Conducting exploratory data analysis (EDA) of engineering colleges and students from various branches in Maharashtra

EPIC 1: Planning and Data Collection

Week 1: Define Objectives and Scope

- Researching on the engineering education in Maharashtra
- Project progress flow, user flow, data flow diagrams and features list.

Week 2: Data Collection

- Identify data sources including government databases, college websites, and surveys.
- Collect relevant data such as college information (location, accreditation, infrastructure), student demographics, academic performance, and placement records.

Week 3: Data Cleaning

- Perform data cleaning tasks such as handling missing values, removing duplicates, and standardizing data formats.
- Validate the integrity of the data and ensure consistency across different sources.

Week 4: Exploratory Data Analysis (Part 1)

- Conduct initial exploratory analysis to understand the distributions and patterns in the data.
- Generate descriptive statistics, histograms, and box plots to summarize the data.
- Identify any outliers or anomalies that may require further investigation.

EPIC 2: Analysis and Reporting

Week 1: Exploratory Data Analysis (Part 2)

- Explore relationships between variables using correlation analysis and scatter plots.
- Conduct segmentations based on demographic factors or college attributes to uncover insights.

Week 2: Advanced Analysis

- Apply advanced techniques such as clustering or principal component analysis (PCA) to identify patterns or groupings within the data.
- Perform hypothesis testing to validate any observed trends or relationships.

Week 3: Visualization and Interpretation

- Create visualizations such as heatmaps, treemaps, and interactive dashboards to communicate key findings.
- Interpret the results of the analysis in the context of the project objectives and provide actionable insights.

Week 4: Documentation and Presentation

- Document the analysis process, including data sources, methodologies, and key findings.
- Prepare a comprehensive report summarizing the insights gained from the EDA.
- Present the findings to stakeholders, highlighting actionable recommendations for improving the quality of education in engineering colleges in Maharashtra.

Database architecture to collect data on various parameters, attributes, and variables for feature extraction in the education domain focusing on engineering colleges and students in Maharashtra:

1. College Information Table:

- **college_id (Primary Key):** Unique identifier for each college.
- **college_name:** Name of the college.
- **location:** Location of the college (city or district).
- **accreditation_status:** Accreditation status of the college (e.g., NAAC, NBA).
- **infrastructure_rating:** Rating of the college infrastructure.
- **courses_offered:** List of engineering branches offered by the college.
- **fees:** Fee structure for different courses.
- **placement_records:** Placement statistics of the college.

2. Student Information Table:

- **student_id (Primary Key):** Unique identifier for each student.
- **college_id (Foreign Key):** Identifier linking to the college they belong to.
- **branch:** Engineering branch of the student (e.g., Computer Science, Mechanical).
- **year_of_study:** Year of study of the student (e.g., 1st year, 2nd year).
- **gender:** Gender of the student.
- **age:** Age of the student.
- **academic_performance:** Academic performance metrics (e.g., GPA, percentage).
- **extracurricular_activities:** Participation in extracurricular activities.
- **internship_experience:** Details of any internship experience.
- **placement_status:** Placement status (e.g., placed, not placed).

3. Demographic Information Table:

- **student_id (Primary Key):** Unique identifier for each student.
- **college_id (Foreign Key):** Identifier linking to the college they belong to.
- **city:** City of residence of the student.
- **state:** State of residence of the student.
- **parental_education:** Education level of the student's parents.
- **family_income:** Family income level.
- **educational_background:** Previous educational background of the student.
- **ethnicity:** Ethnicity or community background of the student.
- **distance_from_college:** Distance of student's residence from college.

4. Course Curriculum Table:

- **course_id (Primary Key):** Unique identifier for each course.
- **course_name:** Name of the course.
- **course_description:** Description of the course curriculum.
- **credits:** Credits associated with the course.
- **syllabus:** Detailed syllabus of the course.
- **recommended_books:** List of recommended books for the course.

5. Placement Records Table:

- **college_id (Primary Key):** Unique identifier for each college.
- **branch:** Engineering branch.
- **year:** Year of placement.
- **average_salary:** Average salary offered.
- **placement_percentage:** Percentage of students placed.
- **top_recruiters:** List of top recruiters.