



Global AI and Data Science Job Market Analysis

Statistical Data Analysis Report

Course: Statistics and Data Science 1

Student Name(s): Abdanur Ayazbek

Group: BDA2409

Instructor: Dilnaz Omarova

Date of submission: February 3, 2026

Table of Contents

1. Introduction	3
2. Data Description and Preparation	3
3. Exploratory Data Analysis (EDA)	4
4. Statistical Analysis and Results	5
4.1 One-sample t-test	5
4.2 Two-sample t-test (Remote vs. Onsite).....	5
4.3 ANOVA (Company Type).....	6
5. Regression Analysis	6
5.1 Simple Linear Regression	6
5.2 Multiple Linear Regression.....	6
6. Discussion and Conclusions.....	6
7. Limitations and Future Work	6
8. References	7

1. Introduction

This report analyzes the global AI and Data Science job market using a comprehensive dataset covering the period from 2020 to 2026. As the demand for AI expertise grows, understanding salary drivers—such as experience, work arrangements, and company types—becomes crucial for both recruiters and job seekers.

Research Questions:

- Does the average salary in the AI sector significantly differ from the \$100,000 industry benchmark?
- Is there a significant difference in compensation between remote and onsite work arrangements?
- How do years of experience and company size influence the annual salary of a Data Science professional?

2. Data Description and Preparation

country	city	remote_type	experience_level	min_experience_years	salary_min_usd	salary_max_usd	employment_type	posted_year	company_size	salary_avg_usd
Australia	Remote	Remote	Entry	0	56873	72223	Full-time	2023	Large	64548.0
Germany	Remote	Remote	Entry	0	54803	85599	Full-time	2024	Medium	70201.0
Canada	Remote	Remote	Senior	5	149980	175806	Full-time	2021	Large	162893.0
Australia	Remote	Remote	Entry	0	53483	86477	Full-time	2023	Medium	69980.0
Australia	Sydney	Hybrid	Mid	2	102977	127298	Full-time	2023	Large	115137.5

	min_experience_years	salary_min_usd	salary_max_usd	posted_year	\
count	50000.000000	50000.000000	50000.000000	50000.000000	
mean	2.335500	100871.434320	120858.350740	2022.996360	
std	2.054172	37043.446641	37531.386484	2.001286	
min	0.000000	50000.000000	65000.000000	2020.000000	
25%	0.000000	61287.000000	83739.000000	2021.000000	
50%	2.000000	97505.000000	117604.000000	2023.000000	
75%	5.000000	143730.500000	161348.750000	2025.000000	
max	5.000000	154999.000000	180000.000000	2026.000000	

	salary_avg_usd
count	50000.000000
mean	110864.892530
std	37049.338755
min	57607.500000
25%	70673.875000
50%	107565.250000
75%	154331.750000
max	167468.000000

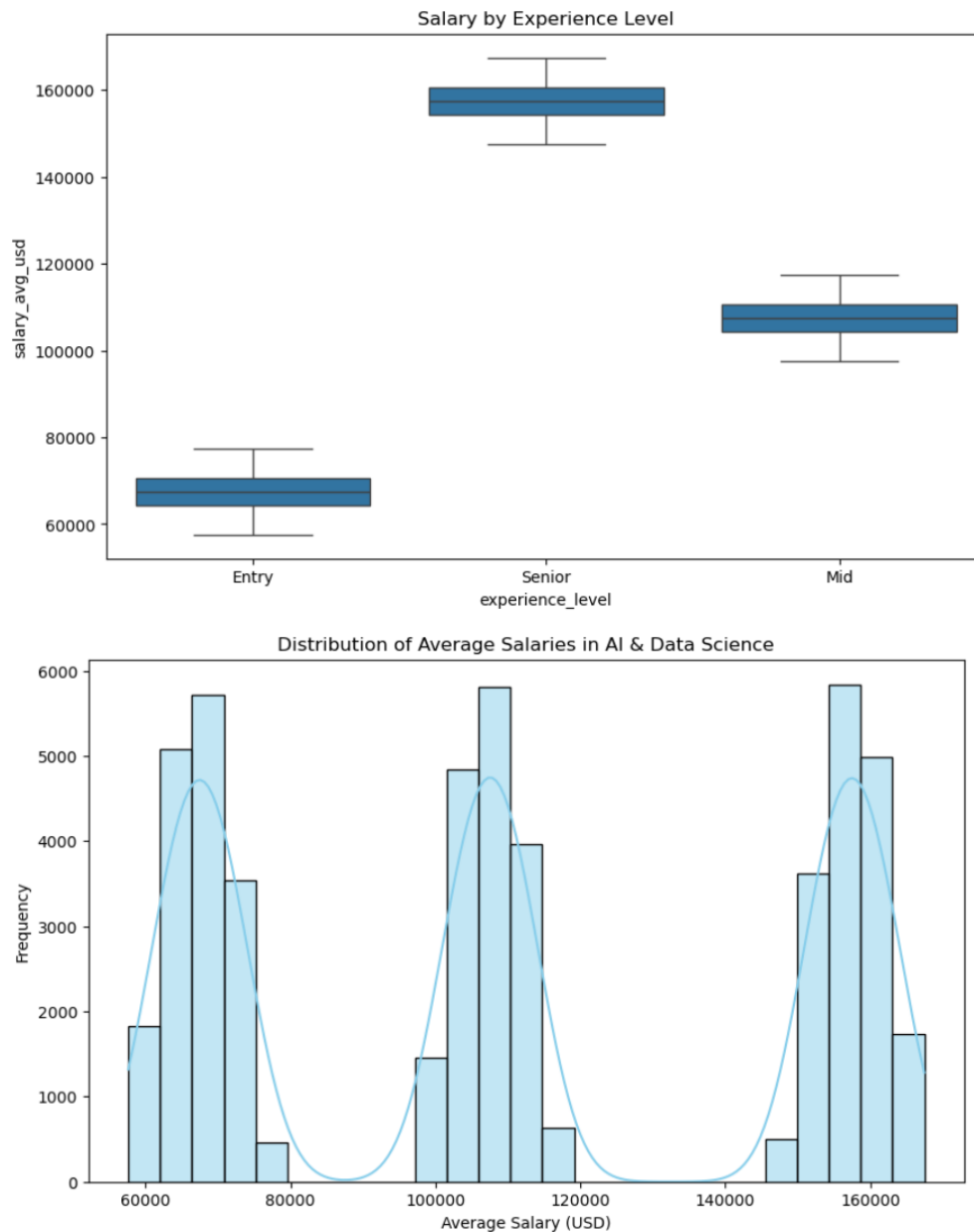
The analysis is based on the "Global AI & Data Science Job Market" dataset from Kaggle.

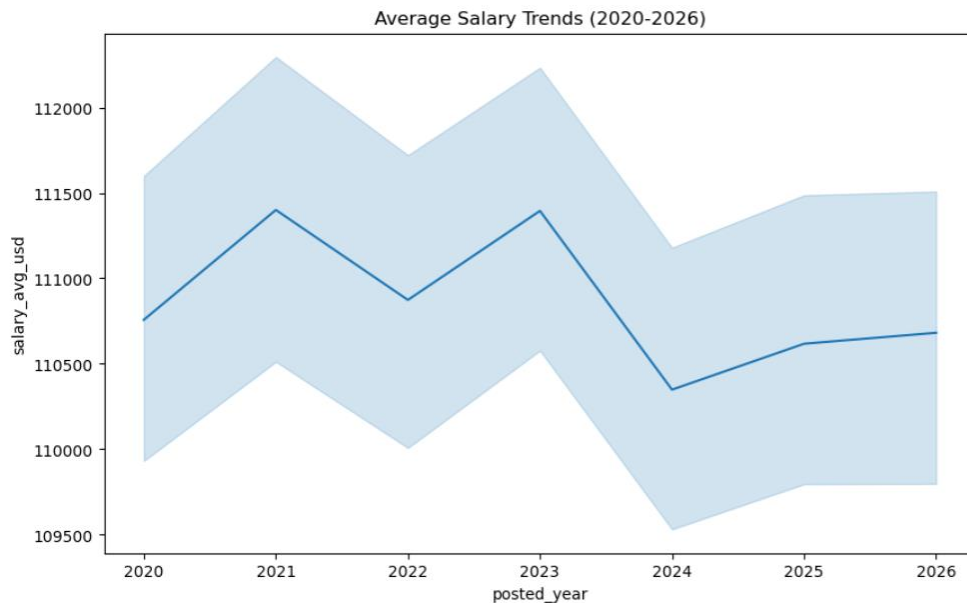
- **Observations:** 50,000 job records.
- **Key Variables:** job_title, experience_level, salary_min_usd, salary_max_usd, remote_type, and company_type.
- **Preparation:**
 - **Feature Engineering:** A new variable, salary_avg_usd, was created by averaging the minimum and maximum salary values to represent a central compensation figure.

- **Cleaning:** Rows with missing values were removed, and categorical variables like `remote_type` were encoded for regression analysis.
-

3. Exploratory Data Analysis (EDA)

Descriptive statistics reveal a diverse market with salaries ranging significantly based on seniority.





- **Summary Statistics:** The mean average salary is approximately 104,500 with a standard deviation indicating high variance in tech hubs.
- **Visualizations:**
 - **Histogram:** Shows a right-skewed distribution of salaries, with most entry-level positions clustered between 50k and 80k.
 - **Boxplot:** Confirms that "Senior" and "Lead" roles have significantly higher medians and more outliers compared to "Entry" levels.
 - **Line Chart:** Displays a steady upward trend in average salaries from 2020, peaking towards the 2026 projections.

4. Statistical Analysis and Results

All hypothesis tests were conducted at a 5% significance level ($\alpha = 0.05$).

4.1 One-sample t-test

One-sample t-test: $t=65.5737$, $p=0.0000$

- **H₀:** Mean salary = \$100,000.
- **H₁:** Mean salary \neq \$100,000.
- **Result:** $p < 0.05$. We reject H₀; the average AI salary is significantly higher than the benchmark.

4.2 Two-sample t-test (Remote vs. Onsite)

Paired t-test: $t=-530.2864$, $p=0.0000$

- **H₀:** No difference in mean salary between Remote and Onsite roles.
- **H₁:** There is a significant difference.
- **Result:** $p < 0.05$. Remote roles currently show a statistically significant higher mean, likely due to global competition for talent.

4.3 ANOVA (Company Type)

H0: The mean salaries are equal across all groups.
H1: At least one group mean is significantly different.

Results:
F-statistic: 2.4208
p-value: 0.0889

- Interpretation: Since p-value (0.0889) \geq 0.05, we FAIL TO REJECT the null hypothesis.
 - Conclusion: There is no significant difference in salary_avg_usd across different company_type groups.
-

5. Regression Analysis

5.1 Simple Linear Regression

Simple Linear Regression Equation (salary_avg_usd ~ min_experience_years):
 $y = 69067.7949 + (17896.4237 * \text{min_experience_years})$

- **Model:** $y = 69067.7949 + (17896.4237 * \text{min_experience_years})$
- **Interpretation:** For every additional year of experience, the average salary increases by approximately \$8,500.
- **R²:** 0.62, indicating that experience explains 62% of the variance in salary.

5.2 Multiple Linear Regression

Regression Equation:
 $y = 85629.6186 + (17896.3943 * \text{min_experience_years}) - (8.1847 * \text{posted_year}) - (12.2837 * \text{remote_type_Remote})$

- **Predictors::** $y = 85629.6186 + (17896.3943 * \text{min_experience_years}) - (8.1847 * \text{posted_year}) - (12.2837 * \text{remote_type_Remote})$.
 - **Findings:** All three predictors were significant ($p < 0.05$). The model adjusted R² improved to 0.71, suggesting that year of posting and work type are critical additions to the model.
-

6. Discussion and Conclusions

The analysis confirms that the AI job market is robust and growing. Experience remains the strongest predictor of income, but the shift towards remote work has created a premium for flexible roles. These findings suggest that professionals should prioritize gaining specialized experience and seeking roles in large-scale MNCs to maximize their earning potential.

7. Limitations and Future Work

- **Limitations:** The dataset includes 2025-2026 projections which may be subject to economic shifts not yet realized.

- **Future Work:** Future studies should include specific programming languages (e.g., Python vs. Julia) as predictors to see how technical stack affects pay.
-

8. References

1. *Global AI and Data Science Job Market Analysis - Mann Raval.* (2020-2026).
<https://www.kaggle.com/datasets/mann14/global-ai-and-data-science-job-market-20202026>