Project Documentation

Glassdoor Insights: Data Science Job Analysis

By: Priscilla Philby Oommen

I. Project Overview

This project analyzes data science job postings from Glassdoor using Excel and Power BI. The goal is to uncover trends in job roles, company information, salary estimates, industry distribution, and more.

II. Tools Used

- Microsoft Excel
- Power BI

III. Dataset

- Source: The csv file “Uncleaned_DS_jobs” is a dataset of data science job posts in glassdoor. The data was scraped from glassdoor's website.

- Data contains:

Index: A unique identifier for each row or job posting in the dataset, typically used for reference or row tracking
Job Title: Title of the job posting
Salary Estimation: Salary range for that particular job
Job Description: This contains the full description of that job
Rating: Rating of that post
Company Name: Name of company
Location: The location of the specific job posting (where the job is based).
Headquarters: The main office address or HQ of the company (not always where the job is).
Size: Total employee in that company
Founded: Year in which the company was established
Type of ownership: Describes the company type i.e non-profit/public/private firm etc
Industry: The specific field or business category the company operates in (e.g., IT Services, Biotech)
Sector: The broader economic segment to which the company belongs (e.g., Information Technology, Healthcare)
Revenue: Total revenue of the company
Competitors: Names of rival companies competing in the same industry or market

Company Category: Company is categorized as Micro, Small, Medium, Large, Enterprise based on Size column
MIN Salary, MAX Salary, AVG Salary: Refers to the minimum, maximum and average salary for that post
Senior Role: Categorizes Job roles as Senior and Not Senior based on Job title.

IV. Steps Followed

- Cleaned data in Excel (e.g., removed blanks, formatted columns)
- Used Excel formulas and Power Query for basic analysis
- Imported cleaned data into Power BI
- Built dashboards using charts, slicers, and KPIs

1. Remove Duplicates
2. Check for missing values. Placeholders value -1 is given for missing values in the dataset. Another Placeholder value "Unknown" is used in categorical and text columns. 0 is used in numerical columns since some of the columns are used for calculations.
3. Columns like Job Title which have many extra values were cleaned using the clean function.
4. Columns like Location and Headquarters were cleaned using Trim function.
5. Extra characters like dollar sign($), billion, million, K(Thousand) were removed using the Find and Replace option in Excel and Replace option in Power Query.
6. In Power Query, the Categorical/Textual column Salary estimate column has been split into numerical columns Min Salary and Max Salary. Then the average salary is calculated.
7. The Revenue column is cleaned by making it consistent and similar to the original Salary Estimate column.
8. The datatype of “Founded” column data type changed to Whole Number.
9. Size column is cleaned by removing unnecessary words "employees". Categorical column "Company Category" is derived.
10. Imported the cleaned dataset into Power BI. Checked for preset relationships and data types.
11. Designed interactive visuals: bar charts, column charts, donut chart, ribbon chart, decomposition tree, treemap, gauge, and KPI cards.
12. Used custom tooltip page for detailed hover insights.
13. Added slicers for interactivity.
14. Designed a responsive, clean dashboard titled "Glassdoor Insights: Data Science Job Analysis Dashboard"

V. Key Insights

- Job Role Trends: Data Scientists make up the majority of postings, but roles like Data Engineers and ML Engineers are rapidly increasing in demand. Specialized roles like Data Architects and AI Engineers command the highest average salaries, reflecting niche skill requirements.

- Location Insights: Major tech hubs like San Francisco, New York,Seattle and Boston dominate job availability. Remote Jobs are increasing, showing industry shift toward flexible work.

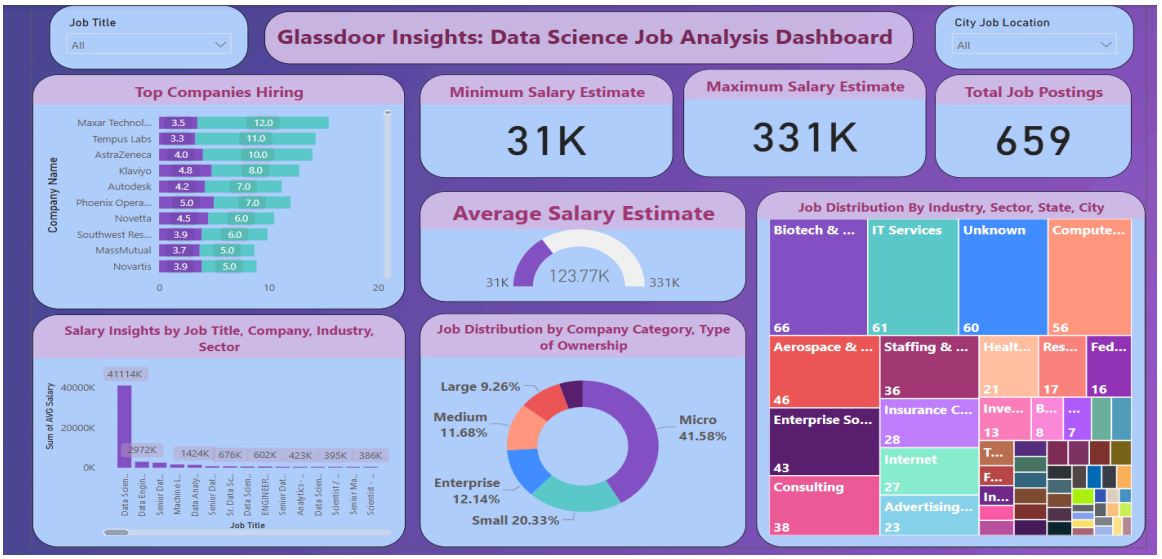
- Company Insights: Companies with higher ratings tend to offer competitive salaries and also have higher retention and competitive benefits. Startups may offer lower base salaries but mention more flexible skills and faster career growth opportunities. Large enterprises tend to offer higher salaries but also demand more experience.

- Salary Distribution: A wide salary range is observed even within the same job title, based on location, company size, and experience level. Entry-level roles have a narrower salary band, while senior roles vary greatly based on the employer. Senior roles are often tied to leadership or hybrid roles like “Lead Data Scientist” or “Principal ML Engineer.”

- Industry Insights: Top hiring industries are Tech & Software, Finance & Banking, Healthcare, E-commerce, etc. Industries like healthcare and finance often demand more regulatory or privacy-related experience (e.g., HIPAA, GDPR).

VI. Screenshots

Power BI Dashboard



VII. Files Included

- “DS_JOBS_CLEAN_DATASET.xlsx” – Cleaned data and basic analysis
- “DS_JOBS_DASHBOARD.pbix” – Power BI dashboard

VIII. How to Use

- Open “DS_JOBS_CLEAN_DATASET.xlsx” to view the cleaned data.
- Open “DS_JOBS_DASHBOARD.pbix” in Power BI Desktop to explore the visuals.

—----------