



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: 04-09-2025

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.
 - Understanding how technological shift effects the field of education

TEAM - 01



• **Koteswar Enamadni**



• **Ifra Naaz Mohammed**



• **Krishnaveni Peesapati**



• **Gnaneswari Vaddepalli**



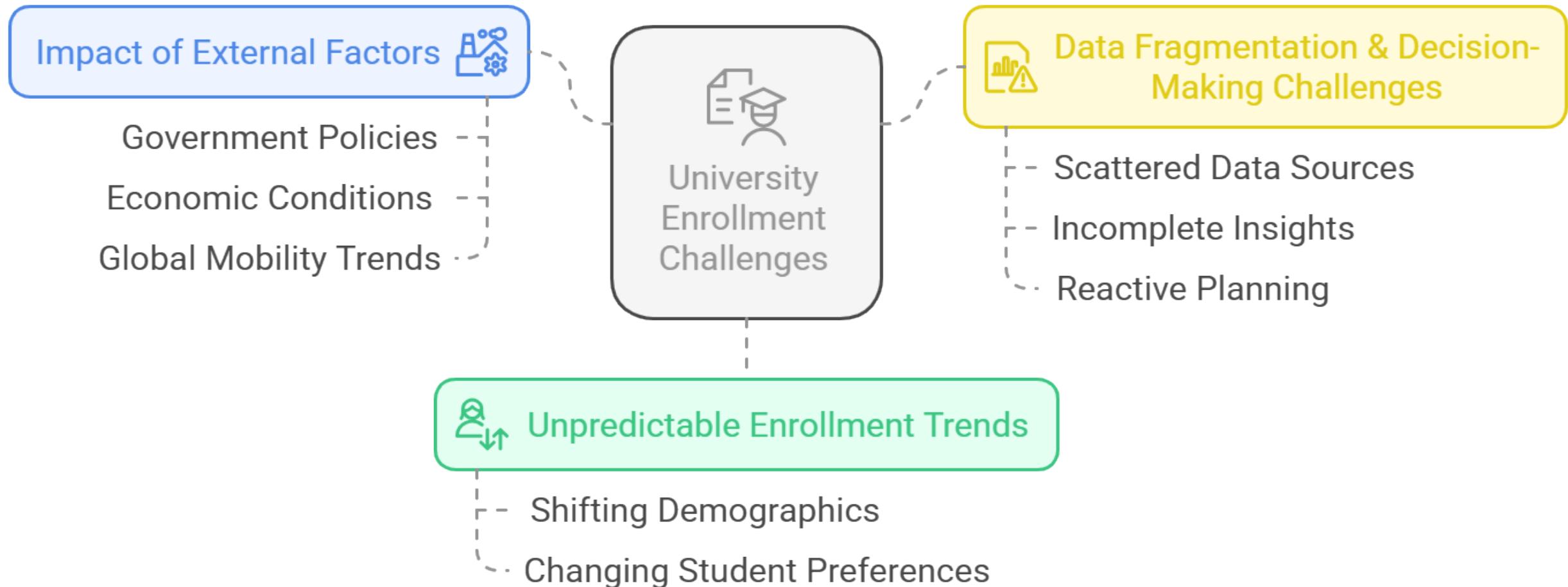
• **Chethan Chakradhar M**



• **Karthik Vinnakota**

The Problem

University Enrollment Challenges: Trends, Factors & Data



Project Goals & Objectives

Primary Goal:

Develop a predictive model that estimates future student enrollments based on multiple factors.

Data Collection & Sources : [Dataset_link](#)

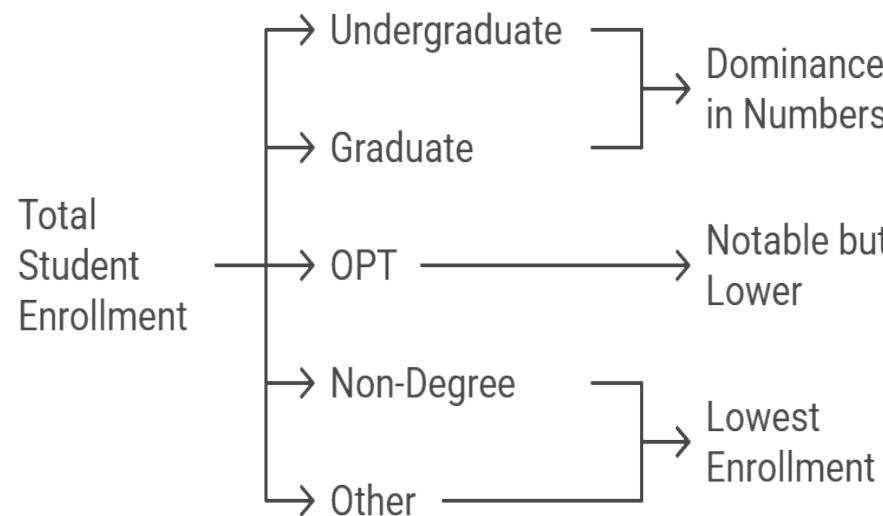
- Dataset Breakdown:
- Student Enrollment Trends (By year, region, academic level).
- Demographics (Gender, marital status, visa type).
- Funding Sources (Self-funded, government-sponsored, institution grants).

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.

Key EDA Visualizations

Student Enrollment Trends and Patterns



Trends in Demographic Categories Over Time

Start of significant upward trend in Full-time and Single categories

2007/08

Stabilization of Full-time and Single categories with slight fluctuations

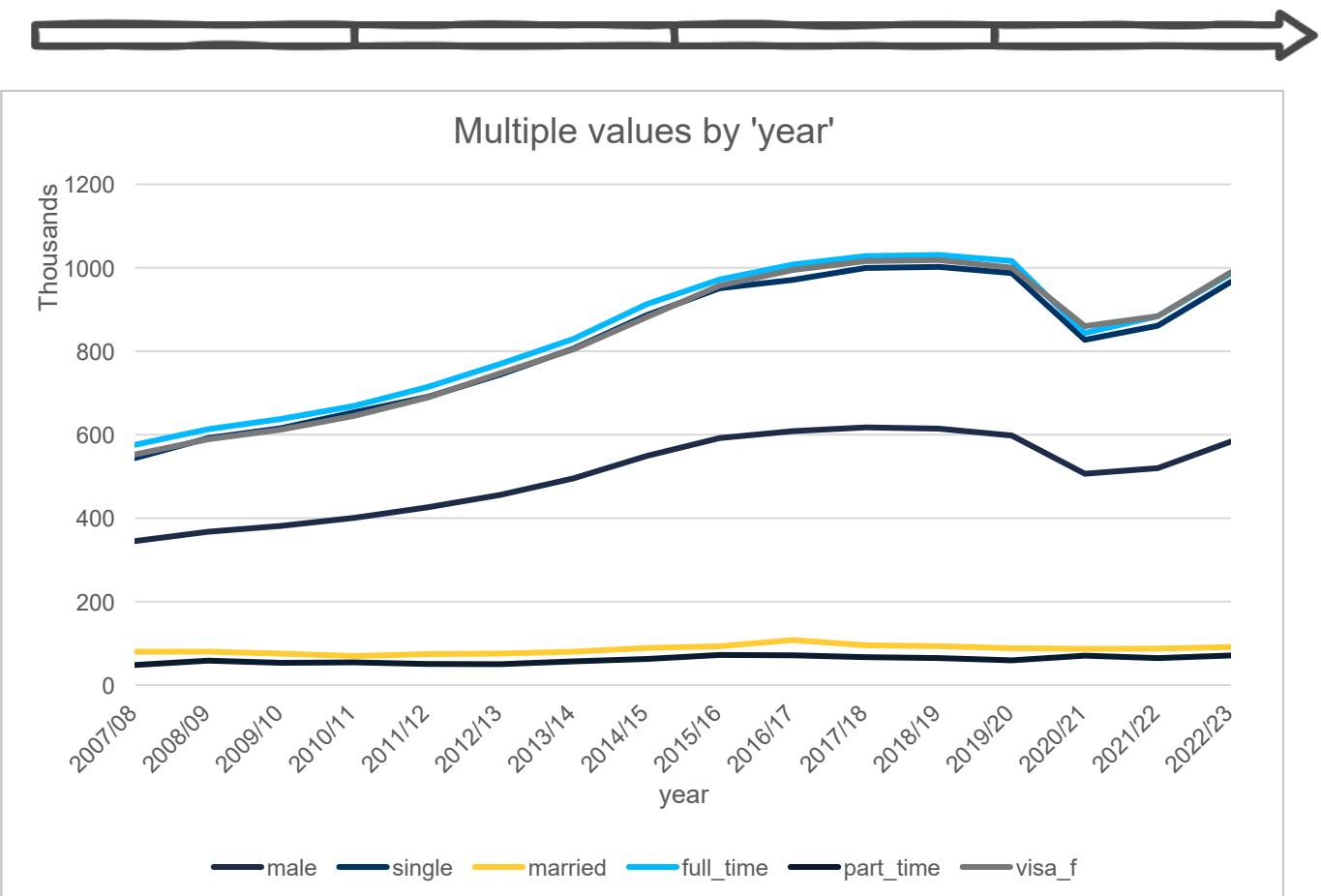
2016/17

Dip in most categories, possibly due to Covid factors

2019/20

Recovery and stabilization of most categories

Post-2019/20



Trends in Employment

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.



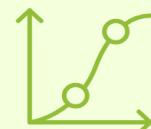
Observations Summary

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

- Challenges:
 - Some datasets had missing values, particularly in **older academic years**.
- Solutions Implemented:
 - **2007 <= data: selected data from and after 2007 to ensure connectivity in data.**
 - **Feature engineering: Cleaned all the data types and created new columns to include in dashboard.**
- Outcome:
 - A well-structured, **high-quality dataset** for training predictive models.

Business Questions for Analysis

Student Demographics & Enrollment

- What is the distribution of students by academic type (Undergraduate, Graduate, Non-Degree, OPT)?
- How has the number of international vs. U.S. students changed over the years?
- What are the trends in student enrollment (full-time vs. part-time, marital status, gender distribution)?

Financial Insights

- What are the primary sources of funding for international students?
- How does funding distribution vary across academic types?

Regional & Visa Analysis

- Which regions and countries contribute the most international students?
- What are the trends in visa types (F, J, and other categories) over the years?

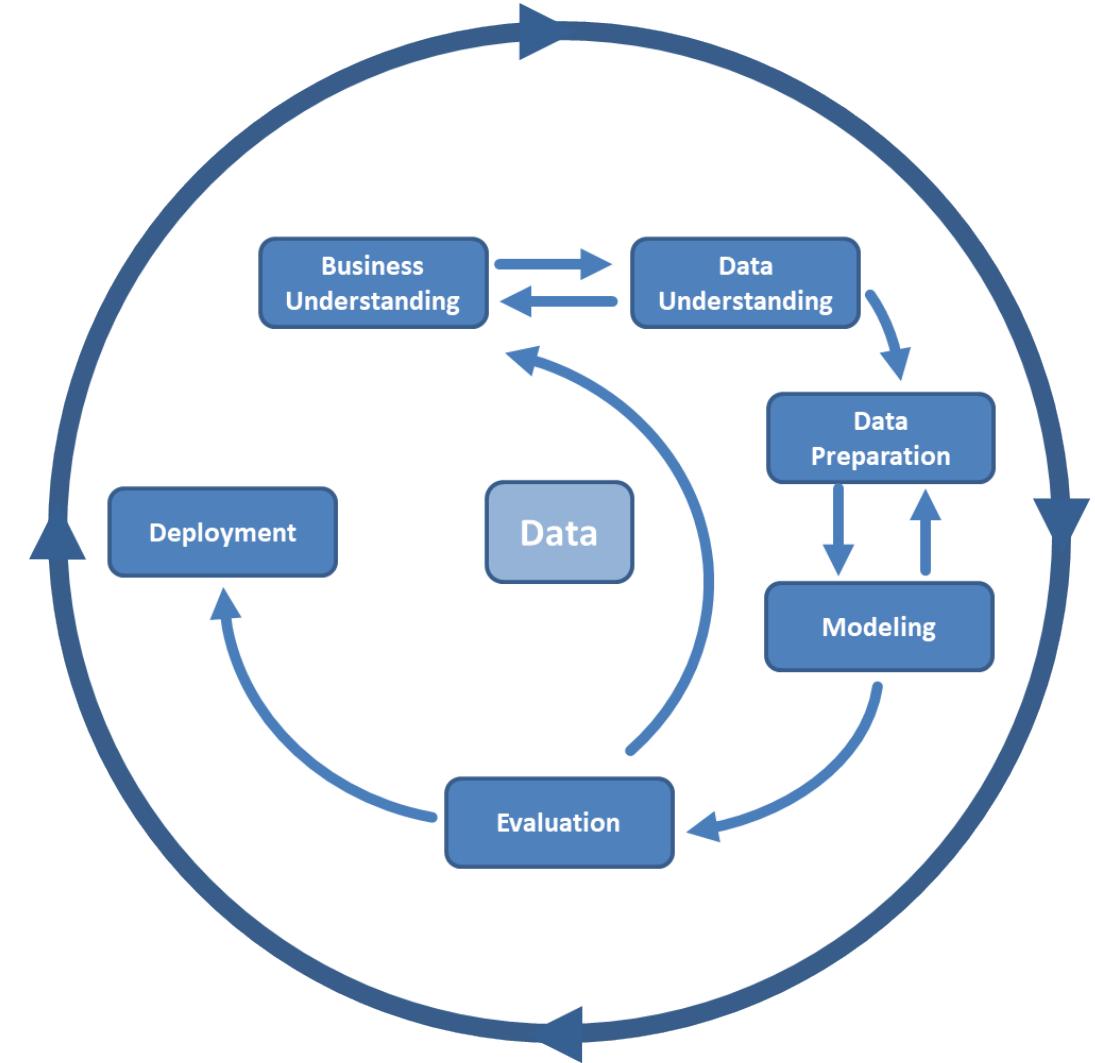
Field of Study Trends

- Which fields of study and majors have the highest enrollment?
- How has the popularity of different majors changed over time?

CRISP DM

Methodology

- **Data-Centric Approach:** Focuses on data understanding and preparation to ensure robust model building.
- **Flexibility:** Its iterative nature allows for updates and refinements with changing data or goals.
- **Stakeholder Focus:** Aligns insights with business objectives for actionable decision-making.
- **Broad Applicability:** Suitable for diverse industries, including forecasting international student demographics.



Model Comparisons

Random Forest:

- We used it as an initial benchmark.
- Accurately predicted past data but **lacked forecasting capability**.
- Served as a reference for evaluating other models.
 - MAE: 6259.04
 - R² Score: 0.8303

Polynomial Regression:

Regression-based approach, but **prone to high variance**.

- Training MAE: 41,133.15
- Testing MAE: 88,058.07
- Training R² Score: 0.8190
- Testing R² Score: 0.5114

ARIMA:

A time-series model that **effectively captures trends** and long-term patterns.

- In-Sample MAE: 38,395.77
- In-Sample R² Score: 0.7513

ARIMA Model

Why ARIMA?

- Specifically designed for time-series forecasting.
- Captures long-term enrollment trends effectively.
- More stable than regression-based models for forecasting.

Strengths of ARIMA:

- Identifies trends without overfitting.
- Provides reliable enrollment projections based on past patterns.
- More interpretable than complex machine learning models for stakeholders.

Final Decision: ARIMA was selected as the best model for student enrollment forecasting due to its ability to provide stable, interpretable, and data-driven predictions over time.

ARIMA Modeling & Forecasting

- **Parameter Selection**

Selected (p, d, q) values using ACF, PACF, and grid search.

- **Model Fitting & Refinement**

Trained ARIMA models on historical data with iterative improvements.

Ensured residuals had no autocorrelation, indicating a well-fitted model.

- **Forecasting & Confidence Intervals**

Generated forecasts for each dataset column.

Included confidence intervals to quantify prediction uncertainty.

Evaluation & Insights

- **Model Performance Assessment**

Evaluated accuracy using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Compared performance across datasets to validate robustness.

Power BI Dashboard Overview

- Interactive dashboards were developed to provide a user-friendly interface for exploring enrollment trends and predictions.
- Key datasets, including student demographics, funding sources, and regional data, were integrated into the dashboards.

Dashboard features:

- Dynamic visualizations (charts, graphs, maps) to represent enrollment patterns.
- Interactive filters to allow users to explore specific subsets of data (e.g., by year, region, academic level).
- KPI cards to highlight key metrics (e.g., total enrollment, growth rate).
- ARIMA-based forecasts presented alongside historical data for trend comparison.
- The dashboards are designed to facilitate data-driven decision-making by providing clear, actionable insights into enrollment trends.

Baseline & Progress [04-02-2025 to 04-13-2025]

Title: What We Have Done So Far

Content:

- Completed web-based dashboard development for all datasets: Status, Origin, Academic, and Source of Fund.
- Integrated ARIMA-based forecasts up to 2032 in each dashboard.
- Enhanced dashboard design with a 3-row layout, KPI cards, and dynamic visualizations.
- Deployed the full project on Streamlit Cloud, ensuring accessibility.
- Enabled GitHub integration for efficient version control and collaboration.

Bi-Weekly Key Milestones Report

- **Week Progress:**

- Finalized the development of interactive dashboards for all key datasets.
- Implemented ARIMA forecasting to provide a clear view of future enrollment trends.
- Optimized dashboard layout and incorporated KPI cards for quick insights.
- Successfully deployed the project on Streamlit Cloud, making it accessible to stakeholders.
- Established GitHub integration to streamline collaboration and updates.

- **Next Steps:**

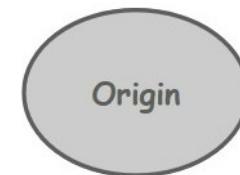
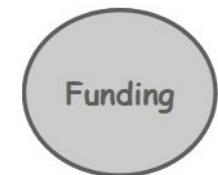
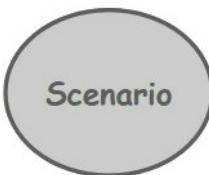
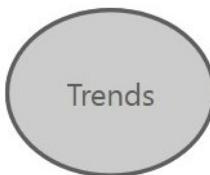
- Begin investigation and selection of alternative platforms (e.g., Python with Dash/Flask, HTML/CSS/JavaScript) for dashboard development.
- Plan the design and implementation of dashboards on the chosen platform to create a unique and enhanced user experience beyond Power BI's capabilities.

Dashboard Analysis

EduPredict: International Student Enrollment Forecast

A Predictive Analysis from 2007 to 2032

EduPredict ML Tool provides insights into past trends and forecasts future international student enrollments. This dashboard enables stakeholders to analyze trends, funding sources, student demographics, and more, helping with strategic planning.



Koteswar's Contributions

- **Data Cleaning and Transformation:** Loads, cleans, and prepares enrollment data (including creating a 'total' enrollment column and handling datetime).
- **Data Validation:** Removed duplicate records and cross-verified totals to maintain data accuracy and reliability for analysis.
- **Feature Engineering:** Modified and enhanced features to improve the dataset's quality and relevance, optimizing it for more effective analysis and modeling.
- **Dashboard Design Contribution:** Created a rough hand-drawn layout to conceptualize the dashboard's structure and functionality, facilitating initial design visualization and team collaboration.

Koteswar's Contributions

- **Data Handling:** Loads, cleans, and prepares enrollment data (including creating a 'total' enrollment column and handling datetime).
- **Stationarity:** Analyzes data stationarity using the Dickey-Fuller test and applies differencing to achieve stationarity if needed.
- **Model Selection:** Employs ACF/PACF analysis to guide ARIMA model order selection and uses RMSE to identify the best-performing model.
- **Forecasting:** Forecasts future total enrollment using the chosen ARIMA model, producing predictions with confidence intervals for a 10-year horizon.
- **Multi-Variable Forecasting & Consolidation:** Forecasts various enrollment categories individually, combines them into a comprehensive table, and visualizes the results.

Ifra's Contributions

Web-Based Interactive Dashboard (MVP):

- Developed a dynamic, real-time dashboard using **Streamlit** for secure data visualization.
- Enabled users to upload their own datasets and instantly view insights across Academic, Origin, Funding, and Status pages.
- Designed intuitive layouts with charts (pie, bar, map, line) tailored for non-technical users.
- Focused on privacy — ensured no data is stored or shared after upload.

Static Power BI Dashboard:

- Built and refined a structured **Power BI dashboard** to answer key enrollment-related business questions.
- Integrated academic, origin, funding, and demographic datasets into a unified visual experience.
- Incorporated **forecasted trends up to 2032** using time series insights.
- Presented insights through KPI cards, breakdowns, and comparative visuals for stakeholder use.

Gnaneswari's Contributions

Power BI Dashboard Enhancement:

- Designed and implemented an interactive Power BI dashboard to facilitate in-depth analysis of student enrollment.
- Leveraged 6 datasets to provide a holistic view of enrollment trends and influencing factors.

Key dashboard functionalities include:

- **Dynamic Filtering:** Enabling users to filter data by various dimensions (year, origin, etc.).
- **Interactive Visualizations:** Employing a range of chart types to present data effectively.
- **Key Performance Indicators (KPIs):** Displaying crucial metrics for quick understanding of enrollment status.
- **Forecast Integration:** Incorporating enrollment predictions to support strategic planning. Focused on creating a user-friendly interface with clear navigation and intuitive controls.

Presentation Outline:

- Started structuring the project presentation to effectively convey the project's objectives, data analysis, and findings.

Chetan's Contributions

Developed and refined the Power BI dashboard for forecasting trends in education fields.

Integrated two key datasets:

- Field of Study Data (2007–2032)
- Student Origin and Visa Status Data (2007–2032)

Created visualizations to analyze:

- Trends by field of study over time
- Regional variations in student enrollment
- Visa-related insights and impact on international student trends
- Applied forecasting techniques to predict future trends up to 2032.
- Implemented dynamic slicers and filters for enhanced interactivity and custom exploration.

Initiated the structuring of the project presentation to clearly and effectively communicate the project's objectives, the methodology behind the data analysis, and the significant findings, ensuring a logical flow and coherent narrative that enhances understanding and engagement for the intended audience.

Krishnaveni's Contributions

Key Contributions:

- Interactive Power BI Dashboard: Designed and built a multi-page dashboard integrating six datasets to provide a holistic view of enrollment data.
- Included button navigation for user-friendly exploration.

Developed visualizations for:

- Distribution by Field of Study
- Annual Enrollment Trends
- Predicted Enrollment Trends from 2022-2032.
- Insights by Status, Origin, Academic Level, and Field of study.

Comprehensive Project Report:

Prepared a detailed report summarizing all key aspects and findings of the entire project.

Future Research:

Continuously refining visualizations and design to enhance data clarity, user experience, and overall understanding.

Karthik's Contributions

Created Power Bi Dashboard Visualizations:

- Enhanced data visualization with comparative trends and year-over-year analysis to highlight changes in student demographics and funding patterns.
- Consolidated diverse datasets—including academic performance, geographic origin, funding sources, and demographic attributes—into a unified data model for consistent reporting.
- Enhanced data visualization with comparative trends and year-over-year analysis to highlight changes in student demographics and funding patterns.

Worked on Documentation:

- Documented dashboard functionality and provided a walkthrough to end users for effective adoption and continued engagement.
- Clearly articulated the rationale behind the ARIMA model selection, outlined the step-by-step modeling pipeline, and integrated findings from EDA into the narrative to provide a seamless flow from data preparation to prediction. Also coordinated with team members to ensure consistency in writing style, formatting, and overall presentation of the final deliverable.



Next steps



Develop dashboards on alternative platforms



Improve UI/UX for better clarity and enhance the user experience.



Refine the dashboard for interactive data visualization based on the layout.



Document insights for reporting and final project submission.



Feedback and further insights: Seeking insights to enhance the dashboard and usability.

THANK YOU

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



University of
New Haven

TAGLIATELA
COLLEGE OF ENGINEERING

**POWER
ON**