University of Management & Technology
Department of Artificial Intelligence
School of Systems and Technology
Spring Semester 2025

PROJECT REPORT

(UNIVERSITY ENROLLMENT AND GRADUTION TRENDS)

Instructor's Name	**Sir Muzzamil Mustafa**	**Subject**	**Data Warehousing and Business Intelligence**
Course Code	DS3137	**Project**	Project 4
Total Marks	15	**Assigned Date**	12th Jun, 2025
1.	**Kaleemullah Younas**	**ID**	F2022332074

Table of Contents

1. INTRODUCTION

The analysis and interpretation of educational data are critical in making informed decisions about institutional policies, strategies, and investments. This report details a structured approach undertaken to analyze university performance data, specifically focusing on student enrollment, graduation rates, and dropout rates. The aim is to provide actionable insights through meticulous data extraction, thorough transformation, and comprehensive visualization using advanced data analytics tools.

2. DATA EXTRACTION

The data utilized for this report originates from the "Most Recent Cohorts Cleaned.csv" file, representing comprehensive information about universities, including student enrollment, graduation rates, dropout rates, and additional key metrics. This dataset is extracted from reputable educational databases, specifically curated to provide insights into trends across various educational institutions.

The extraction process began with identifying credible educational data sources, primarily government and educational portals like the U.S. Department of Education's College Scorecard. This source is highly reliable due to its consistent updates, extensive verification procedures, and standardized data collection practices. We used - Most-Recent-Cohorts-Institution_05192025

3. DATA TRANSFORMATION

After extraction, the dataset underwent rigorous transformation to prepare it for analysis. The initial step involved cleaning the data to handle missing values, outliers, and inconsistent entries. Missing values were addressed using imputation methods; for numerical data such as graduation and dropout rates, mean or median imputation was applied based on distribution characteristics. Categorical missing data were filled using the most frequent or mode-based imputation methods. Next, data formatting involved standardizing the formats across all columns. Numeric fields were verified to ensure consistent decimal precision and units. Dates and identifiers were converted into consistent formats to maintain uniformity across the dataset. For categorical variables, standardized naming conventions were applied to avoid discrepancies due to capitalization or spelling differences.

Calculations constituted an essential part of the transformation process. Key performance indicators (KPIs) were derived to facilitate meaningful analysis. Metrics such as enrollment growth rates, graduation rate percentages, dropout rate percentages, and year-over-year changes were calculated. Advanced calculations involved creating DAX (Data Analysis Expressions) measures in preparation for Power BI visualization. These measures include calculating averages, percentages, and trend analyses crucial for understanding institutional performance over specified periods.

Finally, the cleaned and transformed data was validated through exploratory data analysis (EDA) using Python scripts, employing libraries such as pandas and NumPy. This ensured the data's integrity and readiness for the loading phase.

4. DATA LOADING

The transformed dataset was imported into Power BI for the final analysis phase. The Power BI Desktop application was employed to facilitate the loading and modeling process. Importing was executed using the "Get Data" feature, where the CSV file was uploaded directly into the Power BI environment, providing immediate access to the complete dataset.

Once imported, the data modeling stage commenced by establishing meaningful relationships among tables and creating a robust data model. This involved identifying primary keys, foreign keys, and setting up cardinality (one-to-many, many-to-one, etc.) relationships appropriately to ensure accurate data associations. A star schema model was constructed, featuring a centralized fact table surrounded by dimension tables, optimizing query performance and enhancing clarity in data visualization.

Several DAX measures were developed during the modeling stage to leverage the powerful analytical capabilities of Power BI. These measures included enrollment metrics (e.g., Total Enrollment, Average Enrollment), graduation rate percentages, dropout rate percentages, and comparative metrics highlighting performance variations across years and institutions. Visualization was the culminating step in the data loading process, wherein interactive dashboards were developed to present insights intuitively and effectively. Power BI's visualization toolkit was utilized to create multiple visuals, including bar charts, line graphs, scatter plots, and pie charts. Specific attention was given to visual best practices, ensuring clarity and readability. For instance, scatter plots illustrated enrollment versus graduation rates, bar charts highlighted comparative dropout rates, and line graphs tracked historical enrollment trends over multiple years.

Advanced visualization features such as conditional formatting was applied to highlight critical metrics clearly. Units such as percentages were distinctly indicated to avoid ambiguity, and labels were strategically positioned to enhance readability. Additionally, tooltips were integrated to provide users with detailed information upon interaction, offering deeper insights without overwhelming the initial visual presentation.

5. CONCLUSION

In summary, the systematic approach to data extraction, meticulous transformation processes, and strategic data loading into Power BI resulted in a robust analytical framework capable of yielding comprehensive educational insights. This workflow exemplifies the effective use of data analytics tools and methodologies to uncover actionable trends in student enrollment, graduation rates, and dropout patterns across educational institutions, providing valuable guidance for stakeholders and decision-makers in the educational sector.