



Data Engineering Cohort 1

Module 5

Assignment 5.4

Earning Prediction AWS

Member #1

Name: Muhammad Humza

Roll Number: 2211-018-KHI-DEG

Member #2

Name: Muhammad Ovais

Roll Number: 2211-023-KHI-DEG

Task

Activity tasks/4_data_pipelines/day_4_data_lake/data/output_data/employee_earnings

Using the data manipulation tool of your choice (eg. Python) simulate the earnings predictions for 2 more days. Load it to the Data Lake that you've created today

Rerun queries from Task 3 and Task 4 and see how the results change with this new data.

Create a new query in Athena that calculates the % change in earnings for every employee from a given day compared to the previous day.

Solution

Step 1: using python file to predict earnings

The python file is attached in the .zip file you can find it in
Assignment.ipynb format

Step 2: 5.2 Assignment

All the steps are similar to assignment 5.2 you can refer
to it for your understanding

Step 3: SQL Query to find difference in earnings

QUERY

```
SELECT emp_id, full_name, earnings, earnings_date, (earnings - previous_earnings) / previous_earnings
FROM ( SELECT emp_id, CONCAT(first_name, ' ', middle_initial, ' ', last_name) AS full_name,
LAG(CAST(earnings as double), 1) OVER (PARTITION BY emp_id ORDER BY earnings_date FROM
"ovaissaleem_glue_database"."muhammadhumza_employee_earnings") t WHERE earnings_date
BETWEEN '2022-02-14' AND '2022-02-16'
```

muhammadhumza_employee_earnings

Partitioned

ovaissaleem_athena_employee_earnings

Partitioned

Views (0)

< 1 >

Run again

Explain

Cancel

Clear

Create

Reuse query results
*Athena engine version 3 only

Query results

Query stats

Completed

Time in queue: 151 ms

Run time: 1.328 sec

Data scanned: 24.36 KB

Results (100+)

Copy

Download results

Search rows

< 1 ... >

#	emp_id	full_name	earnings	earnings_date	percentage_change
1	220965	Almeta Y Brookins	9378.0	2022-02-14	39.532807617914
2	220965	Almeta Y Brookins	11243.0	2022-02-15	19.886969503092345
3	220965	Almeta Y Brookins	13108.0	2022-02-16	16.588099261762874
4	316372	Alexander C Goad	6686.0	2022-02-14	71.87660668380462
5	316372	Alexander C Goad	8038.0	2022-02-15	20.2213580616213
6	316372	Alexander C Goad	9390.0	2022-02-16	16.820104503607862
7	597741	Tonya F Wilson	9094.0	2022-02-14	189.24936386768448
8	597741	Tonya F Wilson	11621.0	2022-02-15	27.787552232241037
9	597741	Tonya F Wilson	14149.0	2022-02-16	21.75372171069615
10	622405	Harrison X Hawk	3974.0	2022-02-14	-24.791824375473126
11	622405	Harrison X Hawk	6743.0	2022-02-15	69.67790639154504
12	622405	Harrison X Hawk	9512.0	2022-02-16	41.06480794898413