Exploratory Data Analysis (EDA)

Arjamand Ali
Section 1 - DSAI
Tableau Public Link : here

September 07, 2024

Introduction

This project uses data from the Integrated Postsecondary Education Data System (IPEDS), a key resource for information about colleges, universities, and technical schools in the United States. The data comes from the National Center for Education Statistics and covers things like enrollment numbers, graduation rates, and financial details. The goal of this Exploratory Data Analysis (EDA) is to better understand this data, which will help with future AI and machine learning tasks, like predicting trends in enrollment or analyzing school performance.

Visualization Process

Here's how I went about creating the visualizations:

1. **Data Cleaning:**
    - Data Cleaning is important to make sure the visualizations aer accurate and reliable.Since the data i used was preprocessed i went on and skipped this step.
2. **Choosing Charts:**
    - **TreeMaps:** I used this diagram to visualize the number of students enrolled in Urban Universities in comparison to rural Institutes.
    - **Packed Bubbles Diagram :** This Diagram was great for visualizing the number of enrollments from each county , the highest number having the largest size .
    - **Tree Maps In Tabular Format :** Used this diagram to display multiple dimensional attributes on a single diagram .
    - **Highlighted Tables:** I used this diagram because this diagram has a characteristic of being really neat and simple , showing exact idea very easily .

3. Building the Dashboard:

- I combined these charts into a dashboard, making sure everything was laid out logically so users could easily follow the data story. I also added legends and interactive elements so users could explore the data themselves.

Decision-Making Justification

Here's why I made certain design choices:

- **Clarity:** I chose each chart to best represent the specific data. For example, bar charts were perfect for comparing different types of institutions.
- **Aesthetics:** I picked colors and layouts that are easy on the eyes but still clear. The goal was to make the dashboard look good without sacrificing readability.
- **Interactivity:** I added interactive features like filters and tooltips so users can dig deeper into the data. This makes the dashboard more engaging and useful.

Challenges and Solutions

I faced a few challenges during the project:

- **Data Quality Issues:** As per Project requirement , finding suitable data was a bit research worthy , as many of the datasets online are very large or were just not suitable for this project .
- **Choosing the Right Charts:** It took some experimenting to find the best charts to use. I tried different types before settling on the ones that worked best.

Conclusion :

This EDA project used IPEDS data to uncover key insights about American higher education. Through visualizations like TreeMaps and Packed Bubbles, I presented the data clearly and interactively. Despite challenges in selecting the right data and charts, the final dashboard effectively highlights important trends, supporting future AI and machine learning tasks. This project reinforced the importance of careful design in data analysis.

Dashboard Snip :