Project Proposal: Exploring International Student Trends in US Universities

-Akhil Kaundinya Metlakunta

Introduction

The percentage of international students enrolled in relation to all students in a particular nation is known as international student mobility. In the US, despite a slight decline during the COVID-19 pandemic, the number of international students enrolled is still high—more than a million people, or nearly 6% of the total student body. These students come from more than 227 different countries, but a sizable portion come from China and India.

Objective and scope of work

The project's goal is to investigate trends and patterns pertaining to foreign students enrolled in American universities. We use publicly available datasets and SQL and Python to analyze them to determine what factors affect the enrolment of international students, graduation rates, study fields, and regional preferences. The analysis helps universities customize their curricula to meet the needs of a wide range of students by offering insights into the state of education.

Insights

1. **Recognizing Trends in Enrollment:** Track variations in the number of international students over time and identify the variables that have an impact on these trends, such as adjustments to visa regulations or shifts in the state of the economy.

2. **Regional Preferences:** Determine which states or regions have higher enrolment rates by analyzing regional preferences among international students.

3. **Analysis of Field of Study:** Investigate the most popular fields of study for overseas students and evaluate how academic interests have changed over time.

4. **Graduation Rates:** Assess international students' graduation rates as well as the variables that influence their success or failure in the classroom.

5. **Policy Implications:** Provide input on institutional strategies and policy choices to improve the experience of international students and encourage inclusiveness and diversity on campus.

6. The most common countries of origine for foreign students.

TECHNICAL APPROACH

Analysis: - Evaluation Many aspects need to be considered to better comprehend and apply data analytics for the project. It is also necessary to conduct research on the sources of data availability. It is necessary to weigh the benefits and drawbacks of different software packages before deciding which one would be best for analyzing and interpreting the data. Prior to formulating a method or set of guidelines for determining the gender and age of a given set of data, all these factors must be examined.

Requirements Development: - There will be several categories used to categorize the model requirements, including function, system, input/output, operations, and interface. Once the data is accessed, the initial requirements will be refined to allow for a more thorough formulation of the factors being addressed.

Model Development: - The selection of features to be used as model inputs will be the first stage in the model development process. We'll analyze the data to find features that address the important points. To display patterns, distributions of website visits and rates of use of numerical metrics will be plotted.

The model will be developed further using an agile methodology, whereby brief iterations of design and testing will be used to gradually improve the model. The overall actions to be taken are:

1. **Data Retrieval:** Utilize SQL to extract relevant data from IPEDS and Open Doors datasets, including enrollment figures, graduation rates, and demographic information.

2. **Data Cleaning and Preparation**: Employ Python to clean and preprocess the extracted datasets, handling missing values, standardizing formats, and merging datasets for comprehensive analysis.

3. **Exploratory Data Analysis (EDA):** Utilize Python libraries such as Pandas, Matplotlib, and Seaborn to conduct EDA, visualizing trends in international student enrollment over time, distribution across states,

and popular fields of study.

4. **Statistical Analysis:** Perform statistical tests to identify correlations between factors such as tuition rates, geographical location, and enrollment trends.

5. **Predictive Modeling (Optional):** Employ machine learning algorithms in Python to predict future enrollment trends based on historical data patterns.

6. **Visualization and Reporting:** Create interactive dashboards and visualizations using tools like Tableau or Power BI to present key findings and insights effectively.

Datasets Used:

1. The Integrated Postsecondary Education Data System (IPEDS) offers extensive information on US postsecondary educational institutions, such as enrollment figures, degree completion rates, and demographics of students.

2. Open Doors Data: This annual dataset, which is released by the Institute of International Education (IIE), provides information on trends in international student enrollment as well as the most popular study areas and countries of origin.