

**Mapping the Educational Landscape: A Comprehensive Data Analytics Study of Engineering Colleges in India**

**ACKNOWLEDGMENT**

I want to extend my sincere gratitude to the entire team at “Analytics Space” for granting me the opportunity to work on the “Mapping the Educational Landscape: A Comprehensive Data Analytics Study of Engineering Colleges in India” project. I would also like to express my heartfelt thanks to my academic team, 'Data Trained,' for providing me with this enriching experience. Working on this project not only allowed me to enhance my analytical skills but also exposed me to a plethora of new insights. The invaluable suggestions and guidance from both teams played a pivotal role in the successful completion of this project, and I am truly grateful for the learning and growth it has brought to me.

**CHAPTER 1**

**INTRODUCTION**

**1.1 Project Objective**

The objective of the **"Mapping the Educational Landscape: A Comprehensive Data Analytics Study of Engineering Colleges in India"** project is to conduct an in-depth analysis of engineering colleges across India using data analytics techniques. This study focuses on gathering and analysing data from sources like getmyuni to assess the distribution of colleges, course availability, fee structures, and infrastructure across different states. The analysis aims to provide insights into the density of institutions, variations in tuition fees between government and private colleges, enrolment trends, and the availability of facilities such as laboratories, libraries, hostels, and sports amenities. By leveraging data-driven insights, this project seeks to assist students, parents, and policymakers in making informed decisions about higher education while also facilitating collaboration between institutions and industries to enhance infrastructure and curriculum offerings in India’s engineering education sector.

**Overview of Project**

This project involves multiple stages, including:

* **Web Scraping and Data Collection:** Extracting data on engineering colleges, such as names, available courses, fee structures, and infrastructure details from sources like getmyuni.
* **Data Cleaning and Preprocessing:** Preparing the dataset for analysis by handling missing values, removing duplicates, and standardizing formats.
* **Data Modelling and Analysis:** Applying statistical and machine learning techniques to uncover patterns and insights regarding college distribution, course availability, and fee variations.

**Visualization:** Presenting findings through interactive dashboards and visualizations using Tableau to facilitate decision-making for students, parents, and policymakers.

**1.2 Problem Statement**

The educational landscape of engineering colleges in India is vast, with institutions spread across different states offering a diverse range of courses. However, comprehensive and centralized data regarding these colleges—covering aspects such as course availability, fee structures, and infrastructure—is fragmented and difficult to access.

This lack of easily accessible information poses a significant challenge for students, parents, and policymakers in making informed decisions about higher education. Without clear insights into the quality and affordability of various engineering colleges, students may struggle to find institutions that best suit their needs, leading to potential mismatches in education and career opportunities.

This project aims to bridge this gap by leveraging data analytics to provide a structured and insightful overview of the engineering education sector in India. Through this study, stakeholders will gain a better understanding of course distributions, fee variations, and infrastructure disparities, enabling them to make well-informed decisions.

**1.3 Project Objectives**

The primary objective of this project is to analyse a dataset comprising information about engineering colleges in India. Through this analysis, the project aims to achieve the following specific objectives:

1. **Mapping Engineering Colleges Across States:** Compiling and analysing data on engineering colleges across different states to provide an overview of the distribution and density of institutions.
2. **Course Availability Analysis:** Categorizing and evaluating various undergraduate and postgraduate engineering programs, including specializations and interdisciplinary options.
3. **Fee Structure Analysis:** Investigating tuition fees per year and total course costs, differentiating between government and private institutions and highlighting regional variations.
4. **Government vs. Private Colleges:** Examining the contributions of government and private engineering colleges in terms of enrolment, infrastructure, and academic performance.
5. **Infrastructure Assessment:** Evaluating facilities such as laboratories, libraries, hostels, and sports amenities to assess their adequacy for student learning and research.

By achieving these objectives, this project will provide actionable insights to help students and parents make informed choices, assist policymakers in shaping data-driven education policies, and support colleges in enhancing their offerings.

**CHAPTER 2**

**DATA COLLECTION AND SOURCES**

**Data Source**

The primary data source for this project is getmyuni, a well-known online platform for education-related insights, particularly in engineering education. Getmyuni offers an extensive database of engineering colleges across India, covering key information such as course offerings, fee structures, infrastructure, and institutional categorization.

The dataset for this project focuses on engineering colleges across different states in India, helping in the comprehensive analysis of higher education trends.

The dataset extracted from getmyuni contains the following key variables:

1. **State**: The state where the engineering college is located, aiding in regional analysis and distribution insights.
2. **College Name**: The official name of the engineering college, providing clear identification of the institutions under study.
3. **Course Availability**: The different undergraduate and postgraduate programs offered, including specialization options.
4. **Fee Structure**: The per-year tuition fees and total course fees, helping compare affordability across different institutions.
5. **College Type**: Classification of institutions into government and private colleges, enabling a comparative analysis.
6. **Infrastructure Facilities**: Details about available resources like laboratories, libraries, hostels, and sports facilities, which influence student experience and academic outcomes.

This dataset serves as the foundation for an in-depth analysis of the engineering education landscape in India, offering insights into accessibility, affordability, and academic infrastructure.

**Web Scraping Process**

Given the vast amount of information required, web scraping was used as the primary method for data extraction from the getmyuni website. Python was employed due to its powerful web scraping libraries that facilitate large-scale data collection.

The following Python libraries were utilized in the process:

1. **BeautifulSoup**: For parsing and extracting relevant data from HTML content, such as college names, course details, and fee structures.
2. **Requests**: To send HTTP requests to getmyuni and retrieve the HTML content programmatically.
3. **Pandas**: For structuring, cleaning, and organizing the extracted data into DataFrames, making it easier to analyse.

**Steps in the Web Scraping Process:**

1. **Sending Requests**: HTTP requests were made to getmyuni.com to access data on engineering colleges across all Indian states.
2. **Parsing HTML Content**: BeautifulSoup was used to extract relevant data, identifying specific HTML tags and attributes.
3. **Extracting Data**: Parsed data was stored in structured Pandas DataFrames, ensuring consistency and accuracy.
4. **Saving Data**: The final dataset was saved in CSV format for easy analysis using data visualization tools like Tableau and Python-based analytics.

**Challenges**

Despite the effectiveness of web scraping, certain challenges were encountered:

1. **Data Volume**: Managing a large dataset covering multiple states and colleges required optimized processing techniques.
2. **Website Restrictions**: Challenges such as CAPTCHAs, rate-limiting, and dynamically loaded content were handled using request delays, random user agents, and exception handling.
3. **Data Quality Issues**: Incomplete or inconsistently formatted data required extensive cleaning and preprocessing to ensure usability.

Despite these challenges, the web scraping process successfully generated a robust dataset, providing valuable insights into India's engineering education sector.

### **CHAPTER 3**

### **DATA PREPROCESSING**

Effective data preprocessing is a critical step in ensuring the quality and reliability of the analysis. This stage involves cleaning the raw data, transforming it into a suitable format for analysis, and importing it into Tableau for visualization. Below is a detailed explanation of the steps involved in the data preprocessing phase of this project.

#### **Data Cleaning**

The raw data collected through web scraping contained several inconsistencies and errors that needed to be addressed before proceeding with the analysis. Data cleaning is essential to remove noise and inaccuracies, ensuring that the dataset is accurate, consistent, and ready for analysis. The following steps were taken to clean the data:

1. **Removal of Duplicates:**
   * **Identification of Duplicates:** The first step in the data cleaning process was to identify and remove duplicate entries. Duplicate data can occur when the same hotel is listed multiple times, possibly due to variations in room types or slight differences in hotel name formatting.
   * **Removal Process:** Using the Pandas library in Python, duplicate entries were detected based on unique identifiers such as hotel names and addresses. These duplicates were then removed, leaving only one entry per hotel. This step ensured that each hotel was represented only once in the dataset, preventing any skewing of the analysis.
2. **Handling Missing Values:**
   * **Identifying Missing Data:** Missing values can significantly impact the results of the analysis if not handled properly. In this project, missing data was primarily found in columns related to customer reviews and ratings, where some entries lacked complete information.
3. **Correction of Errors:**

**Data Accuracy:** During the data collection process, some errors were identified, such as incorrect price formatting or misaligned

* + columns. For example, some prices were mistakenly captured in different currencies or formats.
  + **Standardization:** These errors were corrected by standardizing the data. Prices were converted to a consistent currency format (Indian Rupees), and columns were realigned to ensure that all data points were correctly labelled and categorized.

1. **Outlier Detection:**
   * **Identifying Outliers:** Outliers, such as extremely high or low prices, were identified as they could distort the analysis. These outliers were detected using statistical methods like the interquartile range (IQR) and Z-score analysis.
   * **Handling Outliers:** Depending on the nature of the outlier, some were retained if they provided meaningful insights, while others were adjusted or removed if they were determined to be errors or anomalies.

**CHAPTER 4**

**DATA ANALYSIS AND VISUALIZATION**

**4.1 Introduction**

This chapter presents the analysis and visualization of engineering education data in India, focusing on key aspects such as college distribution, fee structures, ratings, and course offerings. The study provides insights into state-wise and city-wise college presence, variations in tuition fees, and institutional ratings, enabling a better understanding of the educational landscape.

**4.2 Engineering Education Landscape: State-Wise Distribution**

The first visualization presents an overview of engineering colleges across India. It includes filters for college names and locations, enabling users to analyze specific institutions and regions. The primary metrics showcased in the dashboard include:

**4.2.1 Key Metrics**

* **Average Fee:** The average tuition fee across the analyzed colleges is ₹728.72K.
* **Average Rating:** The overall rating for the colleges is 2.73 on a scale of 5.
* **Top 10 Colleges by Courses Offered:**
  + *William Carey University* offers the highest number of courses (18).
  + *Welingkar Mumbai* follows with 4 courses.
  + Institutions such as *VVM's Govind Ramnath Kare College of Law, Waikhom Mani Girls College, and Western College of Commerce* offer 3 courses each.

**4.2.2 City-Wise College Distribution**

A bar chart in the dashboard visualizes the number of colleges in various cities:

* Bangalore has the highest number of institutions, followed by Shillong, Ranchi, and Imphal.
* Other major educational hubs include Mumbai, Pune, Ahmedabad, Bhopal, and Kochi.

**4.2.3 State-Wise College Distribution**

A geographical map representation illustrates the distribution of colleges across different states, highlighting key educational centers such as Ahmedabad, Bangalore, and Shillong**.**

**4.3 Fee Structure & Cost Analysis Dashboard**

The second visualization focuses on the financial aspects of engineering education, specifically in Andhra Pradesh. The analysis covers:

**4.3.1 Fee Range Analysis**

* The minimum fee for engineering education starts at ₹216.47K, while the maximum fee extends up to ₹728.72K.
* This indicates a broad variation in tuition costs, depending on the institution and location.

**4.3.2 Top Colleges with Fee Details**

The table in the dashboard lists 14 top colleges along with their respective fee structures:

* *Western College of Commerce & Business Management:* ₹217,000 - ₹1,080,000
* *Xavier Institute of Social Service:* ₹216,000
* *VIT Pune:* ₹211,000 - ₹254,000
* *XLRI Jamshedpur:* ₹180,000
* Other institutions like *Welingkar Mumbai, VVM’s Govind Ramnath Kare College, VNIT Nagpur, XIME Bangalore, and XIME Kochi* have varied fee structures ranging from ₹1,000 to over ₹1,600,000.

**4.3.3 Rating Distribution by Type**

A pie chart represents the breakdown of average ratings based on institution type:

* Public Colleges: Average rating of 2.75.
* Private Colleges: Average rating of 2.67.
* Public institutions have a slightly higher rating than private ones, indicating relatively better perception among students and stakeholders.

**4.4 Conclusion**

The analysis of engineering education in India reveals the following insights:

* Bangalore remains the top educational hub with the highest number of colleges.
* William Carey University offers the most extensive range of courses.
* Tuition fees vary significantly, with some institutions charging over ₹1 million.
* Public colleges generally have better ratings than private ones.

These insights can help students, educators, and policymakers make informed decisions about engineering education in India.

**CHAPTER 5**

**TOOLS AND TECHNOLOGIES**

**Software Requirement:-**

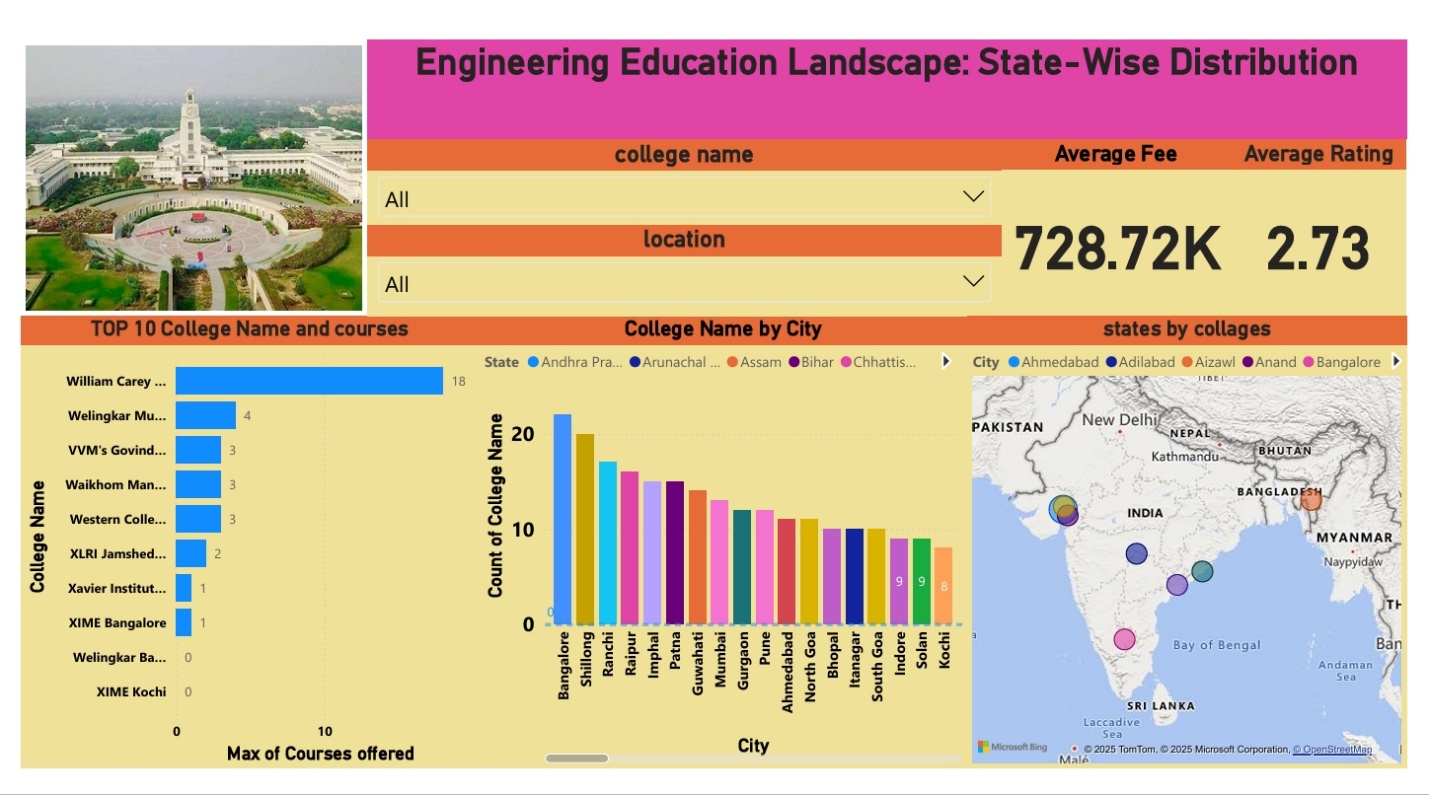
* Python(Jupyter notebook)
* Powerbi
* Excel

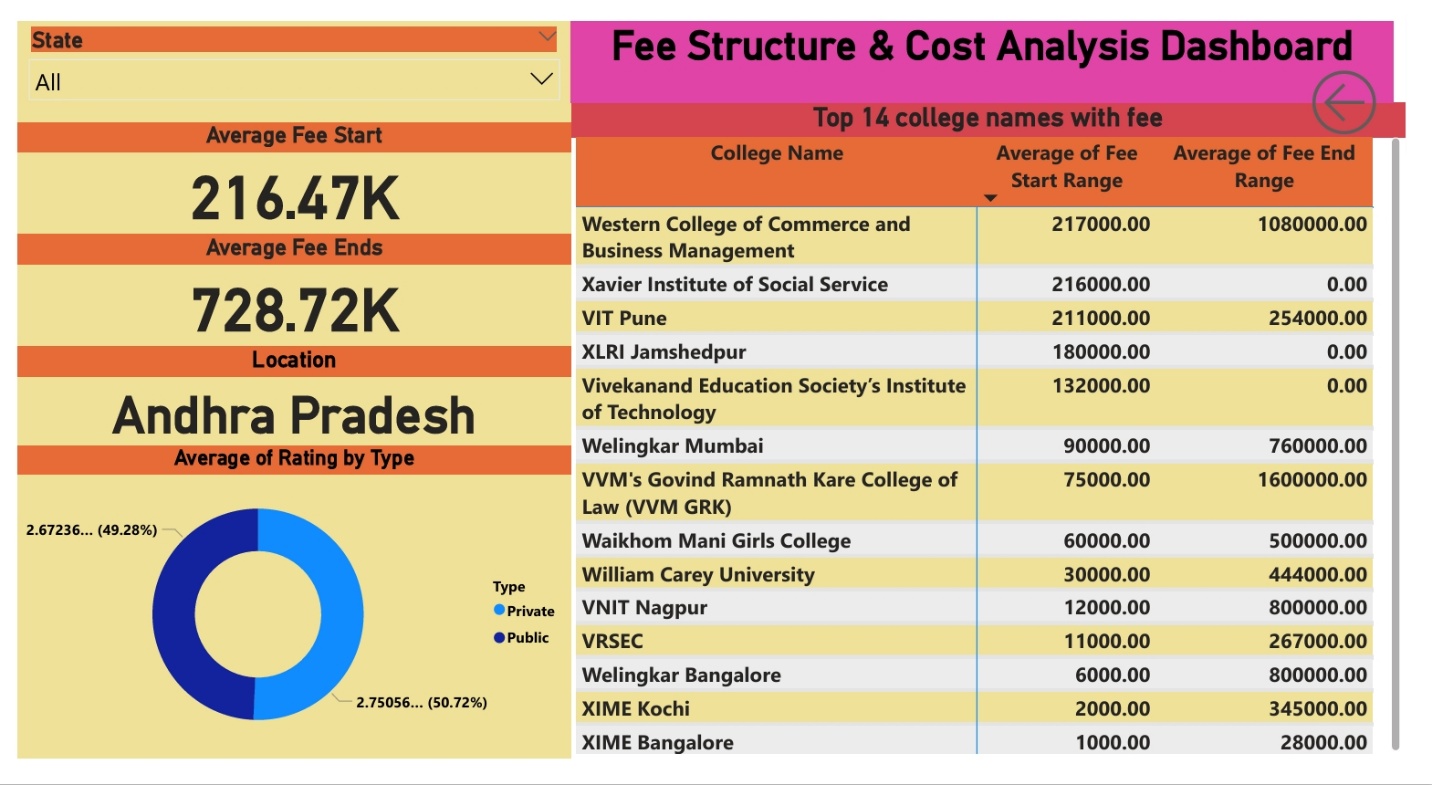
**Main Software Libraries:-**

* Pandas
* Numpy
* Beautiful Soups
* Requests
* CSV

**CHAPTER 6**

**DASHBOARDS**





**CHAPTER 7**

**CONCLUSION**

**Primary Objective**

The goal of this analysis was to utilize data analytics to extract meaningful insights from engineering education data in India. By examining college distributions, tuition fees, course offerings, and institutional ratings, the study provides a comprehensive view of trends shaping the engineering education landscape.

**Key Findings**

**1. College Distribution and Availability**

* Bangalore has the highest concentration of engineering institutions, followed by Shillong, Ranchi, and Imphal.
* Other major educational hubs include Mumbai, Pune, Ahmedabad, and Bhopal.
* A geographical analysis shows that cities like Ahmedabad, Bangalore, and Shillong serve as key engineering education centers.

**2. Course and Specialization Trends**

* William Carey University offers the highest number of courses (18), followed by Welingkar Mumbai (4).
* Several institutions, including VVM’s Govind Ramnath Kare College of Law, Waikhom Mani Girls College, and Western College of Commerce, offer 3 courses each.

**3. Fee Structure and Affordability**

* The average tuition fee across institutions is **₹728.72K**.
* The minimum fee starts at **₹216.47K**, while the highest tuition fee reaches **₹1,600,000**, indicating a broad cost variation.
* Western College of Commerce & Business Management has one of the highest tuition fee ranges, from **₹217,000 to ₹1,080,000**.
* William Carey University has one of the lowest tuition fees, starting at **₹30,000**, while institutions like Welingkar Mumbai and VNIT Nagpur charge significantly more.

**4. Institutional Ratings**

* The overall average rating of colleges is **2.73 out of 5**.
* Public institutions have a slightly higher rating (**2.75**) compared to private institutions (**2.67**), indicating a marginally better perception of government colleges.

**Implications for Stakeholders**

**Students and Parents**

* The data helps students and parents make informed decisions by evaluating institutions based on affordability, infrastructure, and available courses.
* Students interested in modern technological fields like Artificial Intelligence, Machine Learning, and Data Science should focus on institutions that offer specialized programs.

**Educational Institutions**

* Government colleges should consider expanding their course offerings to match the evolving industry landscape.
* Private institutions could improve affordability and financial aid options to cater to a more diverse student population.

**Policy Makers**

* Policies should be introduced to bridge the affordability gap between government and private institutions.
* Increased investment in research facilities and faculty development is essential to improving educational quality.

**Summary**

This analysis presents a data-driven approach to understanding engineering education in India. Insights from tuition fees, college distribution, and institutional ratings provide key takeaways for students, educators, and policymakers. Enhancing infrastructure, ensuring affordability, and aligning course offerings with industry needs will help create a more inclusive and future-ready higher education ecosystem.