



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: 03-19-2025**

# 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**



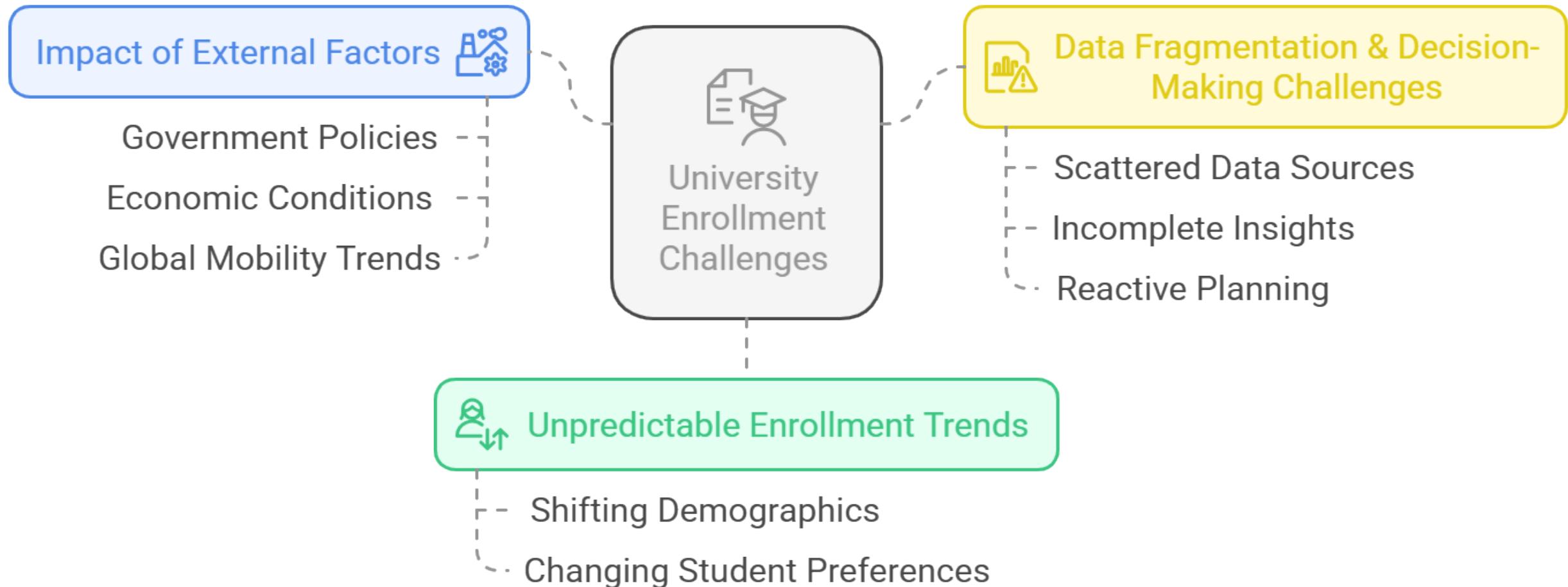
• **Chethan Chakradhar M**



• **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](#)

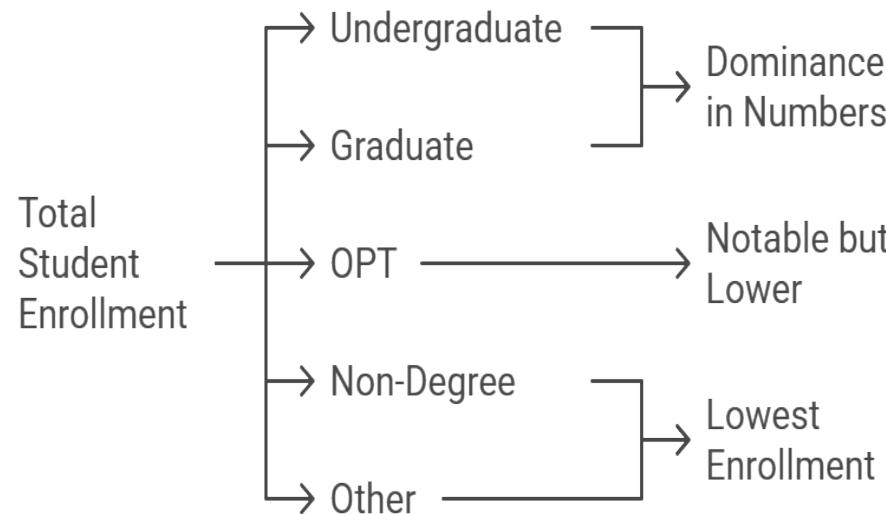
- Dataset Breakdown:
- Student Enrollment Trends (By year, region, academic level).
- Demographics (Gender, marital status, visa type).
- Funding Sources (Self-funded, government-sponsored, institution grants).

## 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

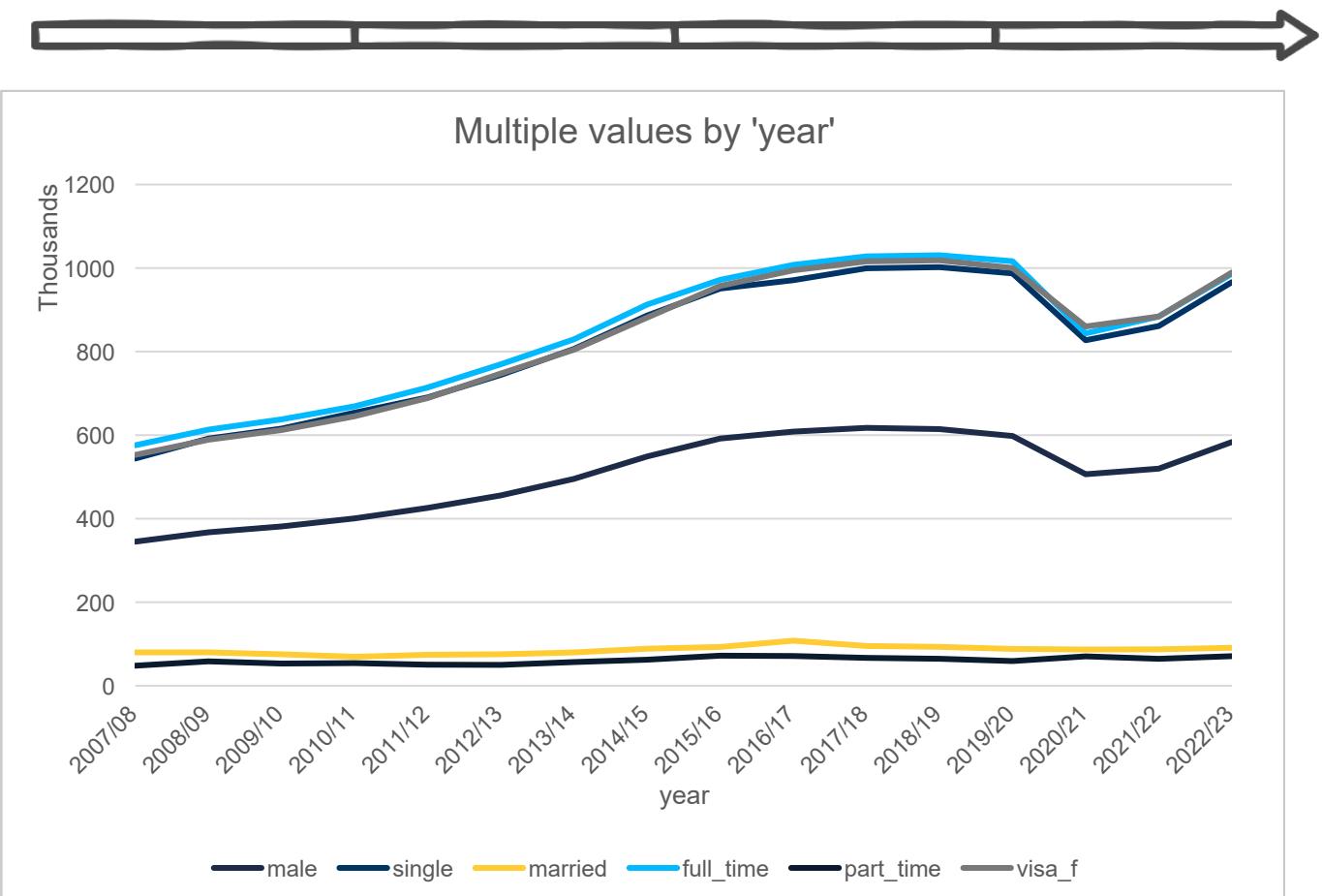
2016/17

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.



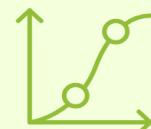
## Observations Summary

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

**Content:**

- **Challenges:**
  - Some datasets had missing values, particularly in **older academic years**.
- **Solutions Implemented:**
  - **Merged datasets:** merged the datasets by year column as the key.
  - **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.

# 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.

# 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?

# Baseline & Progress [03-2-2025 to 03-16-2025]

---

**Title:** What We Have Done So Far

**Content:**

- **Data Cleaning & Transformation:** Merged key datasets (academic, academic\_detail, field\_of\_study) to create a unified structure for analysis. Addressed missing values, invalid entries, and redundancies.
- **Why Use the 2007 Dataset?** Ensures a long-term historical perspective for trend analysis, enabling more accurate forecasting. Captures shifts in enrollment patterns, visa trends, and funding sources over time, improving model reliability and strategic decision-making.
- **Business Insights & Dashboard Design:** Developed intuitive visualizations, computed critical metrics, and structured the dashboard into key sections—Demographics, Program Insights, Funding & Visa Trends.
- **Predictive Modeling & Trend Analysis:** Engineered time-based features, prepared datasets for forecasting, and initiated ARIMA modeling to predict future enrollments.
- **Dashboard Conceptualization:** Created a rough layout to guide dashboard design and stakeholder collaboration.



# Bi-Weekly Key Milestones Report

---

**Title:** Tracking Our Progress

**Content:**

- **Week 3 - 4 Progress:**
  - Conducted **data cleaning & preprocessing**.
  - Built **initial visualization reports** for trend analysis.
  - Built multiple layouts for dashboards.
  - In-depth analysis through summarizing Business Questions.
- **Next Steps:**
  - Enhance **model accuracy** using hyperparameter tuning.
  - Explore **alternative models** for better forecasting.
  - Start integrating insights into **Power BI dashboard**.

# Implementation & Deployment

---

**Feature Engineering:**  
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.

# Koteswar's Contributions

- **Data Cleaning and Transformation:** Combined academic, academic\_detail, and field\_of\_study into a unified dataset with the year column as the key, ensuring the ML model is trained on a streamlined dataset.
- **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.

# Ifra's Contributions

## Data Preparation & Integration:

- Analyzed datasets and ensured data consistency from 2007 onward by cleaning earlier data.
- Removed duplicate records and cross-verified totals.
- Attempted to pivot all datasets into a single sheet for model training & Power BI integration but faced mapping issues.

## Business Questions Development:

- Considered all datasets to define key business questions, ensuring they align with analysis capabilities.

## Power BI Dashboard Planning & Design:

- Strategized and finalized key dashboard pages:
  - Home Screen, About, Demographics, Program-Specific Insights, Funding & Visa Trends, and Business Questions.
- Designed the dashboard template, providing a clear structure for future development.

# Ifra's Contributions

## Ed-Tech ML Tool for Enrollment Predictions - Home Page



The home page of an Ed-Tech ML Tool for Enrollment Predictions. On the left, a sidebar lists navigation options: About, Enrollment, Demographics, Program of Study, Visa & Funding, and Business Questions. The main content area shows enrollment data for the year 2007, with 0.58M total students. A slider allows selecting years from 2007 to 2026.

**Year: 2007**

**Total Students**  
0.58M

**Select Year:**

2007 2009 2011 2013 2015 2017 2019 2021 2023 2024 2025 2026

# Gnaneswari's Contributions

## Data Preparation & Integration:

- Combined all datasets into single dataset & Standardized year columns across datasets to ensure consistency in analysis
- With the final dataset started working on ML model.

**EDA:** focused on student demographics, gender distribution, country of origin, and funding sources by academic type and field.

**Key Insights:** included gender ratio trends across fields and time series analysis of student enrollment.

**Comprehensive Analysis :** combined origin, study field, and funding data to identify top countries, fields, and funding sources, producing structured outputs for further analysis and visualization.

**Dashboard Template :** includes a sidebar navigation with sections like Home, Demographics, Programming, Funding, Visual Analysis, About, Model, Data Source, and Business KPI. The main panel displays total students, enrollment rate, and year, along with widgets for Student Demographics and Enrollment Trends. A Year Slider Bar enables selection between Past Analysis, Present, and Growth/Future. Laying a solid foundation for future advancements.

# Chetan's Contributions

## Data Modeling

- Data modeling used to unify diverse datasets-enrollment, demographics, visa status, and funding into a single, structured framework.
- Conducted thorough data cleaning by addressing missing values, correcting invalid entries, and eliminating duplicate columns to guarantee data precision and uniformity.

## Business insights and Dashboard design

Transforms insights into interactive visuals-charts, maps, and forecasts-offering stakeholders an intuitive tool to monitor key metrics and predict future enrollments.

## Exploratory data analysis (EDA)

Uncovered hidden patterns, validates data quality, and informs modeling by examining historical trends, visa shifts, and program preferences, driving actionable outcomes for institutional growth.

# EduPredict Tool

- Home
- Demographic
- Program Specific
- Funding and Visa analysis
- About
- Future projections

## Total Students

Sort by ▾ This Year

Year	Total Students	Growth Rate (%)
2018	1500	10%
2019	1600	8%
2020	1800	12%
2021	1700	10%
2022	2200	15%

## Analysis

Read More

● Female      ● Male

On average,  
7 out of 10  
Students are  
Female

## Overview

See Details

Total Fields

**45**

► Academic Details

Average Tuition Fee

**\$ 27,943**

► From Tuition Fee

Visa Acceptance

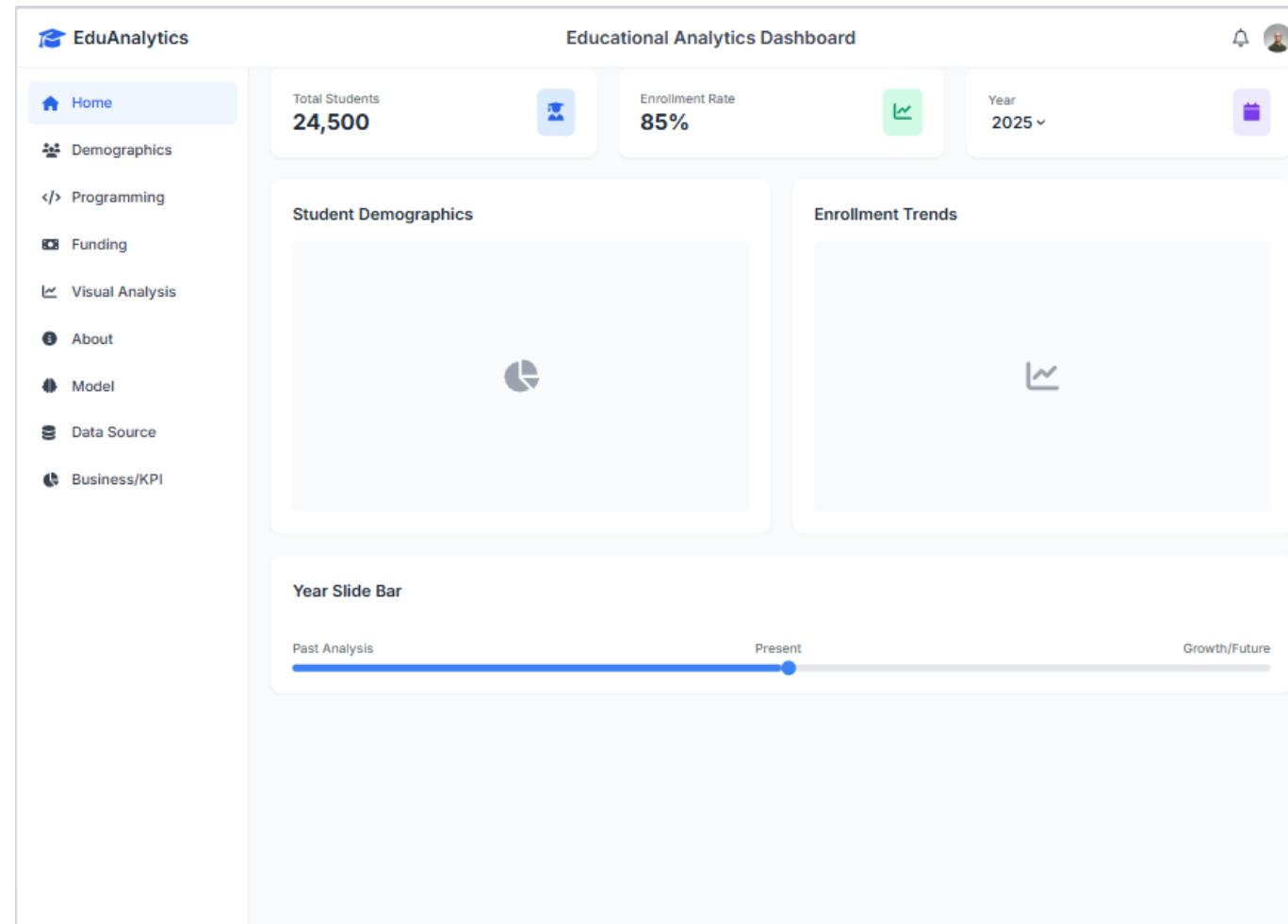
**291**

► From Visa and Funding

# Krishnaveni's Contributions

- **Data Preparation**
  - Performed robust data cleaning by handling missing values, invalid entries, and redundant columns to ensure data accuracy and consistency.
  - Integrated multiple CSV datasets (status, funding, academic details, academic, origin, and field of study) by identifying and linking common fields such as Year2 and academic type.
  - Built a structured foundation via aggregation, pivoting, and merging to enable seamless downstream data modeling and analysis.
- **Business Insights:**
  - Computed essential metrics to uncover critical trends and inform strategic decision-making.
  - Designed an intuitive dashboard layout to visualize key insights such as program-specific trends, funding distributions, and regional student origins.
  - Leveraged data transformation techniques to prepare for clear, impactful visualizations.
- **Predictive Modeling & Trend Analysis:**
  - Engineered time-based features and normalized data to support accurate predictive modeling of future student intake trends.
  - Segmented data into training and testing sets, establishing a robust framework for model development and validation.
  - Set the stage for exploring machine learning models that factor in economic influences and program-specific demand for forecasting admissions through upcoming years.

# Gnaneswari's and Krishnaveni Contribution on Dashboard



# Karthik's Contributions

## Data Preparation & Integration:

- Conducted data cleaning and preprocessing to ensure accuracy and consistency.
- Integrated multiple Excel datasets into a unified dataset by identifying and linking common data fields.
- Created a structured data foundation to enable seamless data modeling and analysis.

## Business Insights & Dashboard Design:

- Developed key business questions based on available datasets to drive data-driven decision-making.
- Planned and designed the dashboard layout, ensuring an intuitive user experience.
- Focused on visualizing key insights, including total admissions, program-specific trends, and visa acceptance rates.

## Predictive Modeling & Trend Analysis:

- Researching on appropriate machine learning models to predict future student intake trends through 2026.
- Incorporated economic factors such as recession impacts on students admissions.
- Analyzed program-specific demand trends to provide strategic insights for educational institutions.

# Karthik's Contributions

- [Home](#)
- [Demographic](#)
- [Program Specific](#)
- [Funding & Visa](#)
- [About](#)
- [Future Projections](#)

## Edu-Predict Dashboard



### Total Students

**889K**

### Year

2017

Select Year:



## Next steps



**Improve model accuracy** with fine-tuned parameters.



**Test different machine learning models** for better performance.



**Build a user-friendly dashboard** for interactive data visualization based on the layout.



**Document insights** for reporting and final project submission.



**Git Link:** provide a git link

# THANK YOU

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



University of  
New Haven

TAGLIATELA  
COLLEGE OF ENGINEERING

**POWER  
ON**