📊 Data Science Job Salaries Analysis (Finance Analyst Domain)

1. Executive Summary

This project analyzes global **Data Science job salaries** using a dataset aggregated from ai-jobs.net.
The goal is to uncover salary trends, identify drivers such as experience level, employment type, company size, and remote work ratio, and provide actionable insights for professionals and recruiters.

Key Findings:

- Executives earn the highest mean salary (~ $200,000), $while entry-level roles average 60,000.
- Contract employment type yields the highest mean salary (~ $180,000), $while part-time roles average 30,000.
- Large companies pay the highest (~$120,000), $compared to 75,000 in small firms.
- Remote jobs offer higher salaries ($120,000) $compared to on-site (~~105,000)~~ ~~and hybrid (~~ $80,000). $- Top-paying roles include * Principal Data Engineer* (328,000), *Financial Data Analyst* ($275,000), $and * Principal Data Scientist* (215,000).
- Highest-paying countries: Russia ($157,500), $USA (144,000), New Zealand ($125,000).
- USA, UK, and Canada provide the most job opportunities.

---

2. Dataset Description

- **Rows:** 607 instances
- **Columns:** 11 variables
- **Key Features:**
  - `work_year`: Year of salary record (2020–2022).
  - `experience_level`: Entry, Mid, Senior, Executive.
  - `employment_type`: Full-time, Part-time, Contract, Freelance.
  - `job_title`: Role title.
  - `salary`: Annual salary in USD.
  - `employee_residence`: Country of employee.

- `job_type`: Remote, onsite, hybrid.
- `company_location`: Country of employer.
- `company_size`: Small, Medium, Large.

---

3. Methodology

- **Data Cleaning:**
  - Removed duplicates (42 rows).
  - Standardized categorical values (experience level, employment type, company size).
  - Converted ISO country codes to names.
  - Dropped redundant columns (`salary_currency`).
- **Exploratory Data Analysis (EDA):**
  - Distribution plots for salary.
  - Grouped means by categorical features.
  - Correlation heatmaps.
  - Boxplots and violin plots for salary comparisons.
- **Modeling:**
  - Linear Regression to predict salary based on experience level, employment type, job type, and company size.
  - Evaluation metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE).

---

4. Insights

Salary Distribution

- Most salaries fall between $50,000$–150,000.
- Few outliers earn >$300,000 annually.

Experience Level

- Entry: ~$61,000 - Mid$: 88,000
- Senior: ~$138,000 - Executive$: 199,000

Employment Type

- Contract: ~$184,000 - Full - time$: 111,000
- Freelance: ~$48,000 - Part - time$: 33,000

Company Size

- Large: ~$118,000 - Medium$: 115,000

- Small: ~$78,000

Job Type

- Remote: ~$120,000 - Onsite: 106,000
- Hybrid: ~$80,000

Top Roles

- Principal Data Engineer: ~$328,000 - FinancialDataAnalyst: 275,000
- Principal Data Scientist: ~$215,000

Geography

- Highest salaries: Russia, USA, New Zealand.
- Most opportunities: USA, UK, Canada.

---

5. Recommendations

- **For professionals:** Target remote roles and large companies for higher pay.
- **For recruiters:** Contract roles attract top talent with higher compensation.
- **For policymakers:** Encourage remote work adoption to improve salary competitiveness.
- **For learners:** Senior and executive roles require long-term skill investment but yield significant salary growth.

---

6. Limitations

- Dataset limited to 607 records.
- Possible reporting bias from self-reported salaries.
- Currency conversion may introduce minor inaccuracies.

---

7. Conclusion

This analysis highlights clear salary drivers in Data Science: **experience, employment type, company size, and remote work**.
Remote roles and large companies consistently offer higher salaries, while executives and specialized roles command premium pay.
Geographic differences show Russia and USA as top-paying countries, while USA, UK, and Canada dominate job opportunities.

---

8. Appendix

- **Tools Used:** Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Streamlit.
- **Domain:** Finance Analyst.
- **Difficulty Level:** Intermediate.

```
1 Start coding or generate with AI.
```