Project Scoping

Name

Analyzing In-Office Salary Trends in AI, Data, and Machine Learning Roles (by André Jardim)

Business Problem

The AI, data, and machine learning industries are experiencing significant growth as more organizations embrace data-driven strategies and automation. However, with a strong preference for in-office roles within these fields, understanding compensation trends is critical to attract and retain skilled talent. This project will analyze salary trends for in-office positions across AI, data, and machine learning roles, focusing on variations by job title, experience level, company size, company location, and changes over time from 2020 to 2024. With a majority of entries originating from the U.S., this analysis will explore how U.S. salary trends align with or diverge from international patterns. Additionally, the dataset lacks a primary key, resulting in duplicate entries that must be carefully managed to preserve data integrity.

Business Impact

By examining in-office salary trends in these fast-growing fields, the analysis will provide companies with benchmarks for competitive compensation strategies, aiding in talent acquisition and retention. This data-driven approach will also benefit industry professionals, offering clear benchmarks that support informed career decisions and negotiations. For policymakers and analysts, the findings can illuminate regional disparities in salary and the relative competitiveness of the U.S. against other regions.

Dataset(s)

- **Source:** AIJobs.net Salary Dataset (2020–2024)
- **Details:** The dataset includes over 59,000 records covering job title, experience level, company size, company location, and employment type.
- **Strengths:** The dataset provides detailed salary information across key fields such as work year, job title, experience level, company size, company location, and employment type.
- **Weaknesses:** There is a high occurrence of duplicate records due to the absence of a primary key, necessitating strategies for managing these duplicates without losing valuable insights. Additionally, the data skews heavily toward U.S. entries, which may impact global comparisons.

Methods

Variables and Comparisons:

- **Relationships between:**
    - Experience level and salary
    - Company size and salary
    - Company location and salary
    - Job title and salary
    - Salary variations over time

Dashboard

Overview:

- Develop an interactive dashboard to visualize in-office trends with filters for company size, company location, job title, and experience level. The dashboard will allow users to dynamically explore how these factors influence salary trends over time.

Mockup:

- **The dashboard will include the following visualizations:**
  - **Bar Chart:** Salary comparison by experience level
  - **Map:** Salary distribution by company location
  - **Bar Chart:** Salary breakdown by company size
  - **Bar Chart:** Salary differences across job titles
  - **Line Chart:** Salary trends over time
- Additionally, scorecards will display key metrics such as average salary by job title and experience level.

Milestones

Data Cleaning and Duplicate Management:

- Remove or handle duplicates appropriately.
- Filter for in-office and full-time positions.

Outlier Detection and Removal:

- Conduct statistical outlier analysis to ensure outliers do not distort average salaries.

Exploratory Data Analysis (EDA):

- Explore salary distributions and identify U.S. versus global trends.

Datafolio and Dashboard Creation:

- Develop a user-friendly dashboard to showcase the findings.

Final Report:

- Summarize insights, recommendations, and implications for companies and professionals in AI, data, and machine learning fields.

Timeline

Week	Tasks
Week 1	Data cleaning, duplicate handling, preparation, and initial descriptive statistics. Outlier detection and Exploratory Data Analysis (EDA).
Week 2	EDA. Begin development of the Datafolio and interactive dashboard.
Week 3	Finalize the Datafolio and dashboard. Prepare the final report, including analysis and recommendations.