STUDENT PROFILE EVALUATION APPLICATION

IDS 521 Final Report

Fall 2015

Ronak Malviya (rmalvi2)

Soham Joshi (sjoshi30)

Sai Shankar Sattari (ssatta3)

Arnold Lee (alee29)

Authors

Ronak Malviya (rmalvi2)
Soham Joshi (sjoshi30)
Sai Shankar Sattari (ssatta3)
Arnold Lee (alee29)

- **Overview**

According to a report from the National Student Clearinghouse[1], only 59 percent of students who begin college as a freshman (first year student) graduate within 6 years. Also, according to the *US News and World Report*, only one out of 3 freshman stays in the same college or university for their second year. Research suggested that the low retention rate and low graduation rate is affected by uninformed or unprepared decisions made by the student or their family. Our software is aim to tackle this problem, and help reducing the cost of attending college or graduate school.

This software is to evaluate profile of a prospective user based on the input, such as standardized test score, GPA, coursework, extracurricular activity, ethnicity, family annual income etc. and determine their probability of getting admitted to a target university. The goal of the system is to save prospective student valuable time and money, by automatically evaluate the student profile against the historical emission data of a particular university, and advise student against those places which has low probability of getting admitted.

Parents can also use this to target which schools their children to apply, and thus it can be used:

- to aid in the decision of buying real properties near the campuses

- to compare potential financial aid options and make necessary adjustment to their asset allocations

With a slight tweaks it can also be used for other academic systems across the globe.

- **Description of Data Sources**

  Our data sources are primarily consist of the following:

- Self-reported test score and other application data directly from students

- Admission data released by the university (for public universities, FOIA requests may be used to get more detailed level data)

- Ranking data from 3rd party applications, such as those coming from *US News & World Report, Financial Times,* and *Bloomberg Businessweek*

  In all cases, data will be pruned (through the ETL layer) prior to import to our warehouse for use.

- **Detailed Enterprise Bus Matrix**

Business Processes	Date	University	Student	Address	Course
Profile Evaluation	X	X	X		X
Acceptance rate	X	X			X
Ranking		X			X
Student geographic distribution			X	X	
Score Requirements	X	X			X
Industry experience of students	X		X		
Course fee	X	X			X

## ● **Dimensional Models & Supporting Arguments**

Accumulating snapshot:
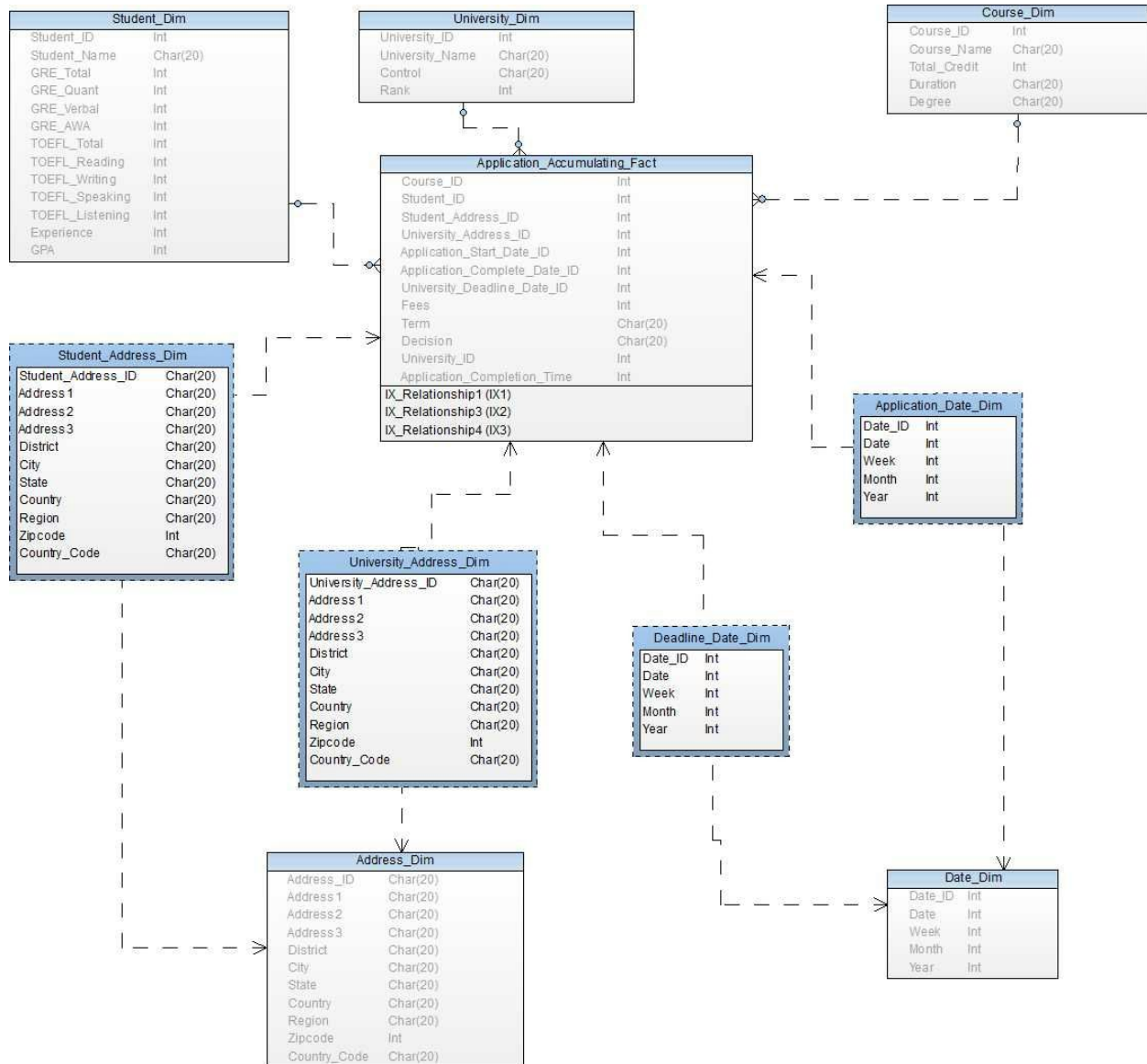
The accumulating snapshot captures information regarding application life cycle. It is continuously updated to capture information about the student input. At the end, we can monitor the time duration used to complete an application by the student, and shared across all users of the application platform.Hence, we can find what is average time taken by a student to finish the

application and the student can plan ahead of time how much time he will require to finish an application.

Periodic Snapshot:
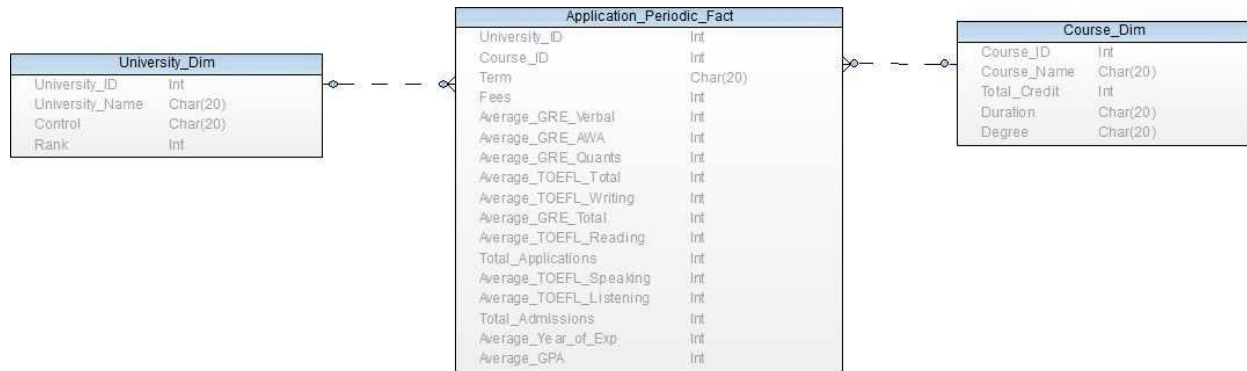
The periodic snapshot records the information for the previous terms for each university and course combination, that can help student to give an overview of the application admittance to a particular course for a particular university.

Fact Table:

The information that is captured in the fact table is basically the application of the student to the university. We have used each application(Transaction-Type) as our grain. This decision was reached because we wanted to store the data at the most atomic level and this would give us the greatest number of dimensions to be associated. To calculate the average fees across universities and the average acceptance rate we have included fees and decision in the fact table.
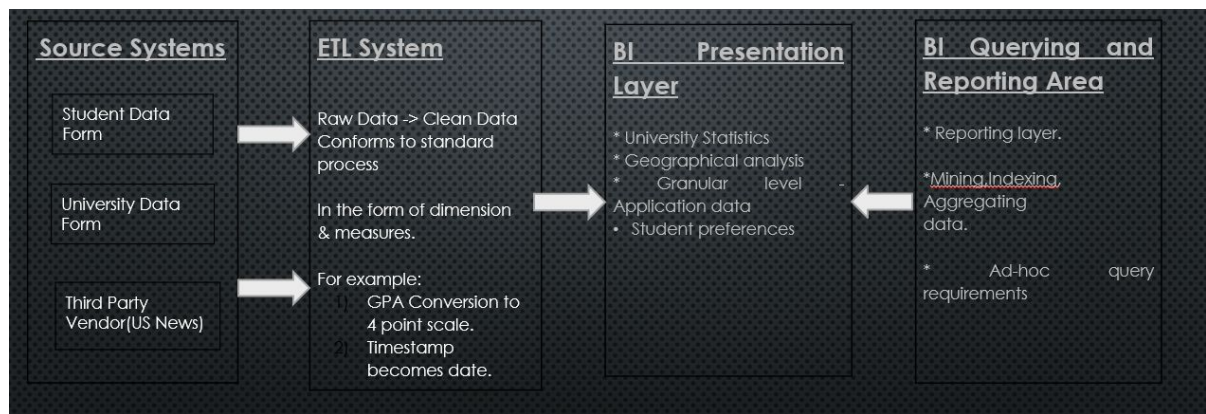
Five dimensions were chosen viz student, address, date, university and course. Each of the dimensions were connected to the fact table by the foreign keys. There address attribute was

common to both the student and the university, so we included it as a common outrigger to both the student and university dimensions.

Dimensions:

The descriptive attributes are placed in separate dimensions. Student dimension records all the exam scores, the experience that a candidate has and his gpa. University dimension records the name, control and ranking of the university. Course dimension records the credits, degree and the duration of the course.

- **Technology & Implementation**

| Source Systems	ETL System	BI Presentation Layer	BI Querying and Reporting Area
Student Data Form	Raw Data -> Clean Data Conforms to standard process	* University Statistics * Geographical analysis * Granular level - Application data • Student preferences	* Reporting layer. *Mining,Indexing, Aggregating data. * Ad-hoc query requirements
University Data Form	In the form of dimension & measures.		
Third Party Vendor(US News)	For example: 1) GPA Conversion to 4 point scale. 2) Timestamp becomes date.		

Above are overview of our warehouse architecture. The source comes from either of student self-reporting, university official sources or through 3rd party publications. ETL engine layer will be powered by Informatica Powercenter. The raw data, along with the processed OLAP cubes, are going to be stored in an Oracle DB format. Lastly visualization on the user

layer will be powered by SAP Business Objects Dashboard. These systems have been chosen

due to familiarity reason as well as avoiding being "locked" to a single vendor.
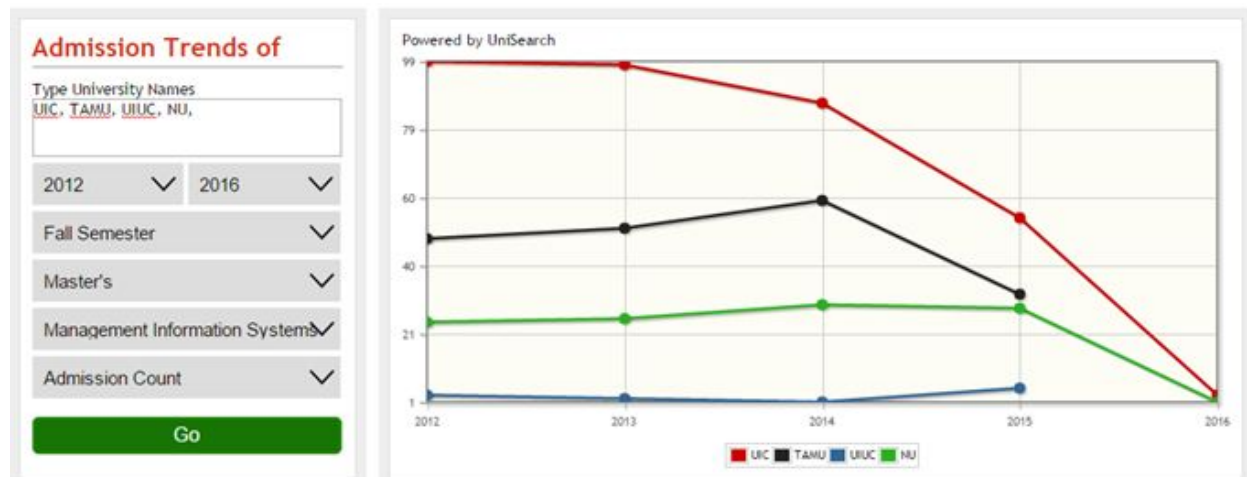
- **Sample Business Questions and BI Reports**

Total Application count
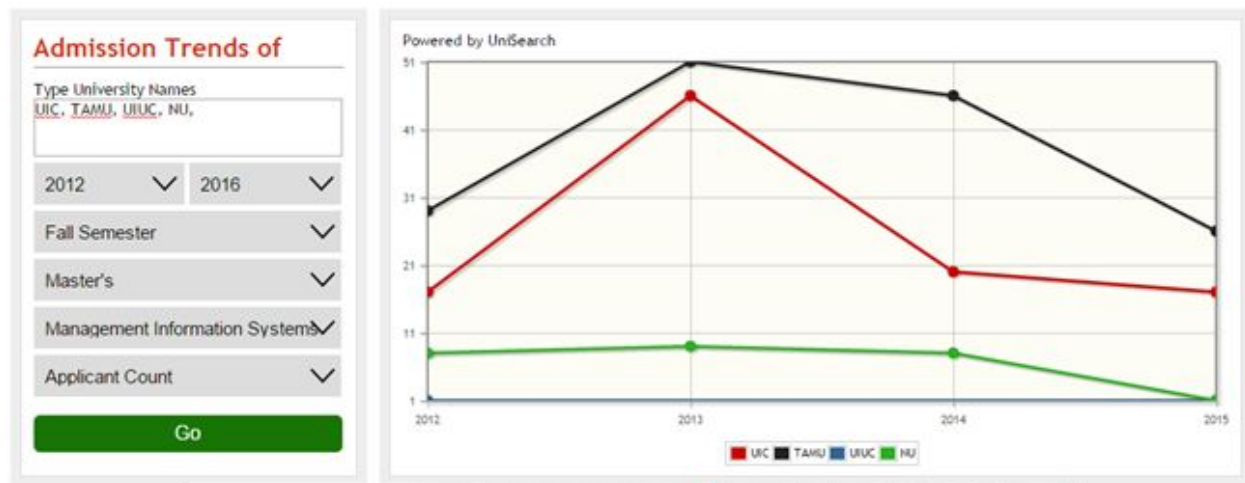
This is a sample output of number of applications received for four different universities

over the four year period (2012-2016) for the Master's in Management Information Systems

degree program. In many cases, the higher the application count, the lower the probability of

getting admitted, since the number of available seats (which determines limit of admissions) are

fixed due to departmental strategic policy or municipal fire code regarding university buildings.

A drastic change in application received also indicates that there are ongoing events that affects

the university brand name, or environmental factors.

Total Admission Count:

This is a sample output of number of admissions offered (not necessarily enrolled) for four different universities over the four year period (2012-2016) for the Master's in Management Information Systems degree program. Admission counts are perhaps the most important metric that can be used to predict the probability of getting admitted. A declining admission count, coupling with flat or increasing number of application received, indicates a tighter admission standard is occurring in a specific college or university. Vice versa, an increasing admission count that is not accompanied by an increase in application received usually indicates loosening admission policy. Many 3rd party rankings, such as *U.S. News & World Report*, factoring in admission ratio in their rankings. Therefore these numbers can be easily subject to manipulation.

Average GRE Score
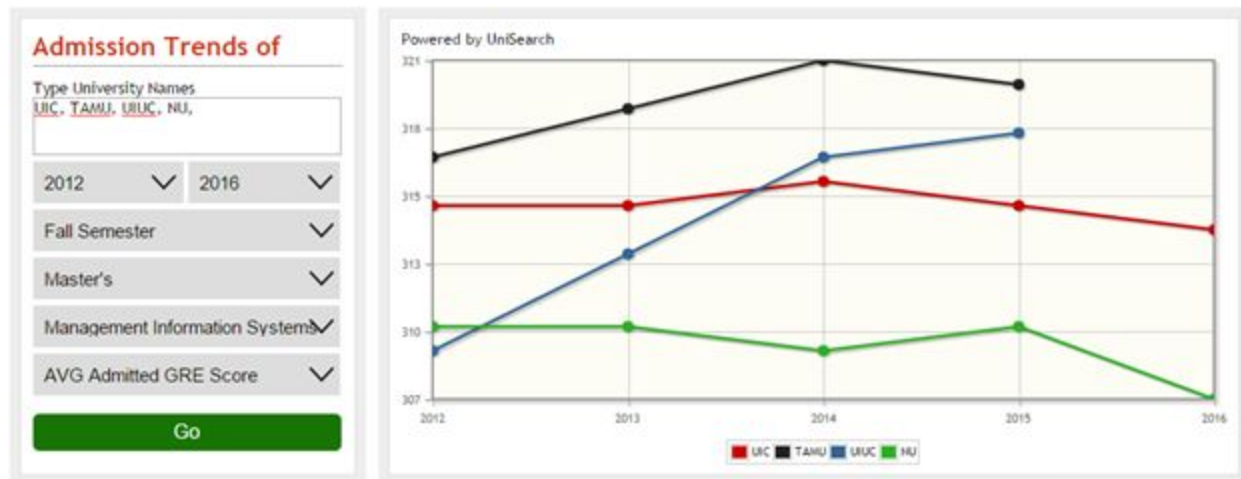
This is a sample output of number of average GRE scores of the admitted cohort, for four different universities over the four year period (2012-2016) for the Master's in Management Information Systems degree program. Much like admission count, the higher the average GRE scores among the admitted cohort, the tighter the admission policy. It is also a major component of the *U.S. News* rankings, and therefore could also be subject of illegal manipulations by insitutions. Furthermore, research have shown that [2] since the revision of the GRE General test in August 2011, the weighted correlation between GRE scores and First-Year Business School grade point average were around $r = 0.4$, which is fairly significant.

Ideally, our working system should be able to answer the following questions given by the user (and the dimension that these question refers to):

- What are the chances for a certain student to get admission to a particular university?

    - A regression analysis report should also be included to see what variables can be improved to increase the probability.

- Which university are suitable for a given student profile

- A list of university will be included based on the calculated probability of

  admission

Suggested University:

- Preferred location / climate for a given student profile

    - Generate maps with density markers indicate where students with similar

      backgrounds are going.

- Aggregate % accepted admission for a given university (reverse engineering the data)

- Which course/major preferred in a particular university

- Application Fees (not tuition) for a particular university (this factor is used along with

  probability of admission)

- Average financial aid package for a particular university, given a student profile

- Job placement percentage for a specified major in a given university

References

[1] Shapiro, D., Dundar, A., Wakhungu, P.K., Yuan, X., Nathan, A. & Hwang, Y. (2015, November). *Completing College: A National View of Student Attainment Rates – Fall 2009*

Cohort (Signature Report No. 10). Herndon, VA: National Student Clearinghouse Research Center. http://nscresearchcenter.org/wp-content/uploads/SignatureReport10.pdf

[2] Young, J. W., Klieger, D., Bochenek, J., Li, C. and Cline, F. (2014), The Validity of Scores from the *GRE*® revised General Test for Forecasting Performance in Business Schools: Phase One. ETS Research Report Series, 2014: 1–10. doi: 10.1002/ets2.12019