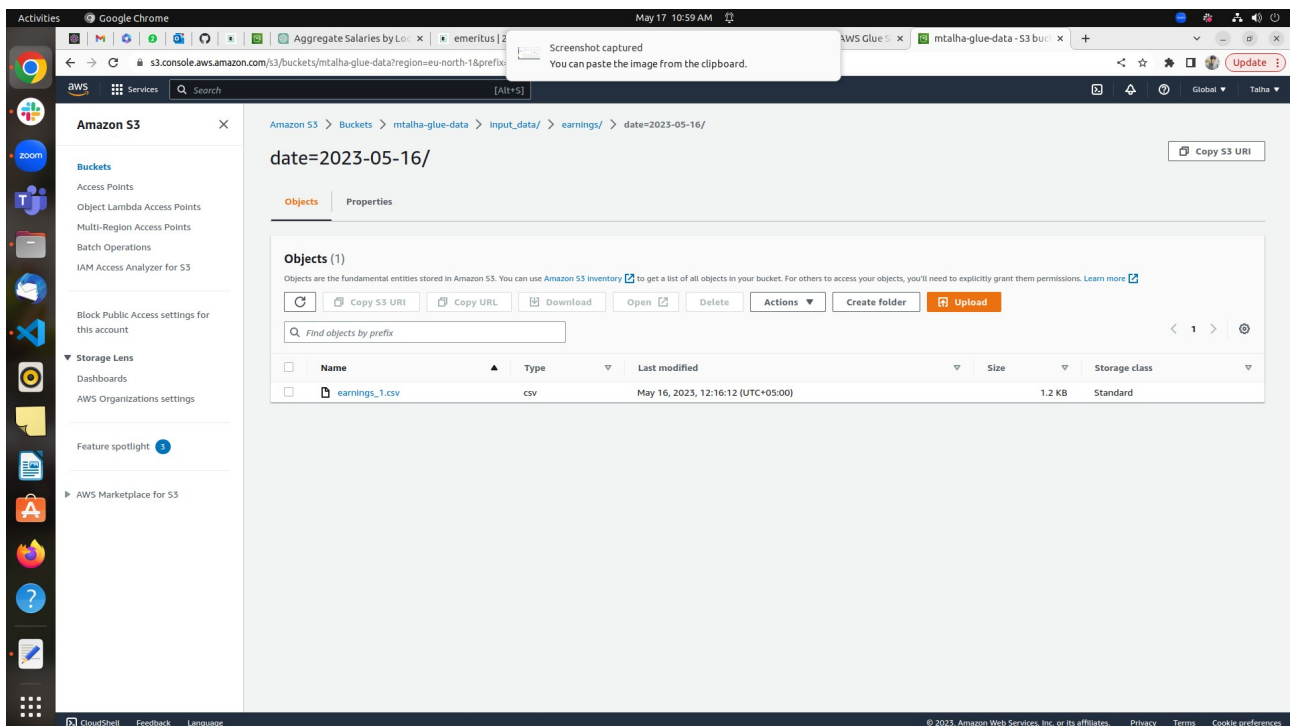
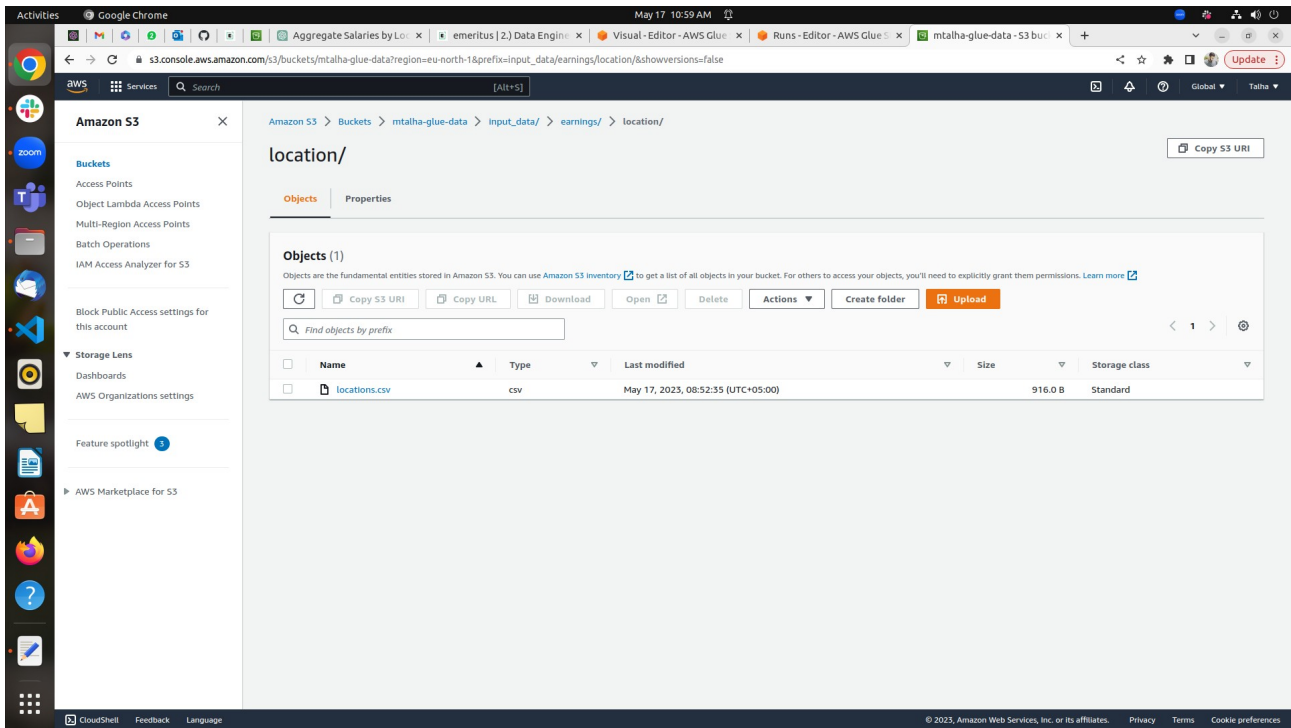
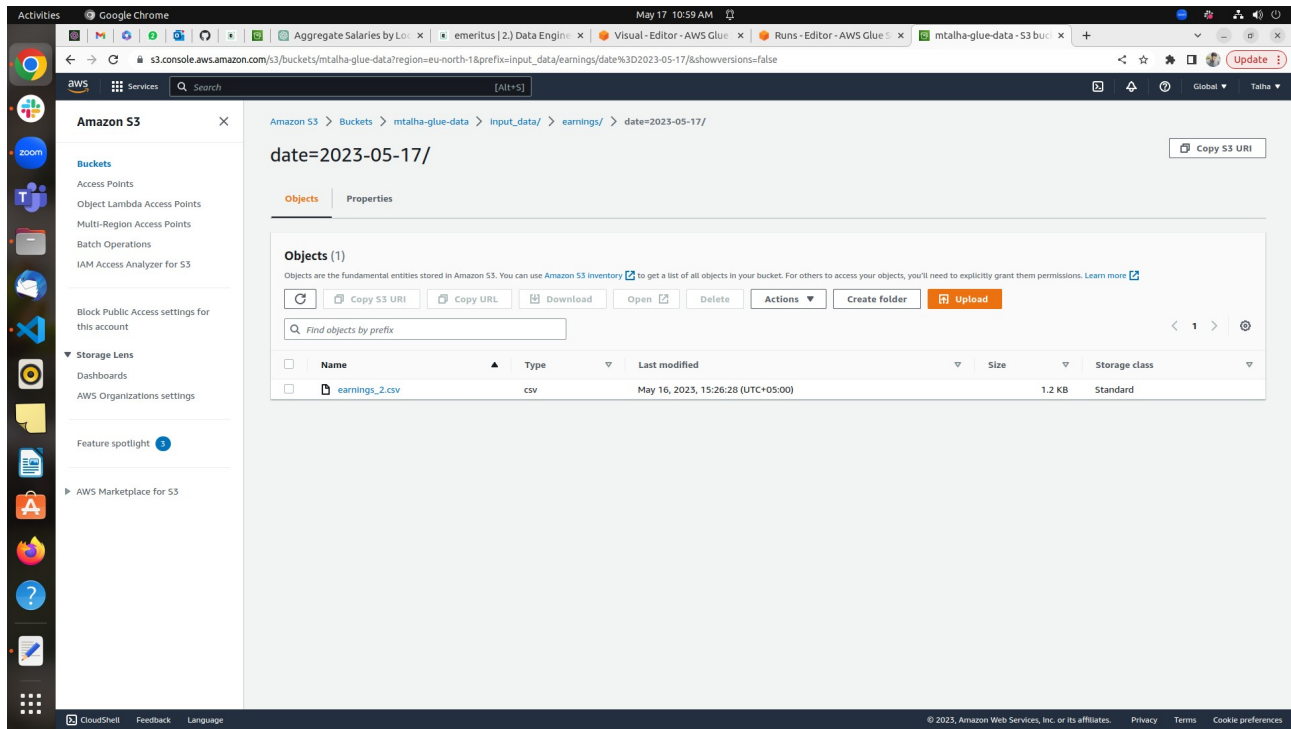


**Using the earnings CSV as a base, prepare a new data file with employees' office locations. Make sure there are 5-6 distinct locations that are shared between employees.**

**Create a Glue job that aggregates the data based on the office location to calculate average salaries and raise percentages for these locations.**

## Uploading data into S3 bucket





Created Crawler

Activities Google Chrome May 17 10:59 AM

eu-north-1.console.aws.amazon.com/glue/home?region=eu-north-1#/v2/data-catalog/crawlers

**AWS Glue**

Getting started  
ETL jobs  
Visual ETL  
Notebooks  
Job run monitoring  
Data Catalog tables  
Data connections  
Workflows (orchestration)

▼ Data Catalog  
Databases  
Tables  
Stream schema registries  
Schemas  
Connections  
**Crawlers**  
Classifiers  
Catalog settings

► Data Integration and ETL  
► Legacy pages

What's New  
Documentation  
AWS Marketplace

Enable compact mode  
Enable new navigation

CloudShell Feedback Language

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

One crawler successfully deleted  
The following crawler is now deleted: "mtalha\_rds\_crawler"

AWS Glue > Crawlers

**Crawlers** info  
View and manage all available crawlers.

Filter crawlers

Table with 8 columns: Name, State, Schedule, Last run, Last run timestamp, Log, Table changes...  
Rows: mतालha\_s3\_earnings\_crawler (Ready, Succeeded, May 16, 2023 at 10:33:09, 1 updated), mतालha\_s3\_location (Ready, Succeeded, May 17, 2023 at 04:05:31, -)

## Created Glue Job

Activities Slack May 17 11:03 AM

eu-north-1.console.aws.amazon.com/gluestudio/home?region=eu-north-1#/editor/job/mtalha\_employee\_earnings\_job/graph

**AWS Glue**

Getting started  
ETL jobs  
Visual ETL  
Notebooks  
Job run monitoring  
Data Catalog tables  
Data connections  
Workflows (orchestration)

▼ Data Catalog  
Databases  
Tables  
Stream schema registries  
Schemas  
Connections  
Crawlers  
Classifiers  
Catalog settings

► Data Integration and ETL  
► Legacy pages

What's New  
Documentation  
AWS Marketplace

Enable compact mode  
Enable new navigation

CloudShell Feedback Language

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

mtalha\_employee\_earnings\_job  
Last modified on 5/17/2023, 11:01:27 AM Try new UI End session Actions

Successfully updated job  
Successfully updated job mतालha\_employee\_earnings\_job. To run the job choose the Run Job button.

Visual Script Job details Runs Schedules Version Control

Source Action Target Undo Redo Remove

Diagram showing data flow: Data source - S3 bucket Amazon S3 → Transform - Join → Transform - SQL Query → Data target - S3 bucket Amazon S3

Transform - SQL Query  
SQL Query

Data preview (5) info  
Previewing 2 of 2 fields  
Filter sample dataset

location	avg_earning
B	6286.75
C	5576.95
A	5926.05
D	5889.7
E	5599.2

# Running the job

The screenshot displays the AWS Glue console interface for the job `mtalha_employee_earnings_job`. The interface includes a left-hand navigation menu with options like 'Getting started', 'Data Catalog', 'Data integration and ETL', and 'Legacy pages'. The main content area shows the 'Runs' tab for the job, displaying a table of job runs. The most recent run is 'Failed' with an error message: 'AnalysisException: Expressions referencing the outer query are not supported outside of WHERE/HAVING clauses:'. The console also shows job details like 'Job name', 'Id', 'Run status', 'Glue version', 'Retry attempt number', 'Start time', 'End time', 'Start-up time', 'Execution time', 'Last modified on', 'Trigger name', and 'Security configuration'.

**Job runs (1/26)**

Run status	Retry	Start time	End time	Duration	Capacity	Worker type	Glue version
Failed	0	05/17/2023 11:01:28	05/17/2023 11:02:40	1 m 5 s	3 DPU's	G.1X	3.0
Failed	0	05/17/2023 10:56:39	05/17/2023 10:57:40	54 s	3 DPU's	G.1X	3.0
Succeeded	0	05/17/2023 10:49:16	05/17/2023 10:50:17	54 s	3 DPU's	G.1X	3.0
Succeeded	0	05/17/2023 10:45:02	05/17/2023 10:46:06	57 s	3 DPU's	G.1X	3.0
Failed	0	05/17/2023 10:43:23	05/17/2023 10:44:18	47 s	3 DPU's	G.1X	3.0

**05/17/2023 10:58:36**

**AnalysisException: Expressions referencing the outer query are not supported outside of WHERE/HAVING clauses:**

**Job name** `mtalha_employee_earnings_job` **Id** `jr_4ecc3ab6bda4cf06513a0a271e08e04951ffc61f89737e20bd16bd595acca` **Run status** `Failed` **Glue version** `3.0`

**Retry attempt number** `Initial run` **Start time** `May 17, 2023 10:58:36 AM` **End time** `May 17, 2023 10:59:28 AM` **Start-up time** `7 seconds`

**Execution time** `45 seconds` **Last modified on** `May 17, 2023 10:59:28 AM` **Trigger name** `-` **Security configuration** `-`

# Sql Query & query output showing the result Avg salary based on the location

**AWS Glue** **mtalha\_employee\_earnings\_job** Last modified on 5/17/2023, 10:58:34 AM

Successfully started job mtalha\_employee\_earnings\_job. Navigate to Run details for more details.

**Visual** Script Job details Runs Schedules Version Control

Source Action Target Undo Redo Remove

Transform - Join Join

Transform - SQL Query SQL Query

Data target - S3 bucket Amazon S3

**Transform** Output schema Data preview

Name: SQL Query

Node parents: Choose one or more parent nodes

Join Join - Transform

Associate an alias with each input source

Input sources: Join SQL aliases: myDataSource

SQL query: Enter a SQL statement to add to your job.

```
1 WITH averages AS (  
2   SELECT  
3     location,  
4     avg(earnings) as avg_earning  
5   FROM  
6     myDataSource  
7   GROUP BY  
8     location  
9 )  
10 SELECT  
11   location,  
12   (avg_earning - (SELECT MIN(avg_earning) FROM myDataSource) / (SELECT MIN(avg_earning) FROM myDataSource))  
13 FROM
```

**AWS Glue** **mtalha\_employee\_earnings\_job** Last modified on 5/17/2023, 11:01:27 AM

Successfully updated job mtalha\_employee\_earnings\_job. To run the job choose the Run Job button.

**Visual** Script Job details Runs Schedules Version Control

Source Action Target Undo Redo Remove

Transform - Join Join

Transform - SQL Query SQL Query

Data target - S3 bucket Amazon S3

**Data preview (5)** Info

Filter sample dataset

location	avg_earning
B	6286.75
C	5576.95
A	5926.05
D	5889.7
E	5599.2

**Query output showing the result raise percentage based on the location**

Activities Slack May 17 11:03 AM

eu-north-1.console.aws.amazon.com/gluestudio/home?region=eu-north-1#/editor/job/mtalha\_employee\_earnings\_job/graph

aws Services Search [Alt+S]

**AWS Glue** × mtalha\_employee\_earnings\_job

Last modified on 5/17/2023, 11:01:27 AM Try new UI End session Actions

Successfully updated job mtalha\_employee\_earnings\_job. To run the job choose the Run Job button.

Visual Script Job details Runs Schedules Version Control

Source Action Target Undo Redo Remove

Visual workflow diagram showing data flow from two Amazon S3 buckets through a Join transform, an SQL Query transform, and finally to an Amazon S3 target bucket.

Transform Output schema Data preview

Data preview (5) Info Previewing 2 of 2 fields

Filter sample dataset

location	avg_earning
B	6286.75
C	5576.95
A	5926.05
D	5889.7
E	5599.2

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

**Saving the data**



Activities

Google Chrome

May 17 11:00 AM

Aggregate Salaries by Loc... x emeritus [2] Data Engine x Visual - Editor - AWS Glue x Runs - Editor - AWS Glue x mta-ha-glue-data - S3 buc x

s3.console.aws.amazon.com/s3/buckets/mta-ha-glue-data?region=eu-north-1&prefix=output\_data/location\_wise\_data/&showversions=false

Update

Services Search [Alt+S]

Amazon S3

Buckets

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

AWS Organizations settings

Feature spotlight

AWS Marketplace for S3

Objects Properties

Objects (45)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	date=2023-05-17/	Folder	-	-	-
<input type="checkbox"/>	run-168429992088-part-block-0-r-00000-snappy.parquet	parquet	May 17, 2023, 10:06:36 (UTC+05:00)	1.2 KB	Standard
<input type="checkbox"/>	run-168429992088-part-block-0-r-00001-snappy.parquet	parquet	May 17, 2023, 10:06:36 (UTC+05:00)	1.3 KB	Standard
<input type="checkbox"/>	run-168429992088-part-block-0-r-00002-snappy.parquet	parquet	May 17, 2023, 10:06:36 (UTC+05:00)	1.2 KB	Standard
<input type="checkbox"/>	run-168429992088-part-block-0-r-00003-snappy.parquet	parquet	May 17, 2023, 10:06:36 (UTC+05:00)	1.5 KB	Standard
<input type="checkbox"/>	run-168429992088-part-block-0-r-00004-snappy.parquet	parquet	May 17, 2023, 10:06:36 (UTC+05:00)	1.3 KB	Standard
<input type="checkbox"/>	run-168429992088-part-block-0-r-00005-snappy.parquet	parquet	May 17, 2023, 10:06:36 (UTC+05:00)	1.3 KB	Standard
<input type="checkbox"/>	run-168429992088-part-block-0-r-00006-snappy.parquet	parquet	May 17, 2023, 10:06:36 (UTC+05:00)	1.3 KB	Standard
<input type="checkbox"/>	run-168429992088-part-block-0-r-00007-snappy.parquet	parquet	May 17, 2023, 10:06:36 (UTC+05:00)	1.4 KB	Standard
<input type="checkbox"/>	run-1684300177066-part-block-0-r-00000-snappy.parquet	parquet	May 17, 2023, 10:09:39 (UTC+05:00)	350.0 B	Standard
<input type="checkbox"/>	run-1684300177066-part-block-0-r-00001-snappy.parquet	parquet	May 17, 2023, 10:09:39 (UTC+05:00)	366.0 B	Standard
<input type="checkbox"/>	run-1684300177066-part-block-0-r-00002-snappy.parquet	parquet	May 17, 2023, 10:09:39 (UTC+05:00)	328.0 B	Standard
<input type="checkbox"/>	run-1684300177066-part-block-0-r-00003-snappy.parquet	parquet	May 17, 2023, 10:09:39 (UTC+05:00)	437.0 B	Standard
<input type="checkbox"/>	run-1684300177066-part-block-0-r-00004-snappy.parquet	parquet	May 17, 2023, 10:09:40 (UTC+05:00)	385.0 B	Standard
<input type="checkbox"/>	run-1684300177066-part-block-0-r-00005-snappy.parquet	parquet	May 17, 2023, 10:09:40 (UTC+05:00)	386.0 B	Standard
<input type="checkbox"/>	run-1684300177066-part-block-0-r-00006-snappy.parquet	parquet	May 17, 2023, 10:09:40 (UTC+05:00)	365.0 B	Standard
<input type="checkbox"/>	run-1684300177066-part-block-0-r-00007-snappy.parquet	parquet	May 17, 2023, 10:09:40 (UTC+05:00)	421.0 B	Standard
<input type="checkbox"/>	run-1684300409278-part-block-0-r-00001-snappy.parquet	parquet	May 17, 2023, 10:13:33 (UTC+05:00)	280.0 B	Standard
<input type="checkbox"/>	run-1684300409278-part-block-0-r-00002-snappy.parquet	parquet	May 17, 2023, 10:13:33 (UTC+05:00)	272.0 B	Standard
<input type="checkbox"/>	run-1684300409278-part-block-0-r-00003-snappy.parquet	parquet	May 17, 2023, 10:13:33 (UTC+05:00)	272.0 B	Standard

CloudShell Feedback Language

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences