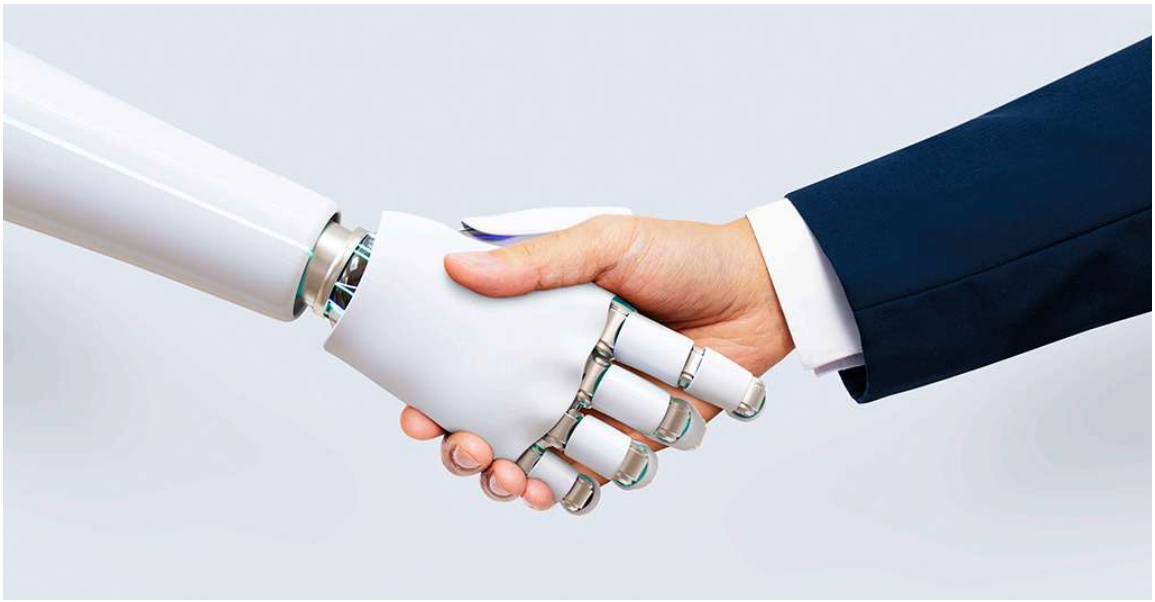
AI Job Market Analysis & Salary Trends SQL & Excel

Prepared by: ***Dina Mohsen Gamal***



AI Job Market Analysis using SQL & Excel

This project explores a dataset of AI-related jobs to uncover insights about salary trends, in-demand skills, education levels, and remote work preferences. The analysis was done using SQL for querying and Excel for data cleaning and preparation.

Table of Contents

---

1. Project Overview

With the rapid growth of the AI industry, job opportunities have exploded across various roles, skills, and experience levels. But what does the AI job market look like in terms of salaries, remote work, education, and in-demand skills?

---

2. Project Goal

To analyze the AI job market and answer key questions related to job types, salaries, skills, and education requirements. This includes exploring the most in-demand roles, salary differences across experience levels and work styles, and identifying trends that matter to both job seekers and companies.

---

3. Key Analytical Questions

General Job Info

1. How many unique job titles are there in the dataset?
2. What are the top 5 most frequent job titles?
3. What is the total number of jobs by employment type?

Salary Analysis

4. What is the average salary for each experience level?
5. What are the top 5 job titles with the highest average salary?
6. Do employees who live in the same country as their company earn higher salaries than those who live in different countries?

Skills Analysis

7. What is the difference in average salaries between employees who work fully remotely and those who work from the office?
8. What are the most common required skills in the dataset?
9. What is the average salary for jobs that require Python?

Education

10. What is the distribution of education levels required?

---

4. Dataset Review & Cleaning

We began by exploring the dataset's structure and understanding the type and meaning of each column. Below is a summary of the dataset's key columns:

Columns Description

Column	Description	Type
job_id	Unique identifier for each job posting	String
job_title	Standardized job title	String
salary_usd	Annual salary in USD	Integer
salary_currency	Original salary currency	String
experience_level	EN (Entry), MI (Mid), SE (Senior), EX (Executive)	String
employment_type	FT (Full-time), PT (Part-time), CT (Contract), FL (Freelance)	String
job_category	ML Engineer, Data Scientist, AI Researcher, etc.	String
company_location	Country where company is located	String
company_size	S (Small <50), M (Medium 50–250), L (Large >250)	String
employee_residence	Country where employee resides	String
remote_ratio	0 (No remote), 50 (Hybrid), 100 (Fully remote)	Integer
required_skills	Top 5 required skills (comma-separated)	String
education_required	Minimum education requirement	String
years_experience	Required years of experience	Integer
industry	Industry sector of the company	String

Dropped Columns

To streamline the analysis, we dropped unnecessary or less relevant columns:

- `posting_date`
- `application_deadline`
- `job_description_length`
- `benefits_score`
- `salary_local`

These columns were excluded because they were irrelevant to the project goals.

Data Cleaning in Excel:

Replaced Abbreviations with Full Forms To make the dataset easier to understand, I replaced several abbreviated codes with their full meanings.

→ This was done using the `SWITCH` function in Excel.

For example:

Experience Levels:

- `EN` → Entry
- `MI` → Mid
- `SE` → Senior
- `EX` → Executive

Employment Type:

- `FT` → Full-time
- `PT` → Part-time
- `CT` → Contract
- `FL` → Freelance

This helped ensure clearer analysis and easier interpretation, especially for non-technical stakeholders.

---

5. General Job Info

```sql
-- How many unique job titles are there in the dataset?
SELECT count(DISTINCT (job_title)) AS count_Job_Titles
FROM AI_Job;
```

There are 20 unique job titles in the dataset

`Output:`

count_Job_Titles
20

```sql
-- What are the top 5 most frequent job titles?
SELECT TOP 5 job_title,count(job_title) AS job_count
FROM AI_Job
GROUP BY job_title
ORDER BY job_count DESC;
```

Output:

job_title	job_count
Machine Learning Researcher	808
AI Software Engineer	784
Autonomous Systems Engineer	777
Machine Learning Engineer	772
AI Architect	771

```sql
-- What is the total number of jobs by employment type?
SELECT  employment_type ,COUNT(employment_type) AS count_employment_type
FROM AI_Job
GROUP BY employment_type
ORDER BY COUNT_employment_type DESC;
```

Output:

employment_type	count_employment_type
Full-time	3812
Freelance	3758
Contract	3721
Part-time	3709

---

6. Salary Analysis

```sql
-- What is the average salary for each experience level?
SELECT experience_level ,AVG(salary_usd) AS AVG_Salary
FROM AI_Job
GROUP BY experience_level
ORDER BY AVG_Salary DESC;
```

Output:

experience_level	AVG_Salary
Executive	187723
Senior	122187
Mid	87955
Entry	63133

```sql
-- What are the top 5 job titles with the highest average salary?
SELECT TOP 5 job_title ,AVG(salary_usd)AS TOP_AVG_SALARY
FROM AI_Job
GROUP BY job_title
ORDER BY TOP_AVG_SALARY DESC;
```

Output:

job_title	TOP_AVG_SALARY
AI Specialist	120570
Machine Learning Engineer	118827
Head of AI	118542
AI Research Scientist	117897
AI Architect	117436

```sql
/* Do employees who live in the same country
   as their company earn higher salaries than those who
   live in different countries?*/

-- Without using a CTE query involves repetition
/*SELECT
  (SELECT AVG(salary_usd)
   FROM AI_Job
   WHERE company_location = employee_residence) AS
avg_salary_same_country,

  (SELECT AVG(salary_usd)
   FROM AI_Job
   WHERE company_location <> employee_residence) AS
avg_salary_different_country,

  (SELECT AVG(salary_usd)
   FROM AI_Job
   WHERE company_location = employee_residence) -
  (SELECT AVG(salary_usd)
   FROM AI_Job
   WHERE company_location <> employee_residence) AS salary_difference;*/

WITH SalaryAvgs AS (
  SELECT
    (SELECT AVG(salary_usd)
     FROM AI_Job
     WHERE company_location = employee_residence) AS
avg_salary_same_country,

    (SELECT AVG(salary_usd)
     FROM AI_Job
     WHERE company_location <> employee_residence) AS
```

```
avg_salary_different_country
)
SELECT
  avg_salary_same_country,
  avg_salary_different_country,
  avg_salary_same_country - avg_salary_different_country AS
salary_difference
FROM SalaryAvgs;
```

`Note:` The negative difference means employees living outside the company's country tend to earn higher salaries.

`Output:`

avg_salary_same_country	avg_salary_different_country	salary_difference
114979	116282	-1303

---

7. Skills Analysis

```
-- What is the difference in average salaries between employees who work
fully remotely and those who work from the office?
SELECT remote_ratio, AVG(salary_usd) AS avg_salary
FROM AI_Job
WHERE remote_ratio IN ('Fully_remote', 'No_remote')
GROUP BY remote_ratio
ORDER BY avg_salary DESC;
```

Output:

remote_ratio	avg_salary
Fully_remote	116160
No_remote	114140

```
-- What are the most common required skills in the dataset?
SELECT skill, COUNT(job_id) AS skill_count
FROM (
  SELECT job_id, LTRIM(RTRIM(value)) AS skill
  FROM AI_Job
  CROSS APPLY STRING_SPLIT(required_skills, ',')   --gives us id an skill
) AS cleaned_skills
GROUP BY skill
ORDER BY skill_count DESC;
```

`Output:`

skill	skill_count
Python	4450

skill	skill_count
SQL	3407
TensorFlow	3022
Kubernetes	3009
Scala	2794
PyTorch	2777
Linux	2705
Git	2631
Java	2578
GCP	2442
Hadoop	2419
Tableau	2341
R	2311
Computer Vision	2284
Data Visualization	2270
Deep Learning	2189
MLOps	2164
Spark	2155
NLP	2145
Azure	2144
AWS	2018
Mathematics	1943
Docker	1862
Statistics	1833

```sql
-- What is the average salary for jobs that require Python?
SELECT AVG(salary_usd) AS avg_salary_python
FROM AI_Job
WHERE required_skills LIKE '%Python%';
```

`Output:`

avg_salary_python
114399

---

8. Education

```sql
-- What is the distribution of education levels required?
SELECT education_required, COUNT(job_id) AS job_count
FROM AI_Job
GROUP BY education_required
ORDER BY job_count DESC;
```

`Output:`

education_required	job_count
Bachelor	3789
Associate	3785
Master	3748
PhD	3678

---

9. Technologies Used

- **SQL Server** – Advanced querying
- **Excel** – Data cleaning, transformation, and preparation

---

10. Final Thoughts

This analysis demonstrates how structured query language (SQL) and spreadsheet tools like Excel can be used effectively to extract actionable insights from job market data. By exploring roles, salaries, and skills, we gain a better understanding of trends shaping the AI workforce.

Connect with me:

- GitHub : https://github.com/dinamohsin?tab=repositories
- LinkedIn : https://www.linkedin.com/in/dinamohsen231102/