NYU TANDON SCHOOL OF ENGINEERING

BA PROJECT 1

Group 1:

Ajith Ganapathi Hegde
Diwei Zhu
Russell M Kaehler
Xiao Fang
Zhuoyang Wu

In partial fulfillment for MG-GY 8413 - Business Analytics course

taught by Prof. J C Bonilla

Project Overview

01	**Problem Introduction**	• Problem Analysis • Client Overview
02	**Initial Data Processing**	• Data Overview • Data Cleaning • Feature Engineering • Data Enhancement & Normalization
03	**Building Solution**	• Building Pipeline • Pipeline application for creating personas
04	**Analysis & Conclusion**	• Selecting Persona • Analysis of Persona Conclusion

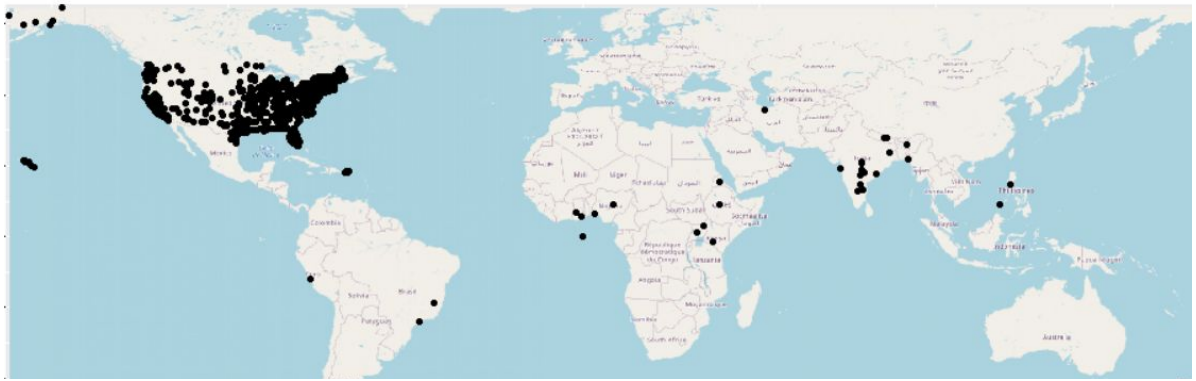
Problem analysis

- Given an **extremely large** dataset containing lifelong data on prospective and admitted students
- We need to **create Personas** based on historical data that will help Marketing Team of Rivier University to make decisions on Future prospective students
- Personas must have **look-alike characters** & we must also look for **external datasets** that can help the analysis

Client Overview : Rivier University

Location :	**Nashua**, **New Hampshire**
Established in :	**1933**
Average Undergraduate Population :	**900-1300**
Average Graduate Population :	**600 - 800**

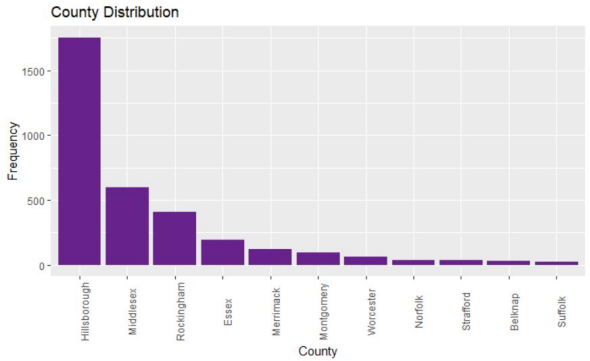
Prospective Students are located throughout the world

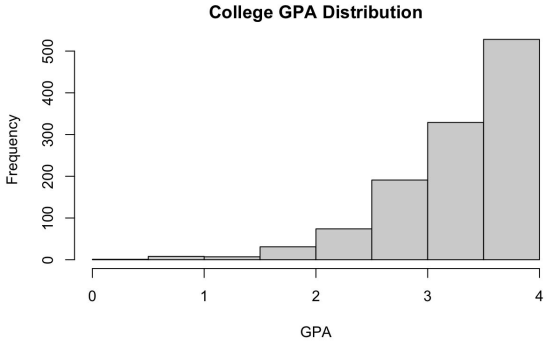


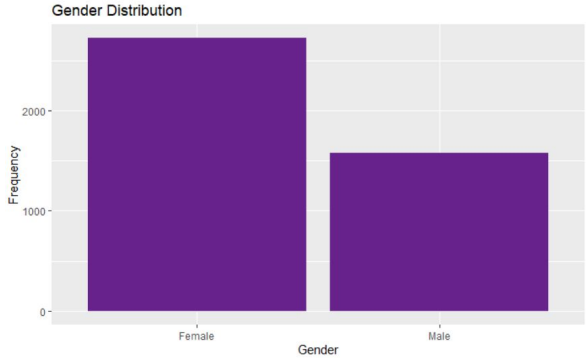
Source: https://www.rivier.edu/about/facts-statistics/

Data Overview

- **Understanding the wireframe** of the given dataset and its metadata files
- Plotting **Histograms** and running various **summary statistics** commands on various columns which had heavy presence of Data
- This helped us understand the columns that are more important; that have **outliers** and correct data entries

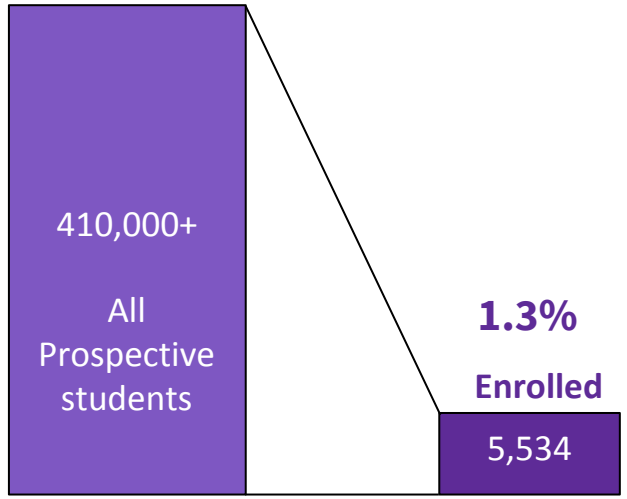


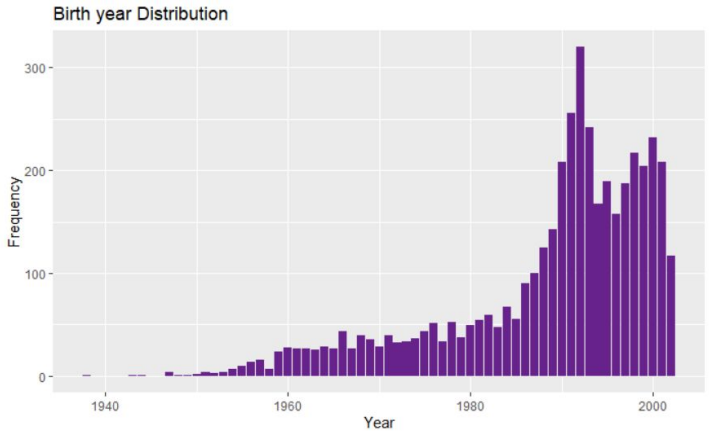




Data Cleaning

- Followed by Data Overview, Filtered Datasets by **milestone Enrolled data** column to focus on enrolled students in the past & removed records where enrollment wasn't present
- Introduced '**NA**' and '**0**' for blanks and empty values
- Removed outliers such as enrolled **before 1933**
- Converted data into standardised formats such as Number, Character, strings and Time





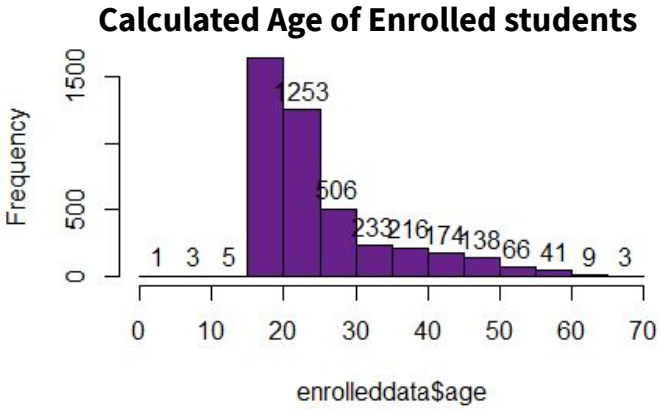
Feature Engineering

- **Inner join** to generate a synthetic column based off of "**Historical Annual median household income by County**" and compared that to household income of students.

Wealthy criteria: Household Income > Median Household Income by County

- Determined **age of applicant** when applying based off of birth year and enrollment year - both extracted from Date of birth and milestone_enrolled_date

Age of Applicant = Milestone_enrolled_year - birth_year



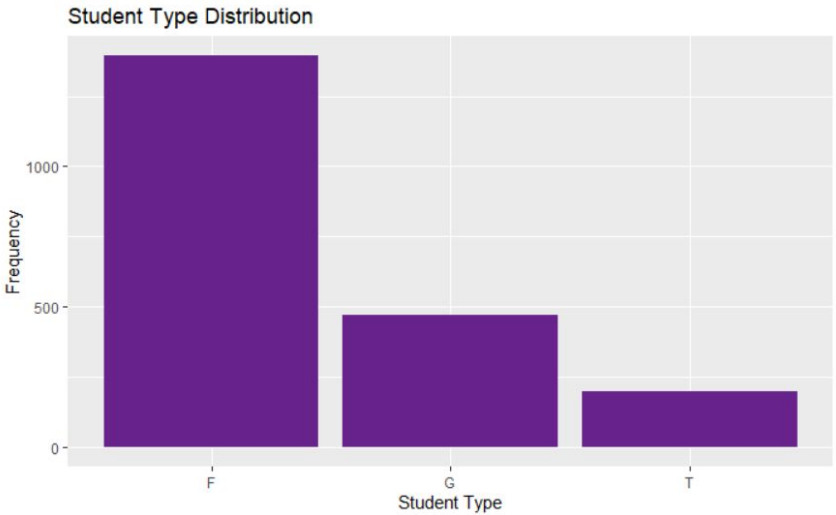
- Calculated Time required for completion of application based on milestone_application_startdate and milestone_application_submitdate

Completion time = milestone_application_submitdate - milestone_application_startdate

https://apps.bea.gov/regional/downloadzip.cfm

Data Enhancement & Normalisation

- **Replaced Duplicate** terms in Student_type by replacing
  - Freshman & First-Time student with **F**
  - Graduate with **G**
- **Filled Blank** Data in Student_type based on High School Gpa, College Gpa, Enrolled Campus, intended_major
- **Referred metadata files** and labels to enhance data quality



Building pipeline

Pipeline functions:

- General

General type (all discrete data)

- Special

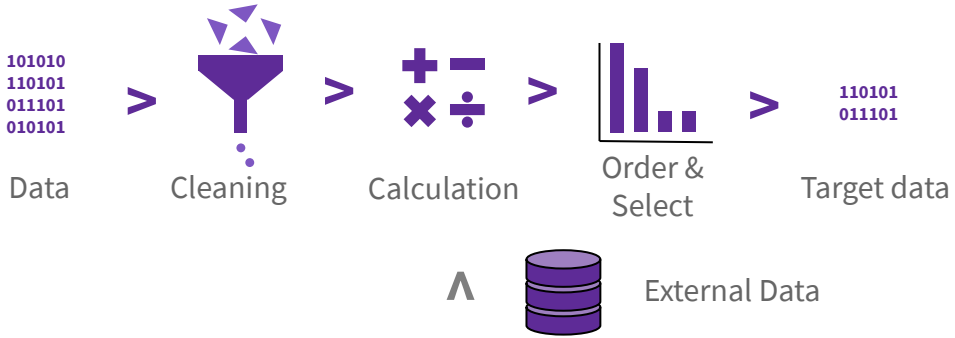
GPA | Label

- Feature engineering

Age | Income

Duration

Pipeline Flow:



Sample usage:

```
testRes<- enrolldate %>%
  pipe_appage() %>%
  pipe_appdur()%>%
  pipe_general(targetCol="intended_major")%>%
  pipe_general(targetCol="address_home_state")
```

Persona Gallery

7.3%

Name: **Charlie**
High School GPA: **3.5**
Applying : **Undergrad**
Submit Applications in **Same Day**

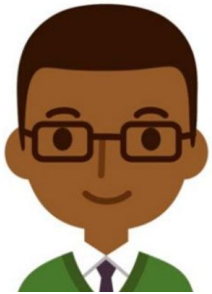
16.3%

Name: **Sophia**
Prefer to Get Admission Information via **Search**
International Student
Willing to Pay for **Deposit**

8.2%

Name: Lily
Gender: **Female**
Live in **New Hampshire**
Get University Information from **Former Students**

Persona Gallery



Name: **Oscar** 9.1%

Gender: **Male**

Intend to learn **Computer Science**

International Student

Name: **Emily** 13.4%

Have **Infrequent Interaction** with Admission Department

Submit Applications in **Same Day**

With an **Affluent** Household Income

A

CONCLUSION

Improve SEO/CTRs & Digital Ad Spends

16%+ : International Students get info through Search method & **98%+** are willing to pay **Deposit**

Same Day Application Completion

Freshman with **3.5+** GPA who complete app on the same day demonstrates Motivation, Interest & Preparation

Select Affluent Family Background

Affluent Family Background & Same Day Application competition is a positive indicator for Enrollment

Info Sessions with Female Ambassadors

Local Female Freshman candidates prefer to stay near their houses and gets their information locally from **former students**

Advertise CS major

70%+ of **International Male** applicants prefer CS Majors





THANK YOU

Any Questions ?