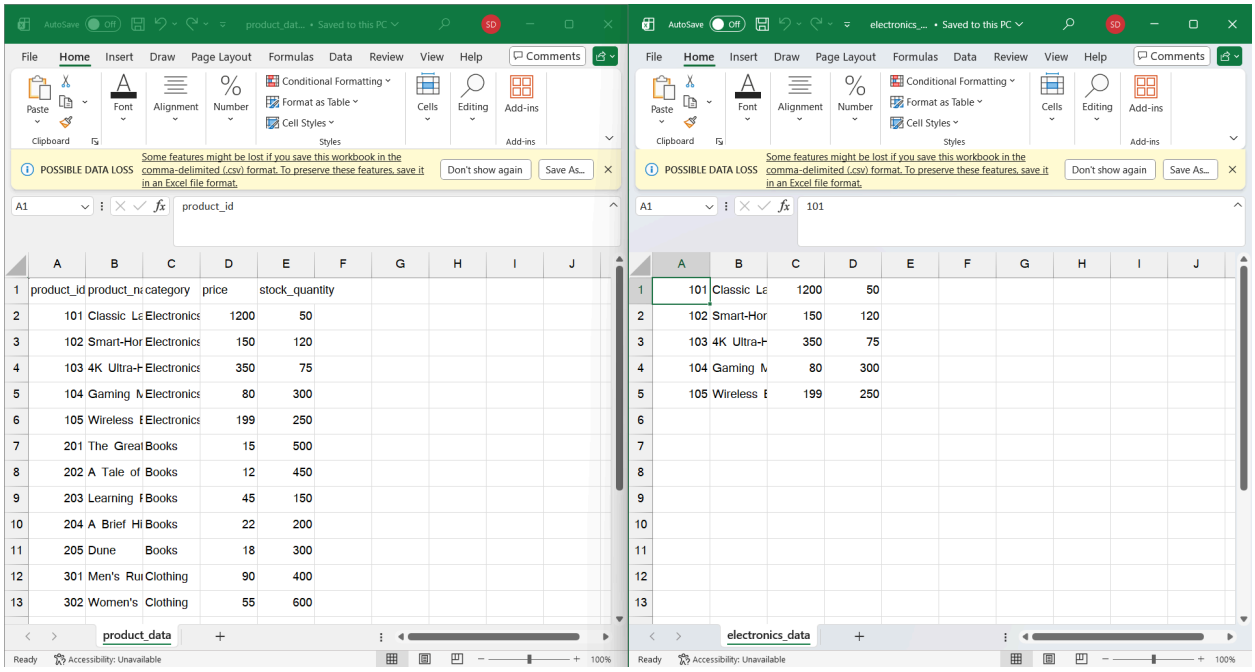
Practical - 6

Name: Sakshi Deshmukh
PRN: 202301040191
Roll no. 154

AWS Athena



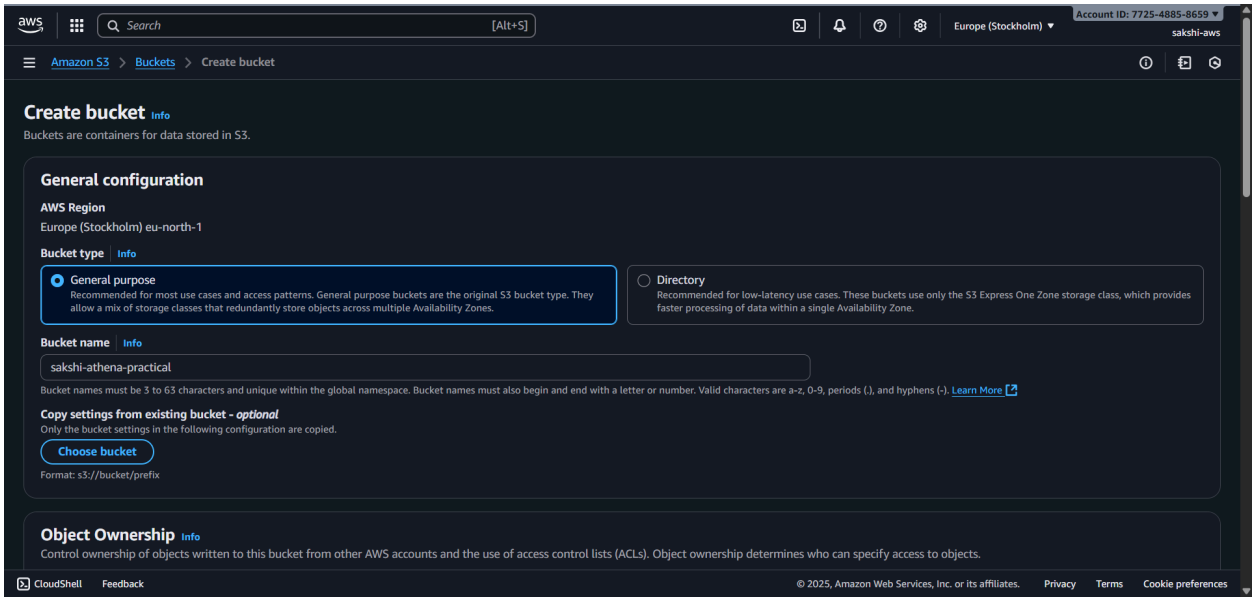
product_data

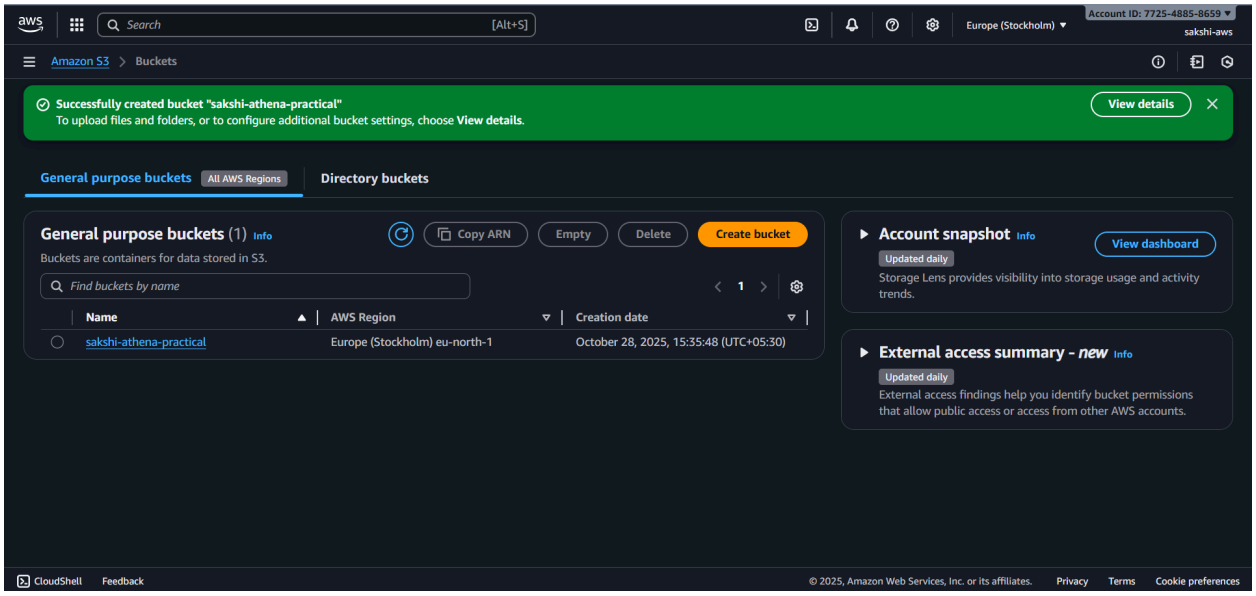
product_id	product_na	category	price	stock_quantity
101	Classic La	Electronics	1200	50
102	Smart-Hor	Electronics	150	120
103	4K Ultra-H	Electronics	350	75
104	Gaming M	Electronics	80	300
105	Wireless E	Electronics	199	250
201	The Great	Books	15	500
202	A Tale of	Books	12	450
203	Learning F	Books	45	150
204	A Brief Hi	Books	22	200
205	Dune	Books	18	300
301	Men's Ru	Clothing	90	400
302	Women's	Clothing	55	600

electronics_data

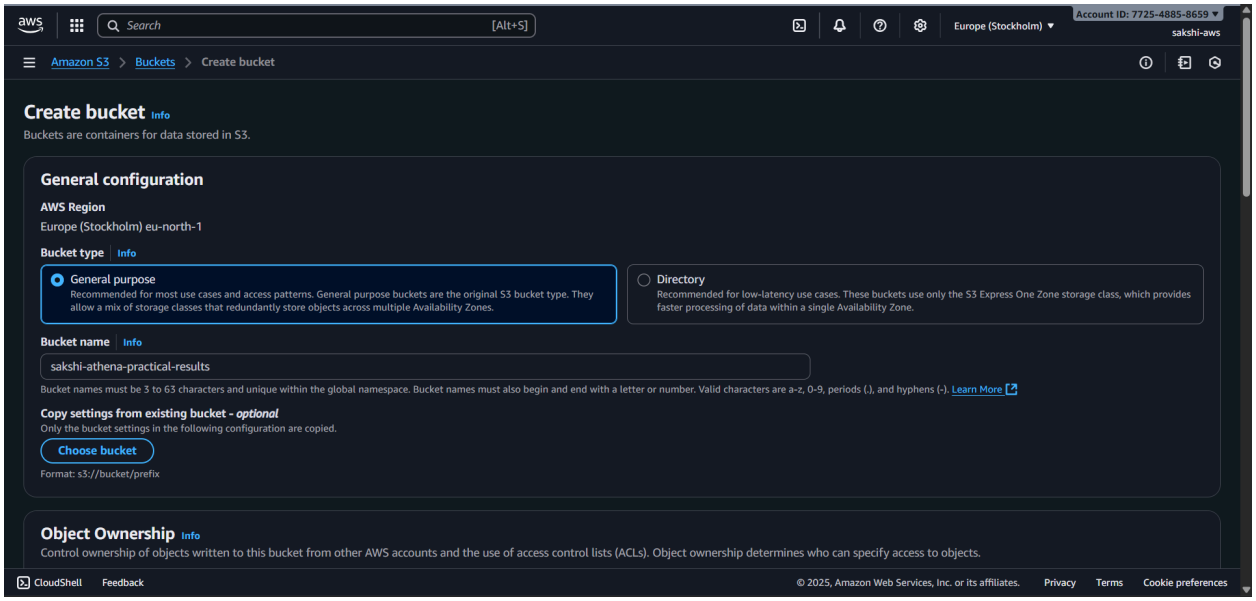
101	Classic La	1200	50	
102	Smart-Hor	150	120	
103	4K Ultra-H	350	75	
104	Gaming M	80	300	
105	Wireless E	199	250	

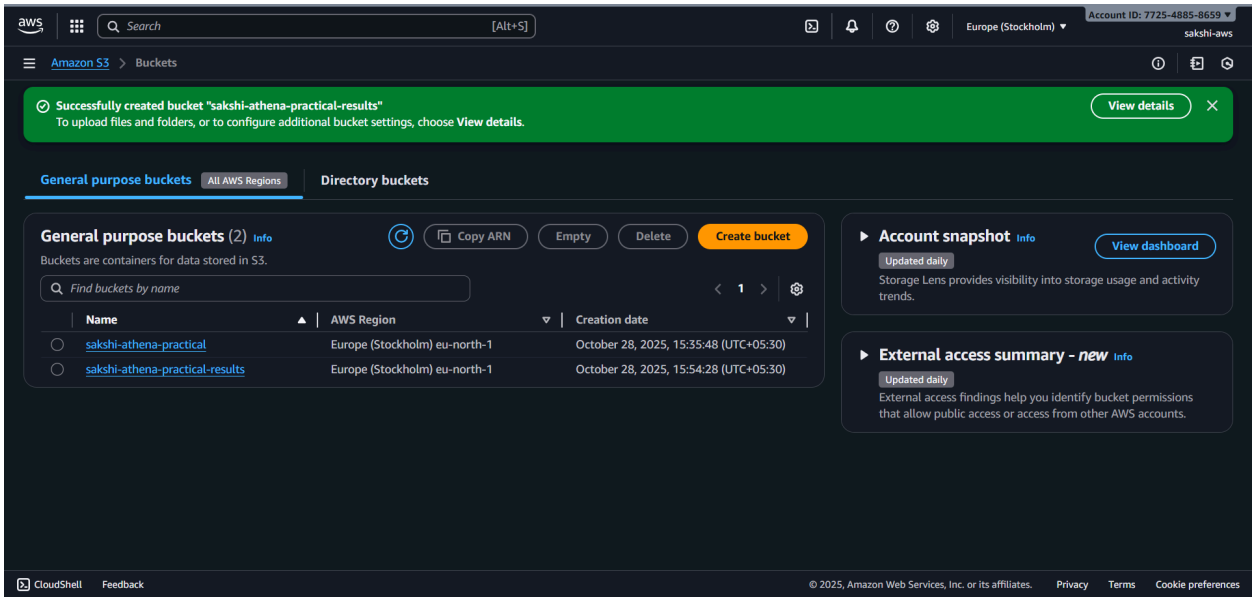
Setting up S3 bucket



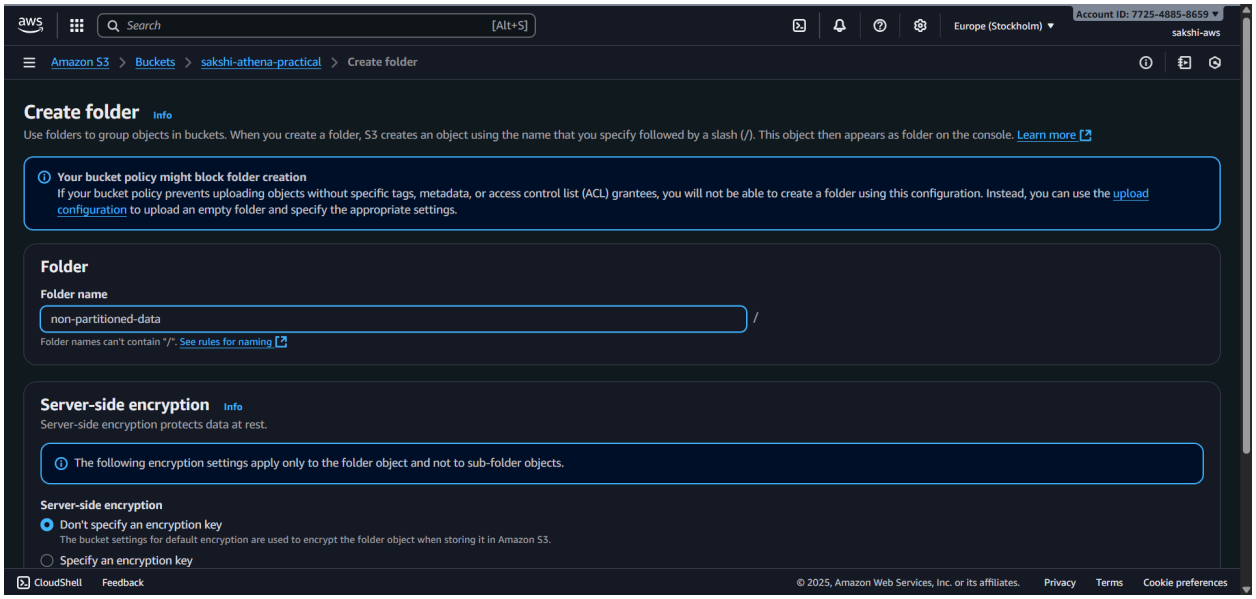


Bucket for storing results

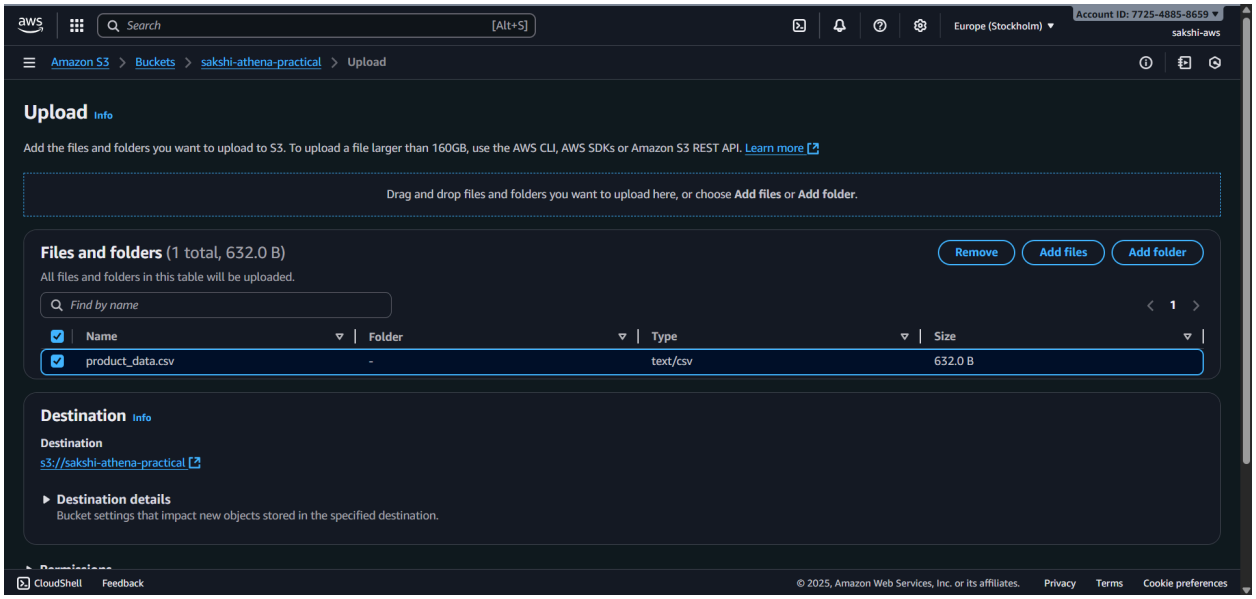


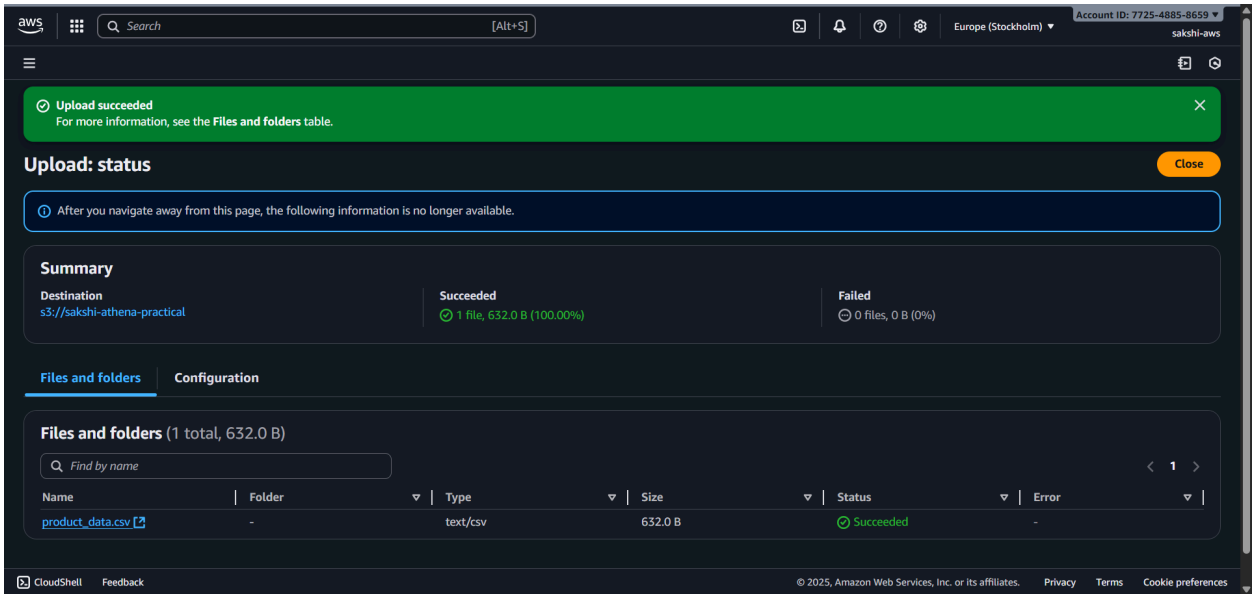


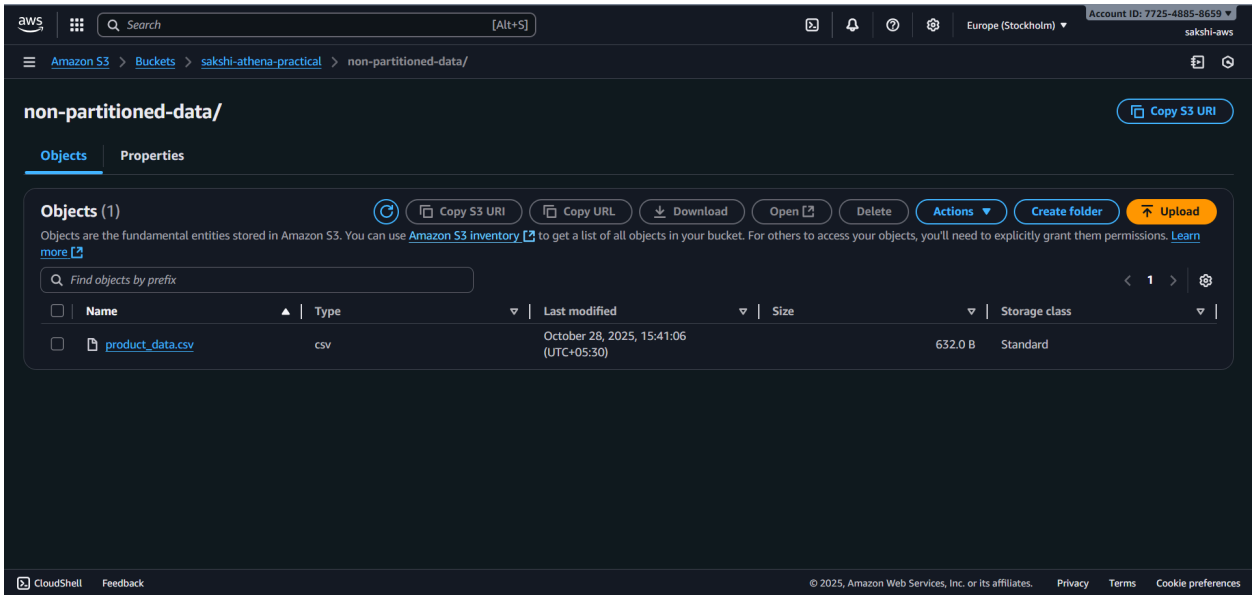
Creating folder for non-partitioned data



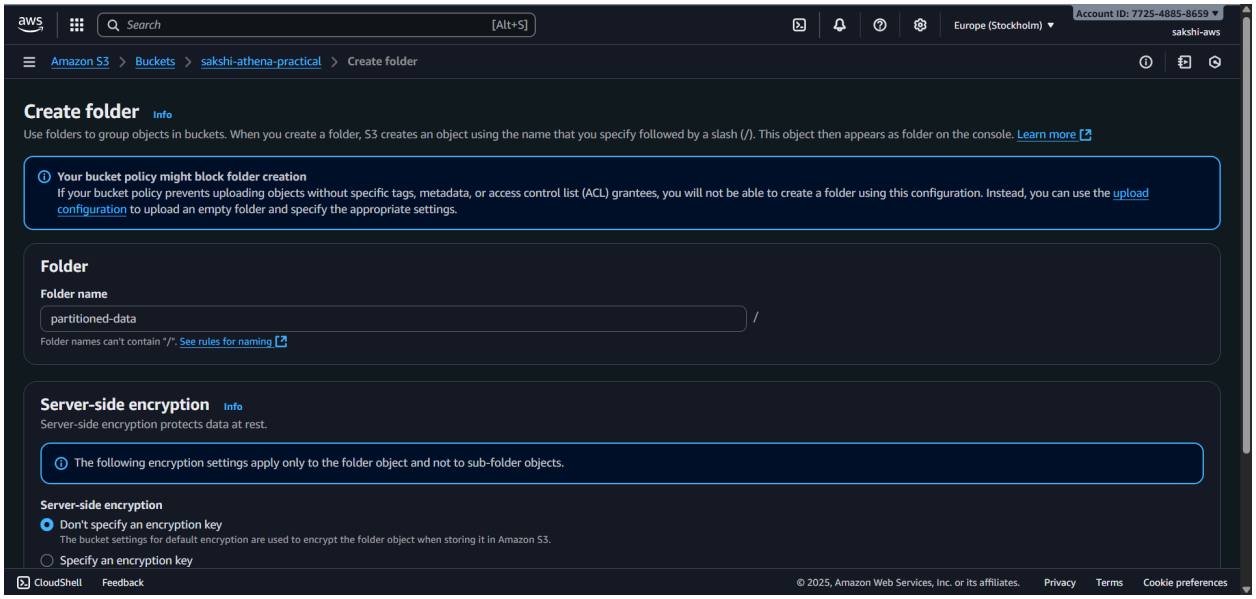
Upload data into folder



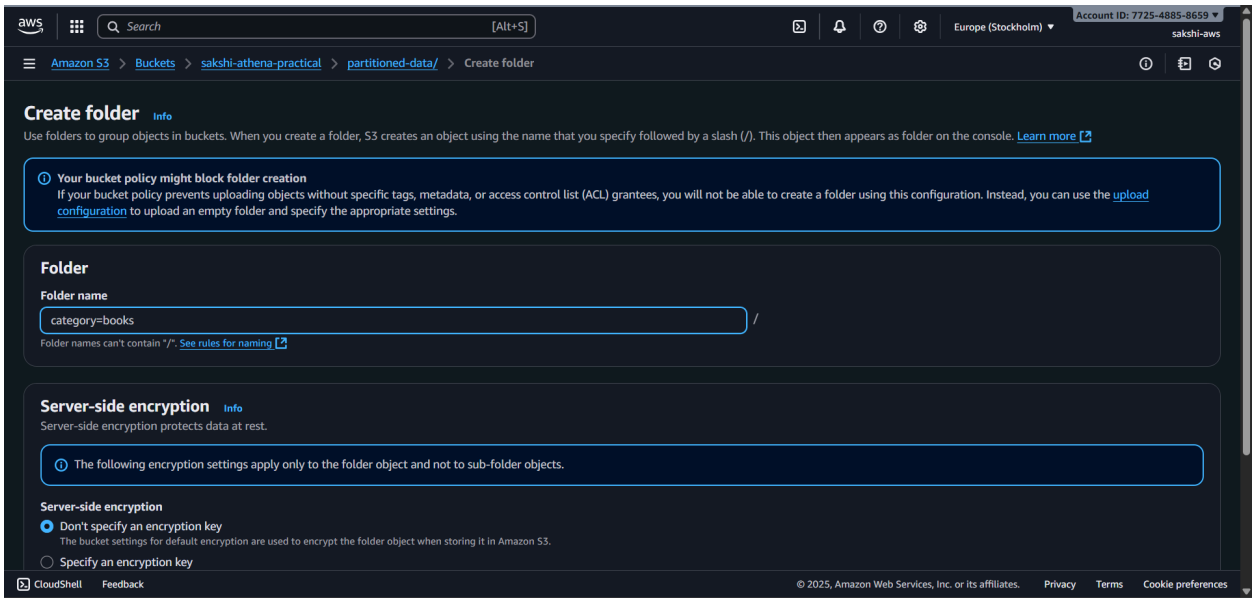

Search [Alt+S]
Europe (Stockholm)
Account ID: 7725-4885-8659
sakshi-aws
Upload succeeded
For more information, see the Files and folders table.
Upload: status
Close
After you navigate away from this page, the following information is no longer available.
Summary
Destination
s3://sakshi-athena-practical
Succeeded
1 file, 632.0 B (100.00%)
Failed
0 files, 0 B (0%)
Files and folders
Configuration
Files and folders (1 total, 632.0 B)
Find by name
Name
Folder
Type
Size
Status
Error
product_data.csv
-
text/csv
632.0 B
Succeeded
-
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

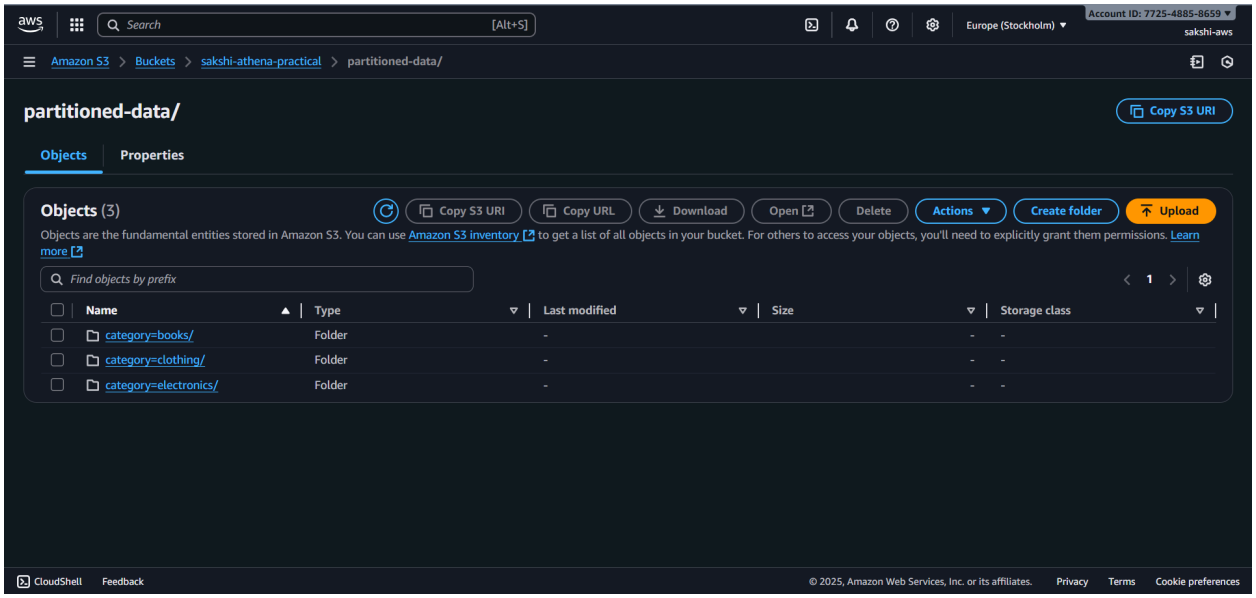

Search [Alt+S]
Europe (Stockholm)
Account ID: 7725-4885-8659
sakshi-aws
Amazon S3
Buckets
sakshi-athena-practical
non-partitioned-data/
non-partitioned-data/
Copy S3 URI
Objects
Properties
Objects (1)
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are the fundamental entities 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
Find objects by prefix
Name
Type
Last modified
Size
Storage class
product_data.csv
csv
October 28, 2025, 15:41:06 (UTC+05:30)
632.0 B
Standard
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Creating folder for partitioned data

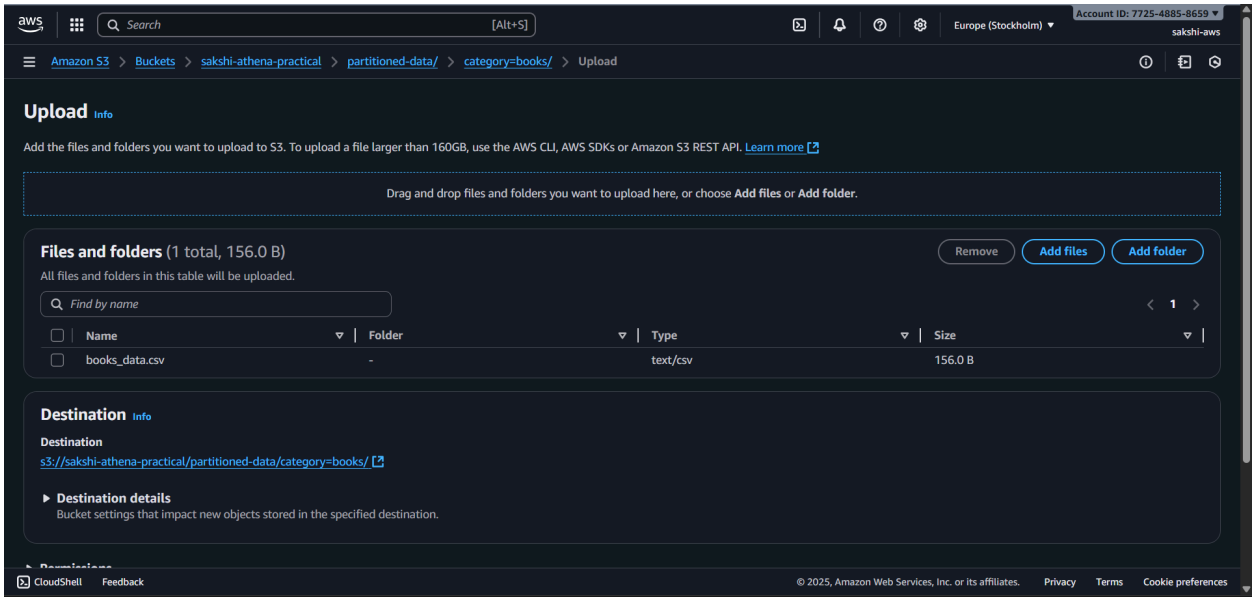


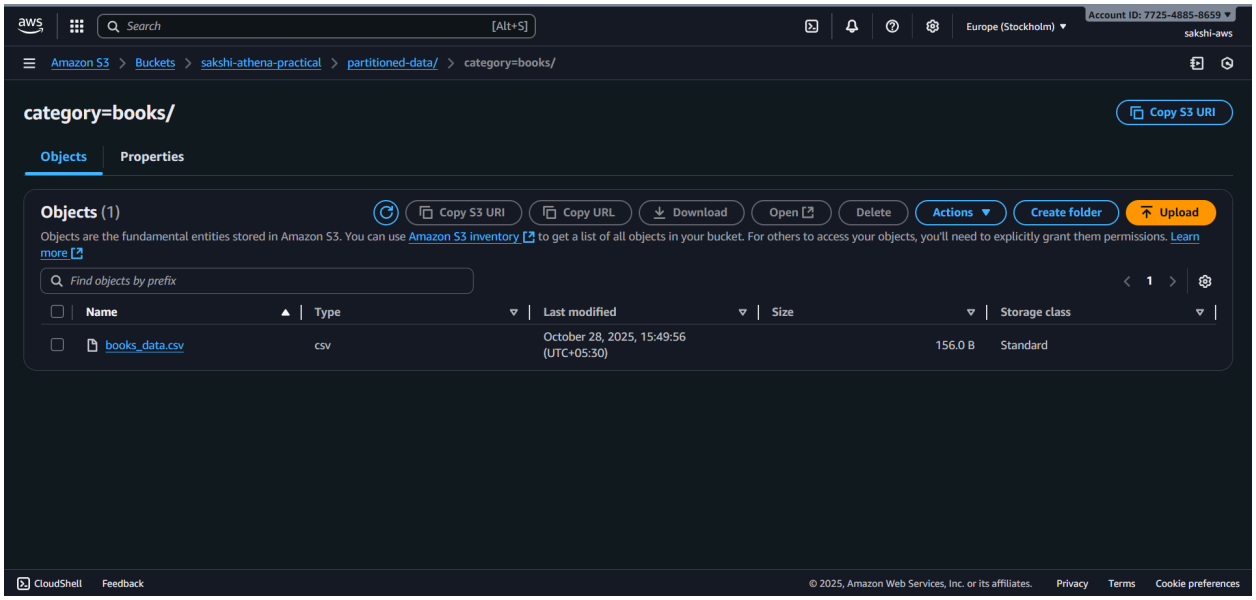
Creating folder for each category

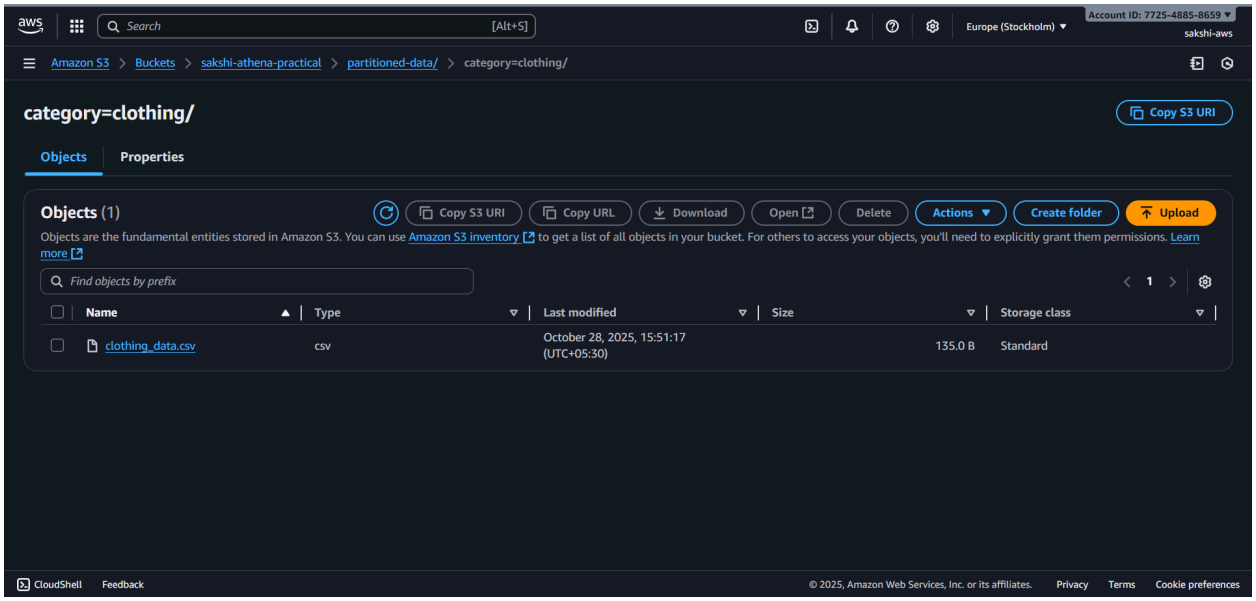


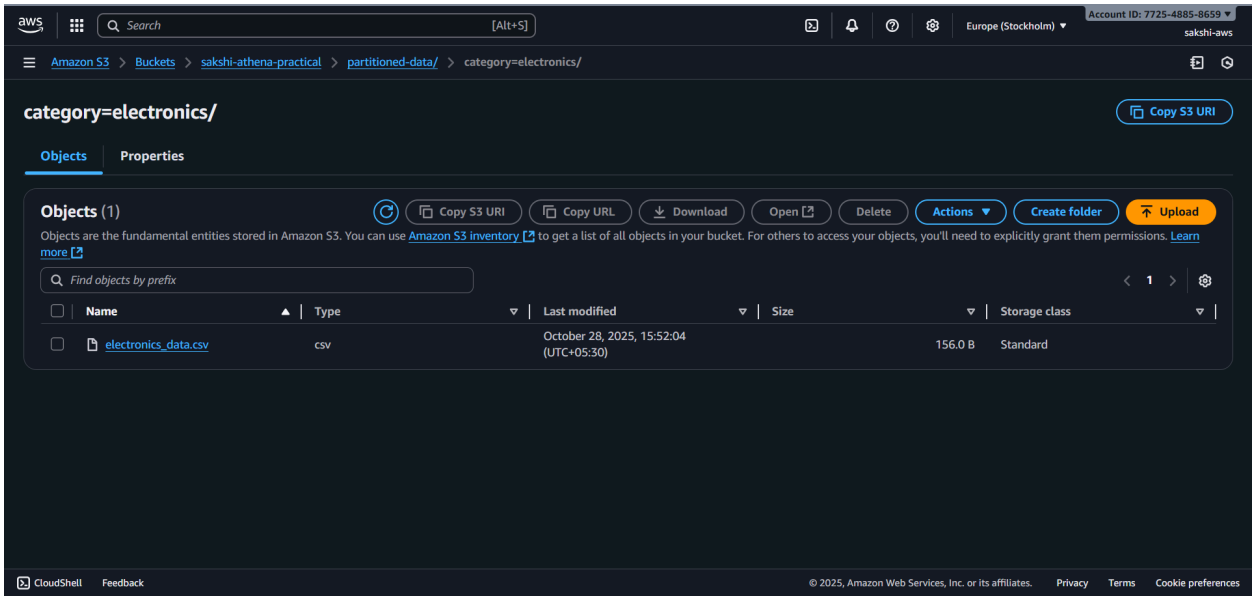


Uploading files in each folder

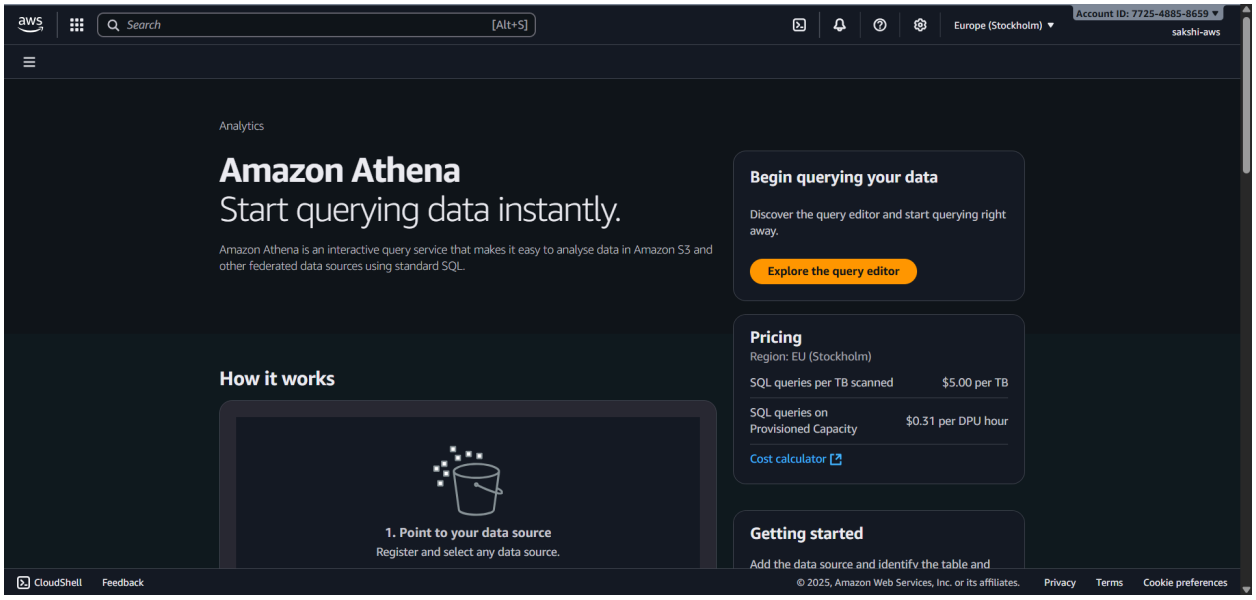



Search [Alt+S]
Europe (Stockholm)
Account ID: 7725-4885-8659
sakshi-aws
Amazon S3 > Buckets > sakshi-athena-practical > partitioned-data/ > category=books/
category=books/
Copy S3 URI
Objects
Properties
Objects (1)
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are the 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
Find objects by prefix
Name
Type
Last modified
Size
Storage class
books_data.csv
csv
October 28, 2025, 15:49:56 (UTC+05:30)
156.0 B
Standard
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

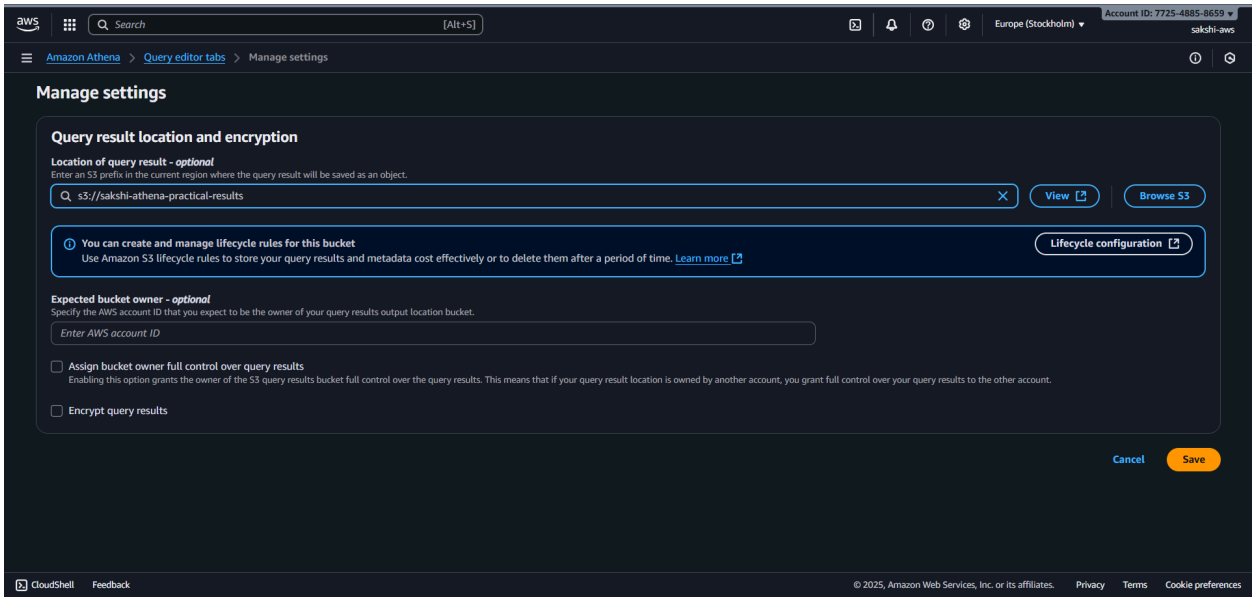

Search [Alt+S]
Europe (Stockholm)
Account ID: 7725-4885-8659
sakshi-aws
Amazon S3 > Buckets > sakshi-athena-practical > partitioned-data/ > category=clothing/
category=clothing/
Copy S3 URI
Objects
Properties
Objects (1)
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are 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
Find objects by prefix
Name
Type
Last modified
Size
Storage class
clothing_data.csv
csv
October 28, 2025, 15:51:17 (UTC+05:30)
135.0 B
Standard
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

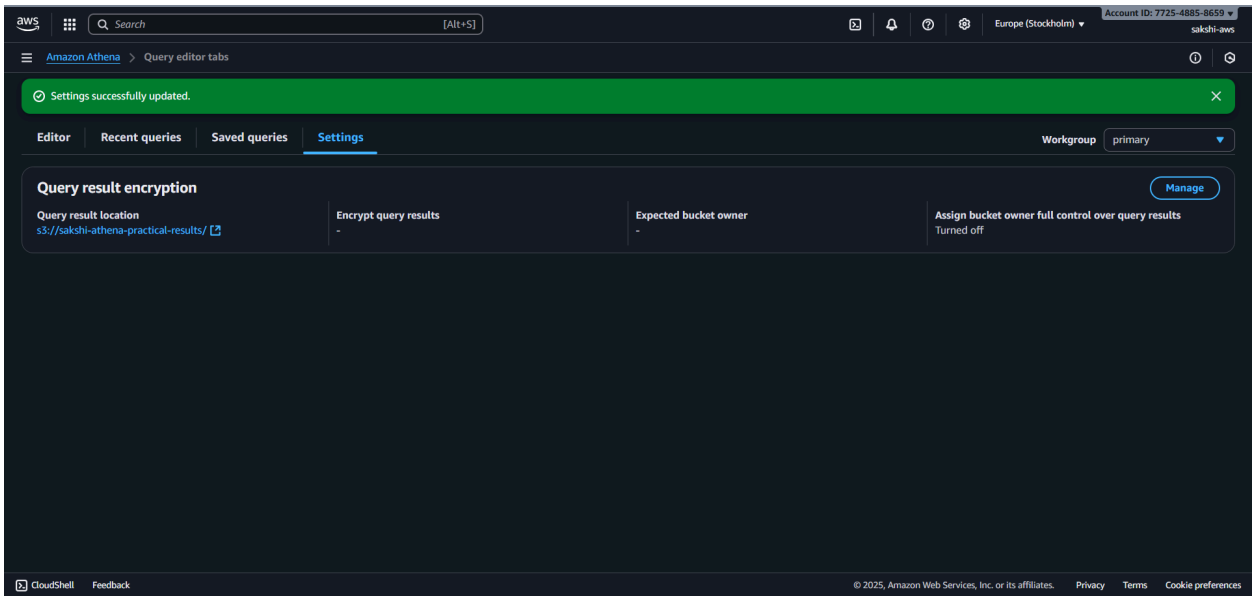


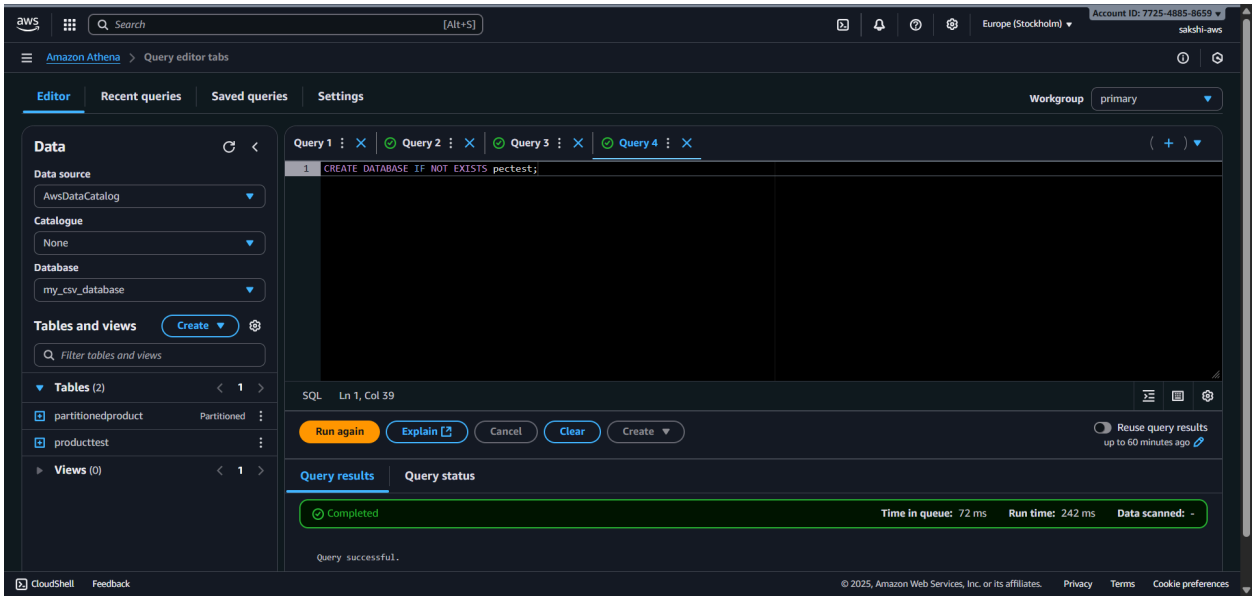
Launching Athena

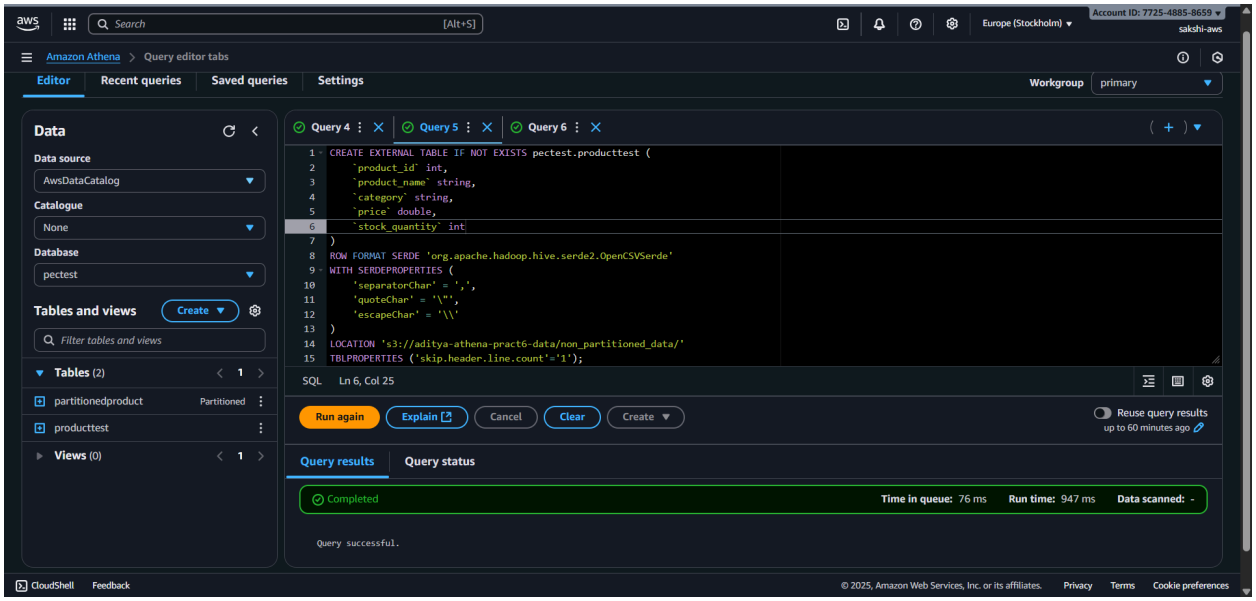


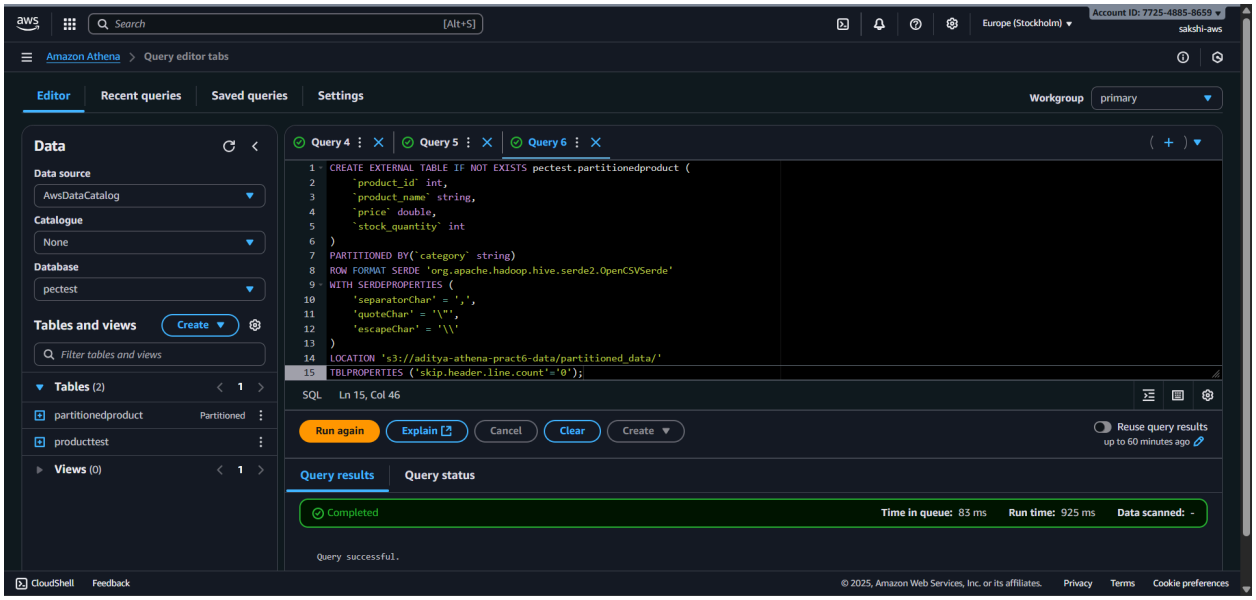
Setting query result location

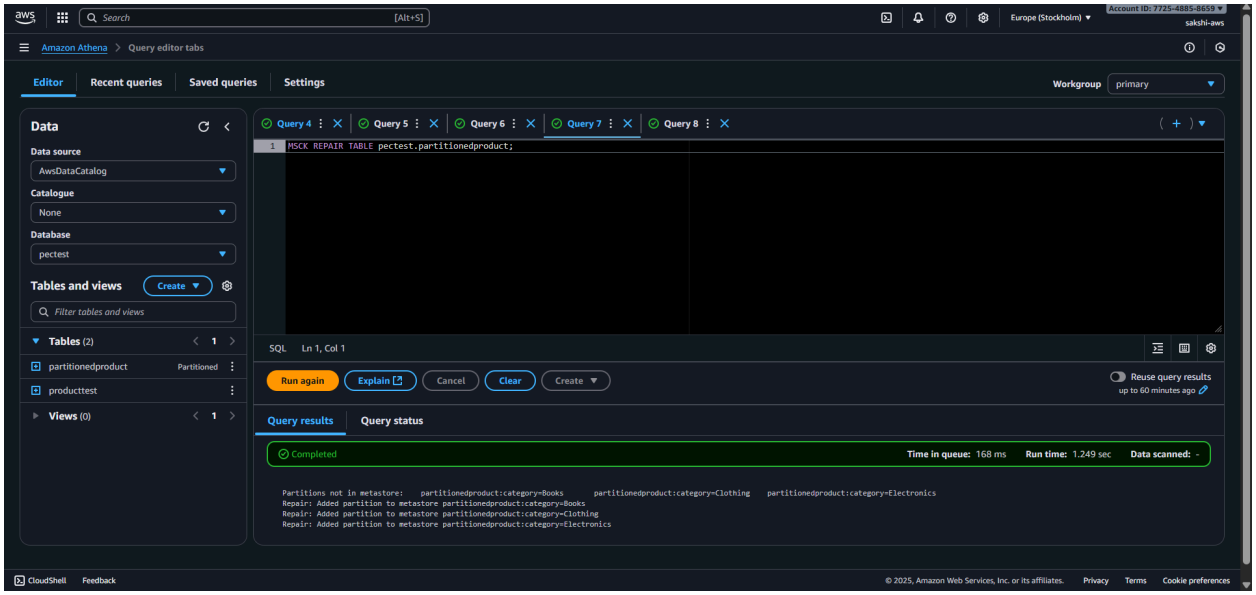




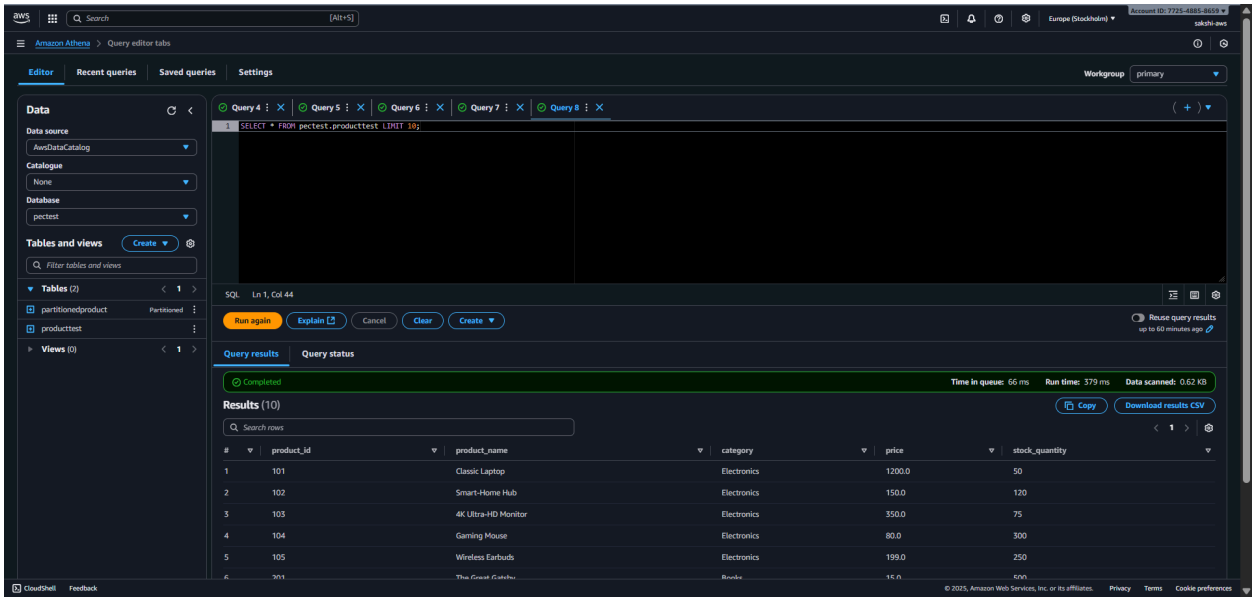

Search [Alt+S]
Europe (Stockholm)
Account ID: 7725-4885-8659
sakshi-aws
Amazon Athena > Query editor tabs
Editor
Recent queries
Saved queries
Settings
Workgroup primary
Data
Data source
AwsDataCatalog
Catalogue
None
Database
my_csv_database
Tables and views
Create
Filter tables and views
Tables (2)
partitionedproduct Partitioned
producttest
Views (0)
Query 1
Query 2
Query 3
Query 4
1 CREATE DATABASE IF NOT EXISTS pectest;
SQL Ln 1, Col 39
Run again
Explain
Cancel
Clear
Create
Reuse query results
up to 60 minutes ago
Query results
Query status
Completed
Time in queue: 72 ms
Run time: 242 ms
Data scanned: -
Query successful.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

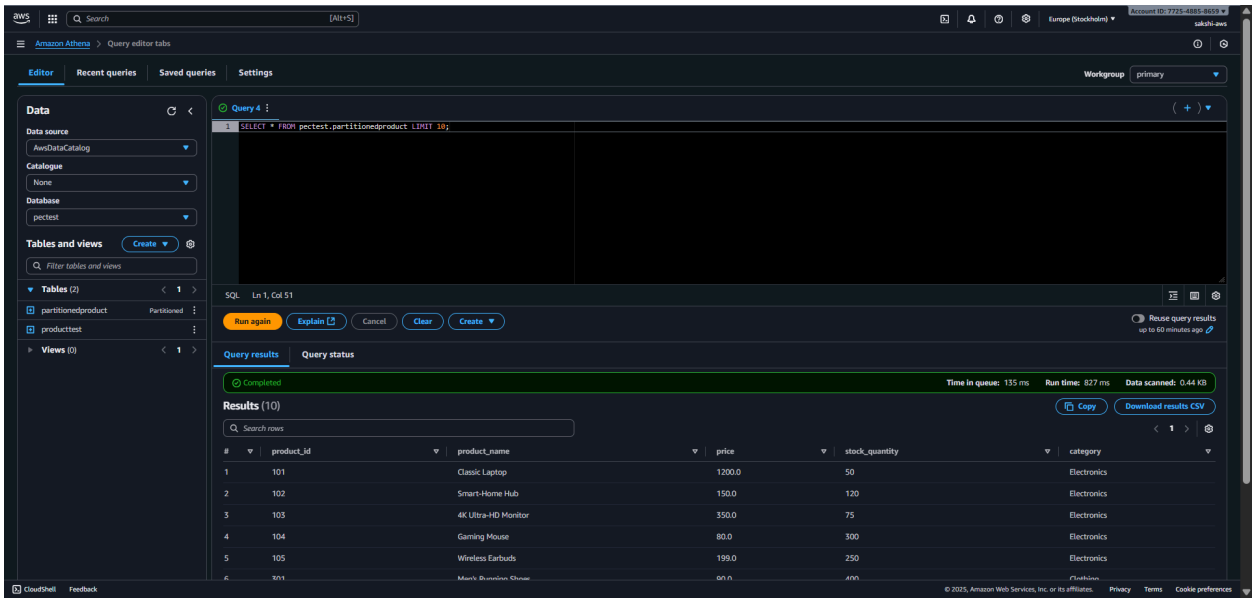

Search [Alt+S]
Europe (Stockholm)
Account ID: 7725-4885-8659
sakshi-aws
Amazon Athena > Query editor tabs
Editor
Recent queries
Saved queries
Settings
Workgroup primary
Data
Data source
AwsDataCatalog
Catalogue
None
Database
pectest
Tables and views
Create
Filter tables and views
Tables (2)
partitionedproduct Partitioned
producttest
Views (0)
Query 4
Query 5
Query 6
1 CREATE EXTERNAL TABLE IF NOT EXISTS pectest.producttest (
2 `product_id` int,
3 `product_name` string,
4 `category` string,
5 `price` double,
6 `stock_quantity` int
7)
8 ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
9 WITH SERDEPROPERTIES (
10 'separatorChar' = ',',
11 'quoteChar' = '\"',
12 'escapeChar' = '\\'
13)
14 LOCATION 's3://aditya-athena-pract6-data/non_partitioned_data/'
15 TBLPROPERTIES ('skip.header.line.count'='1');
SQL Ln 6, Col 25
Run again
Explain
Cancel
Clear
Create
Reuse query results
up to 60 minutes ago
Query results
Query status
Completed
Time in queue: 76 ms
Run time: 947 ms
Data scanned: -
Query successful.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Search [Alt+S]
Europe (Stockholm)
Account ID: 7725-4885-8659
sakshi-aws
Amazon Athena > Query editor tabs
Editor
Recent queries
Saved queries
Settings
Workgroup primary
Data
Data source
AwsDataCatalog
Catalogue
None
Database
pectest
Tables and views
Create
Filter tables and views
Tables (2)
partitionedproduct Partitioned
producttest
Views (0)
Query 4
Query 5
Query 6
CREATE EXTERNAL TABLE IF NOT EXISTS pectest.partitionedproduct (
`product_id` int,
`product_name` string,
`price` double,
`stock_quantity` int
)
PARTITIONED BY(`category` string)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
WITH SERDEPROPERTIES (
'separatorChar' = ',',
'quoteChar' = '\"',
'escapeChar' = '\\'
)
LOCATION 's3://aditya-athena-pract6-data/partitioned_data/'
TBLPROPERTIES ('skip.header.line.count'='0');
SQL Ln 15, Col 46
Run again
Explain
Cancel
Clear
Create
Reuse query results
up to 60 minutes ago
Query results
Query status
Completed
Time in queue: 83 ms
Run time: 925 ms
Data scanned: -
Query successful.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

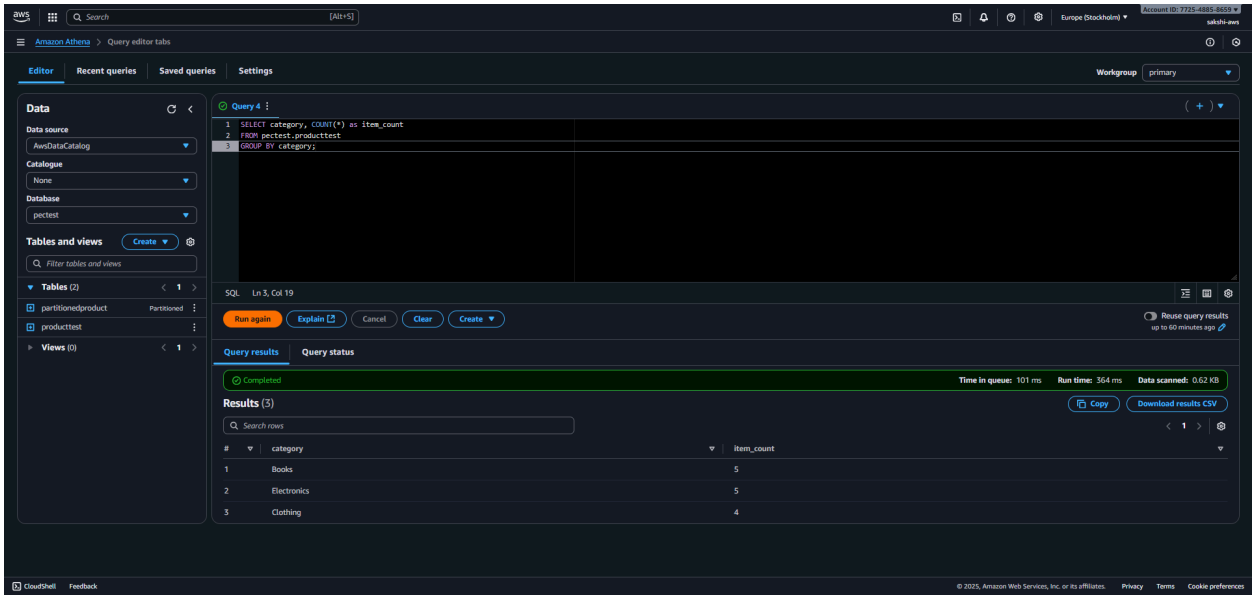

Search [Alt+S]
Europe (Stockholm)
Account ID: 7725-4885-8659
sakshi-aws
Amazon Athena > Query editor tabs
Editor
Recent queries
Saved queries
Settings
Workgroup primary
Data
Data source
AwsDataCatalog
Catalogue
None
Database
pectest
Tables and views
Create
Filter tables and views
Tables (2)
partitionedproduct Partitioned
producttest
Views (0)
Query 4
Query 5
Query 6
Query 7
Query 8
MSCK REPAIR TABLE pectest.partitionedproduct;
SQL Ln 1, Col 1
Run again
Explain
Cancel
Clear
Create
Reuse query results
up to 60 minutes ago
Query results
Query status
Completed
Time in queue: 168 ms
Run time: 1.249 sec
Data scanned: -
Partitions not in metastore: partitionedproduct:category=Books partitionedproduct:category=Clothing partitionedproduct:category=Electronics
Repair: Added partition to metastore partitionedproduct:category=Books
Repair: Added partition to metastore partitionedproduct:category=Clothing
Repair: Added partition to metastore partitionedproduct:category=Electronics
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Previewing the partitioned table

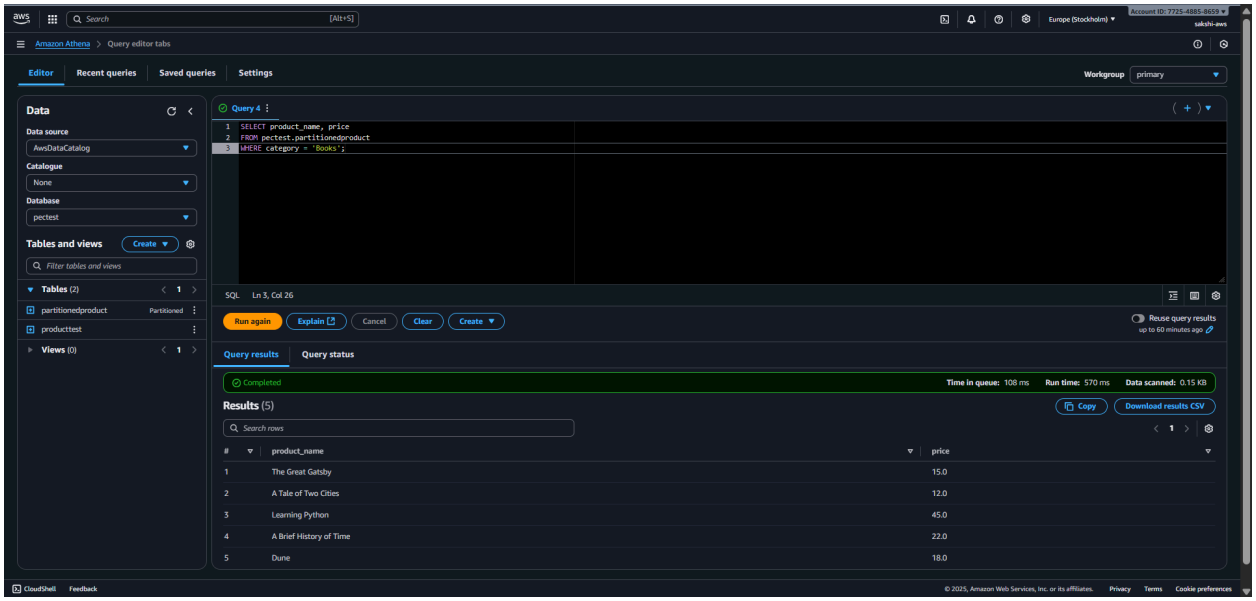




Querying the regular table [data scanned - 0.62 kb]



Querying the partitioned table [data scanned - 0.15 kb]



The non-partitioned table `producttest` required Athena to scan the entire `products_full.csv` file to find rows `WHERE category = 'Books'`. The partitioned table `partitionedproduct` only scanned the data inside the `category=Books/` folder. This is shown in the 'Data scanned' metric, which was significantly lower for the partitioned query. This demonstrates how partitioning improves query performance and reduces cost by pruning data.