



M Intelligence: Test - Solution Consultant and Data Analyst (Marketing Technology/AI), ConnectX

This is a part of Data Analyst

This is a test for the applicant. The dataset consists of salaries and bonuses for each worker and department, obtained from the M-Intelligence dataset. Also, I do this as a report after I already done a query and explain some part to you.

Title

- Coding
- Dashboard

About dataset files

The dataset I got from e-mail, it is including 3 files as a `.CSV`

- worker.csv
- title.csv
- bonus.csv

Tools

The requirement is to perform a test using an SQL query. However, I do not have any experience with local or online SQL tools because I was using SQLite online, which has limitation on the dataset size. Therefore, I have decided to use **VSCODE** and **Python** language for querying instead. By the way, this report I writing on **Notion** and export it to PDF file.

First part

I import libraries such as `pandas`, `numpy`, and `duckdb` for querying as SQL. After that, I import `.csv` files using pandas with `pd.read_csv("")` and convert them to DataFrames.

```
import pandas as pd
import numpy as np
import duckdb

worker = pd.read_csv("worker.csv")
title = pd.read_csv("title.csv")
bonus = pd.read_csv("bonus.csv")

worker = pd.DataFrame(worker)
title = pd.DataFrame(title)
bonus = pd.DataFrame(bonus)
```

Second

I want to verify whether there are any `null values` (`.isnull()`) in the dataset and examine the information for each. While you can check for `null values` using `.info()` and gather information about the dataset columns, I prefer to have a clearer view of the data, so I have decided to take two steps.

```
print("Check null of worker")
print(worker.isnull().sum())
print("-----")

print("Check null of title")
print(title.isnull().sum())
print("-----")

print("Check null of bonus")
print(bonus.isnull().sum())
```

```
Check null of worker
WORKER_ID      0
FIRST_NAME     0
LAST_NAME      0
SALARY         0
JOINING_DATE   0
DEPARTMENT     0
dtype: int64
-----
Check null of title
WORKER_REF_ID  0
WORKER_TITLE   0
AFFECTED_FROM  0
dtype: int64
-----
Check null of bonus
WORKER_REF_ID  0
BONUS_AMOUNT   0
BONUS_DATE     0
dtype: int64
```

Check info

```
worker.info()
print("-----")
title.info()
print("-----")
bonus.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   WORKER_ID       16 non-null    int64
1   FIRST_NAME      16 non-null    object
2   LAST_NAME       16 non-null    object
3   SALARY          16 non-null    int64
4   JOINING_DATE    16 non-null    object
5   DEPARTMENT      16 non-null    object
dtypes: int64(2), object(4)
memory usage: 900.0+ bytes
-----

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  -
0   WORKER_REF_ID   8 non-null      int64
1   WORKER_TITLE    8 non-null      object
2   AFFECTED_FROM   8 non-null      object
dtypes: int64(1), object(2)
memory usage: 324.0+ bytes
-----

...
1   BONUS_AMOUNT    5 non-null      int64
2   BONUS_DATE      5 non-null      object
dtypes: int64(2), object(1)
memory usage: 252.0+ bytes

```

Question Part

As you know, the requirement is to use SQL language. Therefore, I decided to use the `duckdb` library for querying questions.

Question 1: Show the Average salary of each department

```
df1 = duckdb.sql("""
    SELECT
        DEPARTMENT,
        ROUND(AVG(SALARY), 2) avg_salary
    FROM worker
    GROUP BY DEPARTMENT
    ORDER BY avg_salary DESC

    """)

df1
```

The result of Question 1

DEPARTMENT varchar	avg_salary double
Admin	306250.0
HR	233333.33
Account	105000.0

Question 2: Show the number of employees in each department. Arranged from highest to lowest.

```
df2 = duckdb.sql("""
    SELECT
        DEPARTMENT,
        COUNT(WORKER_ID) total_worker
    FROM worker
```

```
GROUP BY DEPARTMENT
ORDER BY total_worker DESC

"""

df2
```

The result of Question 2

DEPARTMENT varchar	total_worker int64
Admin	8
Account	5
HR	3

Question 3: Show employees with the same salary

```
df3 = duckdb.sql("""
    SELECT
        FIRST_NAME || ' ' || LAST_NAME AS full_name,
        SALARY
    FROM worker
    WHERE SALARY IN (
        SELECT
            SALARY
        FROM worker
        GROUP BY SALARY
        HAVING COUNT(1) > 1
    )
    ORDER BY SALARY DESC
    """)
```

df3

The result of Question 3

full_name varchar	SALARY int64
Amitabh Singh	500000
Vivek Bhati	500000
Ami Singh	500000
Viv Bha	500000
Vishal Singhal	300000
Vi Sing	300000
Vipul Diwan	200000
Vipul Diwan	200000
Geetika Chauhan	90000
Mo Ar	90000
Niharika Verma	80000
Ni Ver	80000
Satish Kumar	75000
Satish Kumar	75000
14 rows	2 columns

Question 4: Show the department with the most bonus payments

```
df4 = duckdb.sql("""
    SELECT
        worker.DEPARTMENT,
        COUNT(worker.WORKER_ID) num_worker,
        SUM(bonus.BONUS_AMOUNT) total_bonus
    FROM worker
    INNER JOIN bonus
    ON worker.WORKER_ID = bonus.WORKER_REF_ID
    GROUP BY worker.DEPARTMENT
```



```
ORDER BY total_bonus DESC
""")
```

df4

The result of Question 4

DEPARTMENT	num_worker	total_bonus
varchar	int64	int128
HR	3	13500
Admin	2	6500

Question 5: Show the names of employees with the highest salary plus bonuses in each department

```
df5 = duckdb.sql("""
    SELECT
        worker.DEPARTMENT,
        worker.FIRST_NAME || ' ' || worker.LAST_NAME as name,
        worker.SALARY + bonus.BONUS_AMOUNT as total_
    FROM worker
    INNER JOIN bonus
    ON worker.WORKER_ID = bonus.WORKER_REF_ID
    ORDER BY total_salary DESC
""")
```

df5

The result of Question 5

DEPARTMENT varchar	full_name varchar	total_salary int64
HR	Vishal Singhal	304000
HR	Monika Arora	105000
HR	Monika Arora	104500
Admin	Niharika Verma	83500
Admin	Niharika Verma	83000

Question 6: Show the names of positions with the highest combined salaries in each department

```
df6 = duckdb.sql("""
    SELECT
        title.WORKER_TITLE,
        worker.DEPARTMENT,
        SUM(worker.SALARY) total_salary
    FROM worker
    INNER JOIN title
    ON worker.WORKER_ID = title.WORKER_REF_ID
    GROUP BY worker.DEPARTMENT, title.WORKER_TITLE
    ORDER BY total_salary DESC;
""")

df6
```

The result of Question 6

WORKER_TITLE varchar	DEPARTMENT varchar	total_salary int128
Asst. Manager	Admin	500000
Manager	Admin	500000
Lead	HR	300000
Lead	Account	200000
Executive	Admin	170000
Manager	HR	100000
Executive	Account	75000

Question 7: Extra question Show the names of position with the highest combined salaries and bonus in each department

```
df7 = duckdb.sql("""
    SELECT
        worker.DEPARTMENT,
        title.WORKER_TITLE,
        (SUM(worker.SALARY) + SUM(bonus.BONUS_AMOUNT)) AS total_salary_bonus
    FROM worker
    INNER JOIN bonus
    ON worker.WORKER_ID = bonus.WORKER_REF_ID
    INNER JOIN title
    ON bonus.WORKER_REF_ID = title.WORKER_REF_ID
    GROUP BY worker.DEPARTMENT, title.WORKER_TITLE
    ORDER BY total_salary_bonus DESC;
""")
```

df7

The result of Question 7

DEPARTMENT varchar	WORKER_TITLE varchar	total_salary_bonus int128
HR	Lead	304000
HR	Manager	209500
Admin	Executive	166500

Dashboard Part

On this part, it's about the dashboard that I built using PowerBI on my local computer. This dashboard includes two slides:

- Salaries Expense by Department
- Salaries Expense by Bonus and Work Title

Before you view the dashboard, I want to mention that I'm not very skilled at creating great dashboards, but I've done my best. Lastly, I attach PowerBI files as a **PDF** and `.pbix`, feel free to see it. Thank you.

Salaries Expenses by Department

As you can see in the picture below, the total salary is 3.68 million, which the company expenses to its employees. Therefore, the average is 229K.

Salaries Expense by Department

Account Admin HR

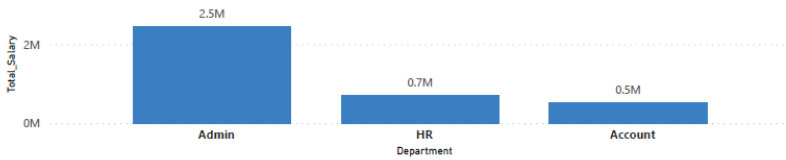
3.68M

Total Salary

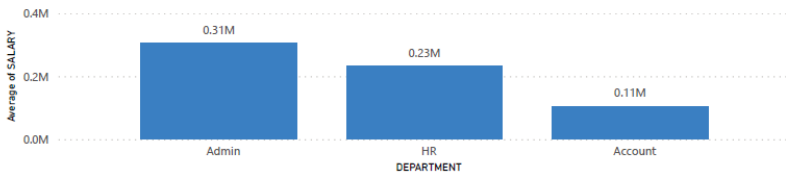
229.69K

Average of SALARY

Total_Salary by Department

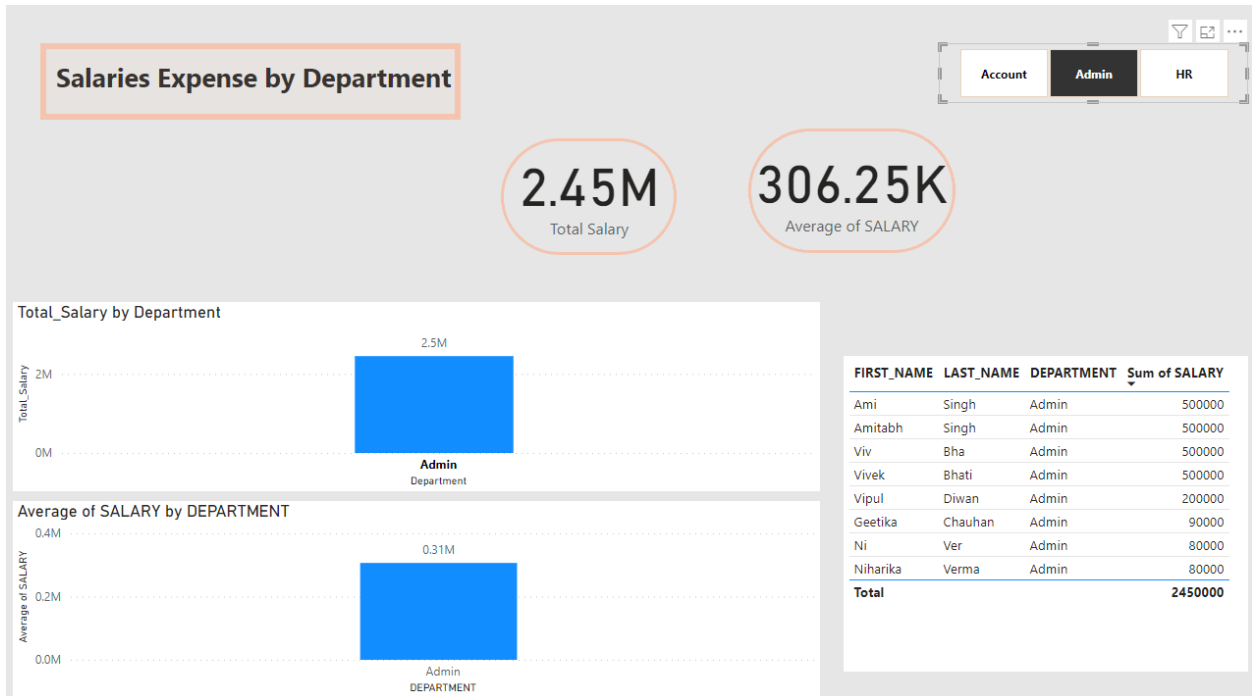


Average of SALARY by DEPARTMENT



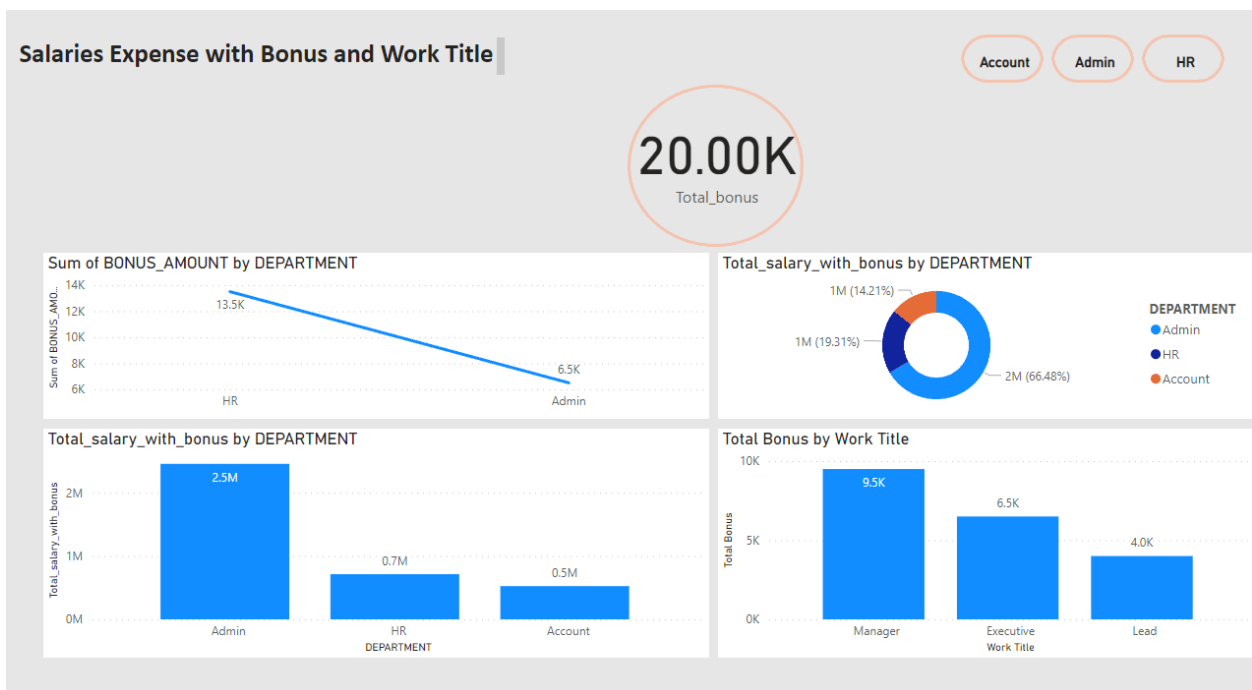
FIRST_NAME	LAST_NAME	DEPARTMENT	Sum of SALARY
Ami	Singh	Admin	500000
Amitabh	Singh	Admin	500000
Viv	Bha	Admin	500000
Vivek	Bhati	Admin	500000
Vi	Sing	HR	300000
Vishal	Singhal	HR	300000
Vipul	Diwan	Account	200000
Vipul	Diwan	Admin	200000
Satish	Kumar	Account	150000
Monika	Arora	HR	100000
Geetika	Chauhan	Admin	90000
Total			3675000

By the way, I want to know which department has the highest salary expenses. In the chart, you can see that it shows the Admin department as having the highest total sum, followed by the Human Resources and Accounting departments, with a combined total of 2.4 million. The average salary is 306K.

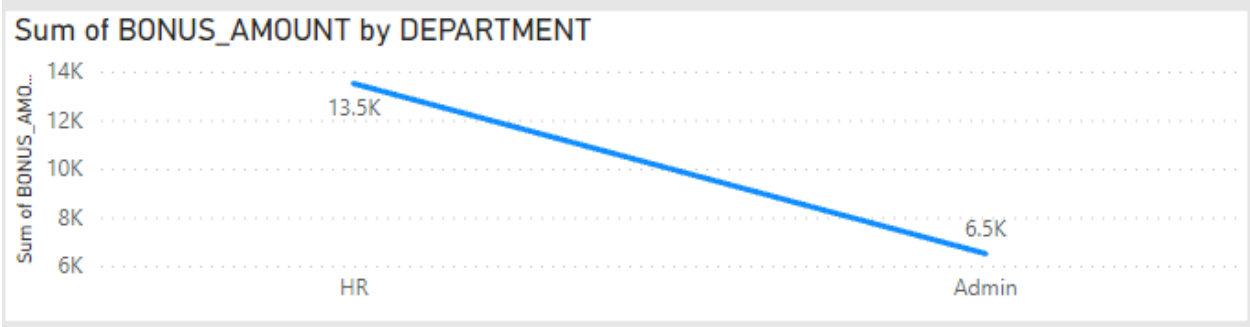


Salaries Expenses with bonus and work title

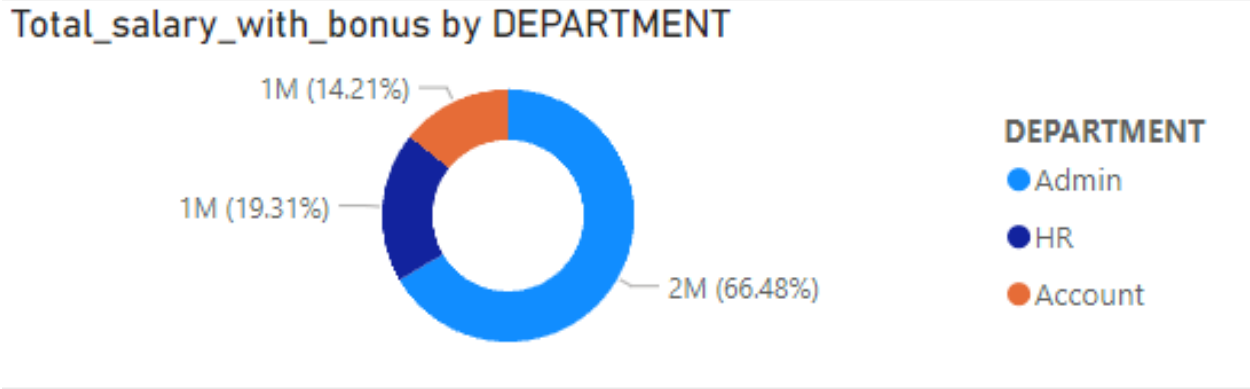
On this dashboard, there are some employees who receive bonuses that vary by department. I want to determine the total bonus amount received by each department.



If you look at the line chart, there are only two departments that receive a bonus: Human Resources and Administration. Among them, the HR department receives the highest total bonus amount, at \$13,500, which is \$6,500 more than the Administration department.



But the HR department does not have the highest total salary with bonus, as shown in the chart below. The total salary in the HR department is only 19.31%, whereas the admin department constitutes 66.48% of the total salary with bonus.



Lastly, if you look at this chart, you will notice that the work titles receiving bonuses are in the top positions, with Manager, Executive, and Lead receiving the highest bonuses.

Total Bonus by Work Title

