

Project Title
Data Science Job Salaries Analysis

*Report submitted in partial fulfillment
For the award of the degree of*

Bachelor of Technology

in

Electronics and Computer Engineering

By

Vaishnavi Jawalkar (Roll No. 1)

Renuka Kamble (Roll No. 2)

Shubham Khanapure (Roll No. 3)

Under the Guidance of Mrs.P.V. Katare



Department of Electronics Engineering
Walchand Institute of Technology, Solapur
(An Autonomous Institute)

Affiliated To
Punyashlok Ahilyadevi Holkar Solapur University, Solapur

Academic Year 2025-26

Phase 1— Project Initiation

Project Title:

Data Science Job Salaries Analysis Dashboard

Objective:

The primary objective of this project is to create an interactive web-based analytical dashboard that allows users to explore, filter, and visualize salary data for various data science job roles across different countries, experience levels, and remote work ratios. The dashboard helps in understanding salary distribution, identifying key factors influencing compensation, and generating career recommendations for aspiring professionals in the field of data science.

Scope:

- The project provides an interactive Streamlit dashboard accessible via any browser.
- It supports custom filters for detailed analysis:
 - Country/Region (employee_residence)
 - Job Role (job_title)
 - Experience Level (experience_level)
 - Remote Work Ratio (remote_ratio)
- It includes multiple graphical representations such as boxplots, bar charts, and summary tables to simplify data interpretation.
- Finally, it offers career insights by highlighting the top-paying job roles and experience levels globally.

Significance:

This dashboard is useful for:

- Students and professionals exploring salary expectations.
- Recruiters and companies analyzing market compensation.
- Researchers and analysts studying salary trends across demographics.

Technology Used:

Category	Tools/Technologies
Programming Language	Python
Data Handling	Pandas, NumPy
Visualization	Matplotlib, Seaborn
Dashboard Framework	Streamlit
Dataset	ds_salaries.csv
IDE / Platform	VS Code, Jupyter Notebook

Phase 2 — Data Exploration and Preprocessing

Dataset Overview:

The dataset `ds_salaries.csv` contains salary data for various data science-related jobs across the world. It includes multiple parameters such as:

- Job Title (`job_title`) — defines the professional role (e.g., Data Scientist, ML Engineer, Analyst).
- Experience Level (`experience_level`) — categorized as Entry (EN), Mid-level (MI), Senior (SE), and Expert (EX).
- Employment Type (`employment_type`) — such as full-time, part-time, or contract.
- Employee Residence (`employee_residence`) — indicates the country of residence.
- Remote Ratio (`remote_ratio`) — defines the level of remote work (0%, 50%, 100%).
- Salary in USD (`salary_in_usd`) — provides a standardized measure for comparing salaries globally.

Data Cleaning and Processing Steps:

1. Handling Missing Values:
The dataset was checked for missing or null entries. Minor inconsistencies were removed using Pandas functions like `dropna()` and `fillna()`.
2. Data Type Standardization:
Columns such as `salary_in_usd` were converted to numeric format for easier analysis.
3. Categorical Normalization:
Country names and experience levels were standardized for uniform representation across all visualizations.
4. Outlier Detection:
Using Seaborn's boxplot, unusually high or low salary entries were identified and analyzed to ensure realistic insights.
5. Data Validation:
Random samples were verified to confirm data correctness before visualization.

Exploratory Analysis Insights:

- There is a direct positive correlation between experience level and salary.
- US-based and Western European professionals earn significantly more than others.
- Fully remote employees often receive competitive salaries comparable to on-site employees, highlighting global hiring trends.
- Job roles like Machine Learning Engineer and Data Scientist consistently rank at the top in terms of compensation.

Phase 3 — Dashboard Design and Implementation

Framework Overview:

The dashboard was developed using Streamlit, a modern Python library for building data applications with minimal code. Streamlit automatically refreshes visualizations when users adjust filters, making it ideal for interactive analytics.

Dashboard Components:

1. Title and Description Section:
Displays the dashboard's title, description, and purpose using `st.title()` and `st.markdown()`.
2. Sidebar Filters:
Allows dynamic filtering by:
 - Country
 - Job Role
 - Experience Level
 - Remote Work RatioThese controls update the entire dashboard in real-time.
3. Data Summary Section:
Presents key statistics such as average, median, minimum, and maximum salary grouped by job role and country.
4. Visualization Section:
 - Boxplots: Salary distribution by experience level, job role, and remote ratio.
 - Bar Charts: Comparative salary analysis by job and by country.
 - Combination Graphs: Salary trends across multiple variables.
5. Career Recommendations:
The dashboard automatically lists the top 5 job roles and top 3 experience levels based on the filtered dataset.
6. Filtered Dataset Preview:
A data table allows users to view the specific subset they analyzed.

Backend Functionality:

The code uses Pandas for filtering and aggregation:

```
filtered_df = df[
    df['employee_residence'].isin(selected_countries) &
    df['job_title'].isin(selected_jobs) &
    df['experience_level'].isin(selected_experience) &
    df['remote_ratio'].isin(selected_remote)
]
```

Aggregations and averages are computed using `.groupby()` and `.mean()`.

Matplotlib and Seaborn generate plots that Streamlit renders directly with `st.pyplot()`.

Phase 4 — Visualization and Analysis

Analytical Insights and Interpretations:

1. Salary vs Job Role

- The highest average salaries were found in roles like **Machine Learning Engineer, Data Scientist, and AI Researcher**.
- Roles such as **Data Analyst** and **Data Engineer** showed stable but comparatively lower pay scales.

2. Salary vs Country

- Professionals residing in the **United States, Germany, and Canada** earn the highest salaries globally.
- Asian and South American countries showed lower salary averages, indicating regional pay disparities.

3. Salary vs Experience Level

- Entry-level professionals (EN) earned significantly less than senior (SE) or expert (EX) levels.
- Senior-level salaries were approximately **2.5x higher** than entry-level, proving the importance of skill advancement and experience accumulation.

4. Salary vs Remote Ratio

- Fully remote jobs often offered competitive pay, sometimes exceeding in-office roles, due to **global talent acquisition trends**.
- Hybrid (50%) remote roles also provided balanced salaries, indicating flexibility's growing role in compensation.

Graphical Representations:

1. **Boxplots** displayed salary variation across experience levels and remote ratios.
2. **Bar Graphs** compared average salaries between countries and job titles.
3. **Heatmaps (optional)** could be used to visualize correlation between multiple attributes.

Outcome of Phase 4:

This phase successfully revealed key relationships between job characteristics and salary patterns. It validated that factors like **location, experience, and remote flexibility** strongly influence compensation in the data science industry.

Phase 5 — Conclusion and Future Scope

The Data Science Job Salaries Analysis Dashboard successfully fulfills its objective of providing a comprehensive, interactive visualization platform for salary analysis. The insights derived from the dashboard empower professionals to make data-driven career decisions, identify high-paying roles, and understand market trends globally. The project demonstrates the powerful combination of Streamlit for web interactivity and Pandas/Seaborn for analytics, showcasing how data visualization can bridge the gap between raw data and decision-making.

Key Findings:

- Job roles and experience levels are primary determinants of salary.
- Country and remote ratio also play significant roles in compensation.
- Data visualization helps simplify complex information for easy understanding.

Recommendations:

1. Professionals should pursue advanced technical roles like ML Engineer or AI Researcher for higher pay.
2. Continuous upskilling and participation in real-world projects enhance experience levels.
3. Considering remote work opportunities can improve salary potential and work flexibility.

Future Scope:

- Integration of time-series data to track salary trends over multiple years.
- Addition of interactive maps to display salary heat zones.
- Use of Machine Learning models to predict future salary trends based on current data.
- Deployment of the dashboard on the web for public access.

Outcome of Phase 5:

This final phase consolidates all findings, emphasizing how visualization can be used to interpret data intuitively. The project can be extended into a full-scale analytical platform in the future.