MYSQL PROJECT

ANALYSIS OF DATA SCIENTISTS SALARY

From:
Nethravathi K Hangarki

WHAT IS MYSQL ?

Structured Query Language is a Programming language for string & processing information in a relational database. A Relational Database stores information in tabular form, with rows & columns representing different data attributes and the various relationship b/w the data values.

ADVANTAGES OF MYSQL

- Easy to learn
- Easy syntax
- Interactive language
- Creative several views
- Portable
- High performance
- Data security

PROJECT NAME

ANALYSIS OF DATA SCIENTISTS SALARY

SUMMARY

The name of this study is to investigate the factors influencing the salaries of Data Scientists. To achieve this, a dataset containing various relevant variables was utilized. This report describes the exploratory analysis conducted to understand the relationshio b/w these factors and Data Scientists salaries.

INTRODUCTION

Data Science is a rapidly growing field and Data Scientists play a crucial role in analyzing and interpreting large volumes of data. As this profession becomes increasingly in demand, It is important to understand the factors that may influence Data Scientists salaries. This analysis focuses on investigating these factors & their impact on salaries.

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 curreny code.

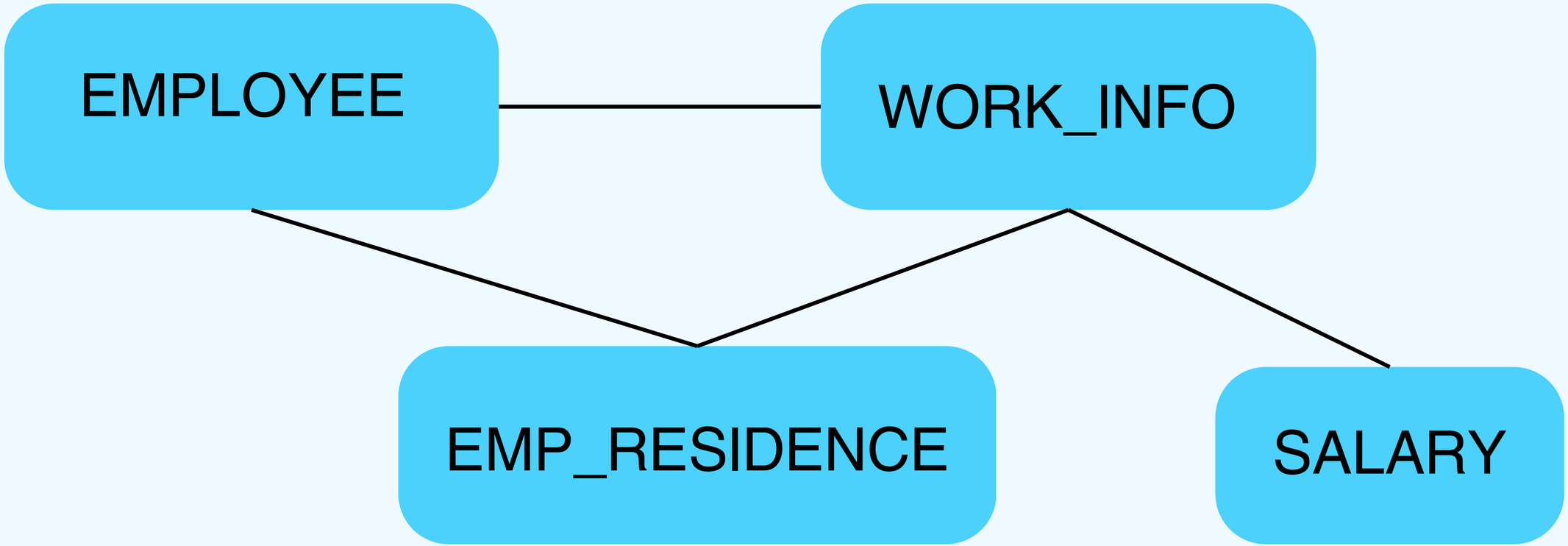
Salary_USD: The salary in USD.

Emp_Residence: Employee's primary country of residence during the work year as an ISO 3166 country code.

Remote_Ratio: The overall amount of work done remotely.

Company_Location: The country of the Employer's main office or contracting branch.

Entity Relationship Diagram



ENTITY RELATIONSHIP DIAGRAM

EMPLOYEE

Emp_id

Work_Year

Experience_Level

Job_Title

Work_Info

Emp_id

Emp_Type

Salary

Emp_Residence

Emp_id

Emp_Residence

Remote_Ratio

Company_Location

Salary

Salary

Salary_Currency

salary_USD

Employee

Emp_id	Work_Year	Experience_Level	Job_Title
1	2023	SE	Principle_Data_Scientist
2	2023	MI	ML_Engineer
3	2023	MI	ML_Engineer
4	2023	SE	Data_Scientist
5	2023	SE	Data_Scientist
6	2023	SE	Applied_Scientist
7	2023	SE	Applied_Scientist
8	2023	SE	Data_Scientist
9	2023	SE	Data_Scientist
10	2023	SE	Data_Scientist
11	2023	SE	Data_Scientist
12	2023	SE	Data_Scientist
13	2023	SE	Data_Scientist
14	2023	EN	Applied_Scientist
15	2023	EN	Applied_Scientist
16	2023	SE	Data_Modeler
17	2023	SE	Data_Modeler
18	2023	SE	Data_Scientist
19	2023	SE	Data_Scientist
20	2023	MI	Data_Analyst

Work_Info

Emp_id	Emp_Type	Salary
18	PT	19700
3	CT	25500
2	CT	30000
1	FT	80000
11	FT	90700
13	FT	100000
12	PT	130000
15	PT	130700
16	PT	130760
7	FT	136000
9	FT	141000
17	PT	146000
10	Ft	147000
20	FT	150000
19	FT	170000
4	FT	175000
5	FT	185000
14	FT	213660
8	FT	219000
6	Ft	222200

Emp_Residence

Emp_id	Emp_Residence	Remote_Ratio	Company_Location
1	ES	100	ES
2	US	100	US
3	US	100	US
4	CA	100	CA
5	CA	100	CA
6	US	0	US
7	US	0	US
8	CA	0	CA
9	CA	0	CA
10	US	0	US
11	US	0	US
12	US	100	US
13	US	100	US
14	US	0	US
15	US	0	US
16	US	0	US
17	US	100	US
18	US	100	US
19	US	0	US
20	US	100	US

Salary

Salary	Salary_Currency	Salary_USD
80000	EUR	85847
30000	USD	30000
25500	USD	25500
175000	USD	175000
185000	USD	185000
222200	USD	222200
136000	USD	136000
219000	USD	219000
141000	USD	141000
147000	USD	147000
90700	USD	90700
130000	USD	130000
100000	USD	100000
213660	USD	213660
130700	USD	130700
130760	USD	130760
146000	USD	146000
19700	USD	19700
170000	USD	170000
150000	USD	150000

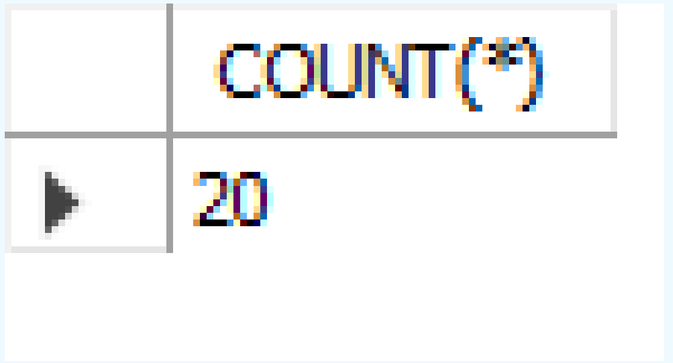
Write a query to join Employee, Work_Info,Emp_Residence and Salary tables.

```
Select
Employee.Emp_id,Employee.Work_Year,Employee.Experience_L
evel,Employee.Job_Title,
Work_Info.Salary,
Salary.Salary,Salary.Salary_Currency,Salary.Salary_USD,
Emp_Residence.Emp_Residence,Emp_Residence.Remote_Ratio,
Emp_Residence.Company_Location
From Employee
LEFT JOIN Work_Info ON
Employee.Emp_id=Work_Info.Emp_id
LEFT JOIN Salary ON
Work_Info.Salary=Salary.Salary
LEFT JOIN Emp_Residence ON
Employee.Emp_id=Emp_Residence.Emp_id;
```

Emp_id	Work_Year	Experience_Level	Job_Title	Salary	Salary	Salary_Currency	Salary_USD	Emp_Residence	Remote_Ratio	Company_Location
1	2023	SE	Principle_Data_Scientist	80000	80000	EUR	85847	ES	100	ES
2	2023	MI	ML_Engineer	30000	30000	USD	30000	US	100	US
3	2023	MI	ML_Engineer	25500	25500	USD	25500	US	100	US
4	2023	SE	Data_Scientist	175000	175000	USD	175000	CA	100	CA
5	2023	SE	Data_Scientist	185000	185000	USD	185000	CA	100	CA
6	2023	SE	Applied_Scientist	222200	222200	USD	222200	US	0	US
7	2023	SE	Applied_Scientist	136000	136000	USD	136000	US	0	US
8	2023	SE	Data_Scientist	219000	219000	USD	219000	CA	0	CA
9	2023	SE	Data_Scientist	141000	141000	USD	141000	CA	0	CA
10	2023	SE	Data_Scientist	147000	147000	USD	147000	US	0	US
11	2023	SE	Data_Scientist	90700	90700	USD	90700	US	0	US
12	2023	SE	Data_Scientist	130000	130000	USD	130000	US	100	US
13	2023	SE	Data_Scientist	100000	100000	USD	100000	US	100	US
14	2023	EN	Applied_Scientist	213660	213660	USD	213660	US	0	US
15	2023	EN	Applied_Scientist	130700	130700	USD	130700	US	0	US
16	2023	SE	Data_Modeler	130760	130760	USD	130760	US	0	US
17	2023	SE	Data_Modeler	146000	146000	USD	146000	US	100	US
18	2023	SE	Data_Scientist	19700	19700	USD	19700	US	100	US
19	2023	SE	Data_Scientist	170000	170000	USD	170000	US	0	US
20	2023	MI	Data_Analyst	150000	150000	USD	150000	US	100	US

Write a query to count the number of Employees

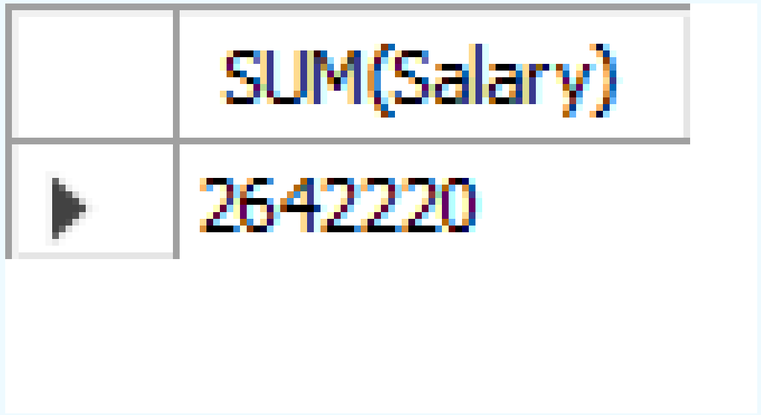
Select COUNT(*) from Employee



	COUNT(*)
▶	20

Write a query to get sum of salary of the employees

```
Select SUM(Salary) from
        Work_Info;
```



	SUM(Salary)
▶	2642220

Write a query to list the distinct Job Title of the employees

```
Select DISTINCT Job_Title from
        Employee;
```

Job_Title
Applied_Scientist
Data_Analyst
Data_Modeler
Data_Scientist
ML_Engineer
Principle_Data_Scientist

Write a query to list Salary in descending order

```
Select * from Work_Info
ORDER BY Salary DESC
```

Emp_id	Emp_Type	Salary
6	Ft	222200
8	FT	219000
14	FT	213660
5	FT	185000
4	FT	175000
19	FT	170000
20	FT	150000
10	Ft	147000
17	PT	146000
9	FT	141000
7	FT	136000
16	PT	130760
15	PT	130700
12	PT	130000
13	FT	100000
11	FT	90700
1	FT	80000
2	CT	30000
3	CT	25500
18	PT	19700

Write a query to display count of the specific job title in an ascending order

```
Select Job_Title, COUNT(Emp_id)
    FROM Employee GROUP BY
Job_Title ORDER BY Job_Title
```

Job_Title	COUNT(Emp_id)
Applied_Scientist	4
Data_Analyst	1
Data_Modeler	2
Data_Scientist	10
ML_Engineer	2
Principle_Data_Scientist	1

Write a query to extract the salary where salary is greater than equal to 1 lakh

Emp_id	Emp_Type	Salary
13	FT	100000
12	PT	130000
15	PT	130700
16	PT	130760
7	FT	136000
9	FT	141000
17	PT	146000
10	Ft	147000
20	FT	150000
19	FT	170000
4	FT	175000
5	FT	185000
14	FT	213660
8	FT	219000
6	Ft	222200

Write a query to display SE(Senior Level) who are Data Scientist

```
Select * from Employee WHERE
    Experience_level='SE' AND
     Job_Title='Data_Scientist';
```

Emp_id	Work_Year	Experience_Level	Job_Title
4	2023	SE	Data_Scientist
5	2023	SE	Data_Scientist
8	2023	SE	Data_Scientist
9	2023	SE	Data_Scientist
10	2023	SE	Data_Scientist
11	2023	SE	Data_Scientist
12	2023	SE	Data_Scientist
13	2023	SE	Data_Scientist
18	2023	SE	Data_Scientist
19	2023	SE	Data_Scientist

Write a query to display Employees salary between 1L to 1.5L

Select * from Work_Info WHERE Salary BETWEEN 100000 and 150000;

	Emp_id	Emp_Type	Salary
▶	13	FT	100000
	12	PT	130000
	15	PT	130700
	16	PT	130760
	7	FT	136000
	9	FT	141000
	17	PT	146000
	10	Ft	147000
	20	FT	150000


THANK
YOU