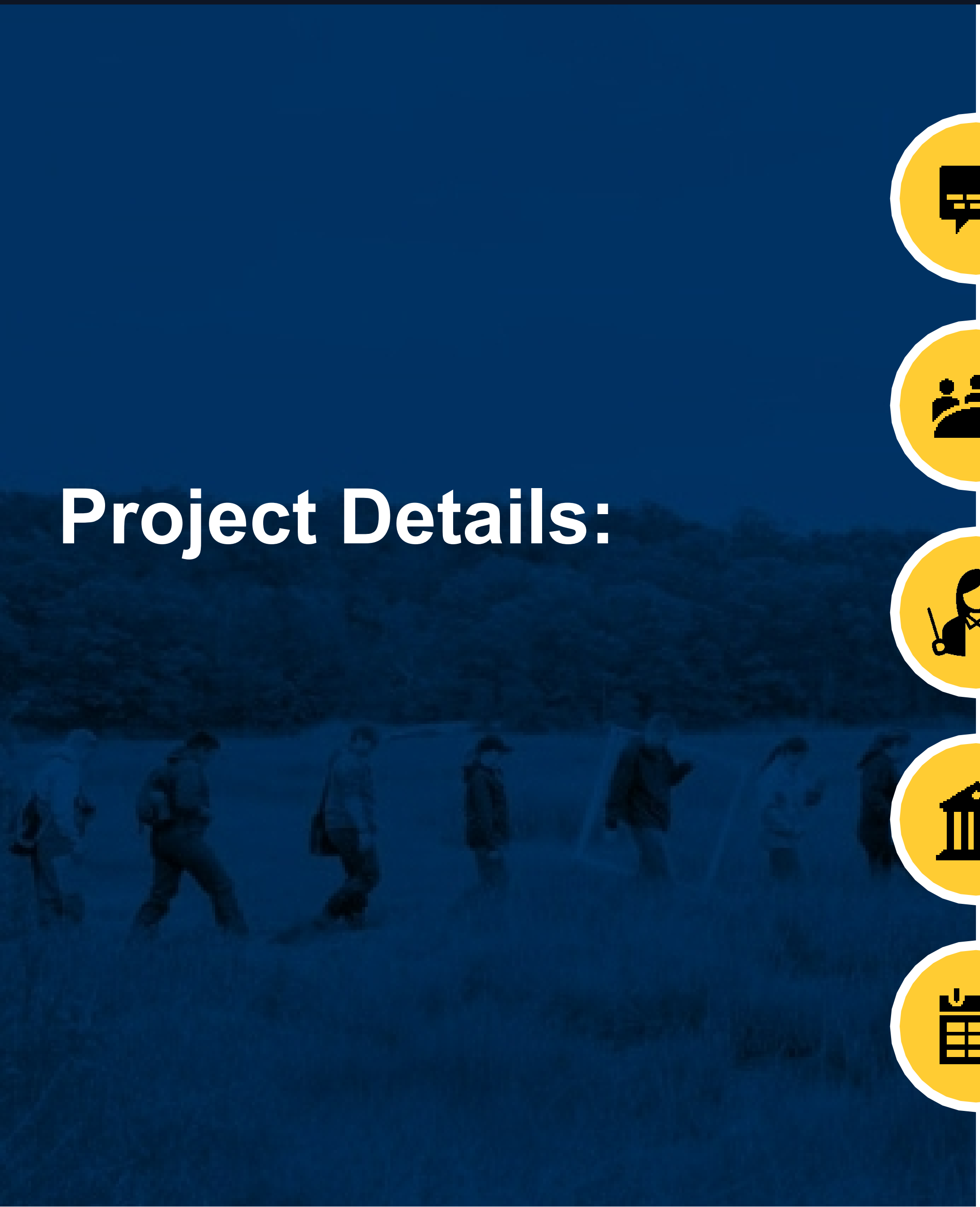
University of New Haven

TAGLIATELA COLLEGE OF ENGINEERING

EDU PREDICT TOOL
PREDICTING ENROLLMENT TRENDS IN HIGHER EDUCATION IN THE USA



Project Details:

Project Title: EDU PREDICT TOOL

Project Team: 01

Project Advisor: Dr. Ardiana Sula

MSDS Capstone – SP 25 DSCI 6051-07

Date of Presentation: 02-28-2025

EduPredict – Forecasting Higher Education Enrollment Trends

- **Goal:** Analyze and predict future enrollment patterns for international students in U.S. universities using data-driven techniques.
- **Outcome:** A **Power BI-based interactive dashboard** that provides enrollment trend forecasts under different scenarios (**Baseline, Growth, and Decline**).
- **Core Features:**
  - Machine learning-driven insights for **strategic academic planning**.
  - Customizable filters for **region, study level, and time period** to explore data dynamically.
- **Significance:**
  - Equips university officials with insights to **optimize resource planning and policy decisions**.
  - Supports institutions in **adapting to demographic and economic shifts** in student enrollment.

TEAM

Koteswar Enamadni

Ifra Naaz Mohammed

Krishnaveni Peesapati

Gnaneswari Vaddepalli

Chethan Chakradhar M

Karthik Vinnakota

The Problem

1

Unpredictable Enrollment Trends

- Universities struggle to accurately forecast student enrollments due to shifting demographics, changing student preferences, and external influences.

2

Impact of External Factors

- Enrollment is heavily influenced by **government policies, economic conditions, and global mobility trends**, making long-term planning difficult.

3

Data Fragmentation & Decision-Making Challenges

- Enrollment data is scattered across multiple sources, leading to incomplete insights and reactive rather than proactive strategic planning.

Project Goals & Objectives

Primary Goal: Develop a **predictive model** that estimates future **student enrollments** based on multiple factors.

Key Objectives:

- Integrate **historical datasets** on enrollment, demographics, and funding sources.
- Identify **patterns & trends** in student admissions across various academic levels.
- Build a **forecasting model** using machine learning.
- Deploy results via a **user-friendly Power BI dashboard** for dynamic analysis.

Data Collection & Sources

Title: Where Our Data Comes From

Content:

- **Dataset Breakdown:**
  - **Student Enrollment Trends** (By year, region, academic level).
  - **Demographics** (Gender, marital status, visa type).
  - **Funding Sources** (Self-funded, government-sponsored, institution grants).
- **Data Quality:**
  - Missing data handled using **forward fill & interpolation** techniques.
  - Cleaned, structured, and prepared for **model training**.

Exploratory Data Analysis (EDA)

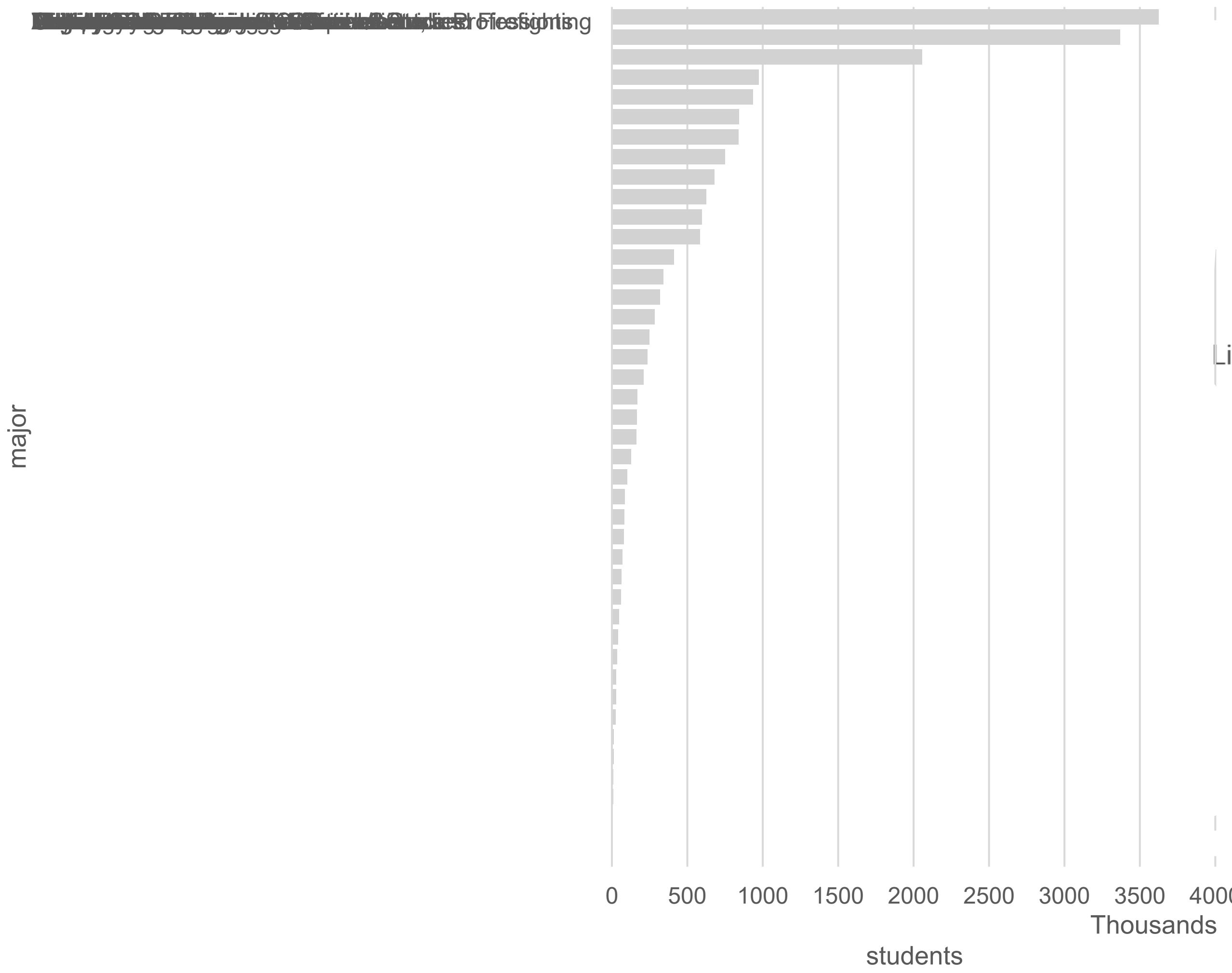
Title: Making Sense of the Data

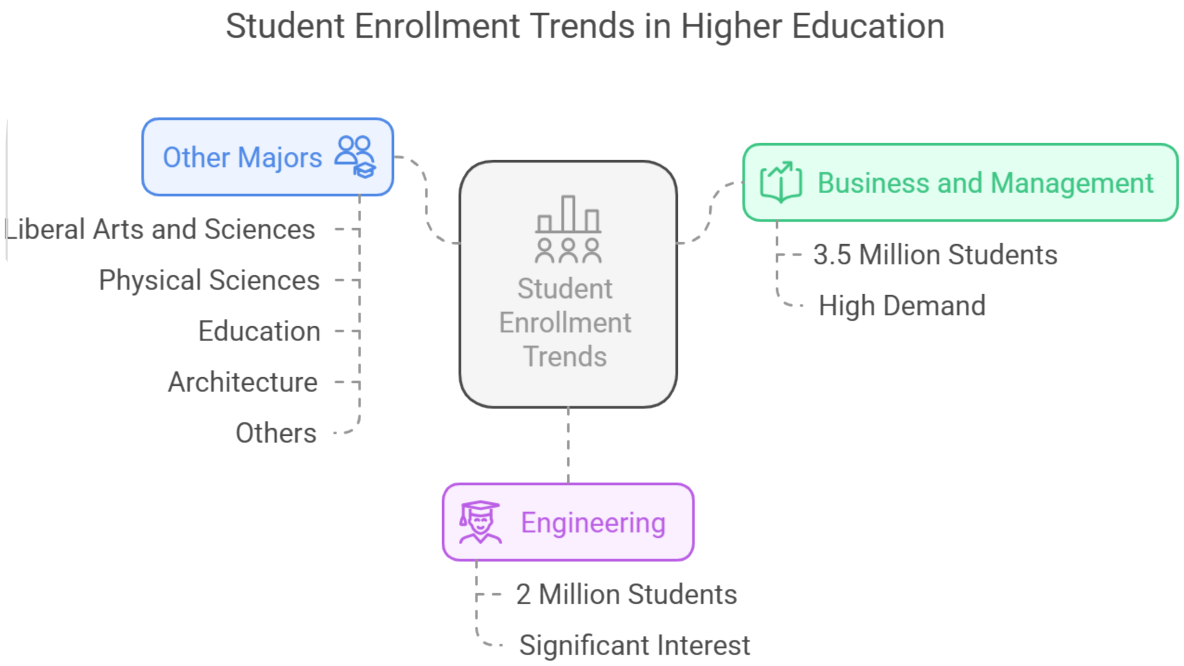
Content:

- **What We Did:**
  - Examined trends in **student enrollment across different years**.
  - Analyzed **regional student distribution** to identify dominant source countries.
  - Explored funding types to understand **how students finance their studies**.
  - Correlation analysis to detect **relationships between key factors**.
- **Key Insights:**
  - **Business & STEM fields** are the most popular.
  - **Asia contributes the highest number** of international students.
  - Enrollment growth has **gradually increased over the years**.

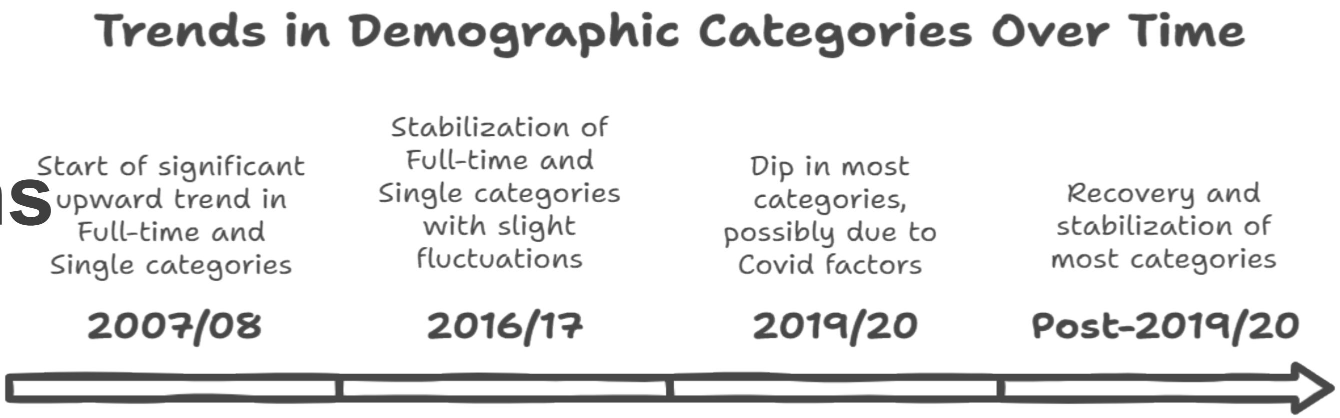
Key EDA Visualizations

'major': Business and Management and Engineering have noticeably higher 'students'.

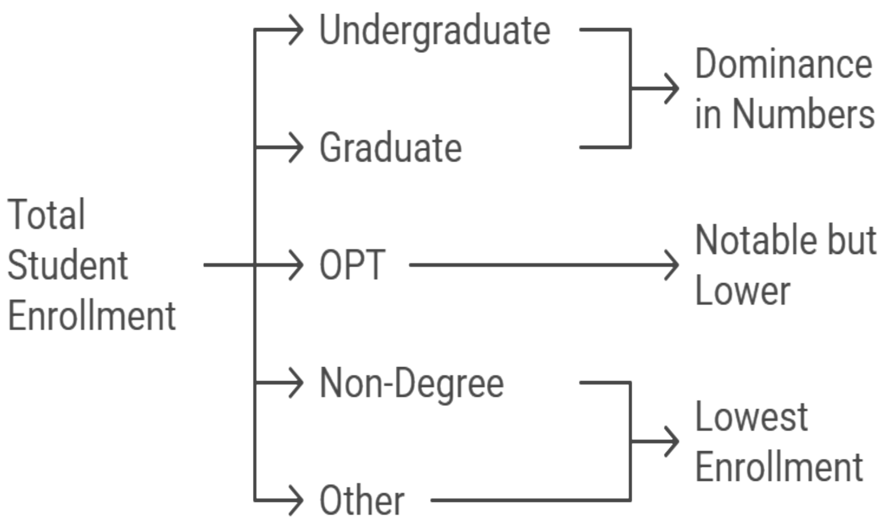


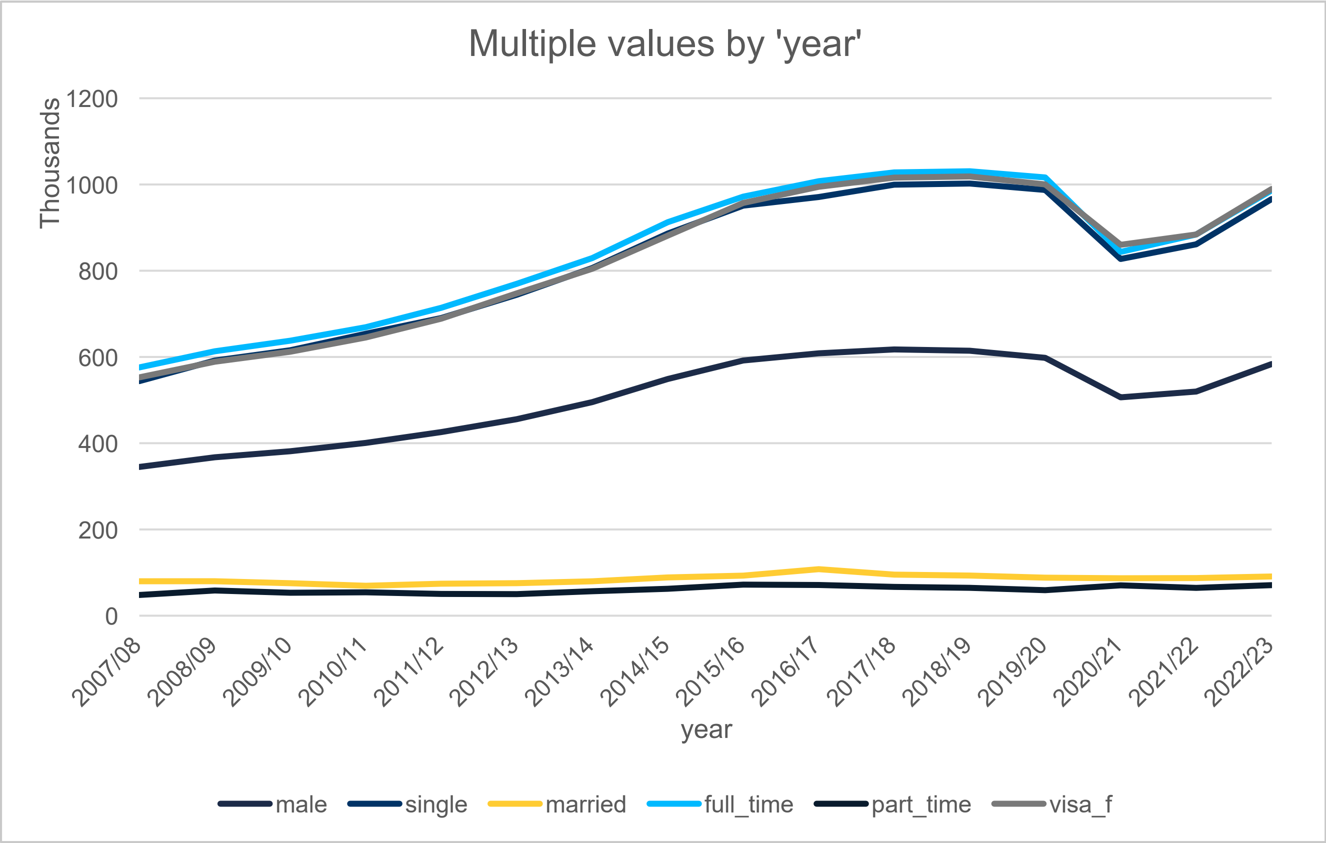


Key EDA Visualizations



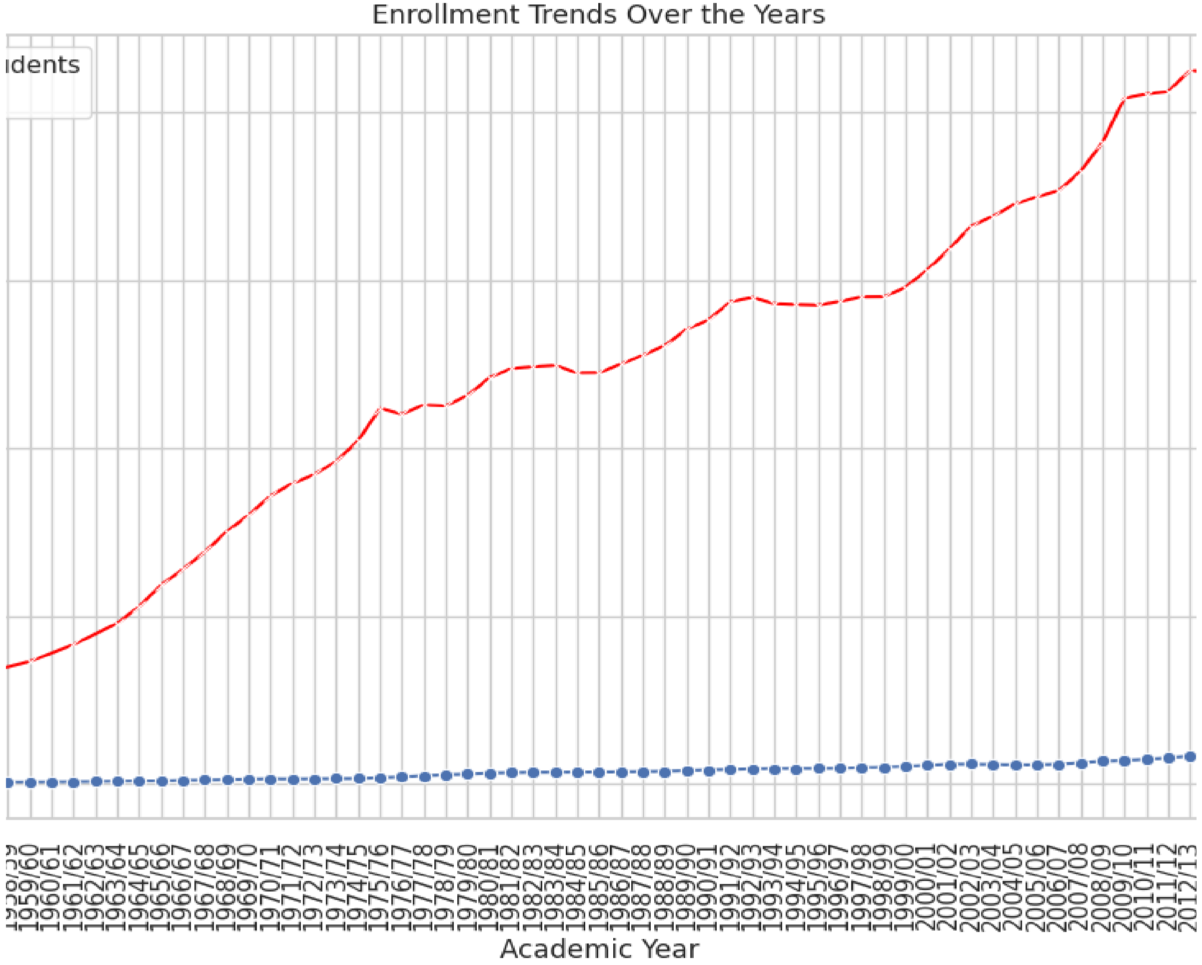
Student Enrollment Trends and Patterns



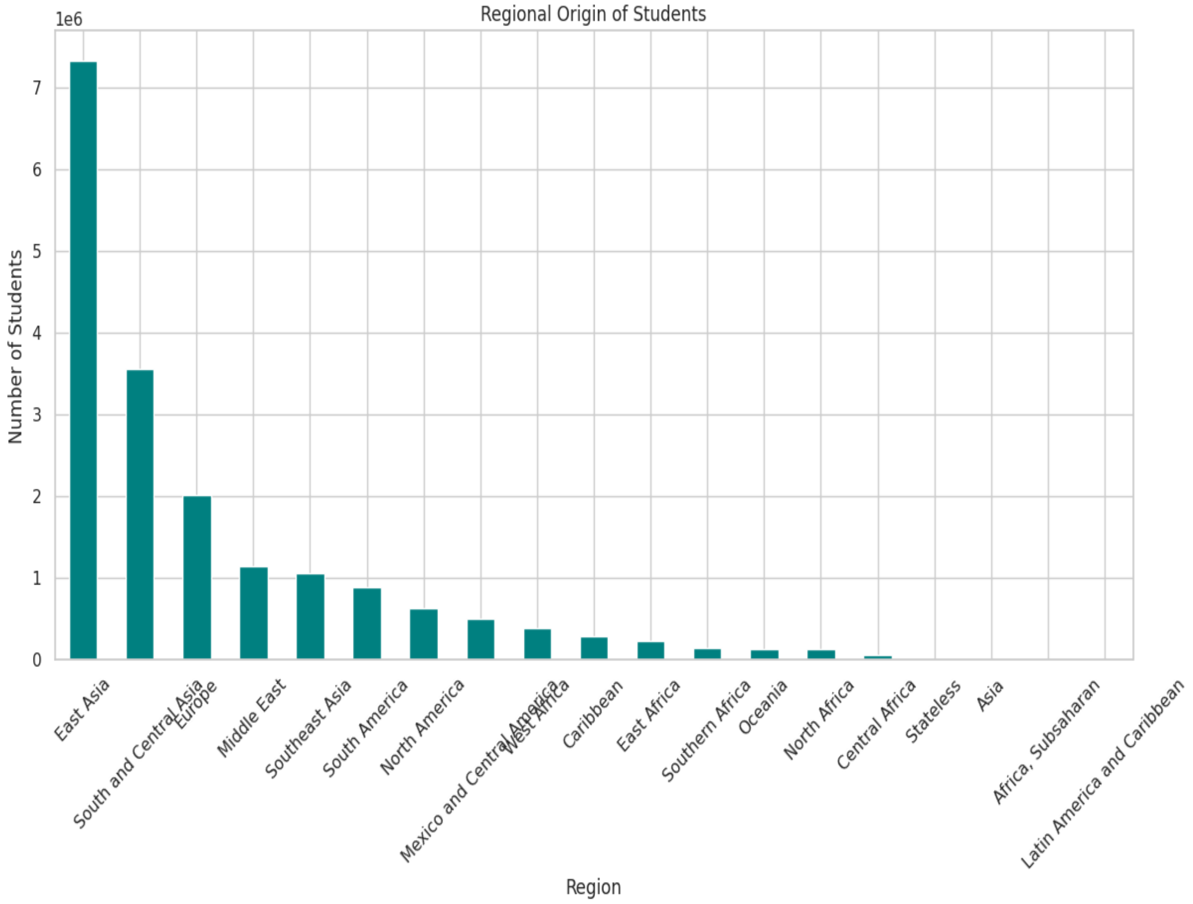


Key EDA Visualizations

Trend Analysis: Growth in **international vs. U.S. students** over time.

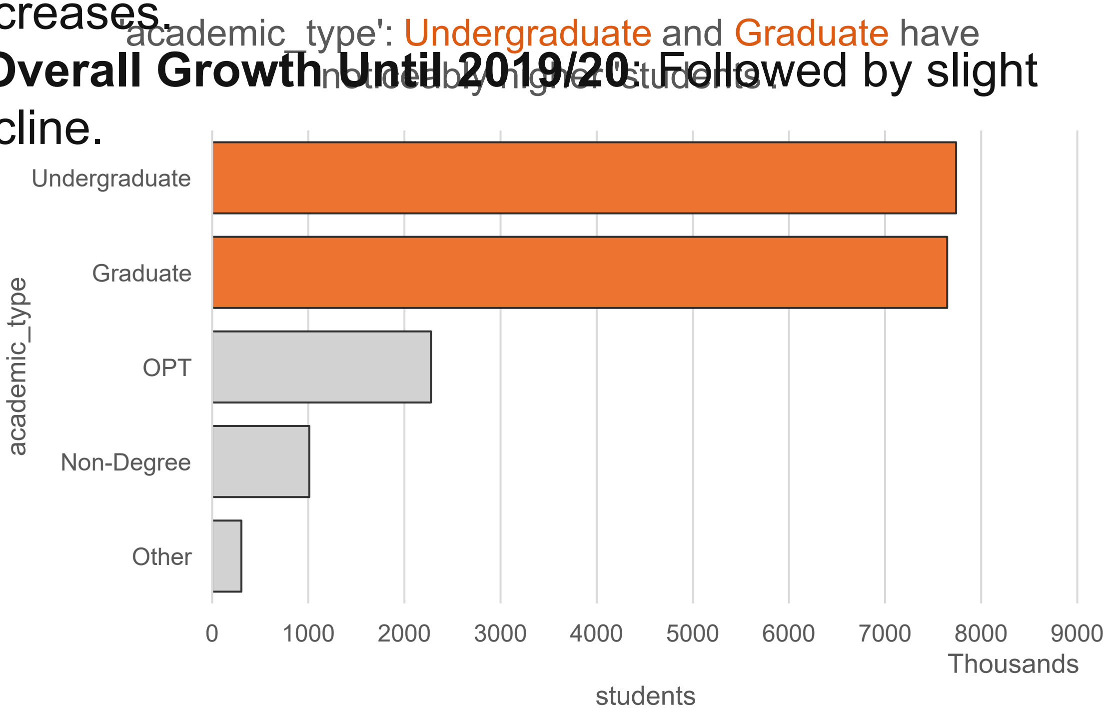


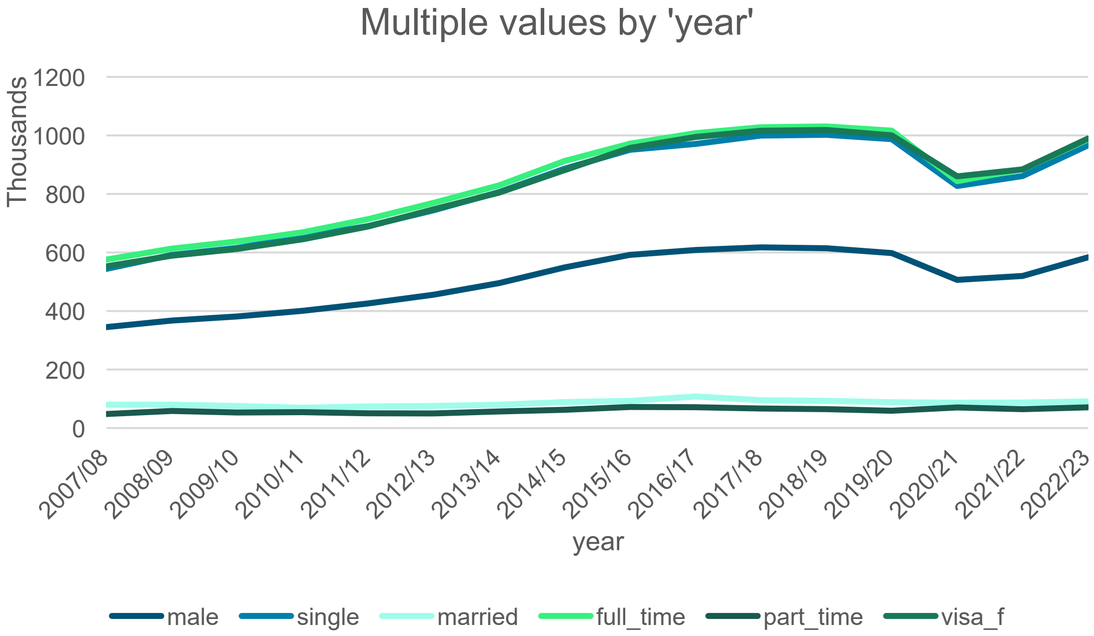
Regional Breakdown: Top contributing countries & continents for international students.



Key EDA Visualizations

1. **Full-time, Married Increase**: These categories show upward trends.
2. **Male, Part-time, Visa_f Stable**: Little to no change over time.
3. **Single Declines**: Peaks around 2017/18, then decreases.
4. **Overall Growth Until 2019/20**: Followed by slight decline.





1. **Full-time, Married Increase**: These categories show upward trends.
2. **Male, Part-time, Visa_f Stable**: Little to no change over time.
3. **Single Declines**: Peaks around 2017/18, then decreases.
4. **Overall Growth Until 2019/20**: Followed by slight decline.

Trends in Employment Categories

Full-time Growth

Significant upward trend observed from 2007/08 to 2016/17.

Male Growth

Slower upward trend compared to full-time and single categories.

Married Stability

Remains relatively flat throughout the observed period.

Observations Summary

Highlights notable growth and external factors affecting trends.

Single Growth

Notable increase from 2007/08 to 2016/17, stabilizing afterward.

Part-time Growth

Shows an upward trend but at a slower rate.

Visa_f Increase

Slight increase over the years, lowest among all categories.

Handling Missing Data

Title: Ensuring Data Accuracy

Content:

- **Challenges:**
    - Some datasets had missing values, particularly in **older academic years**.
- **Solutions Implemented:**
    - **Forward Fill:** Used to maintain **continuity in time-series data**.
    - **Linear Interpolation:** Applied to estimate missing records based on known values.
    - **Outlier Detection:** Cleaned inconsistent data to avoid skewing results.
- **Outcome:**
    - A well-structured, **high-quality dataset** for training predictive models.

Modeling Approach

Title: Choosing the Right Prediction Models

Content:

- **Forecasting Techniques Used:**
  - **ARIMA**: For traditional time-series forecasting.
  - **Prophet**: Captures seasonality and event-driven changes.
  - **XGBoost**: Handles complex interactions in **demographic and financial factors**.
- **Trade-Offs:**
  - ARIMA works well for **time-dependent trends** but lacks flexibility.
  - Prophet helps with **seasonal patterns** but needs additional tuning.
  - XGBoost is **highly accurate** but computationally expensive.

Baseline & Progress [02-26-2025 to 03-02-2025]

Title: What We Have Done So Far

Content:

- **Reviewed past submissions** and finalized project scope.
- **Conducted detailed EDA** to understand **key trends & data gaps**.
- **Handled missing data** using **forward-fill and interpolation**.
- **Developed the first baseline forecasting model (ARIMA)** to test initial predictions.
- **Refined feature selection** to improve model accuracy.

Bi-Weekly Key Milestones Report

Title: Tracking Our Progress

Content:

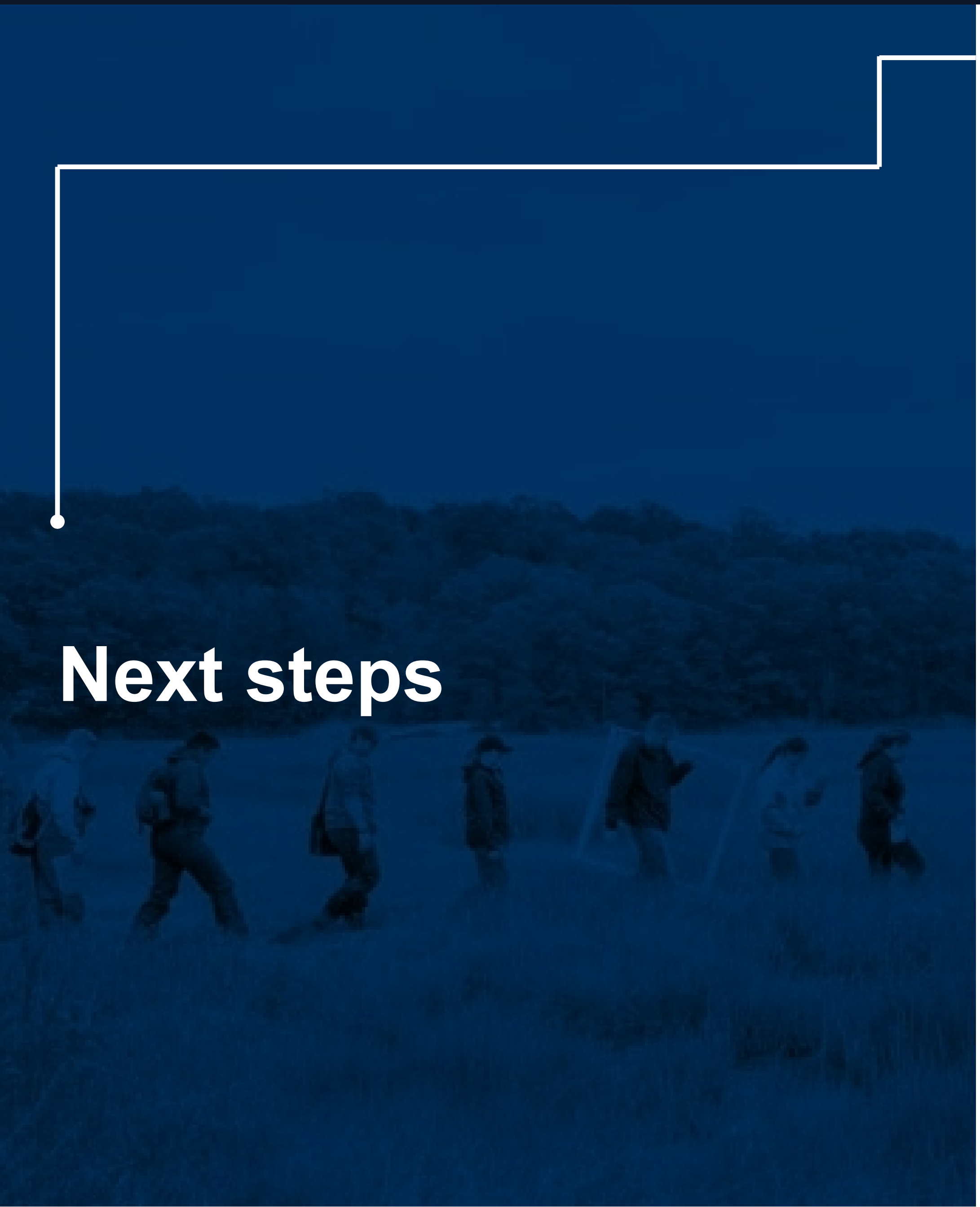
- **Week 1 Progress:**
  - Reviewed **project scope and data sources**.
  - Conducted **data cleaning & preprocessing**.
  - Built **initial visualization reports** for trend analysis.
- **Next Steps:**
  - Enhance **model accuracy** using hyperparameter tuning.
  - Explore **alternative models** for better forecasting.
  - Start integrating insights into **Power BI dashboard**.

Implementation & Deployment

FeatureEngineering: Categorical encoding, normalizing numerical features.

Model Optimization: Hyperparameter tuning to **increase predictive accuracy**.

Deployment Strategy: Integrating forecasts into an **interactive Power BI dashboard &** Enabling **user-driven filtering** for customized trend analysis.



Next steps

Refine our EDA to enhance trend interpretation.

Improve model accuracy with fine-tuned parameters.

Test different machine learning models for better performance.

Build a user-friendly dashboard for interactive data visualization.

Document insights for reporting and final project submission.

THANK YOU

Please let us know if you have any further suggestions or improvements.

University of New Haven

TAGLIATELA
COLLEGE OF ENGINEERING

POWER ON