CSE 564: Visualization

Project Prelim Report: Examining patterns in student drop out rate to improve education system

---

Introduction:

By graphically representing the dataset containing student records, we can identify patterns and trends that may indicate risk factors for student dropout, thereby enabling proactive measures to boost retention rates. Through data visualization, educational institutions can gain insights into students' academic trajectories, pinpoint areas for enhancement at both individual and institutional levels, and design specific interventions or programs to foster academic success more effectively.

Abstract:

This exploration leverages visualization techniques to dissect the multifaceted influences of social and economic factors on student achievement. By employing graphs, charts, and maps, the study delves into the interconnectedness of variables like parent's qualification,parent's occupation,marital status, inflation rate, unemployment rate, etc . Analyzing data encompassing diverse demographics, educational contexts, and economic landscapes, the research aims to visually elucidate patterns and correlations. This data visualization approach seeks to inform policymakers and educators about the critical role these socioeconomic factors play in shaping educational opportunities and outcomes, ultimately fostering evidence-based strategies for equitable and inclusive education systems.

Tools and Technologies:

Languages: Frontend- D3.js, ReactJS
Backend-Python and R for preprocessing
IDE- VS Code and Google Colab for data preprocessing

Dataset:

- *Data Points:* The dataset comprises a substantial sample size of 4424 points, potentially exceeding 500 data points, given its coverage of diverse variables related to undergraduate students.
- *Dimensionality:* With 29 variables, the dataset surpasses the requirement for having more than 15 dimensions, offering a comprehensive view of student-related factors and macroeconomic indicators.
- *Variable Types:* It provides a balanced mix of numerical and categorical features, including numerical variables like application order, age at enrollment, and academic performance metrics. Categorical features encompass a wide range of socio-economic and demographic factors. Additionally, macroeconomic indicators contribute to the dataset's richness.

Progress:

Selection of Visualization Techniques:

- Identified suitable visualization techniques based on the research objectives and characteristics of the data.

- Since the dataset has plenty of features, we would drop some using dimensionality reduction techniques like PCA.
- Choose visualization methods such as bar plots, pie charts, scatter plots, and radar charts to effectively convey insights into the relationships between social, economic, and academic factors.

1. Stacked Bar Plot:

- We used interactive stacked bar plots to visualize comparisons of academic performance across different socioeconomic groups or educational levels.
- Selecting a bar plotted the pie chart for the selected categorical variable.
- Categorical data were represented with distinct bars, facilitating easy comparison and identification of disparities.

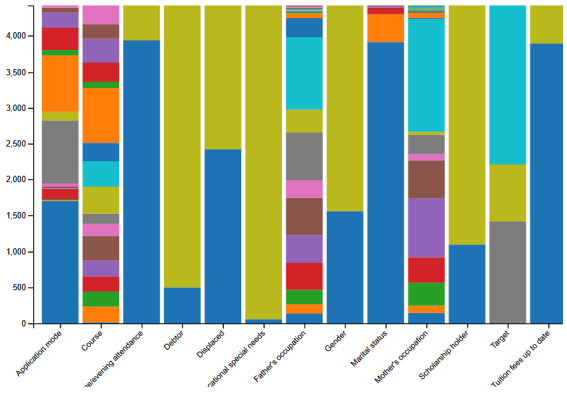

Fig: Bar Plot(Need to make changes)

2. Pie Chart:

- We employed pie charts to illustrate the distribution or proportion of various socioeconomic factors within the student population.
- Highlighting the relative significance of different factors and their contributions to academic outcomes, this visualization will be adjusted depending on the bar plot, highlighting subcategories for variables with the highest dropout percentage.
- Clicking on the slice(class) of the pie for the given variable, will display the distribution of graduated, enrolled, and dropped out students.

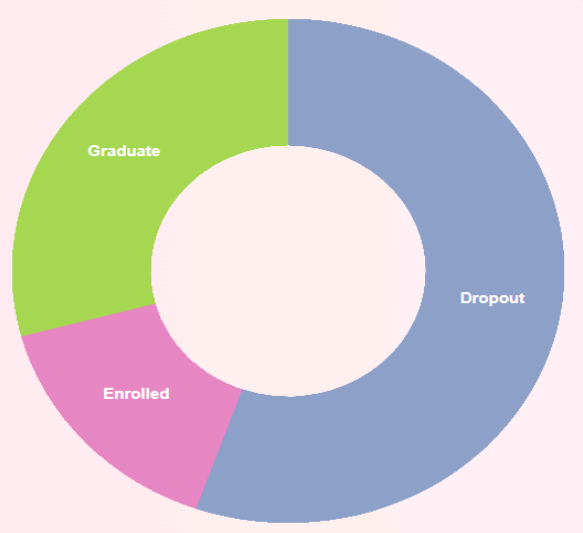

Fig: Pie Chart

3. Radar Chart:

- The radar chart depicted the correlation between various attributes within the dataset, offering a comprehensive view of their interrelationships.
- This visualization served as a complementary tool to other charts like scatterplots, providing additional insights into the correlation structure of the data.
- By allowing user interactions, such as selecting and deselecting attributes, the radar chart enabled dynamic exploration of correlations, facilitating the identification of meaningful patterns.
- Insights gleaned from the radar chart aided in understanding the complex interactions between socioeconomic indicators and academic performance,

informing targeted interventions and support strategies.

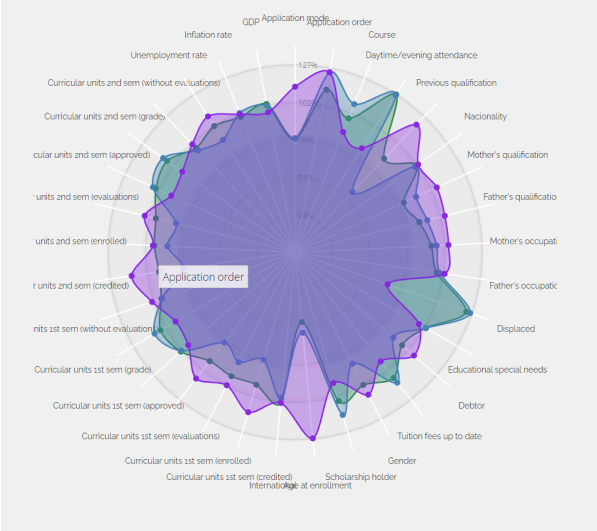

Fig: Radar Chart

4. MDS Plot:

- MDS is a powerful method for visualizing high-dimensional data in a lower-dimensional space while maintaining the pairwise distances between data points as accurately as possible.
- Selecting the numerical variable will plot a radar plot for it. Similarly deselecting it will remove it.

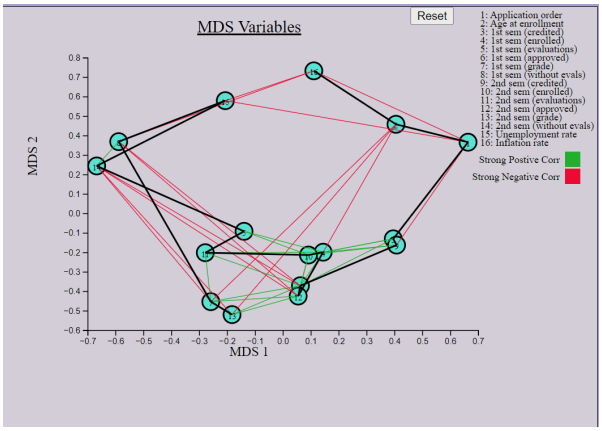

Fig: MDS plots

User Interaction:

- User interaction points were integrated into the visualization pipeline to facilitate dynamic exploration and engagement.
- On the multidimensional scaling (MDS) plot, users could select numerical variables of interest.
- Upon selecting a numerical variable, a radar plot of the chosen variable was dynamically generated.
- Users could click on the variable again to remove it from the radar plot, enabling iterative exploration of relationships between different variables.

Future Implementation:

- Scatter plot to interactively visualize both categorical and numerical variables together to draw insights and tell a story.
- Refine existing implementations, to improve readability and interpretability.
- Research and Development

References:

Dataset:https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention

Used Images from previous reports as dataset is the same as used for Labs.