INTRODUCTION TO DATA SCIENCE
MILESTONE 1

Title: Student Job and Course Recommendation System

Objective

This project aims to develop a personalized recommendation system for students, helping them find suitable job opportunities and courses based on their individual profiles. The system will suggest jobs by analyzing factors like a student's academic branch, 12th-grade percentage, skills, and career goals. It will also recommend courses that help bridge skill gaps or provide additional learning to help students achieve their career aspirations.

Tool and Technology Stack

- Tech Stack: Python, HTML, CSS, Flask, JavaScript, TypeScript.
- Programming Language: Python
- Libraries: Pandas, Matplotlib, Seaborn for data analysis, Scikit-learn
- Tool Type: Classification models such as Support Vector Classifier, Decision Tree Classifier, Random Forest Classifier (Ensemble)
- Platform: Visual Studio Code

Data to Be Used

- Datasets: We are using three datasets for this project Studentdata.csv, Online_Courses.csv, Job_Postings.csv
- Features in each dataset used:
  - Student data: Branch, Percentage_12th, Skills, Career_Goal.
  - Courses data: Title, Category, Course Type, Skills.
  - Job data: Job Posting Date, Job Title, Job Title Full, Job Skills, Job Location, Company Name
- Data Source: Kaggle
- Data Structure: CSV

Project Timeline

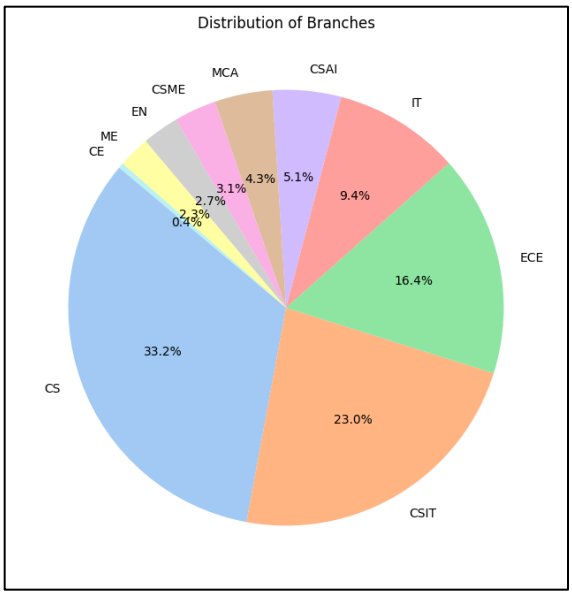
Task	Date
Task 1: Data Collection & Preprocessing	02/07/2025 - 02/15/2025
Task 2: Exploratory Data Analysis (EDA)	02/16/2025 - 02/20/2025
Task 3: Model Selection & Training	02/21/2025 - 03/16/2025
Task 4: Model Evaluation & Fine-tuning	03/24/2025 - 04/10/2025
Task 5: Final Report and Analysis	04/11/2025 - 04/21/2025

Data Preprocessing: Check Preprocess.ipynb in Scripts Folder to see how we preprocessed the data.
Exploratory Data Analysis: Check EDA.ipynb in Scripts Folder for a better understanding.

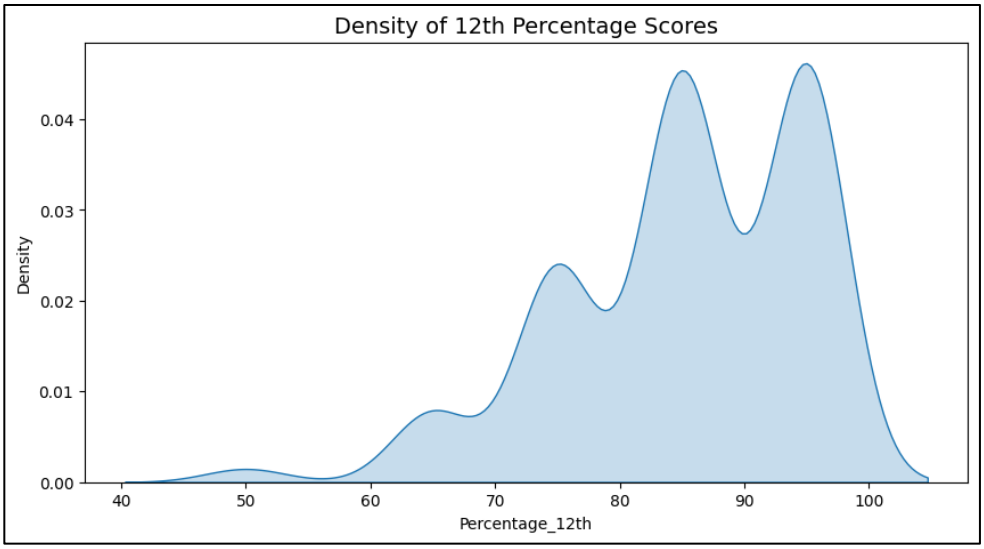
Visualizations from Exploratory Data Analysis

- **Student Data**

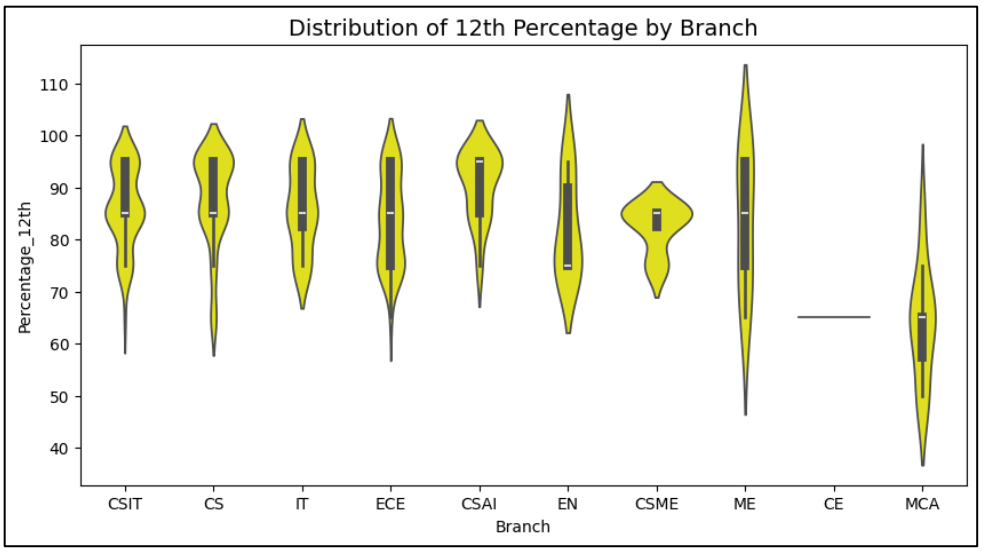
  1. Distribution of Branches enrolled



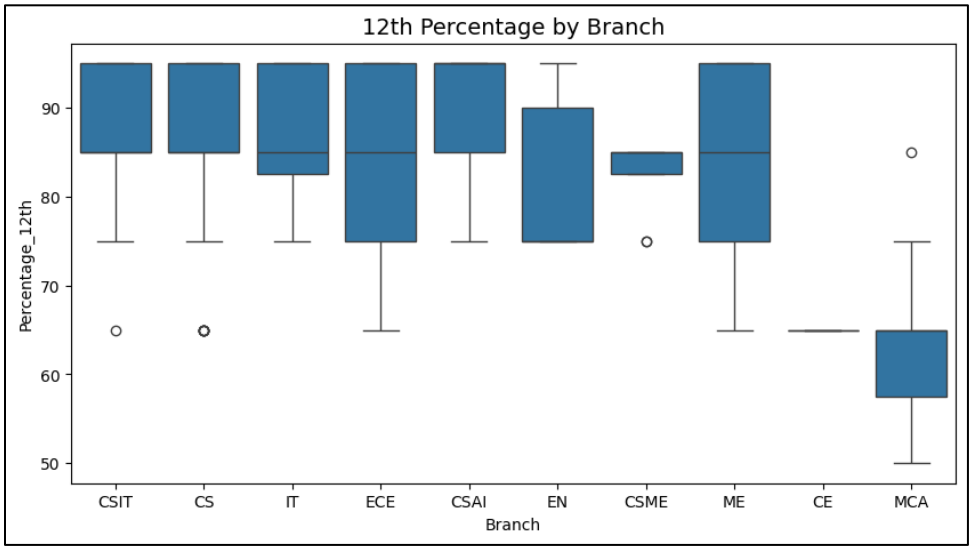
2. Density of 12th Percentage Scores



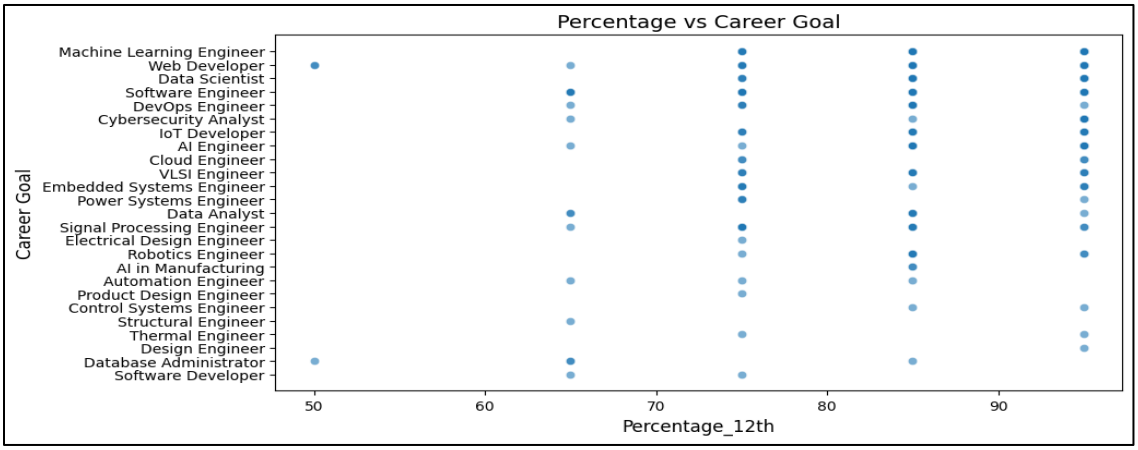
3. Violin Plot for 12th Percentage



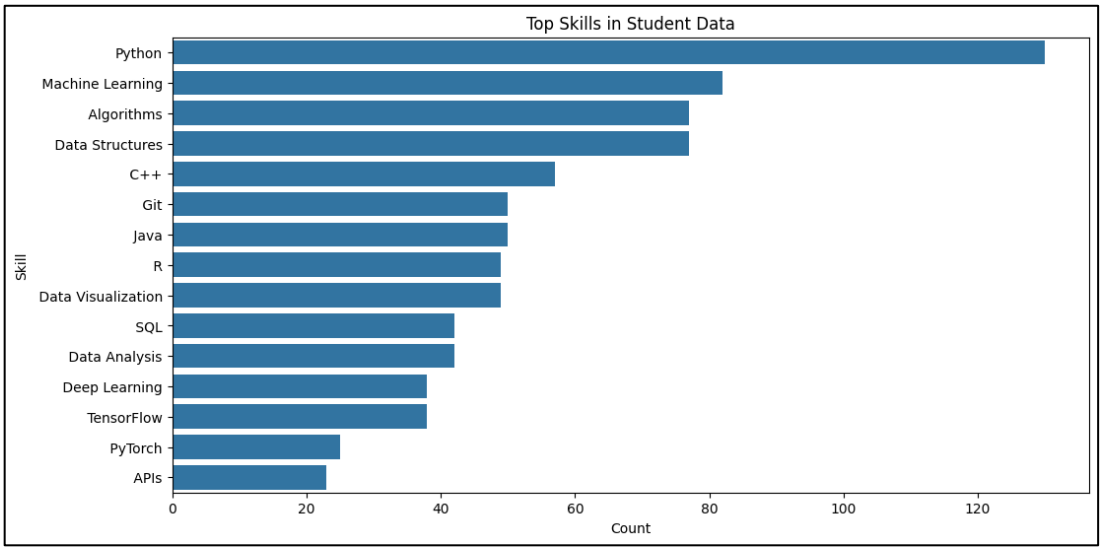
4. Box plot for 12th Percentage by Branch



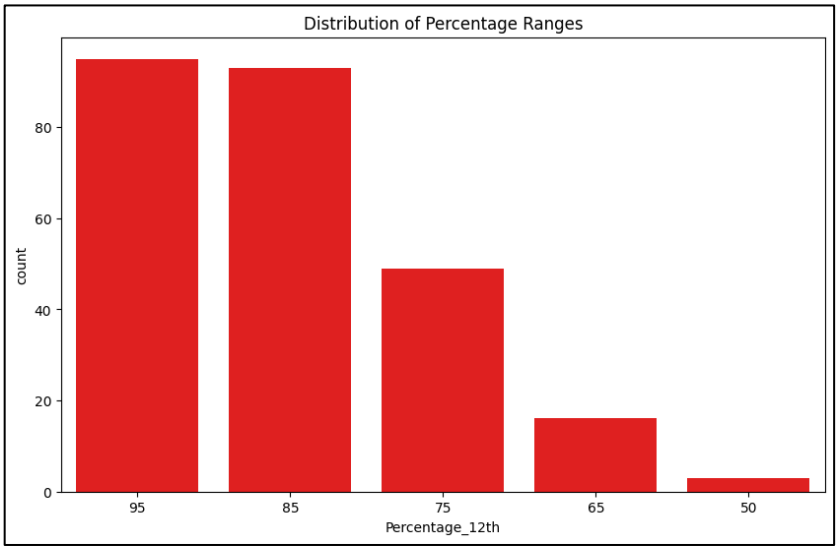
5. Scatter plot for Percentage vs Career Goal



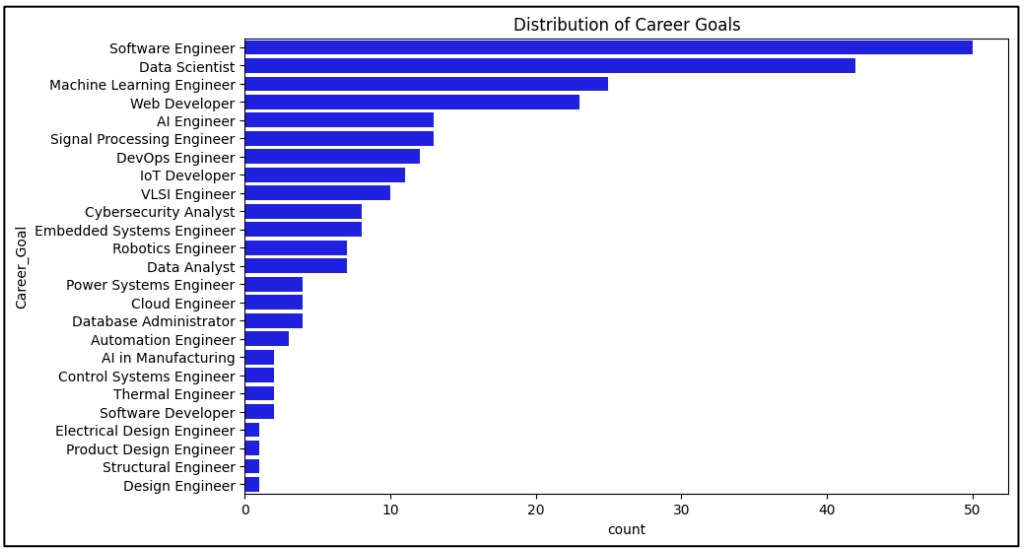
6. Bar plot for Skills



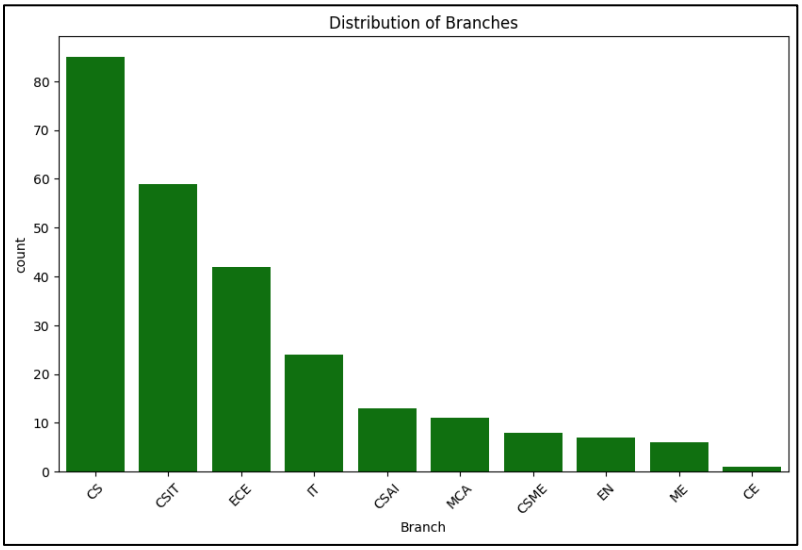
7. Count plot for 12th Percentage



8. Count plot for Career goals

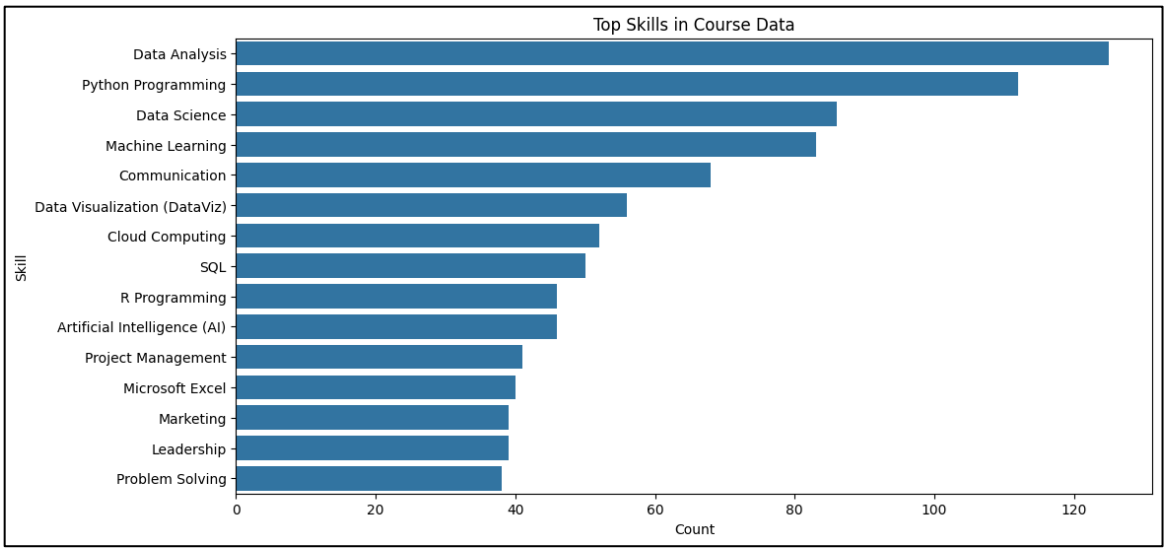


9. Count plot for distribution of Branches

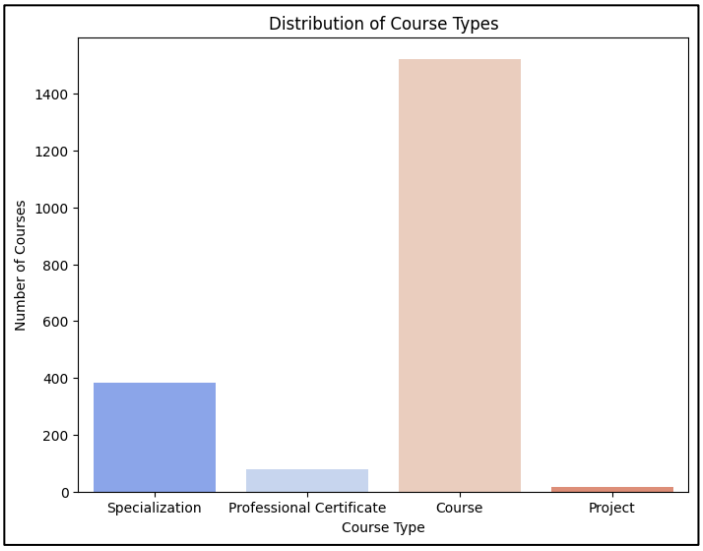


- **Course Data**

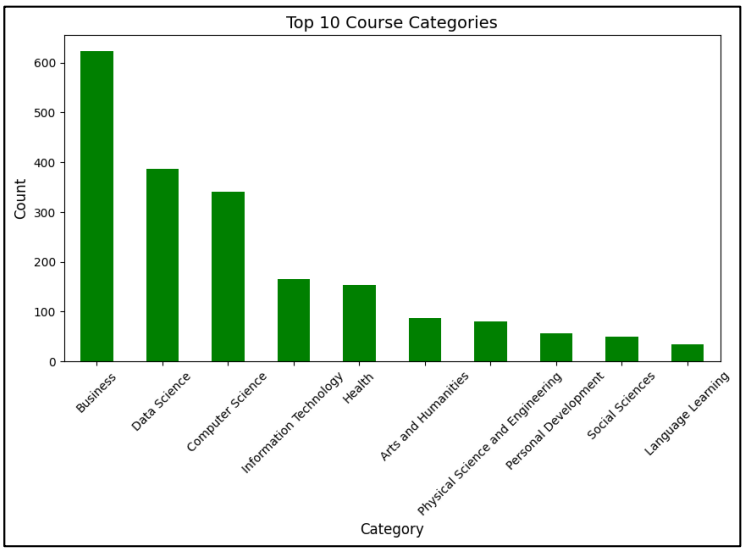
1. Bar plot for Top Skills



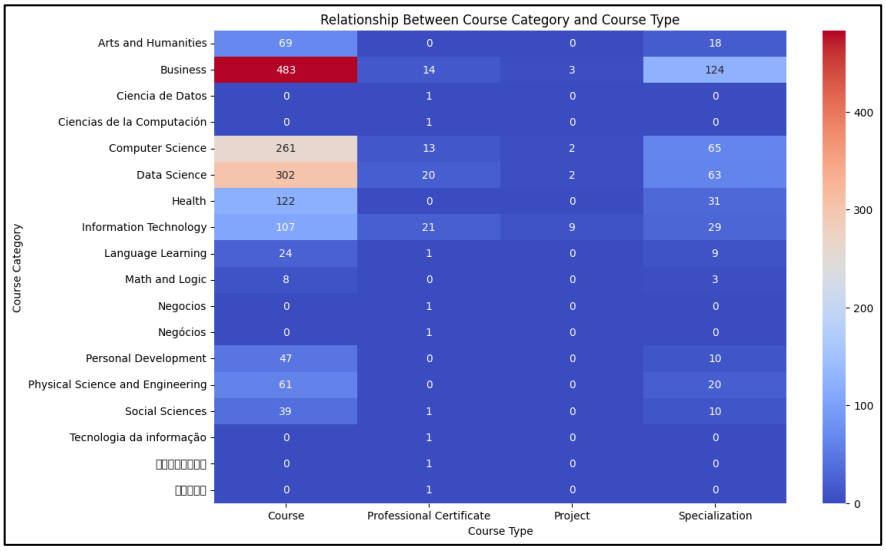
2. Count plot for Course types



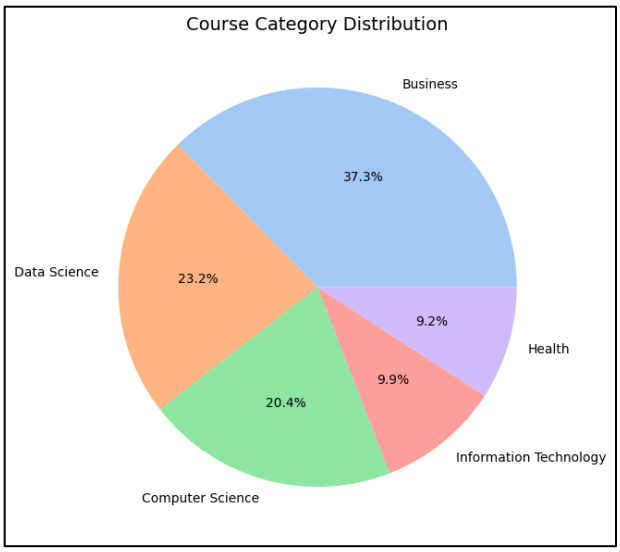
3. Bar plot for Course Categories



4. Heatmap for correlation between Course Category and Type

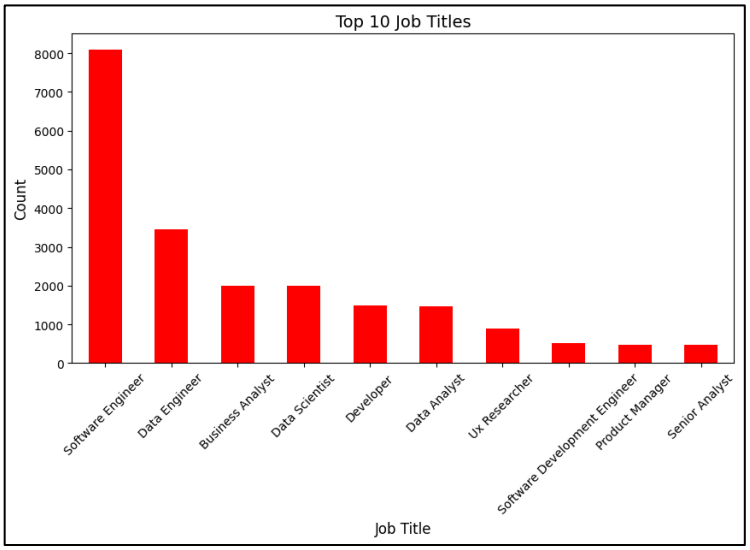


5. Pie chart on Course Category

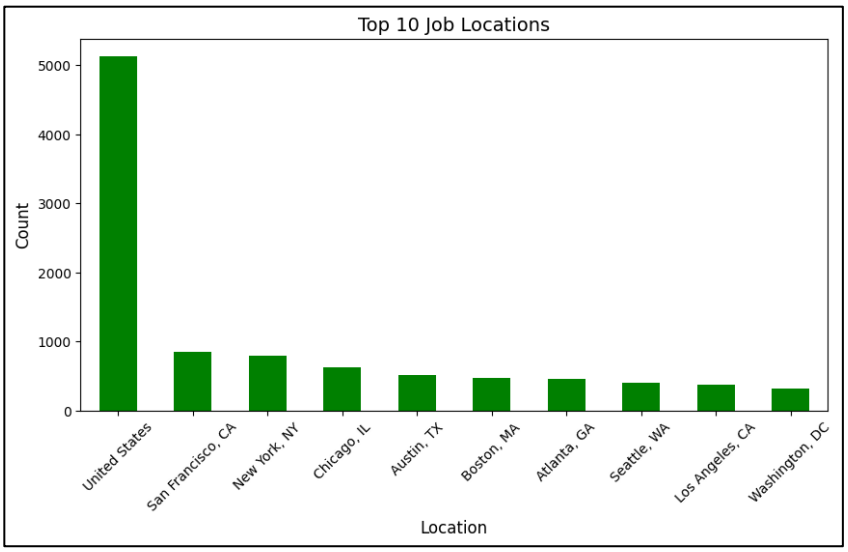


- **Job Data**

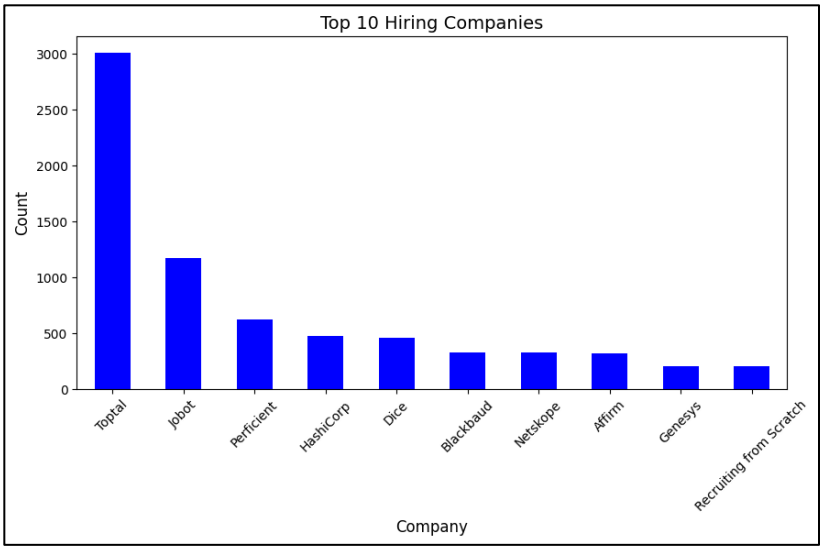
1. Bar plot on Top 10 Job Titles



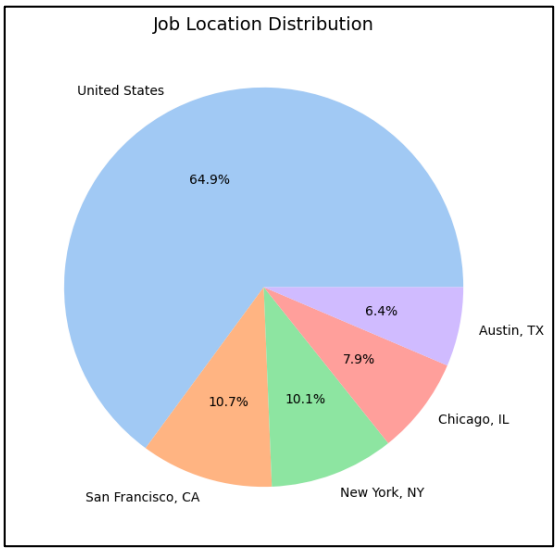
2. Bar plot on Top 10 Job locations



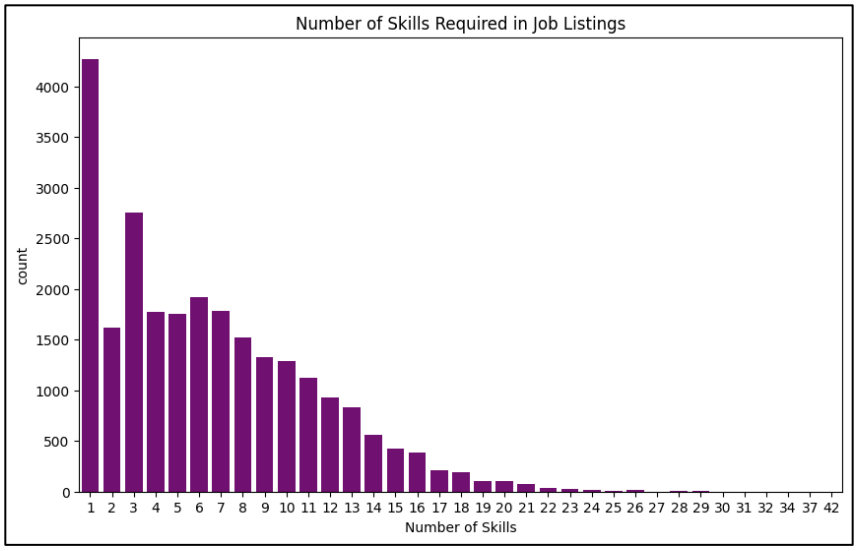
3. Bar plot on Top 10 Hiring Companies



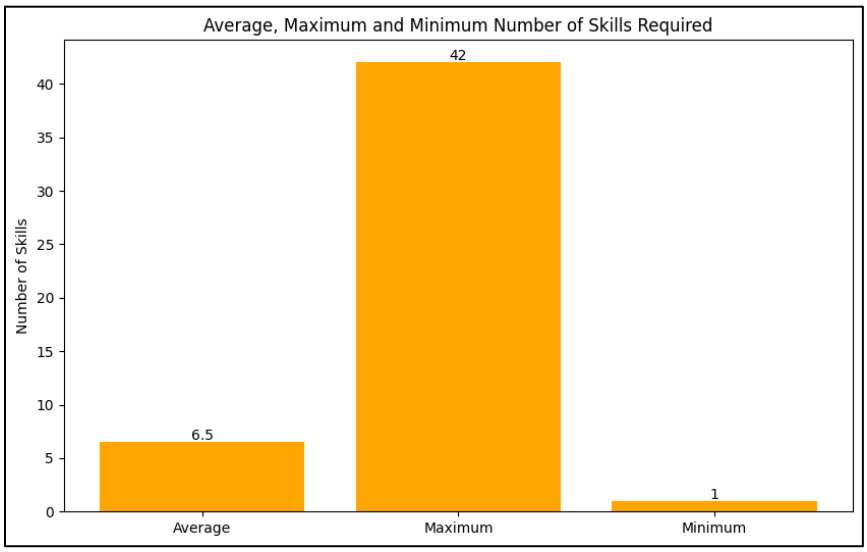
4. Pie chart on Job location Distribution



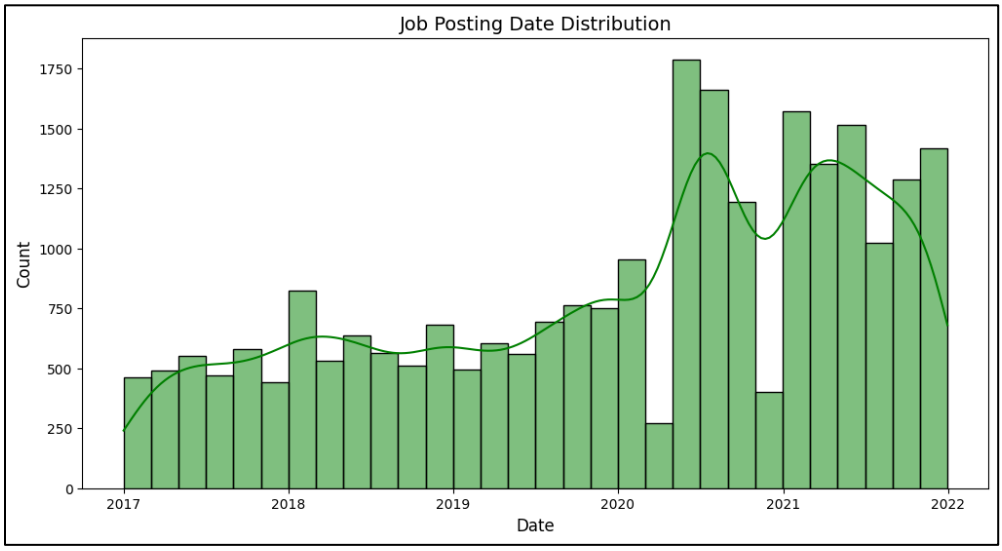
5. Count plot on Number of Skills



6. Bar plot on Avg, Max and Min Skills Required



7. Histogram on Job Postings



EDA Observations

- **Observation 1**: High School Percentage Distribution
  Insight: Many students have high school percentages between 75% and 95%.

- **Observation 2:** Career Goal Trends
  Insight: The most common career aspirations include Machine Learning Engineer, Web Developer, and Data Scientist.

- **Observation 3:** Top Student Skills
  Insight: The most common skills include Python, Java, SQL, HTML, CSS, JavaScript, Machine Learning, and Deep Learning.

- **Observation 4:** Job Titles with Highest Demand

Insight: The most frequently posted job titles include Software Engineer, Data Engineer, Business Analyst, and Data Scientist.

- **Observation 5:** Type of Courses offered
  Insight: The Course type 'Course' is mostly enrolled by students compared to course types like Specialization, Professional Certificate and Project.

- **Observation 6:** Top Skills in Courses offered
  Insight: Courses with skills like Data Analysis, Python Programming, Data Science, Machine Learning are more in number.

- **Observation 7:** Top Course Categories
  Insight: Courses related to Business, Data Science lead the chart. While courses related to Social Science and Language Learning are at the bottom of the chart.

- **Observation 8:** Top Hiring Companies
  Insight: Companies like Toptal, Jobot, Perficient are leading the charts in hiring.

- **Observation 10:** Job Locations
  Insight: Most job postings are in the United States, with states like California, New York, Chicago and Texas leading.