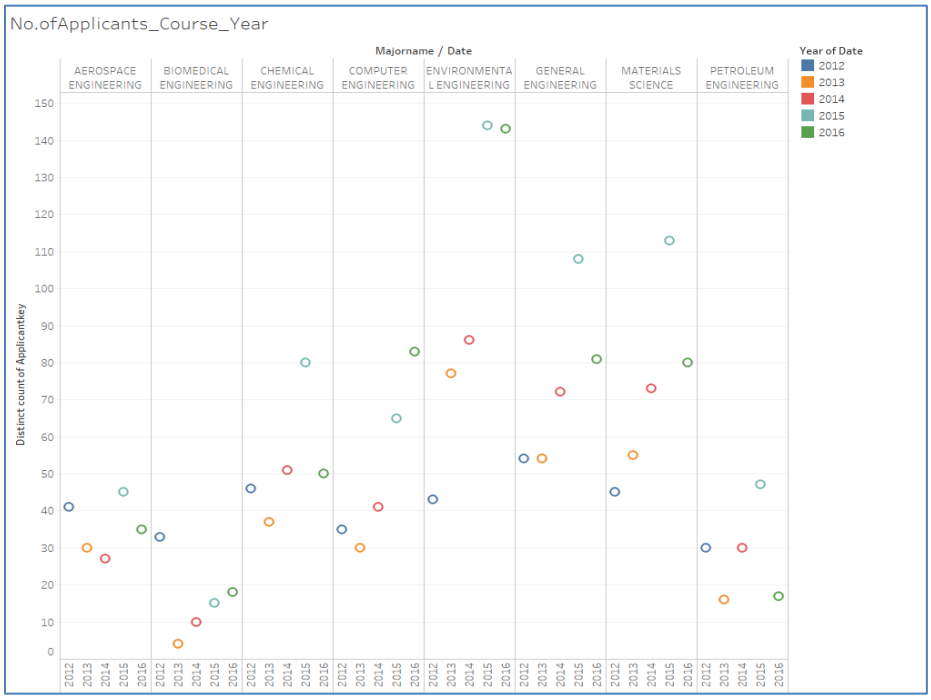

Figure: Number of Applications for each major over the years

This report gives the management a clear view of the number of applications that were received for each major of the University over the year. This report is highly flexible as it lets the user choose the year the user wants to see the applicant trends in and also choose the majors for which he/she wants to see the trends in. This report helps us analyze number of applicants for each major. For example, in the visualization below we see that number of applicants for Environmental Engineering major has been increasing steadily from 2012 to 2016. There may be several reasons such as industry's increased demand for students of this major. So, based on this analysis university may decide to increase its intake for this course.

Figure: Number of Scholarships Offered versus Accepted over the years

This report gives the management trends related to the number of applicants that were offered scholarships versus the number of applicants who accepted the scholarships and accepted the admission offer to the university. This report is highly flexible as it lets the user choose the year the user wants to see the applicant trends in and also choose the majors for which he/she wants to see the trends in. This information helps the management at University decide whether scholarship is one of the criteria to pull bright applicants toward their University. We see below those students who accepted scholarships was very less as compared to the total number of scholarships given out by the University in 2015 and 2016 and hence the University offered only 42 scholarships to eligible students in 2017 out of which 38 accepted it. The University might continue with this trend in 2018 as well.

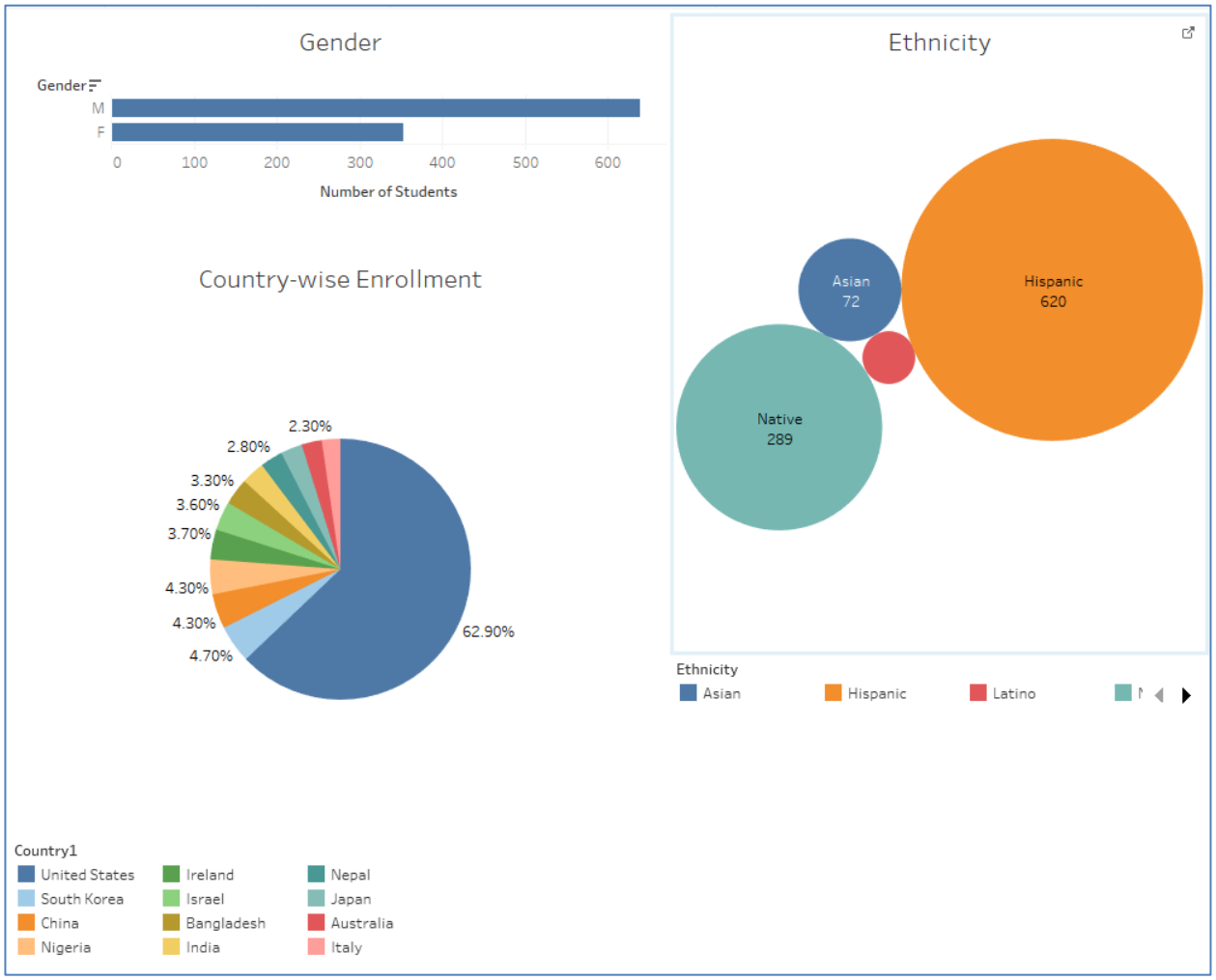

Figure: Student Demographic Dashboard

This dashboard is used to show the demographics of our university. The dashboard consists of three sections. The first one shows the number of students enrolled at our university based on the gender. The second part shows the ethnicity of the students. Finally, the last section shows the countries to which the student belong to. All three sections of the dashboard are connected and can be used as a filter to drilldown through the data. For example, if the user selects 'M' in the first section, the ethnicity and country wise enrollment sections will only show data for Males. Similarly, if a user selects 'Asian' in the ethnicity section of the dashboard, the other sections will display data for students who are of Asian Ethnicity.

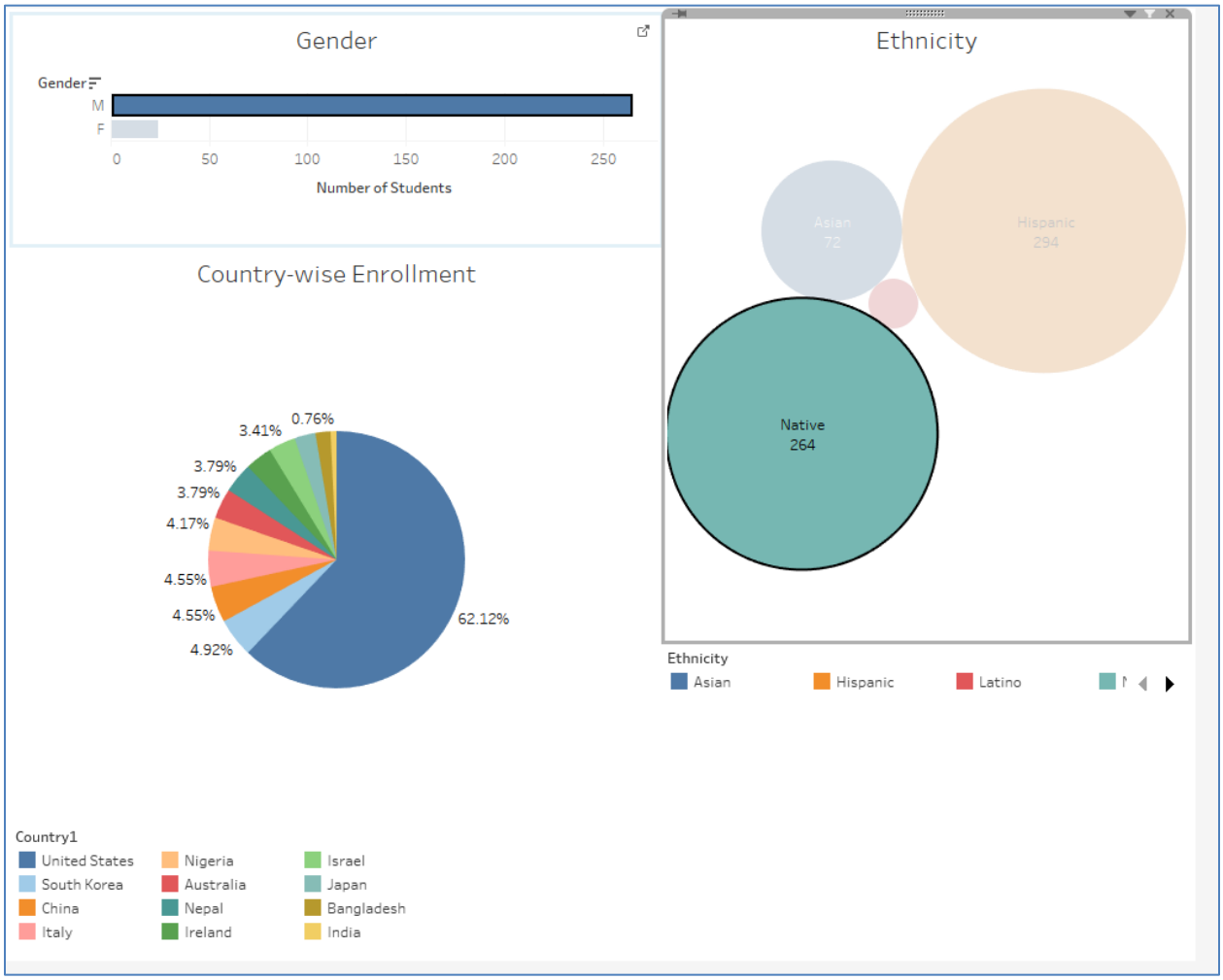

Figure: Drill-down functionality of Student Demographics dashboard

6.2 **Reports related to students' GPA and academic performance**

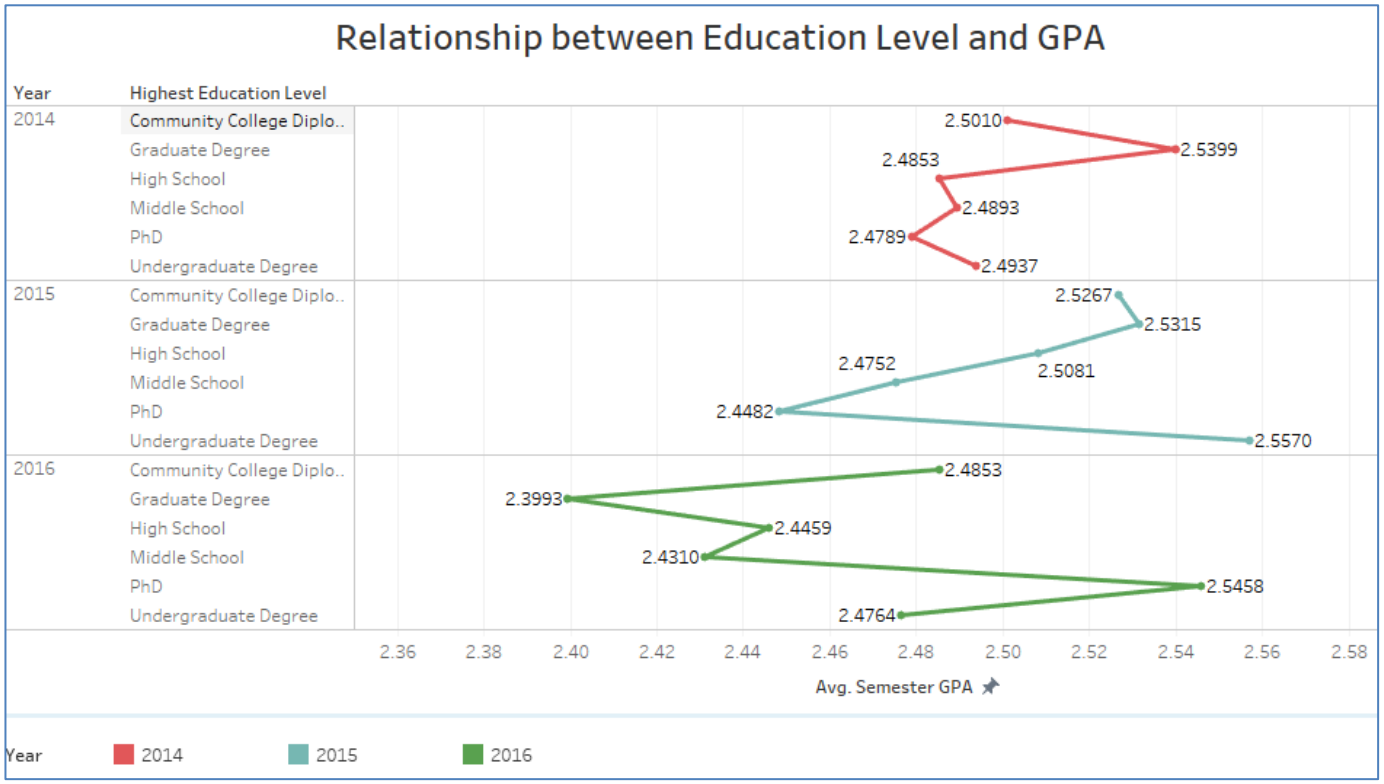

Figure: GPA variance of students as per highest education level

This report provides the management with an overview of how the average GPA of students with different highest education level varies. The number of years and the highest education level can be easily filtered and be chosen. The user can choose the specific year he/she wants to see the trends in or the students with the level of highest education he/she wants to analyze. The university admit the students hailing from different backgrounds. The open education system provides the opportunity to students from different backgrounds and education level to take admission in completely different area of education. We examine the prediction of individual performance of each candidate based on his previous education level. This report also provides the insight about likelihood of a student choosing majors and minors based upon the previous educational qualification.

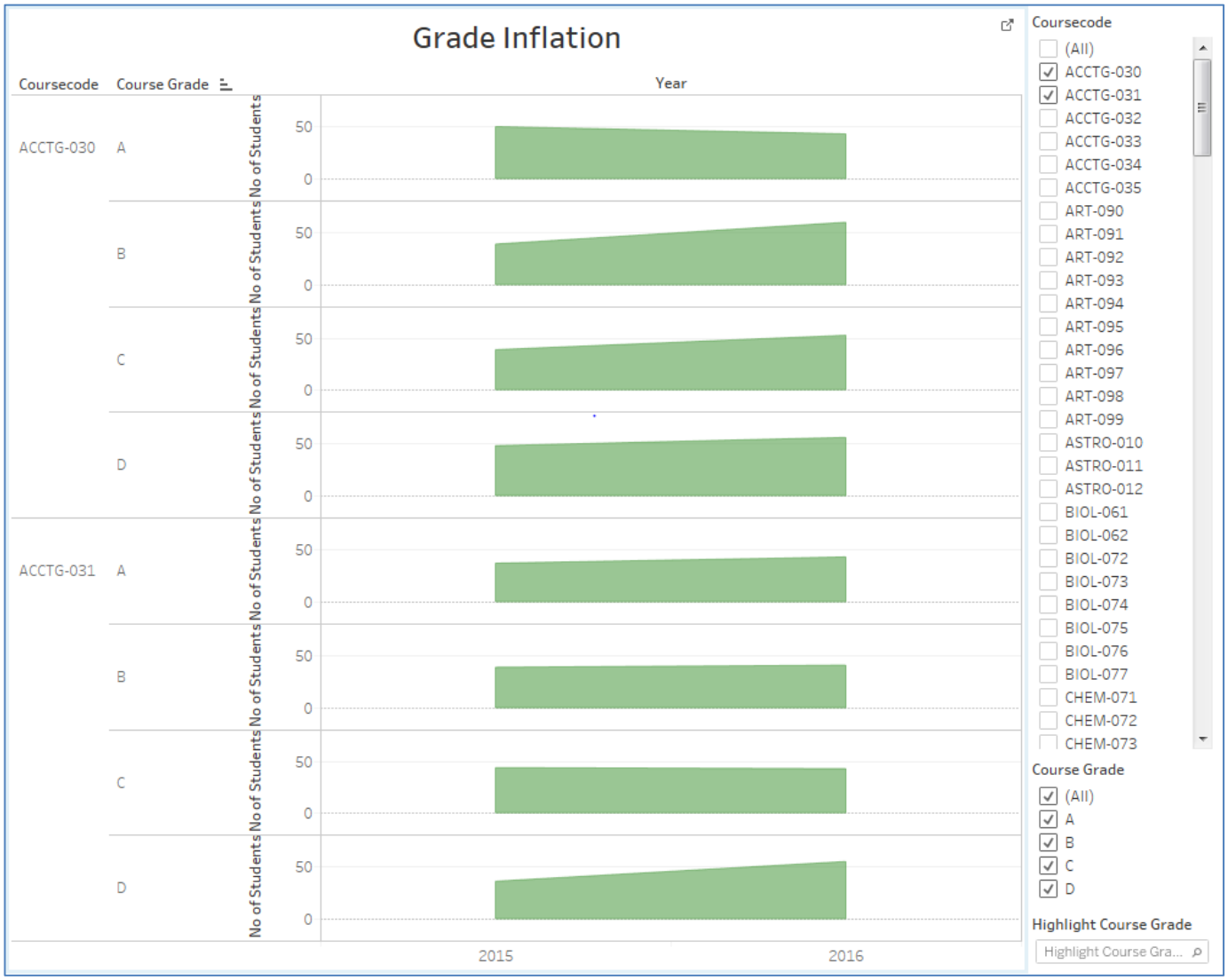

Figure: Performance of Students in a course over the years

Grade inflation is also an important aspect of grading system. In grading system sometimes defining grades and corresponding scaling measure become a nightmare. In order to provide a quantitative analysis of grades awarded in a particular course, we have designed this report. This report will provide a bird's eye view about a certain course and its corresponding division of grades over the years. This can also be utilized while deciding the new grading scheme during the curve fitting of a particular class. This report is highly flexible in nature and can be used by the user to choose the specific course that the user wants to check the grade inflation for.

6.3 **Reports related to Placements of the students**

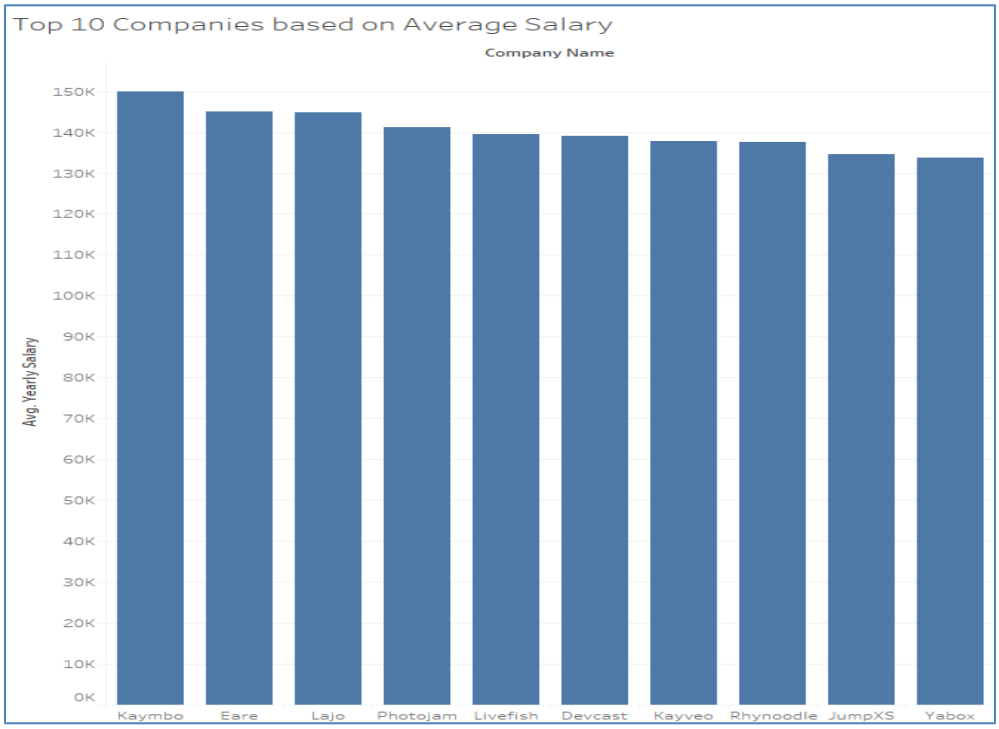

Figure: Top 10 Companies based on average salary

Based on the above report, we see the top 10 companies University wide that offer the highest compensation to the students of the Wales University. The University could contact these companies to invite them to become sponsors for departments at Wales University or even to ask for donations they would want to make towards certain departments at the University. The management could contact the 10 companies identified in the above report to send their representatives for discussions or guest lectures etc.

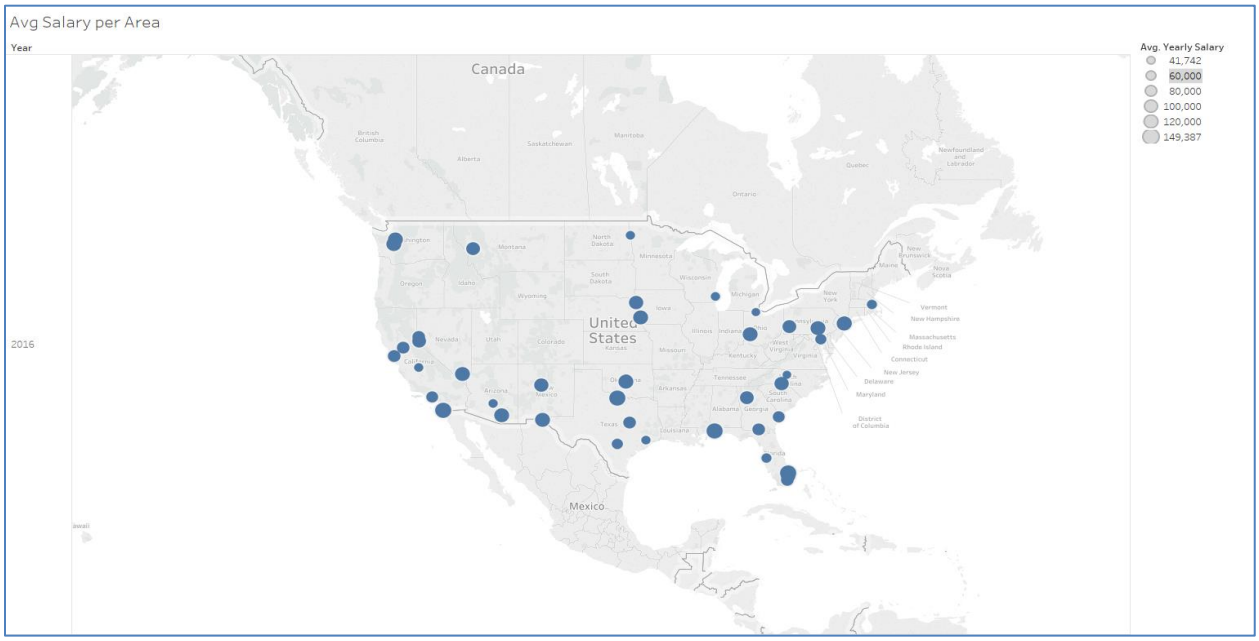

Figure: Average Salary of students per states in United States

It is important for the management at the university to understand which companies hire the most candidates and in which states these candidates are hired for. This will help us identify companies who can potentially be sponsors or donors to our university. This information will also help us promote our university in these states among students assuming that students would prefer working in their home state after completing their education at our university.

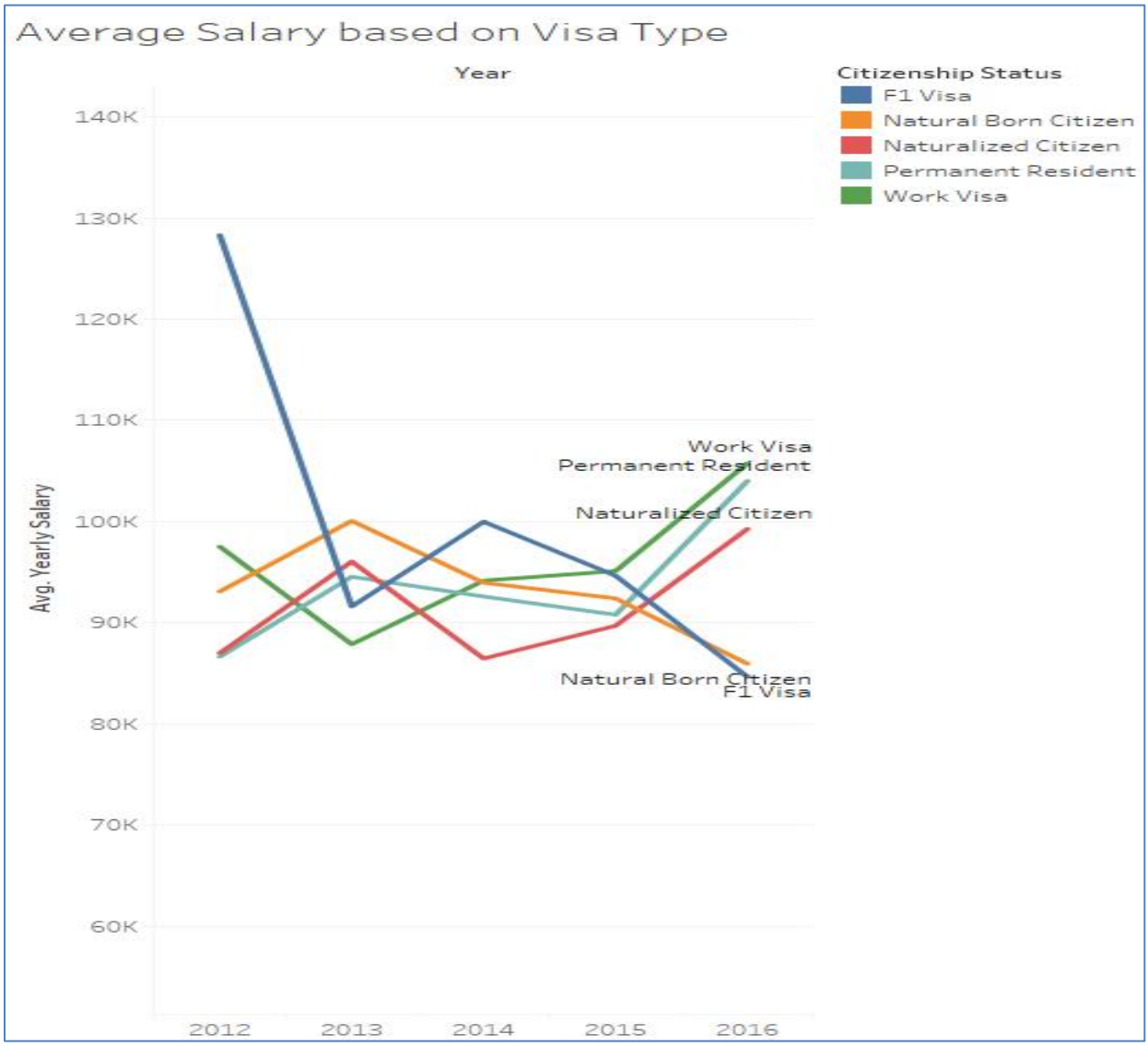

Figure: Average Salary of Students based on Visa Type

The information about average salary of students by their visa status will help identify if students of a particular visa type command higher annual salaries than others. This will help us promote our university to those students thus leading to higher enrollment and income. This report can also indicate to the management of the University whether students having a certain Visa Type are facing difficulties in getting high-paying jobs and can concentrate on increasing the skills of those students to help them.

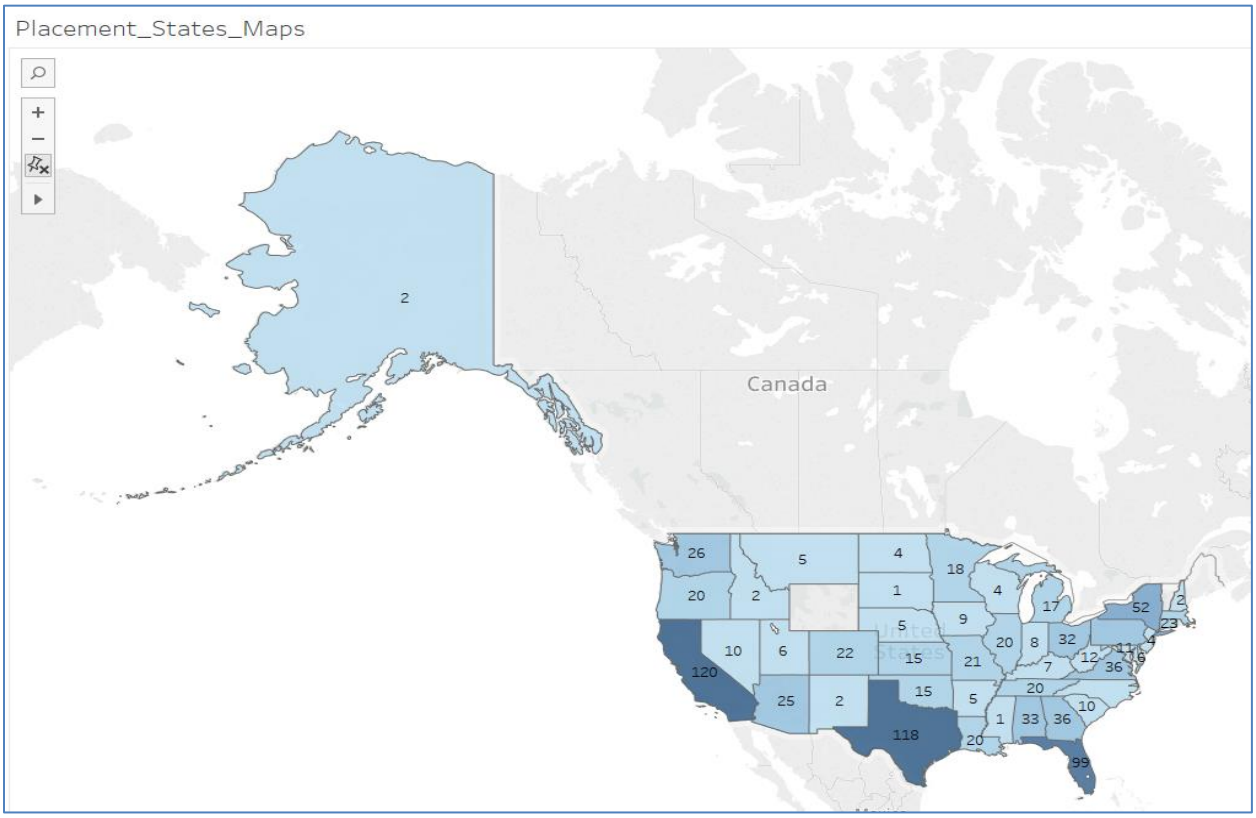

Figure: Placement statistics by states

The above report shows number of placement of students based on the states in which they are placed in. This information will help us find more companies in states which have fewer students placed into. It will also help us identify companies in states which can be invited to the university career fair.

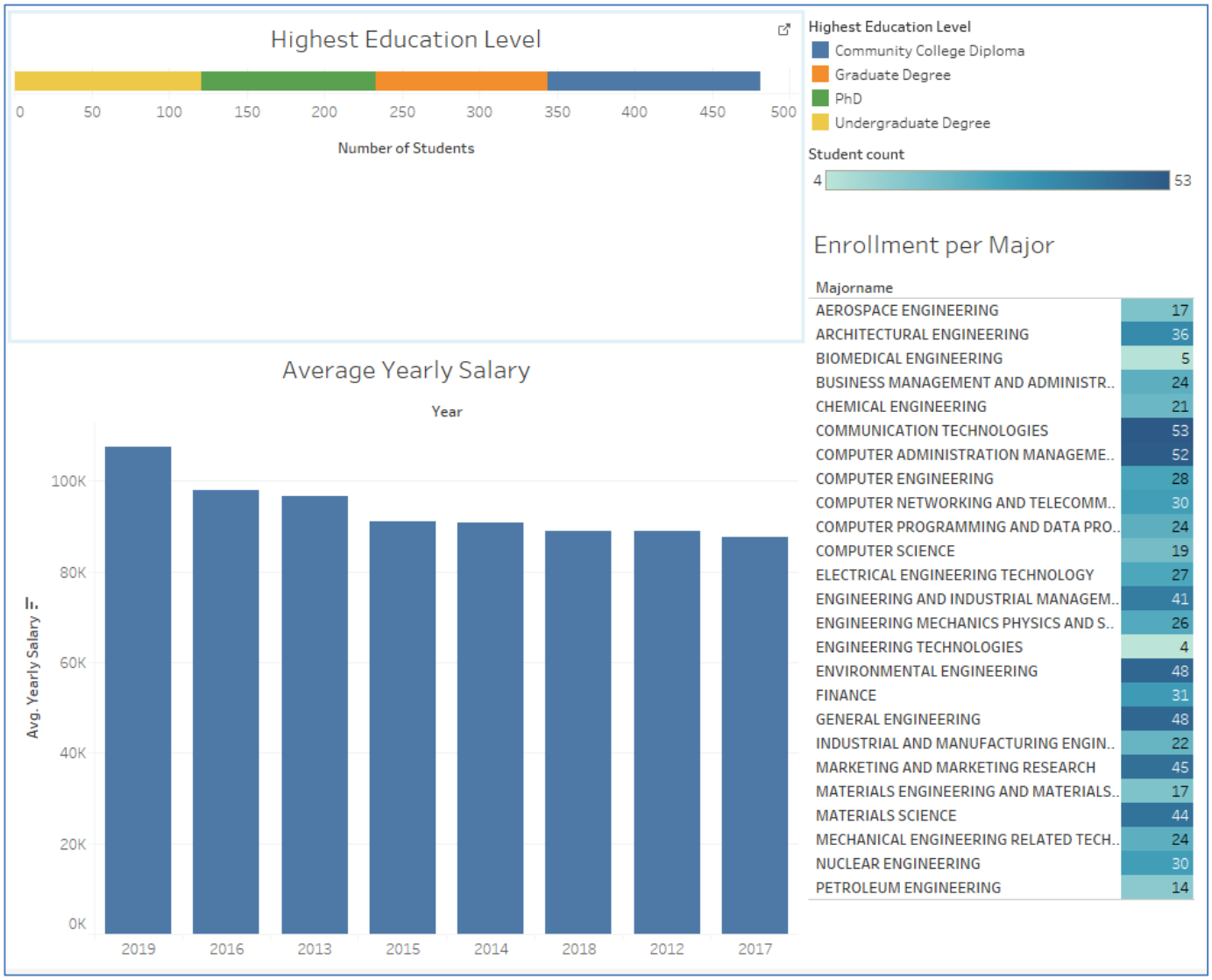

Figure: Placement Statistics Dashboard

The Placement Statistics dashboard shows the placements stats of the university students. There are three sections in the dashboard; Highest Education level, Enrollment count by major, and average yearly salary of the placed students. A user can drill-down on any of the sections of the dashboard to display more detailed information. For example, if a user selects 'Graduate Degree' in the first section, the other two sections of the dashboard will display information only for graduate students. This dashboard helps users find trends in placements over the years by students' major or by highest education level.

7. Meta data

7.1 Meta Data for Facts

Fact Table : Placements				
Type of Fact Table		Factless Fact Table		
Relevant Dimensions		Company, Company_Location, Student		
Fact / Measures Details:				
Name	Description	Type	Format	Load Frequency
Search_Start_Date	Date since when student is looking for jobs	Non-Additive	Date / Time	Bi-Weekly
Placement_Date	Date when the student accepts a job offer	Non-Additive	Date / Time	Bi-Weekly
Current_Status	Status of the student - hired or not	Non-Additive	Text	Bi-Weekly
Salary	Salary of the student at the company	Additive	Number	Bi-Weekly
Position	Position of the student at the company	Non-Additive	Text	Bi-Weekly

Fact Table : StudentCourseMapping				
Type of Fact Table		Factless Fact Table		
Relevant Dimensions		Student, Course, Instructor, Section, Date		
Fact / Measures Details:				
Name	Description	Type	Format	Load Frequency
Course_grade	Name of the applicant	Non-Additive	Text	At the start of each term
Letter_grade	Verbal Score of the applicant on the entrance exam	Additive	Number	At the start of each term

Fact Table : Applications				
Type of Fact Table		Factless Fact Table		
Relevant Dimensions		Applicants, Exam, Major, Date		
Fact / Measures Details:				
Name	Description	Type	Format	Load Frequency
Name	Name of the applicant	Non-Additive	Text	Weekly during Nov 1 - Jan 31
VerbalScore	Verbal Score of the applicant on the entrance exam	Additive	Number	Weekly during Nov 1 - Jan 31
QuantScore	Quant Score of the applicant on the entrance exam	Additive	Number	Weekly during Nov 1 - Jan 31
Status	Status of the student's application	Non-Additive	Text	Weekly during Nov 1 - Jan 31
Registered	Status of the student's registration	Non-Additive	Text	Weekly during Nov 1 - Jan 31

Fact Table : SectionWiseGPA				
Type of Fact Table		Periodic Snapshot Table		
Relevant Dimensions		Instructor, Date, Course, Section		
Fact / Measures Details:				
Name	Description	Type	Format	Load Frequency
no_of_A	No. of As student received	Non-Additive	Number	At the end of each term
no_of_B	No. of As student received	Additive	Number	At the end of each term
no_of_C	No. of As student received	Additive	Number	At the end of each term
no_of_D	No. of As student received	Non-Additive	Number	At the end of each term
enrollment_count	Status of the student's registration	Non-Additive	Number	At the end of each term
avg_gpa	Status of the student's registration	Non-Additive	Number	At the end of each term
max_gpa	Status of the student's registration	Non-Additive	Number	At the end of each term

Fact Table : SWCumulativeGPA				
Type of Fact Table		Accumulating Snapshot Fact Table		
Relevant Dimensions		Student, Date		
Fact / Measures Details:				
Name	Description	Type	Format	Load Frequency
Semester_gpa	Semester's GPA of the student	Semi-Additive	Number	At the end of each term
Cumulative_gpa	Cumulative GPA of the student	Semi-Additive	Number	At the end of each term
creditsearned	Credits earned up till now by the student	Additive	Number	At the end of each term
graduated	Status of the student's graduation	Non-Additive	Text	At the end of each term

7.1 Meta Data for Dimensions

Dimension: Applicants		
Description	This dimension stores all the data of the applicants that apply to the Wales University	
Conformed Dimension	N/A	
Load Frequency	Weekly during Nov 1 - Jan 31	
Attribute Details:		
Name	Description	SCD Type
applicant_id	Applicant's unique id for storing their record in the data warehouse	Type 0
applicant_name	Applicant's name	Type 0
phone	Applicant's Phone Number	Type 0
gender	Applicant's Gender	Type 0
dob	Applicant's Date of Birth in MM/DD/YYYY format	Type 0

Dimension: Course		
Description	This dimension stores data related to the various courses offered to students of the Wales University	
Conformed Dimension	Yes	
Load Frequency	At the start of each term	
Attribute Details:		
Name	Description	SCD Type
course_id	Unique id for storing the course record in the data warehouse	Type 0
course_name	Name of the course	Type 0
course_code	Course Code	Type 0
course_credit	Credits earned by taking the course	Type 0
course_strength	The number of students attending the course	Type 1
isLatest	To reflect if the course is being offered in the current term	Type 0

Dimension: Student		
Description	This dimension stores data related to the various students of the University	
Conformed Dimension	Yes	
Load Frequency	At the start of each term	
Attribute Details:		
Name	Description	SCD Type
student_id	Unique id for storing the student record in the data warehouse	Type 0
majorkey	Major of the student	Type 0
UIN	Unique Identifcation number of student	Type 0
applicantkey	This key links student data with data used during application to university	Type 0
citizenship_status	Citizenship status for student	Type 3
highest_education_level	Highest education level of student	Type 0

Dimension: Company		
Description	This dimension stores data related to the companies that hire students of the Wales University at various positions	
Conformed Dimension	N/A	
Load Frequency	Quarterly	
Attribute Details:		
Name	Description	SCD Type
company_id	Company's unique id for storing their record in the data warehouse	Type 0
company_name	Company's name	Type 0

Dimension: CompanyLocation		
Description	This dimension stores data related to the location of the companies that hire students of the Wales University at various positions	
Conformed Dimension	N/A	
Load Frequency	Quarterly	
Attribute Details:		
Name	Description	SCD Type
location_id	Unique id for storing the companies record in the data warehouse	Type 0
city	City the company is located in	Type 1
state	State the company is located in	Type 1
country	Country the company is located in	Type 1

Dimension: Department		
Description	This dimension stores data related to the various departments of the Wales University	
Conformed Dimension	N/A	
Load Frequency	Yearly	
Attribute Details:		
Name	Description	SCD Type
department_id	Unique id for storing the department record in the data warehouse	Type 0
department_name	Name of the department	Type 0
department_code	Department Code	Type 0
department_dean	Name of the dean of the department	Type 3
department_address	The address of the department building in the university	Type 1

Dimension: Exam		
Description	This dimension stores data related to the various exams that students take to apply to the Wales University	
Conformed Dimension	N/A	
Load Frequency	Yearly	
Attribute Details:		
Name	Description	SCD Type
exam_id	Unique id for storing the student's exam record in the data warehouse	Type 0
exam_name	Name of the exam	Type 0
exam_code	Exam Code	Type 0
no_of_sections	Number of sections in the exam	Type 1
company	company conducting the exam (ETS, PearsonVue etc.)	Type 1

Dimension: Instructor		
Description	This dimension stores data related to the various Instructors teaching at the University	
Conformed Dimension	Yes	
Load Frequency	Quarterly	
Attribute Details:		
Name	Description	SCD Type
instructor_id	Unique id for storing the instructor record in the data warehouse	Type 0
instructor_name	Name of the instructor	Type 0
phone	Phone number of instructor	Type 1
designation	Designation of instructor	Type 3
office_address	Office address of instructor	Type 3

Dimension: Major		
Description	This dimension stores data related to the various majors offered by departments of the University	
Conformed Dimension	N/A	
Load Frequency	Quarterly	
Attribute Details:		
Name	Description	SCD Type
major_id	Unique id for storing the major record in the data warehouse	Type 0
major_name	Name of the major	Type 0
major_code	Major Code	Type 0
credits	Credits required for completing the major	Type 0

Dimension: Section		
Description	This dimension stores data related to the different sections for courses.	
Conformed Dimension	Yes	
Load Frequency	At the start of each term	
Attribute Details:		
Name	Description	SCD Type
section_id	Unique id for storing the section record in the data warehouse	Type 0
section_code	Code of the section	Type 0
section_name	Name of the section	Type 0

Dimension: Date		
Description	This dimension stores data related to date	
Conformed Dimension	Yes	
Load Frequency	Yearly	
Attribute Details:		
Name	Description	SCD Type
date	Date	Type 0
day_of_week	Numeric value for day of week	Type 0
day_name	Name of day	Type 0
day_of_month	Numeric value for day of month	Type 0
day_of_year	Numeric value for day of year	Type 0
month_name	Name of month	Type 0
week_of_year	Numeric value for week of year	Type 0
month_of_year	Numeric value for month of year	Type 0
year	Year	Type 0
term_number	Numerical value for term of year	Type 0
term_name	Name of term	Type 0
term_year	Combination of term name and year	Type 0
isWeekDay	Flag to notify if it is a weekday	Type 0

8. Appendices

8.1 Appendix A: OLAP definitions

OLAP (Online Analytical Processing) is a process of creating and managing multidimensional enterprise data and viewing by the user who seeks an understanding of what the data is saying. OLAP provides analytical tools to visualize the data. The benefits of using OLAP are listed as follows:

- Multidimensional presentation can create an understanding of relationships not previously realized
- OLAP creates a single platform for all the information and business needs; planning, budgeting, forecasting, reporting and analysis
- It allows a manager to pull down data from an OLAP database in broad or specific terms
- No matter how much or how fast data is processed through OLAP software or servers, the reporting that results is presented in a consistent presentation, so analysts and executives always know what to look for where

8.2 Appendix B: SSIS Terminology

Microsoft Integration Services is a platform for building enterprise-level data integration and data transformations solutions. Integration Services includes a rich set of built-in tasks and transformations; tools for constructing packages; and the Integration Services service for running and managing packages. One can use the graphical Integration Services tools to create solutions without writing a single line of code; or to program the extensive Integration Services object model to create packages programmatically and code custom tasks and other package objects.

Connections

A connection includes the information necessary to connect to a particular data source. Tasks can reference the connection by its name, allowing the details of the connection to be changed or configured at run time.

Event handlers

A workflow can be designed for a number of events in the different scopes where they might occur. In this way, tasks may be executed in response to happenings within the package — such as cleaning up after errors.

Tasks

A task is an atomic work unit that performs some action. The data transformation task copies data; it implements the ETL features of the product

Parameters

Parameters allow you to assign values to properties within packages at the time of package execution. You can have project parameters and package parameters.

9. Glossary

Terms	Definitions
Business Model	An object-oriented model that captures the kinds of things in a business or a business area and the relationships associated with those things (and sometimes associated business rules, too). Note that a business model exists independently of any data or database. A data warehouse should be designed to match the underlying business models or else no tools will fully unlock the data in the warehouse.
Data Dictionary	A collection of Meta Data. Many kinds of products in the data warehouse arena use a data dictionary, including database management systems, modeling tools, middleware, and query tools.
Data Model	The road map to the data in a database. This includes the source of tables and columns, the meanings of the keys, and the relationships between the tables.
Meta Data	It is basically the description of what kind of information is stored where, how it is encoded, how it is related to other information, where it comes from, and how it is related to your business.
Star Schema	A standard technique for designing the summary tables of a data.
Snowflake Schema	A layering of Star Schema that scales that technique to handle an entire warehouse.
Data Visualization	Techniques for turning data into information by using the high capacity of the human brain to recognize visually recognize patterns and trends.
Data Mart	A subset of a data warehouse that focuses on one or more specific subject areas. The data usually is extracted from the data warehouse and further denormalized and indexed to support intense usage by targeted customers.

10. Bibliography

[1] https://en.wikipedia.org/wiki/SQL_Server_Integration_Services

[2] http://searchsqlserver.techtarget.com/definition/Microsoft-SSIS-SQL-Server-Integration-Services

[3] https://en.wikipedia.org/wiki/Online_analytical_processing

[4] http://www.jatit.org/volumes/Vol69No1/7Vol69No1.pdf

[5] http://www.oppaga.state.fl.us/MonitorDocs/Reports/pdf/0931rpt.pdf

[6] http://www.texasstudentdatasystem.org/TSDS/Education_Data_Warehouse/

[7] http://blog.trueinteraction.com/big-data-trends-in-the-education-sector

[8] http://ijcsit.com/docs/Volume%205/vol5issue03/ijcsit20140503412.pdf

11. Project Work Breakdown

Team Member Name	Tasks Performed	Signature
Anand Iyer	Dimensional modelling, Creation of tables, Data generation, SSIS data integration, Tableau report generation	
Anshul Prakash	Dimensional modelling, Data warehouse Architecture, Creation of tables, Data generation, Tableau report generation, Final report	
Brian Fernandes	Dimensional modelling, Creation of tables, Data generation, SSIS data integration, Tableau report generation	
Mallika Malhotra	Dimensional modelling, Data warehouse Architecture, Creation of tables, Data generation, Tableau report generation, Final report	
Sumit Singh	Dimensional modelling, Creation of tables, Data generation, SSIS data integration, Tableau report generation	