JIGSAW ACADEMY
THE ONLINE SCHOOL OF ANALYTICS

Students' Early Attrition Modelling for Clearwater State University

Project Description: *Clearwater State University* offers a wide variety of degree programs, from online degrees to a doctorate in education.

Programs are offered in the streams of the arts, education, business & nursing.

Some key strategic goals of the University are:

- Increase enrolment of students
- Improve retention, progression and graduation rates
- Recruit better academically qualified undergraduate and graduate students
- Increase external funding and recognition

Business Questions: Leverage data on student demographic profile, course preferences, performance record, grades, financial background, financial aid and other application information to:

1. Identify key drivers of early student attrition
2. Build a predictive model to identify students with higher early attrition risk
3. Recommend appropriate interventions based on the analysis

Note: Early attrition is defined as student drop out within one year of successfully joining the program (completed admission process)

Deliverables: Besides submitting the well documented code, it is expected that main takeaways and insights gained through the analysis will be presented in the form of a PowerPoint report. This report should outline and summarize the following:

- Exploratory Data Analysis showing key associations
- Segment Analysis showing key drivers of attrition
- Results of statistical model along with business interpretations and recommended interventions

JIGSAW ACADEMY
THE ONLINE SCHOOL OF ANALYTICS

Project Approach Notes:

Approaching business questions

1. Business question 1 can be addressed by performing Exploratory Data Analysis:
   Here we suggest that you perform various exploratory data analysis to understand student profiles, early attrition rates in various programs, how does attrition change across various combinations of student characteristics (demographic, performance, choice of programs etc.). You may need to perform objective segmentation or discriminant analysis to better understand key drivers of early attrition i.e. student profiles that have much higher propensity to attrite.

2. Business question 2 can be answered by building a statistical model:
   In this phase you will build need to build models to predict early attrition. You may need to explore different statistical/ML algorithm based models and decide the one most suited for this phenomenon. The next step would be to go through the model development and validation process to build most appropriate model that best predicts early attrition, validate this model to ensure robustness of patterns identified.

   Suggested approaches: Logistic Regression, Extreme Gradient Boosting, GBM, Random Forest.

3. Business question 3 should be answered based on the statistical model built.

Data:

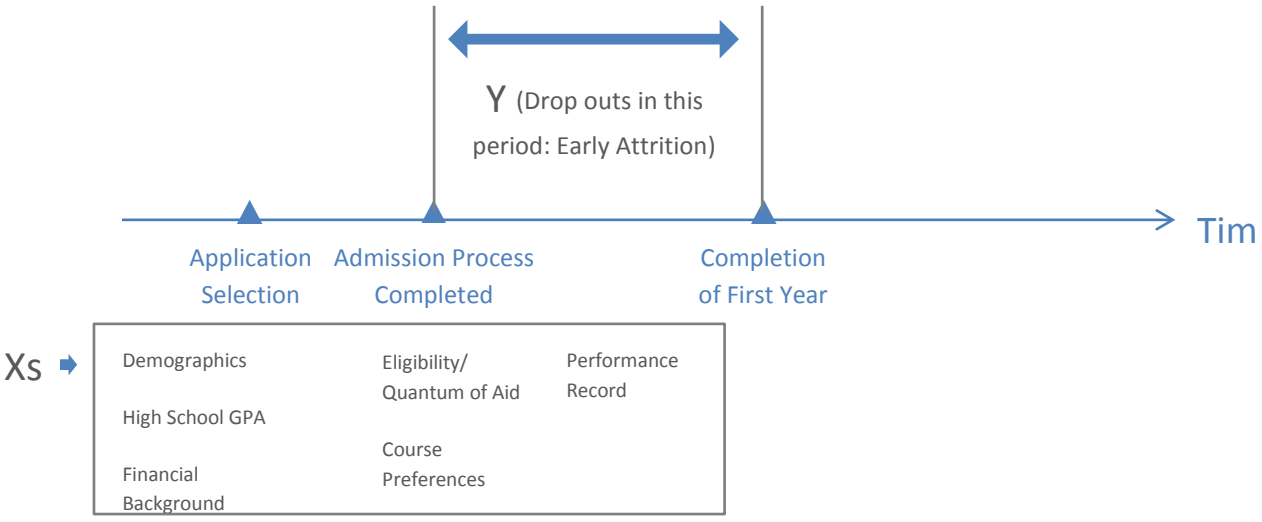
Primarily, data will be of the following high level categories:

a. Student Application data (demographics, High school GPA)
b. Financial Indicators (financial background, eligibility, quantum of aid)
c. Course Preferences
d. Performance Record

JIGSAW ACADEMY
THE ONLINE SCHOOL OF ANALYTICS

Miscellaneous Notes:

Ideally, we will define our overall framework for analysis



a) Define key dependent variables (Y), independent variables (Xs), operating definition of how they will be calculated, appropriate cut-off points and indicators

b) Baseline performance over last few years - % early attrition across different courses. This will provide an existing benchmark, with focus on identifying ways to improve this through the project.

c) Next, we suggest you work through understanding the data, dictionary and contents. We also recommend performance of outlier and sanity checks on data to identify attributes that are more robust and more suited for analysis.

d) The next step should Be exploring the data and then using the suitable model to reach the desired outcome.