# Data Science Salary Analysis Project

**Introduction**

This project analyzes data science salaries across various job positions, aiming to uncover insights on factors such as experience level, employment type, company size, and remote work. Using SQL queries on a structured salary dataset, we investigate salary trends, disparities, and key patterns that influence compensation in the field of data science. This analysis provides valuable information for aspiring data professionals to understand salary expectations and plan their careers accordingly.

**Key Questions Analyzed**

* How many unique job titles are included in the dataset?
* What are the top 10 highest average salaries for different data science job positions?
* Does employment type significantly affect salary levels?
* How does experience level impact salary trends?
* How have salaries evolved over the years?
* What is the distribution of employees and average salary by remote work ratio?
* How does company size influence average salaries?
* What are the top 3 highest-paying job titles per year?
* How has salary growth changed over time by experience level?
* How do salary trends look when pivoted by experience level over the years?

**Data**

The dataset we are using was downloaded from Kaggle site available at:

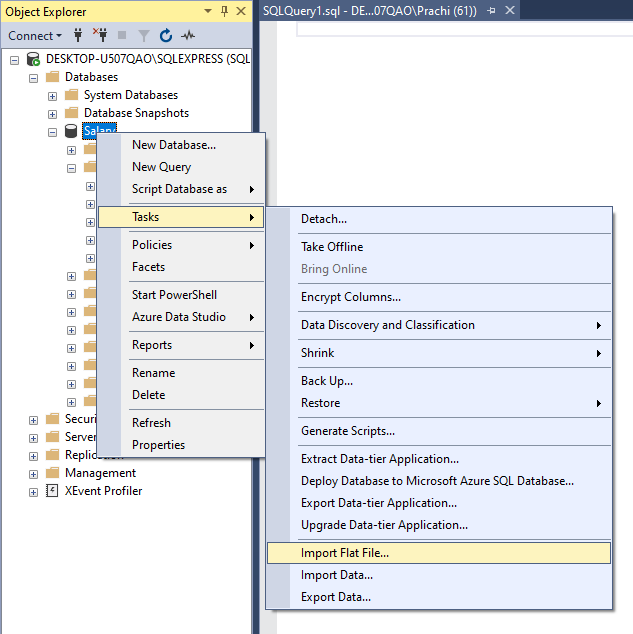
<https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries>

The salary dataset used in this project contains the following columns:

* **Work\_year**: The year the salary was paid.
* **Experience\_level**: The experience level in the job during the year (EN - Entry-level, MI - Mid-level, SE - Senior-level, EX - Executive-level).
* **Employment\_type**: The type of employment for the role (PT - Part-time, FT - Full-time, CT - Contract, FL - Freelance).
* **Job\_title**: The role worked in during the year.
* **Salary**: The total gross salary amount paid.
* **Salary\_currency**: The currency of the salary paid as an ISO 4217 currency code.
* **Salary\_in\_usd**: The salary in USD.
* **Employee\_residence**: Employee's primary country of residence during the work year.
* **Remote\_ratio**: The amount of work done remotely (0 - No remote work, 50 - Partially remote, 100 - Fully remote).
* **Company\_location**: The country of the employer's main office.
* **Company\_size**: The average number of employees in the company (S - Small, M - Medium, L - Large).

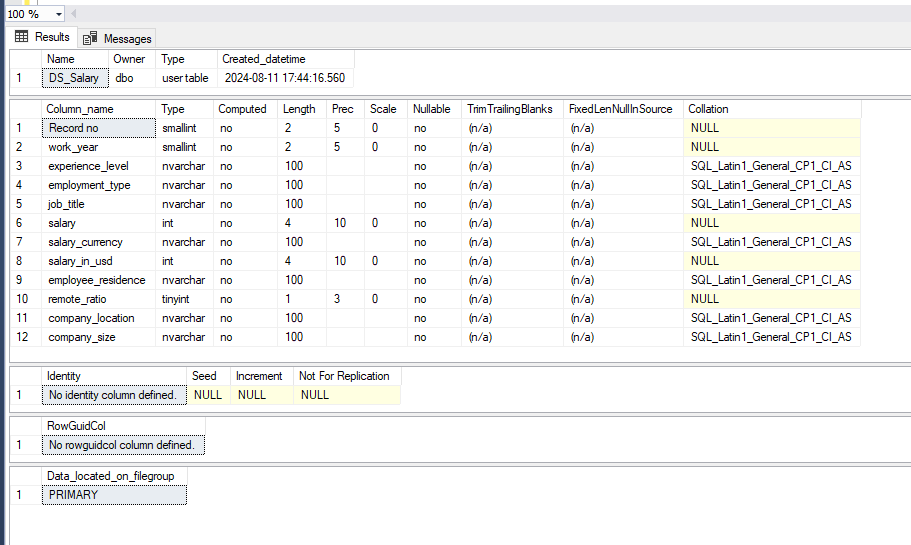
**Data Import**

The dataset was in CSV format and was imported into **Microsoft SQL Server Management Studio** using the **Import Flat File** functionality.



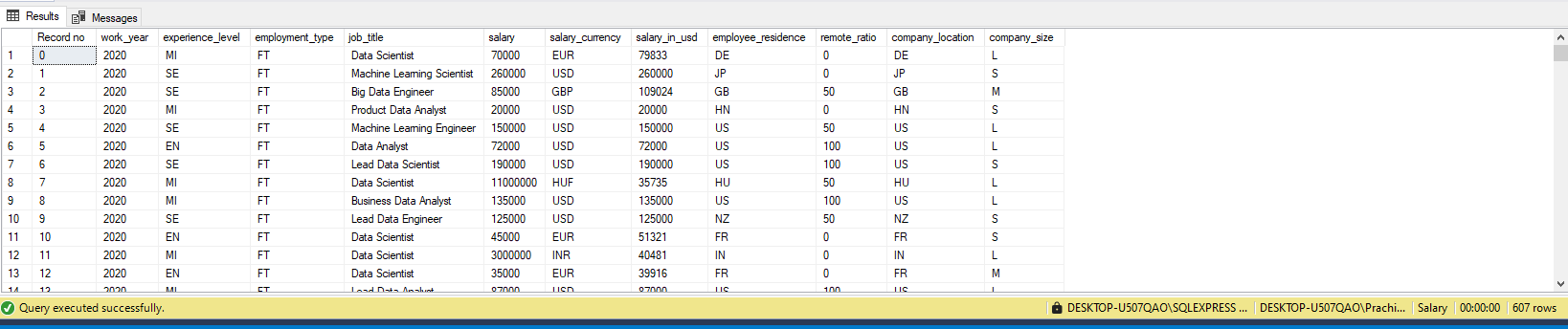
sp\_help DS\_Salary

gives us the detailed information about the table likes the table creation date and time, column names in the table with its datatype, if there is any constraints on the table etc.



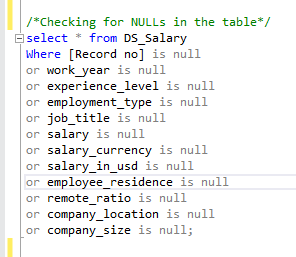
select \* from DS\_Salary

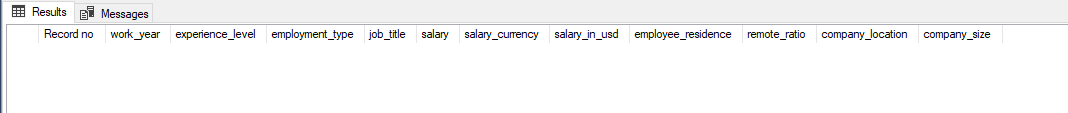
Shows us all the records in the tables, there are 12 columns and 607 rows in the DS\_Salary table.



**Data cleaning.**

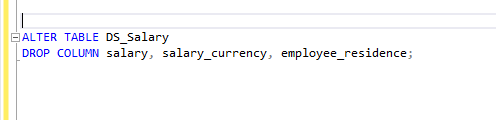
Now that we have imported our data got an overview of the data, we will check for any nulls in our data.



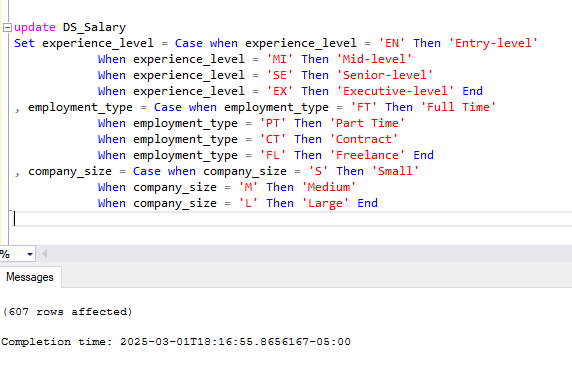


The blank output confirms that there are no nulls in the data.

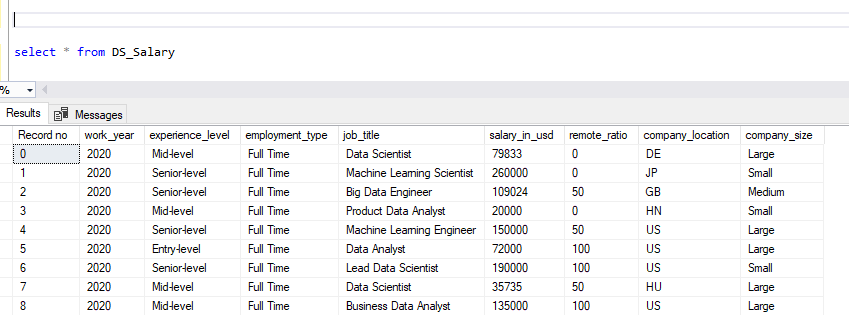
Next, there are two salary fields in our table salary and salary\_in\_usd, for analysis purpose we will use the salary\_in\_usd to be on same scale and avoid any futher calculations. As a result of which we will be deleting the salary and salary\_currency fields. Also we wont be using the employee\_residence since our analysis focuses on company attributes.



To get better understanding from the query output, we will replace the short forms with the full forms for the experience\_level, employment\_type and company\_size field



Our data now looks much better and self explanatory at a quick glance and we are ready to start with our analysis.



**Data Analysis and Insights**

For analysis purpose we will use average rather than maximum or minimum salary\_in\_usd field for comparing and analyzing because.

* **Represents Overall Trends**

The average gives a general idea of the central tendency of a dataset, making it useful for understanding the typical value.

MAX and MIN only show the highest or lowest values, which might be outliers and not representative of the overall trend.

* **Reduces Impact of Outliers**

The mean considers all data points, making it a more reliable measure for comparisons.

MAX and MIN can be heavily influenced by extreme values, leading to misleading insights.

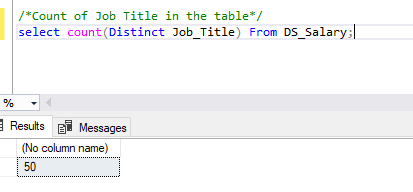
* **Better for Comparisons and Forecasting**

When comparing trends over time (e.g., sales performance, customer spending), the average provides a smoother insight rather than focusing on peak or low moments.

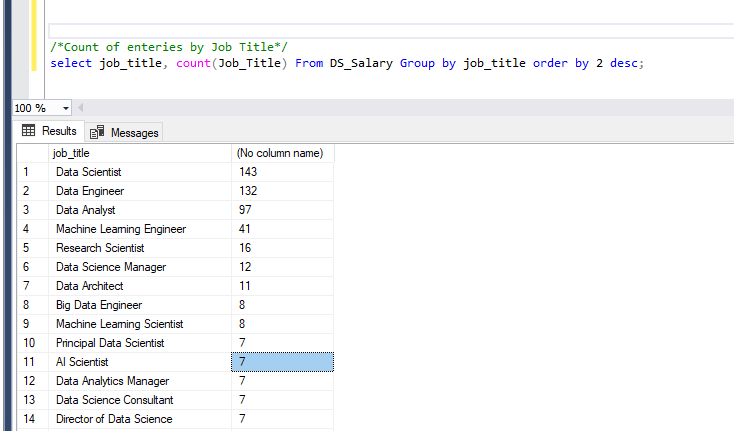
In forecasting, using averages helps in trend analysis, while MAX/MIN may cause fluctuations.

1. **How many job title was included in the dataset?**

We start by checking how many distinct job title are present in our dataset, using count with distinct will serve this purpose.

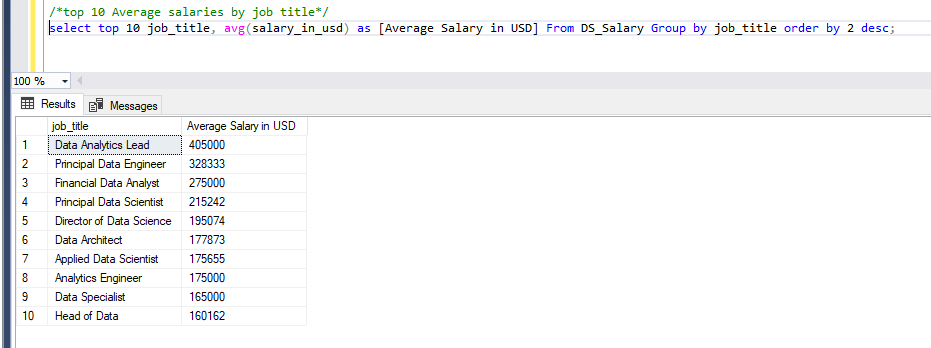


We have 50 distinct job titles in our dataset.



Further analysis showed the most common job titles, with **Data Scientist** and **Data Engineer** having the highest occurrences.

1. **What are the top 10 average salaries for the different data science job positions?**

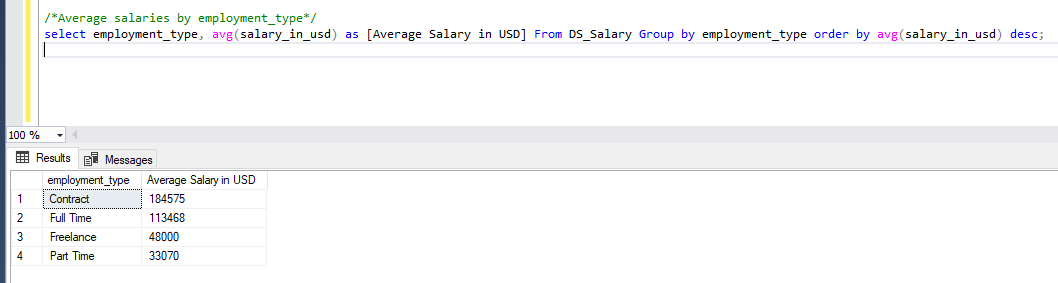


Using **AVG()** function with GROUP BY job title, we identified the top 10 job titles with the highest salaries.

**Key Insight:** The highest-paying role is **Data Analytics Lead**, followed by **Principal Data Engineer** and **Financial Data Analyst**.

Next we will check if and how different variable affect the salary.

1. **Does Employment type affect the salary**

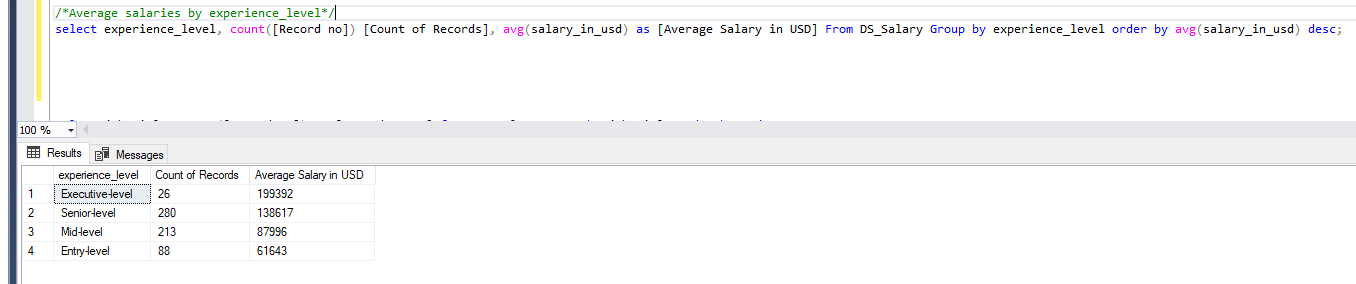


**Finding:** **Contract-based employees earn more on average than full-time employees.**

**Explanation:** Contractors typically receive higher salaries due to **lack of benefits** (e.g., insurance, paid leave) compared to full-time roles.

**Observation:** Part-time employees earn the least, as expected.

1. **Does experience level affect the salary**

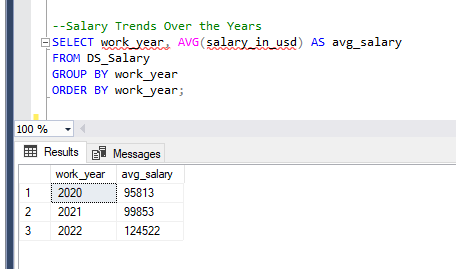


**Finding:** Salaries increase significantly with experience.

**Executives (EX) earn 2.5x more** than Entry-Level (EN) professionals.

**Key Insight:** Moving from **Mid-Level to Senior-Level shows the largest jump in salary.**

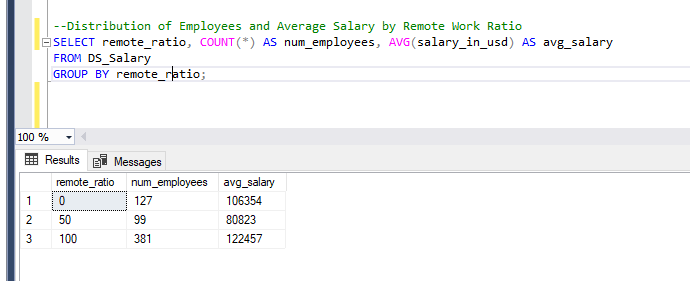
1. **Salary Trends Over the Years**



**Finding:** The **average salary increased year over year, with a significant spike in 2022**.

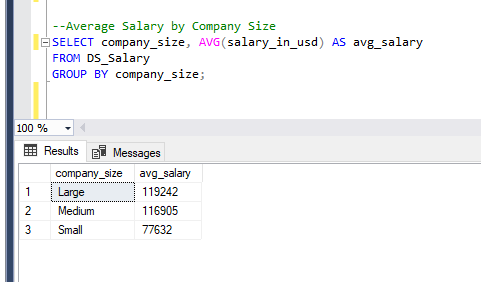
**Possible reasons:** Increased demand for data science professionals and market adjustments post-pandemic.

1. **Distribution of Employees and Average Salary by Remote Work Ratio**



100% remote jobs have the highest average salary, followed by onsite jobs

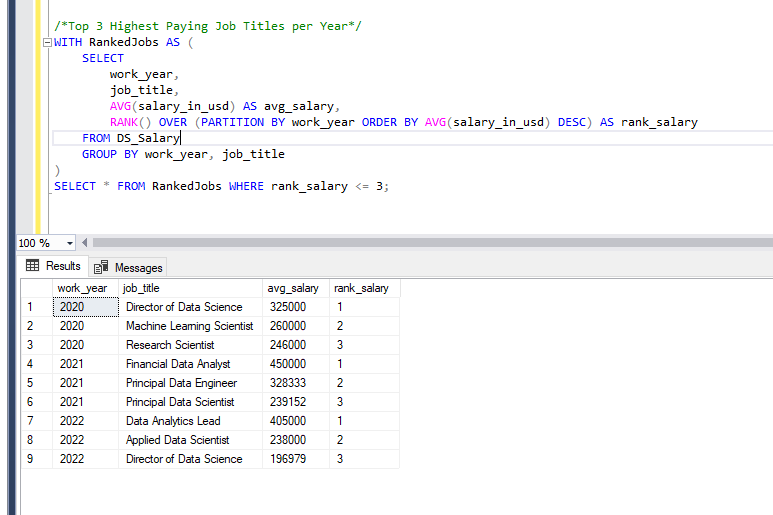
1. **Average Salary by Company Size**



**Finding:** Large companies offer **higher average salaries** than small and medium-sized firms.

**Possible reason:** Larger firms have **higher budgets** and more structured compensation models.

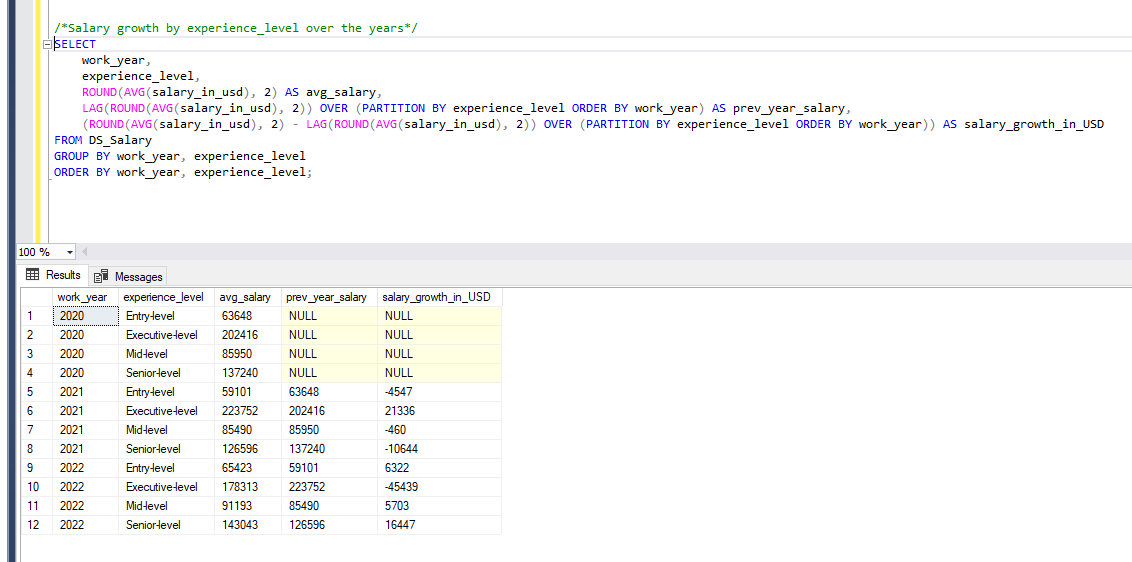
1. **Top 3 Highest Paying Job Titles per Year**



**Key Finding:** The **Director of Data Science** was the highest-paid job in **2020**, but dropped to **third place in 2022**.

**Machine Learning Engineers** saw a **consistent rise in salary rankings** over the years.

1. **Salary growth by experience level over the years**

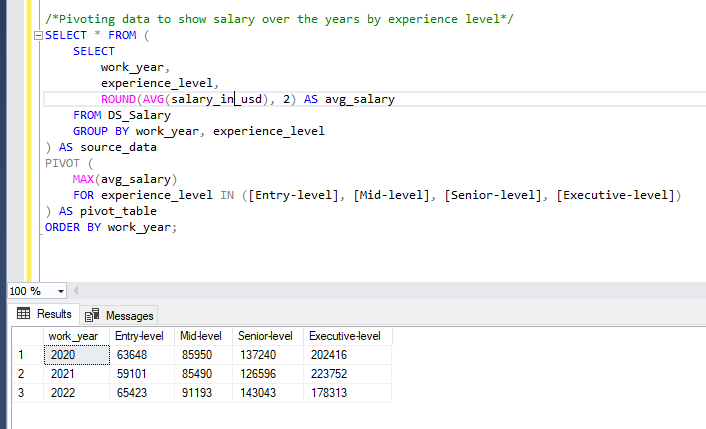


**2021:** Salaries declined for all experience levels **except executives**.

**2022:** Executive-level salaries were the **only ones to decrease**, while other levels saw an increase.

.

1. **Pivoting data to show salary over the years by experience level**



**Using pivot tables**, we visualized salary trends across experience levels.

**Key Insight:** **Mid-level salaries grew the most**, followed by **Senior-level salaries**.

**Conclusion & Key Takeaways**

* **Experience Level Matters:** Higher experience results in significantly higher salaries, with executives earning the most.
* **Employment Type Has an Effect:** Contractors earn more than full-time employees but lack benefits.
* **Remote Work Pays More:** Fully remote roles tend to offer higher salaries than on-site jobs.
* **Salaries Are Increasing Over Time:** The data indicates a growing demand for data science roles, reflected in salary trends.
* **Company Size Plays a Role:** Large companies generally pay better salaries than small or medium-sized firms.

**Future Recommendations**

* Conduct further **regional salary analysis** to understand variations by country.
* Analyze how **certifications and skills** influence salary growth.
* Investigate **gender pay gaps** in data science roles.