Deliverable 4
Phase 3

CSI4142 : Fundamentals of Data Science
winter 2023

uOttawa

Faculté de génie
Faculty of Engineering

School of Electrical Engineering and Computer Science
University of Ottawa

Professeur: Yazan Otoum, Ph.D

Group members :
Souleymane Wilfried Sankara, #300100940
Marie Fabienne Sawadogo; #300101795
Tata Saidatou Berte, #300100935

PART A1) OLAP QUERIES

DRILL DOWN

1) Drill down query from Job Classification to Job Title

SELECT job_classification, CAST(AVG(min_salary) AS numeric(10,3))
FROM job_tableAverage minimum salary in all data scientist job classifications

GROUP BY job_classification;

- Average minimum salary in data scientist jobs

SELECT job_title, CAST(AVG(min_salary) AS numeric (10,3))
FROM job_table
WHERE job_classification = 'data scientist
GROUP BY job_title;

- Average minimum salary in Senior data scientist jobs

SELECT job_title, CAST(AVG(min_salary) AS numeric (10,3))
FROM job_Table
WHERE job_title = 'Senior Data Scientist'
GROUP BY job_title;

2) Drill down query from Industry to company_name

- Average of average salary in the industries that hire data scientist

SELECT industry, (avg_salary)
FROM job_table
GROUP BY industry;

- Average of average salary in the Security Services industry

SELECT company_name, MAX(avg_salary)
FROM job_table
WHERE industry = 'Security Services'

GROUP BY company_name;

- Average of average salary in the Praetorian 4.7 company

```
SELECT company_name, MAX(avg_salary)
FROM job_table
WHERE company_name = 'Praetorian  4.7'
GROUP BY company_name;
```

ROLL UP

3) Roll up query from Job title to Job Classification

- Average minimum salary in Senior data scientist jobs

```
SELECT job_title, CAST(AVG(min_salary) AS numeric (10,3))
FROM job_Table
WHERE job_title = 'Senior Data Scientist'
GROUP BY job_title;
```

- Average minimum salary in data engineer jobs

```
SELECT job_title, CAST(AVG(min_salary) AS numeric (10,3))
FROM job_table
WHERE job_classification = 'data engineer'
GROUP BY job_title;
```

- Average minimum salary in all data scientist job classifications

```
SELECT job_classification, CAST(AVG(min_salary) AS numeric(10,3))
FROM job_table
GROUP BY job_classification;
```

4) ROLL UP query from company_name to Industry

- Average of average salary in the Praetorian 4.7 company

```
SELECT company_name, MAX(avg_salary)
```

FROM job_table
WHERE company_name = 'Praetorian 4.7'
GROUP BY company_name;

- Average of average salary in the Security Services industry

SELECT company_name, MAX(avg_salary)
FROM job_table
WHERE industry = 'Security Services'
GROUP BY company_name;

- Average of average salary in the industries that hire data scientist

SELECT industry, (avg_salary)
FROM job_table
GROUP BY industry;

SLICE

1) Sum of average salary of data scientist jobs in New York

SELECT job_title , SUM (avg_salary)
FROM job_table
WHERE city = 'New York'
GROUP BY job_title;

2) Sum of min salary of data scientist jobs with company size between 5001 and 10000

SELECT job_title , SUM (min_salary)
FROM job_table
WHERE size_range = '5001-10000''
GROUP BY job_title;

DICE

Sum of average salary of data scientist jobs in New York with company size is between 5001
and 10000

```
SELECT job_title , SUM (avg_salary)
FROM job_table
WHERE city = 'New York'
AND size_range = '5001-10000'
GROUP BY job_title;
```

COMBINING QUERIES

This query slices the data by Type_of_ownership = 'Private', drills down by City and Industry, computes the average of the Avg_Salary column, and returns only those rows where the average salary is greater than 75. The result is then ordered by City in descending order and Industry in ascending order

```
SELECT city, industry, AVG(avg_Salary) AS avg_Salary
FROM job_table
WHERE type_of_ownership = 'Company - Private'
GROUP BY city, industry
HAVING AVG(Avg_Salary) > 75
ORDER BY City DESC, Industry ASC;
```

PART A2)

a- Iceberg Query

The 5 companies with the highest sum of average salary

```
SELECT company_name, SUM(avg_salary) AS total_salary
FROM job_table
GROUP BY company_name
ORDER BY total_salary DESC
LIMIT 5;
```

b- Windowing queries

Rank the companies according to the average of their average salaries

```
SELECT company_name, AVG(avg_salary) AS avg_salary, RANK() OVER (ORDER BY
AVG(avg_salary) DESC) AS rank
FROM job_table
GROUP BY company_name
ORDER BY avg_salary DESC
```
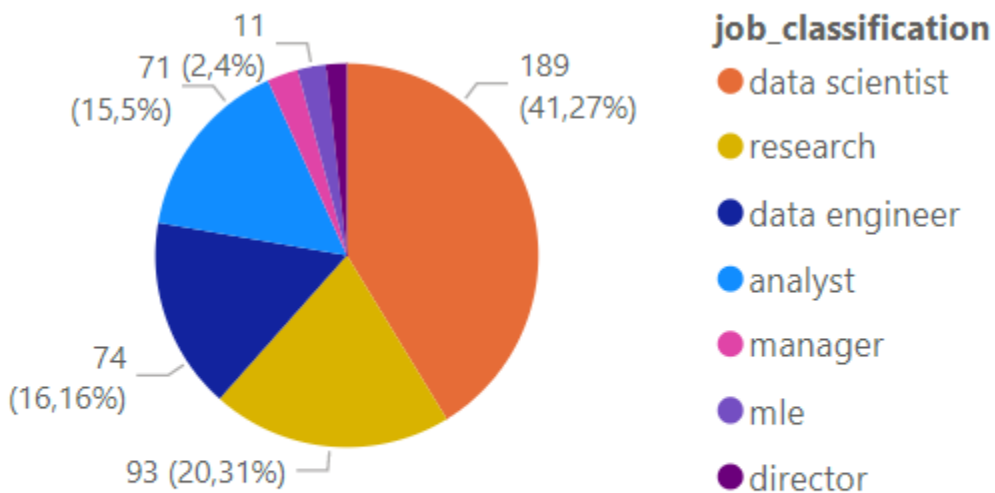
c- Using the window clause

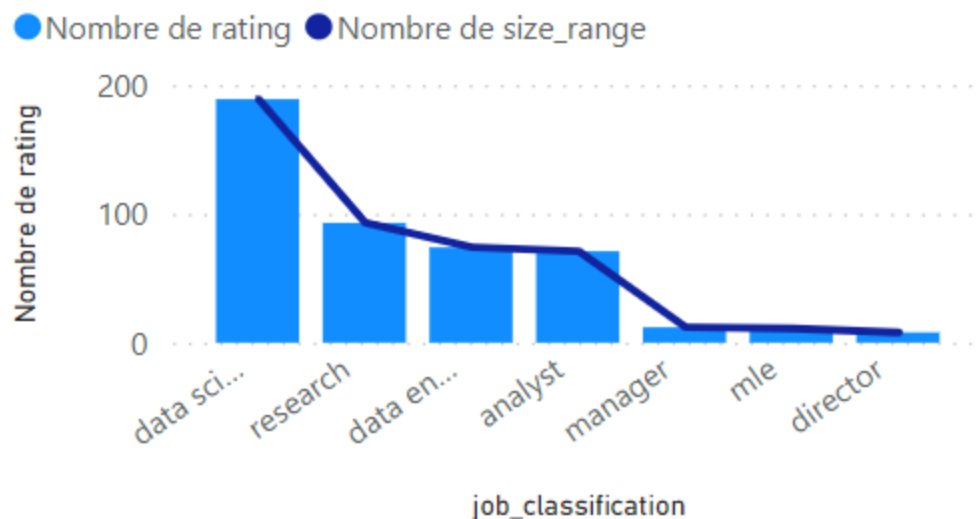Comparison of average of minimum salary and average of average salary

```
SELECT avg(min_salary) OVER w, avg(avg_salary) OVER w
  FROM job_table
  WINDOW w AS (PARTITION BY job_classification ORDER BY avg_salary DESC);
```
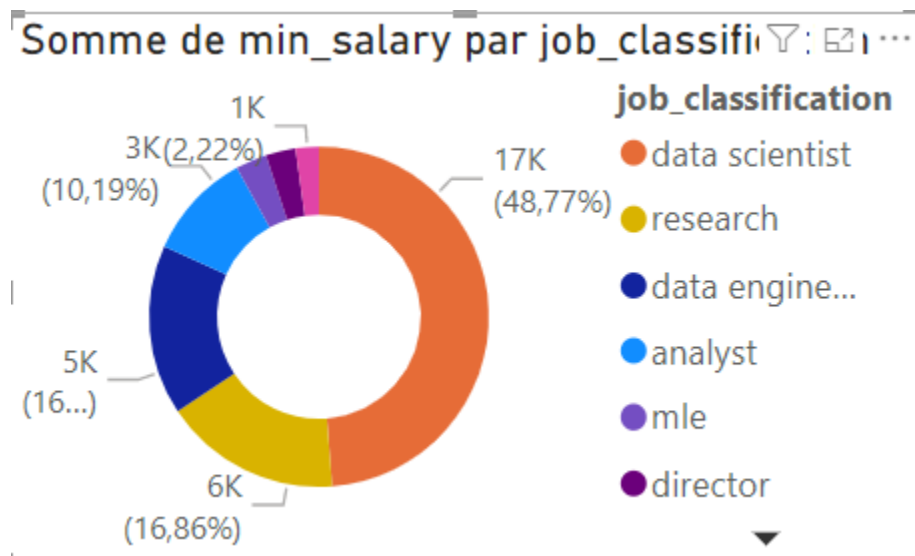
Part B) BI dashboard

Exploration figures

Nombre de rating et Nombre de size_range par job_classification
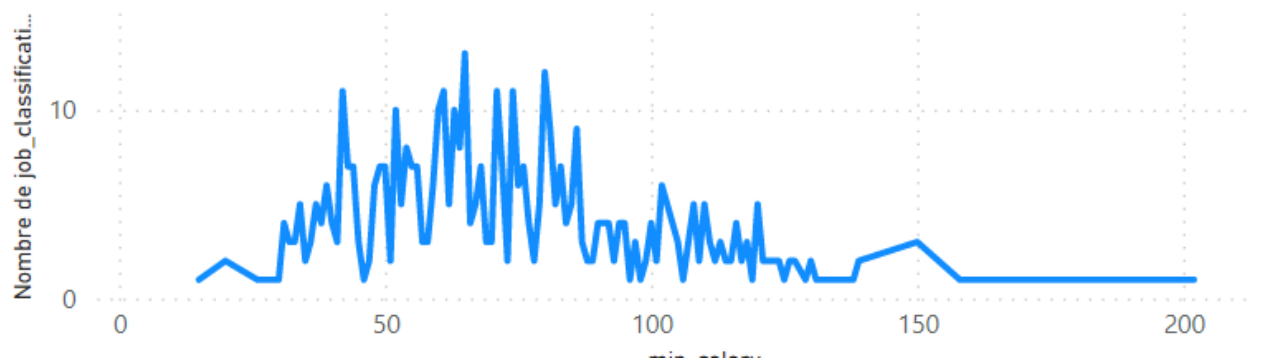
Drill down Query: Average minimum salary in all data scientist job classifications

Somme de min_salary par job_classification

Slice
Sum of min salary of data scientist jobs with company size between 5001 and 10000

Nombre de job_classification par min_salary

Dice

Sum of average salary of data scientist jobs in New York with company size is between 5001 and 10000

Nombre de job_classification par min_salary

Roll Up

- Average minimum salary in all data scientist jobs

Nombre de avg_salary par job_classification

Bar chart showing job_classification on the y-axis and Nombre de avg_salary on the x-axis:
- data scientist: ~190
- research: ~90
- data engineer: ~70
- analyst: ~65
- manager: ~12
- mle: ~12
- director: ~8

Iceberg Queries

The 5 companies with the highest sum of average salary

Somme de max_salary et Somme de min_salary par company

● Somme de max_salary ● Somme de min_salary

Area chart with company_name on the x-axis:
- Takeda Pharmaceut... 3.7
- MassMutual 3.6
- Liberty Mutual Insurance 3.3
- Novetta 4.0
- AstraZeneca 3.9