**Prog6280a**

**(Jobs in Data Science)**

**Program Description:** This lab program will guide you through using Pandas and Matplotlib in Python (or equivalent dataframe/plotting libraries in R, Julia, etc.) to analyze and visualize the “[Jobs and Salaries in Data Science](https://www.kaggle.com/datasets/hummaamqaasim/jobs-in-data)” dataset. The goal is to help you learn data wrangling and visualization techniques while uncovering interesting insights from the dataset.

1. Introduction to Dataframes and Plots

* **Pandas**: Basics of data manipulation (reading CSV, data cleaning, filtering, grouping).
* **Matplotlib**: Fundamentals of data visualization (basic plots, customization, saving plots).

2. Dataset Overview

* Introduction to the "Jobs and Salaries in Data Science" dataset.
* Loading the dataset using Pandas.
* Basic data exploration (head, describe, info).

3. Data Wrangling Exercises

* Handling missing values (if any).
* Converting data types (e.g., converting work\_year to datetime format).
* Normalizing salary data for different currencies (if applicable).

4. Data Visualization Tasks

Each task should guide you to tell a “story” with the data using various types of plots:

**Task 1: Salary Trends Over Time**

* Plot Type: Line plot.
* Objective: Show how average salaries in data science have changed over the years.
* Columns Used: work\_year, salary\_in\_usd.

**Task 2: Salary Distribution by Job Title**

* Plot Type: Box plot.
* Objective: Compare the salary ranges and medians across different job titles.
* Columns Used: job\_title, salary\_in\_usd.

**Task 3: Experience Level vs. Salary**

* Plot Type: Bar plot.
* Objective: Analyze how salaries vary with experience level.
* Columns Used: experience\_level, salary\_in\_usd.

**Task 4: Geographic Salary Distribution**

* Plot Type: Histogram (horizontal; or geospatial plot if GIS skills are included).
* Objective: Understand the distribution of salaries across different countries.
* Columns Used: employee\_residence, salary\_in\_usd.

**Task 5: Company Size and Salary Structure**

* Plot Type: Box or Violin plot.
* Objective: Explore how company size relates to salary distributions.
* Columns Used: company\_size, salary\_in\_usd.

**Task 6: Impact of Work Setting on Salary**

* Plot Type: Scatter plot.
* Objective: Investigate if there's a correlation between work setting and salary.
* Columns Used: work\_setting, salary\_in\_usd.

5. Analysis and Interpretation

* Interpret the plots (one sentence each).
* Discuss factors influencing salary trends in the data science field.

6. Additional Challenges (Optional)

* Multi-variable analysis (e.g., job title vs. experience level vs. salary).
* Time-series analysis for specific job titles.
* Advanced data visualizations (interactive plots, heatmaps).

**Data Location:** jobs\_in\_data\_science.csv

**Sample Output (not including interpretations):**

Task 1: Salary Trends Over Time

A graph with a line going up

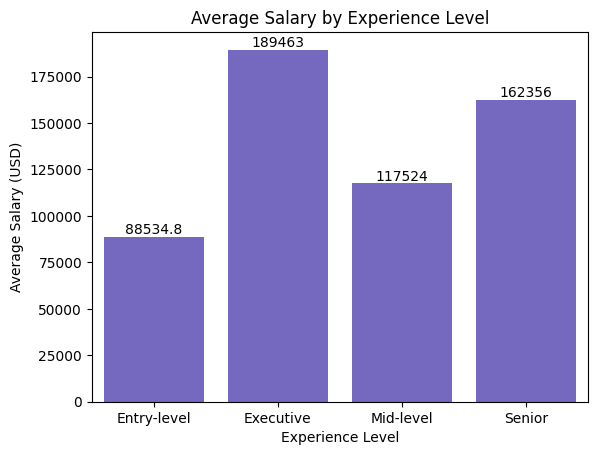
Description automatically generated

Task 2: Salary Distribution by Job Title

A graph with numbers and lines

Description automatically generated

Task 3: Experience Level vs. Salary



Task 4: Geographic Salary Distribution

data:image/png;base64,

Task 5: Company Size and Salary Structure

A diagram of a company size

Description automatically generated