PROJECT REPORT

CAMP BI V3 Batch 1 Group 4

INDEX

Supervised by: Pritam Gorain

Group Members:

- Aryan Kaundal
- Kajol Gupta
- Saurabh Pratap Singh
- Shivam Gupta
- Shubham Dixit
- Vrishti Saxena

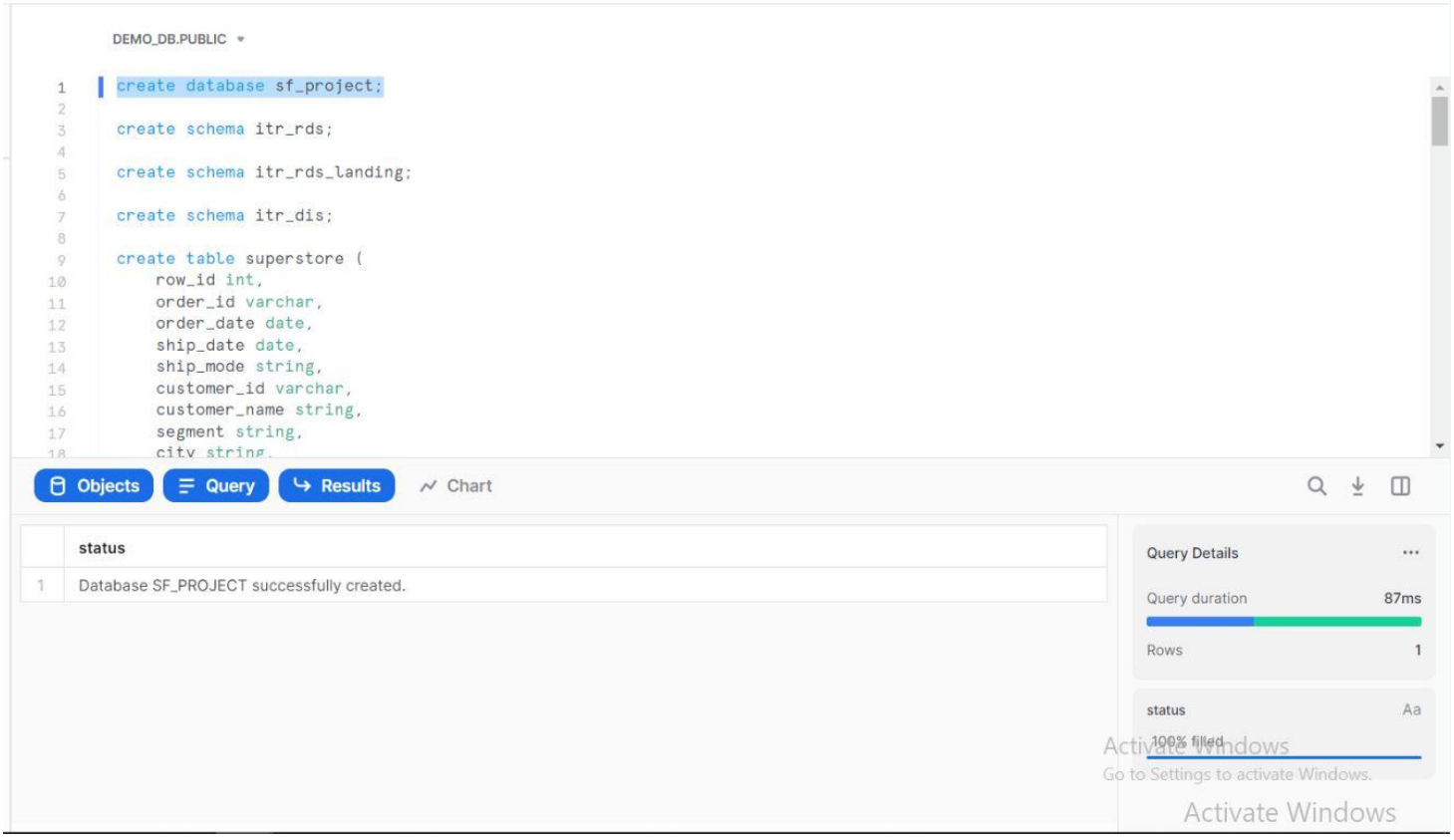
- ***PROJECT DESCRIPTION***

The project involves data ingestion and analysis from public datahub Kaggle link.

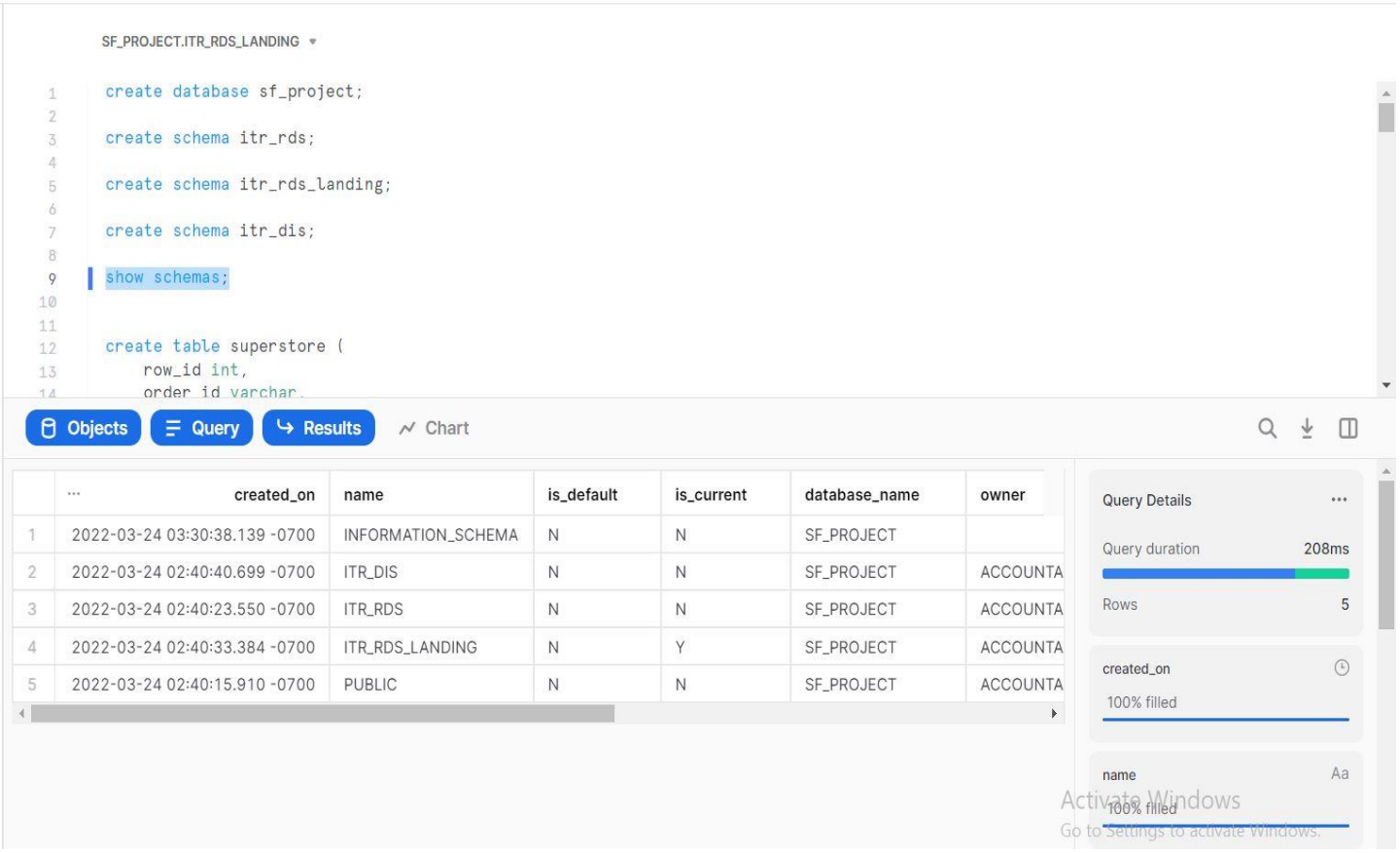
- Steps involved in performing the data ingestion:
    a. Loading data to external stage
    b. Ingesting data into the landing schema
    c. Ingesting data into the consumer table
    d. Perform analysis on the given dataset

UNIT TEST REPORT

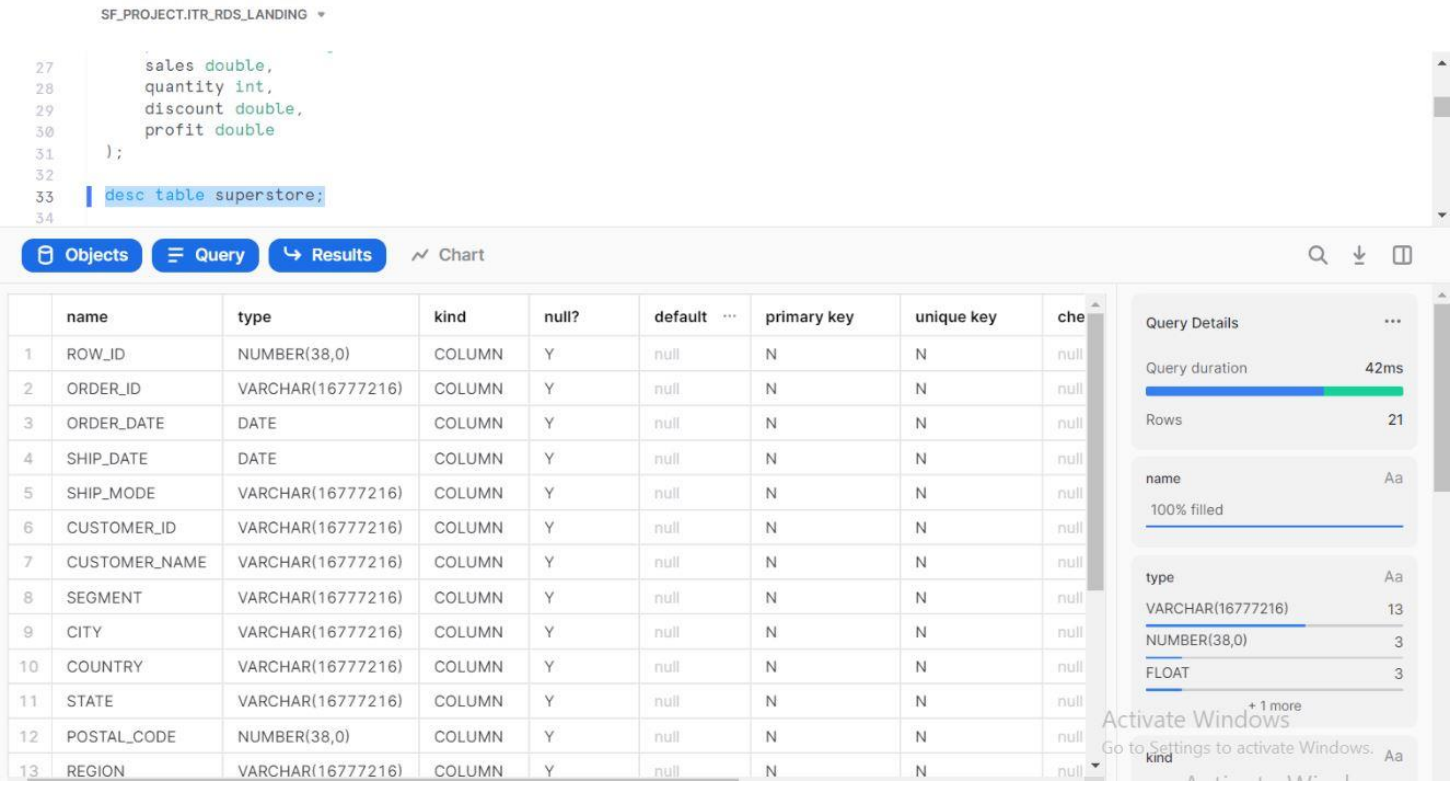
1. Created a database named SF_PROJECT;



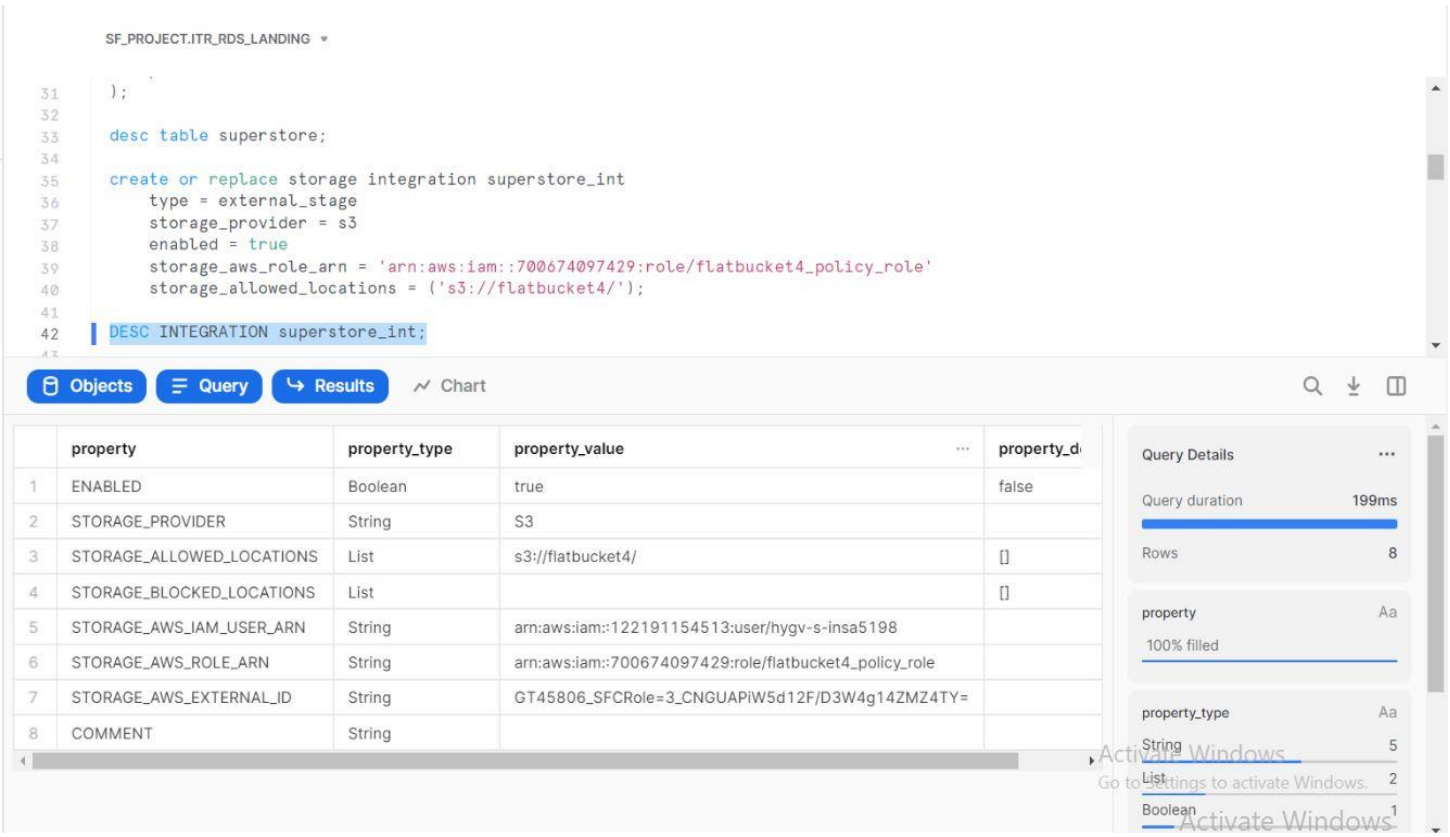
2. Created three schemas named ITR_RDS, ITR_RDS_LANDING and ITR_DIS



3. Created a table named SUPERTORE as per the data set



4. Created Integration object named SUPER_STORE



SF_PROJECT.ITR_RDS_LANDING

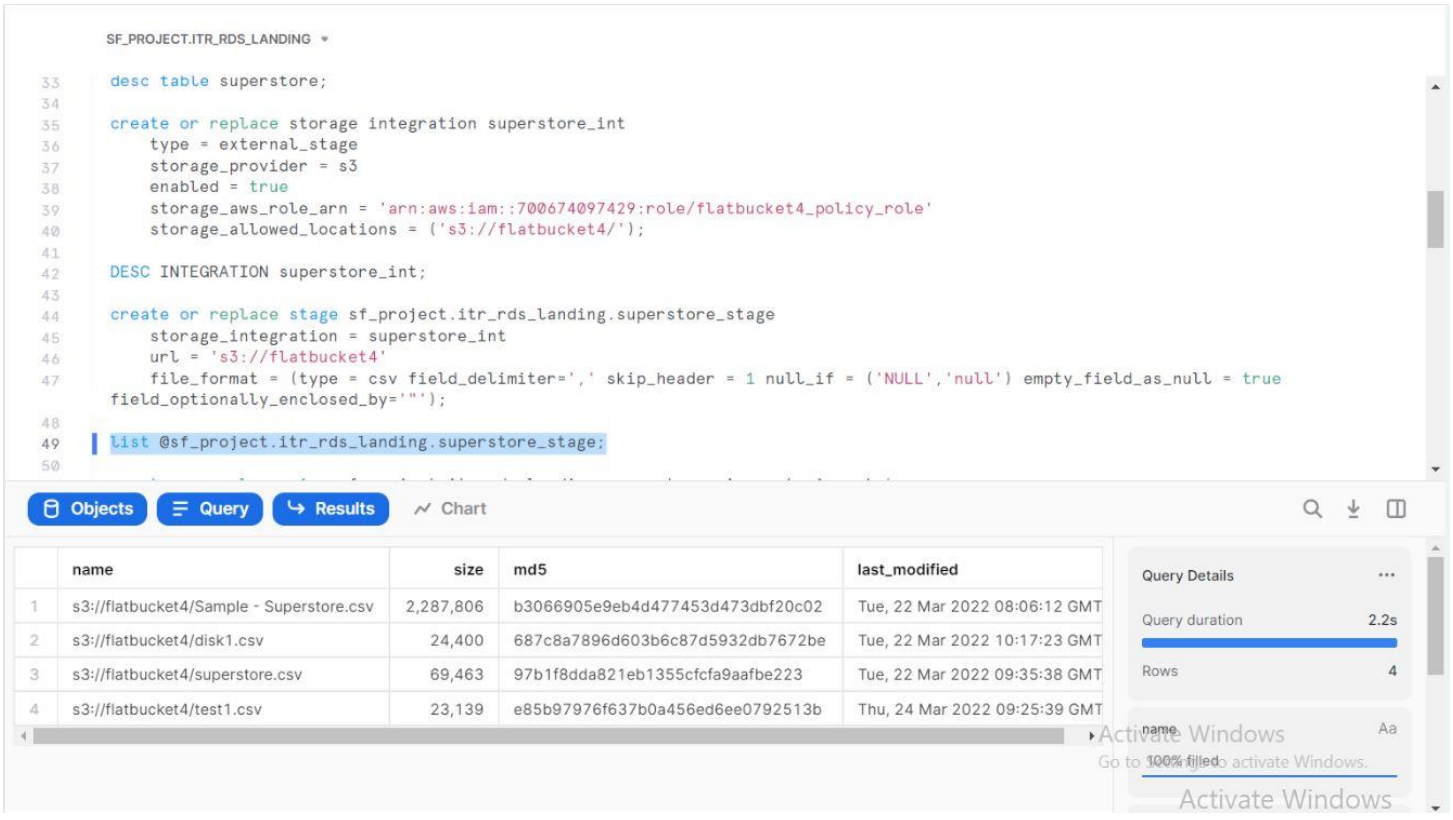
```
);

desc table superstore;

create or replace storage integration superstore_int
    type = external_stage
    storage_provider = s3
    enabled = true
    storage_aws_role_arn = 'arn:aws:iam::700674097429:role/flatbucket4_policy_role'
    storage_allowed_locations = ('s3://flatbucket4/');

DESC INTEGRATION superstore_int;
```

Objects | Query | Results | Chart

	property	property_type	property_value	property_d
1	ENABLED	Boolean	true	false
2	STORAGE_PROVIDER	String	S3	
3	STORAGE_ALLOWED_LOCATIONS	List	s3://flatbucket4/	[]
4	STORAGE_BLOCKED_LOCATIONS	List		[]
5	STORAGE_AWS_IAM_USER_ARN	String	arn:aws:iam::122191154513:user/hygv-s-insa5198	
6	STORAGE_AWS_ROLE_ARN	String	arn:aws:iam::700674097429:role/flatbucket4_policy_role	
7	STORAGE_AWS_EXTERNAL_ID	String	GT45806_SFCRole=3_CNGUAPiW5d12F/D3W4g14ZMZ4TY=	
8	COMMENT	String		

Query Details
Query duration 199ms
Rows 8
property Aa
100% filled
property_type Aa
String 5
List 2
Boolean 1

5. Created external stage named SUPERSTORE_STAGE for loading data structures



SF_PROJECT.ITR_RDS_LANDING

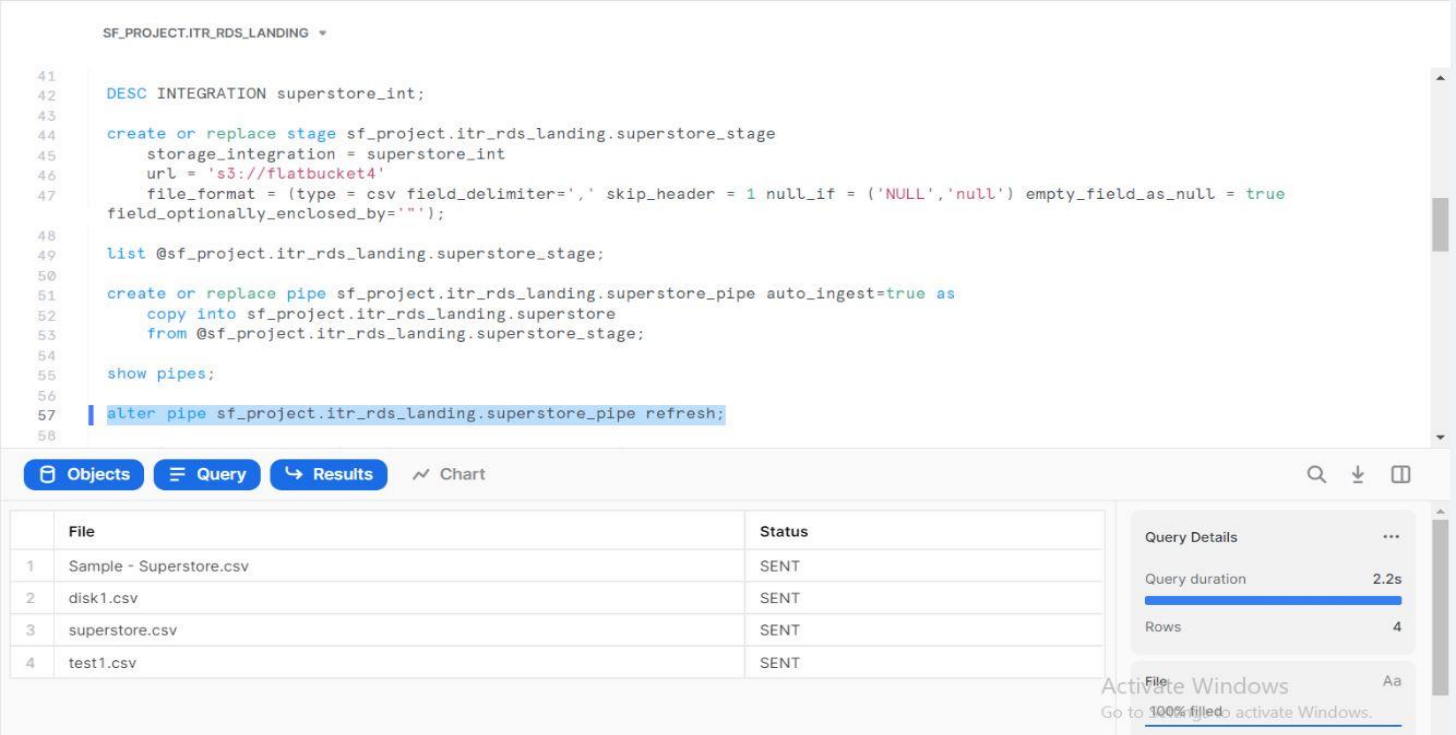
```
desc table superstore;

create or replace storage integration superstore_int
    type = external_stage
    storage_provider = s3
    enabled = true
    storage_aws_role_arn = 'arn:aws:iam::700674097429:role/flatbucket4_policy_role'
    storage_allowed_locations = ('s3://flatbucket4/');

DESC INTEGRATION superstore_int;

create or replace stage sf_project.itr_rds_landing.superstore_stage
    storage_integration = superstore_int
    url = 's3://flatbucket4'
    file_format = (type = csv field_delimiter=',' skip_header = 1 null_if = ('NULL','null') empty_field_as_null = true
field_optionally_enclosed_by='"');

List @sf_project.itr_rds_landing.superstore_stage;
```

Objects | Query | Results | Chart

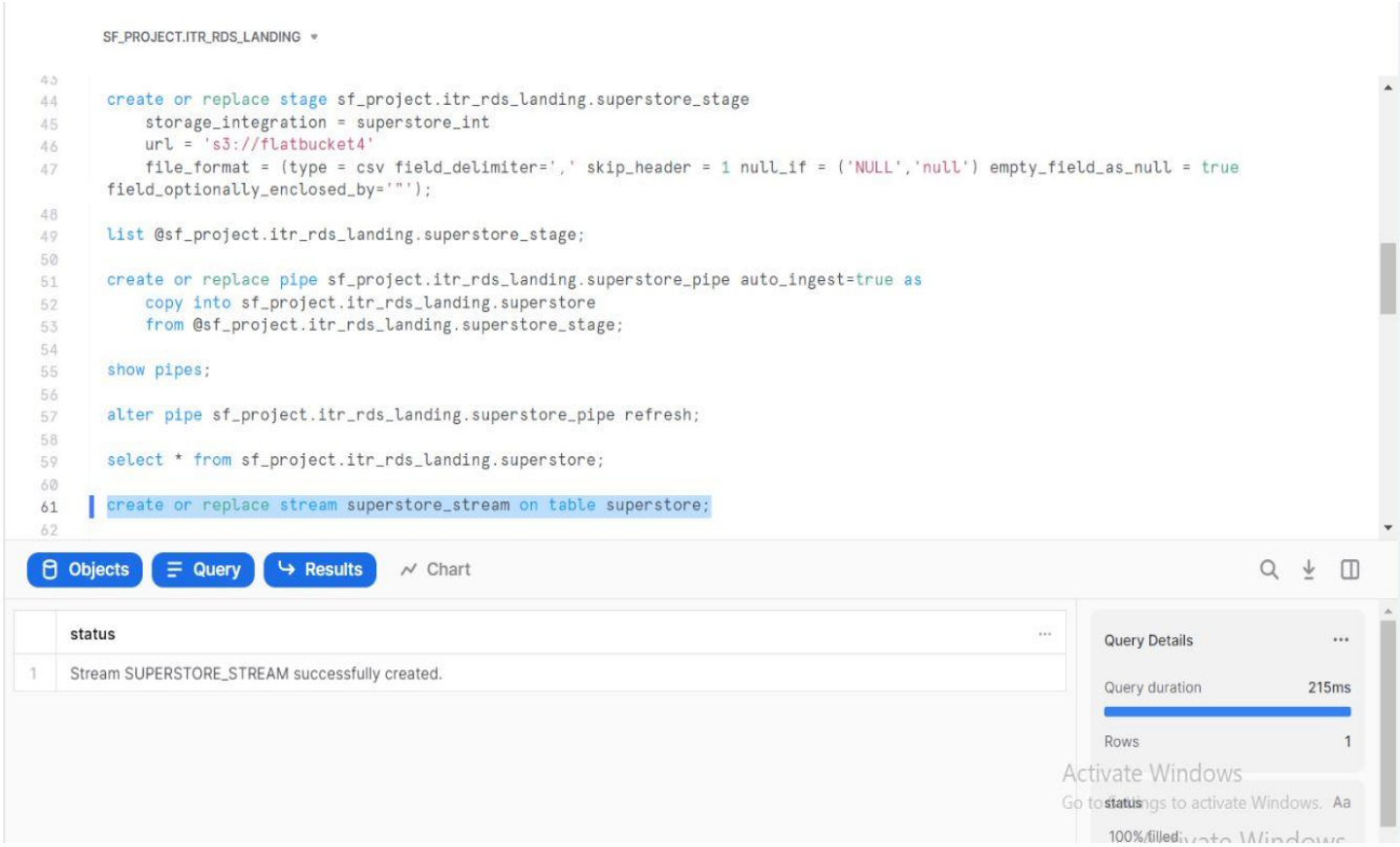
	name	size	md5	last_modified
1	s3://flatbucket4/Sample - Superstore.csv	2,287,806	b3066905e9eb4d477453d473dbf20c02	Tue, 22 Mar 2022 08:06:12 GMT
2	s3://flatbucket4/disk1.csv	24,400	687c8a7896d603b6c87d5932db7672be	Tue, 22 Mar 2022 10:17:23 GMT
3	s3://flatbucket4/superstore.csv	69,463	97b1f8dda821eb1355cfcfa9aafbe223	Tue, 22 Mar 2022 09:35:38 GMT
4	s3://flatbucket4/test1.csv	23,139	e85b97976f637b0a456ed6ee0792513b	Thu, 24 Mar 2022 09:25:39 GMT

Query Details
Query duration 2.2s
Rows 4
name Aa
100% filled

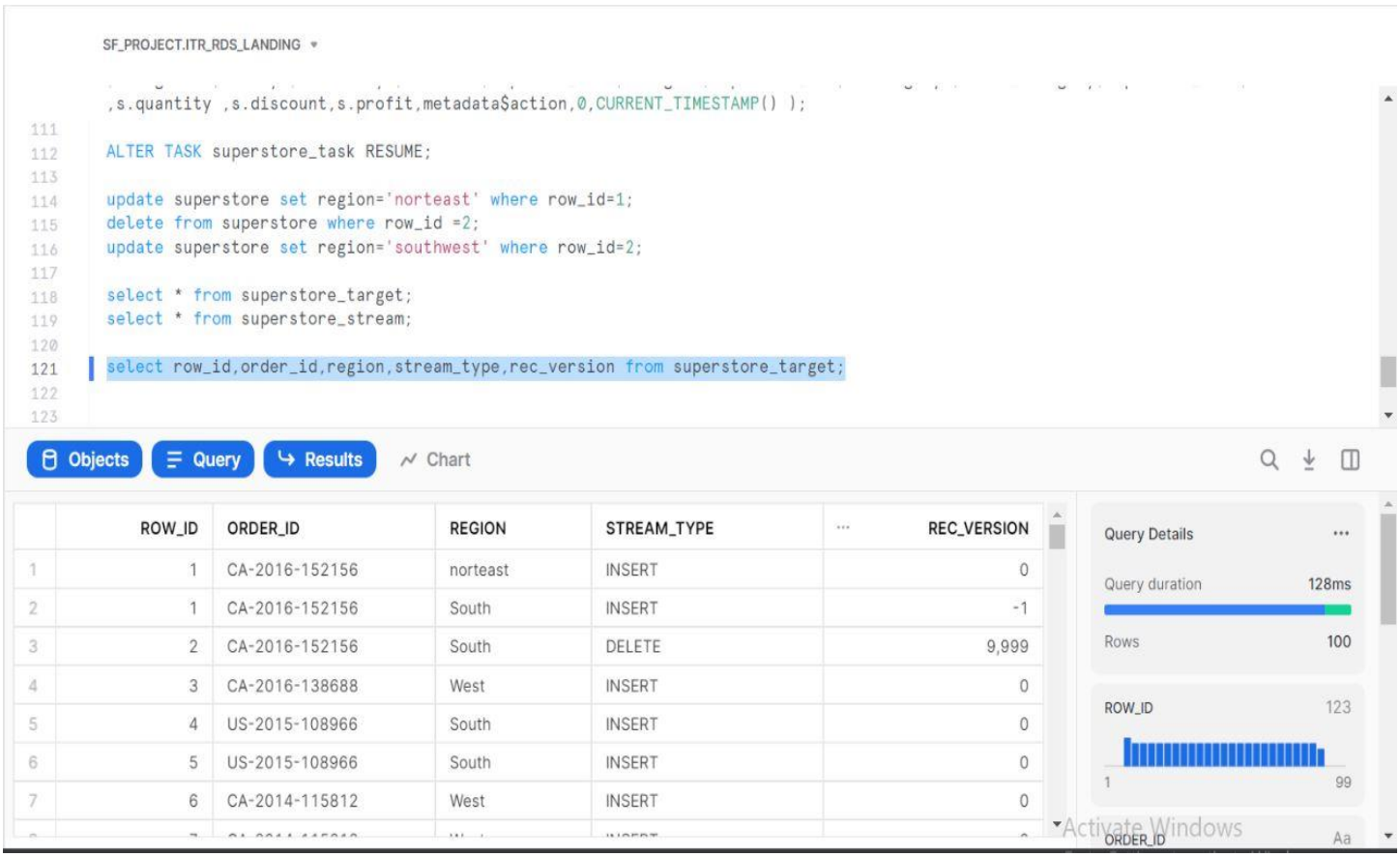
6. Created a snowpipe named SUPERSTORE_PIPE for autoingesting the data from S3 bucket – flatbucket4



7. Created stream SUPERSTORE_STREAM on table SUPERSTORE



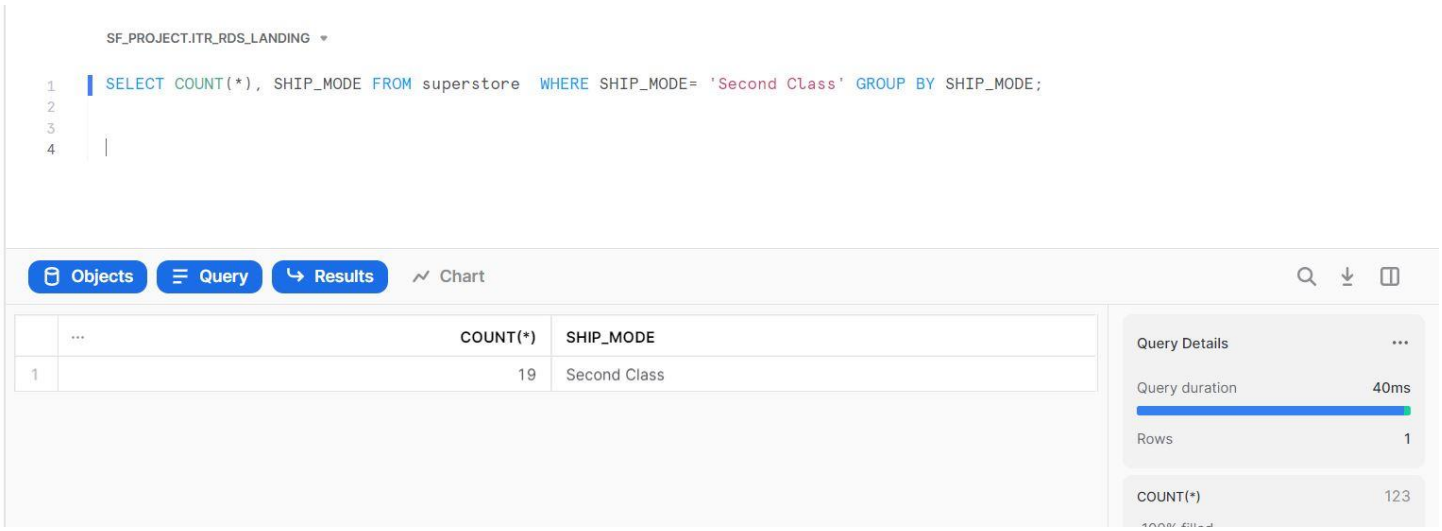
8. Performed SCD 2 operation on consumer table SUPERSTORE_TARGET as per changes that happen in the source table SUPERSTORE



- ***DATA ANALYSIS ON THE GIVEN DATASET***

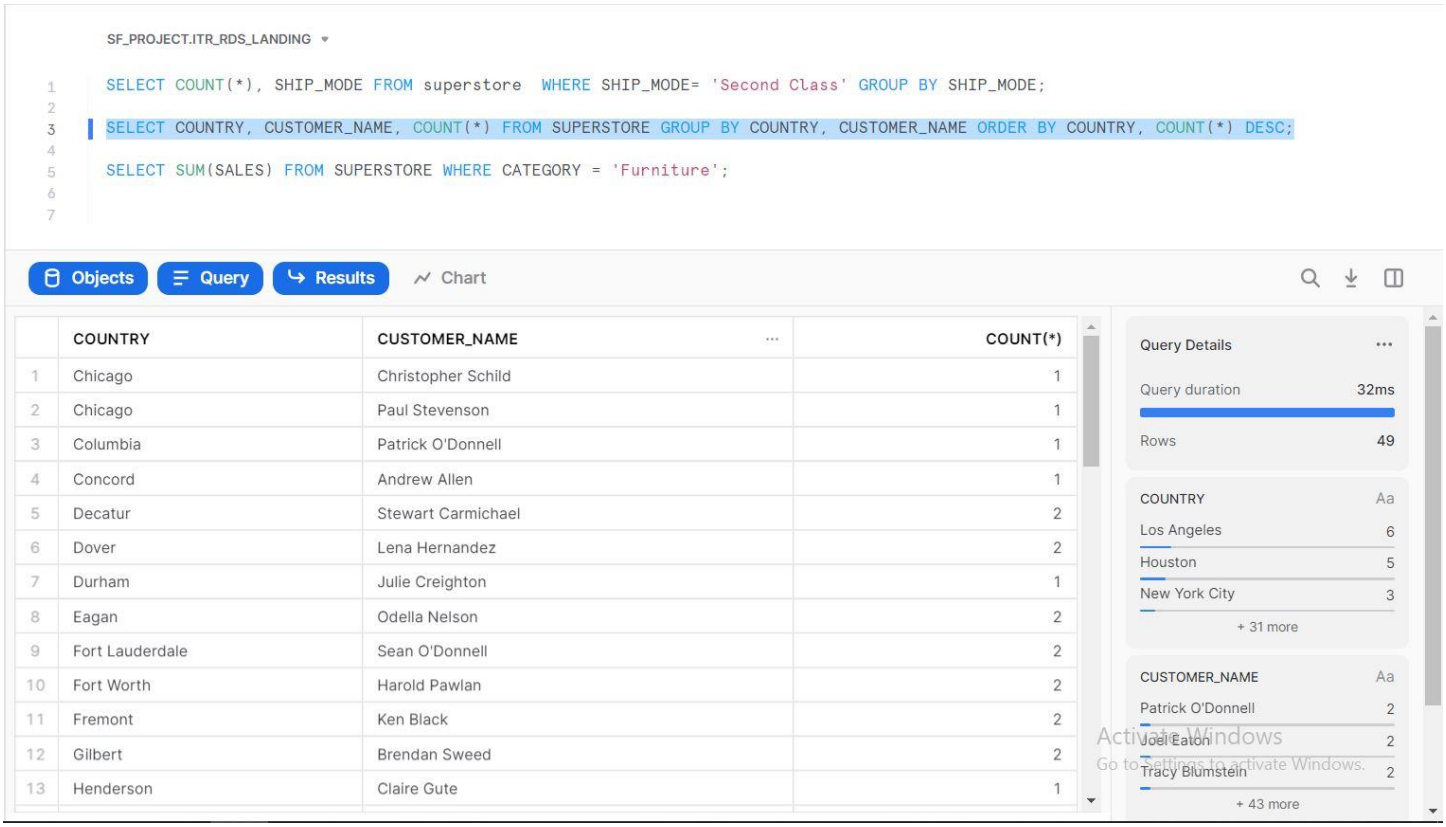
--1. Calculate the number of Orders those with Ship Mode as 'Second Class'

SELECT COUNT(*), SHIP_MODE FROM superstore WHERE SHIP_MODE= 'Second Class' GROUP BY SHIP_MODE;



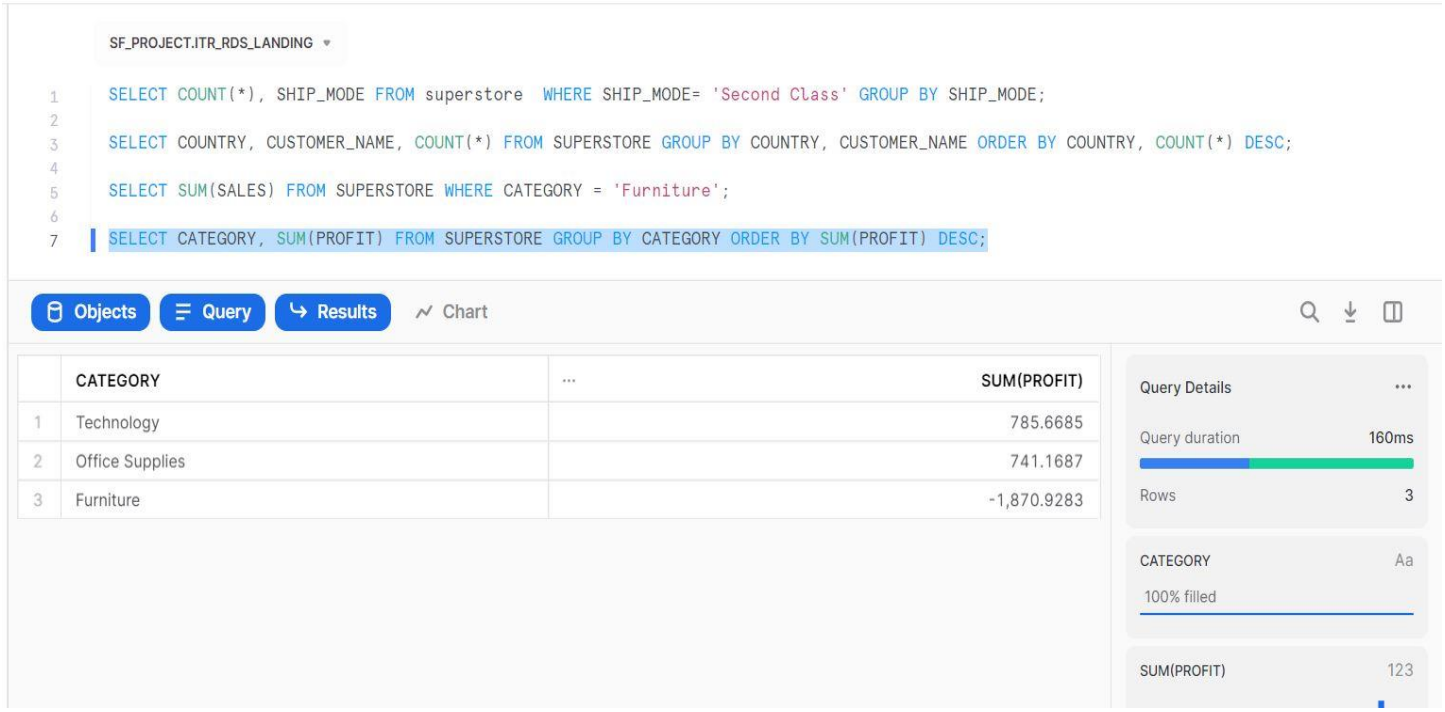
--2. List down the most valuable customers Country wise

SELECT COUNTRY, CUSTOMER_NAME, COUNT(*) FROM SUPERSTORE GROUP BY COUNTRY, CUSTOMER_NAME ORDER BY COUNTRY, COUNT(*) DESC;



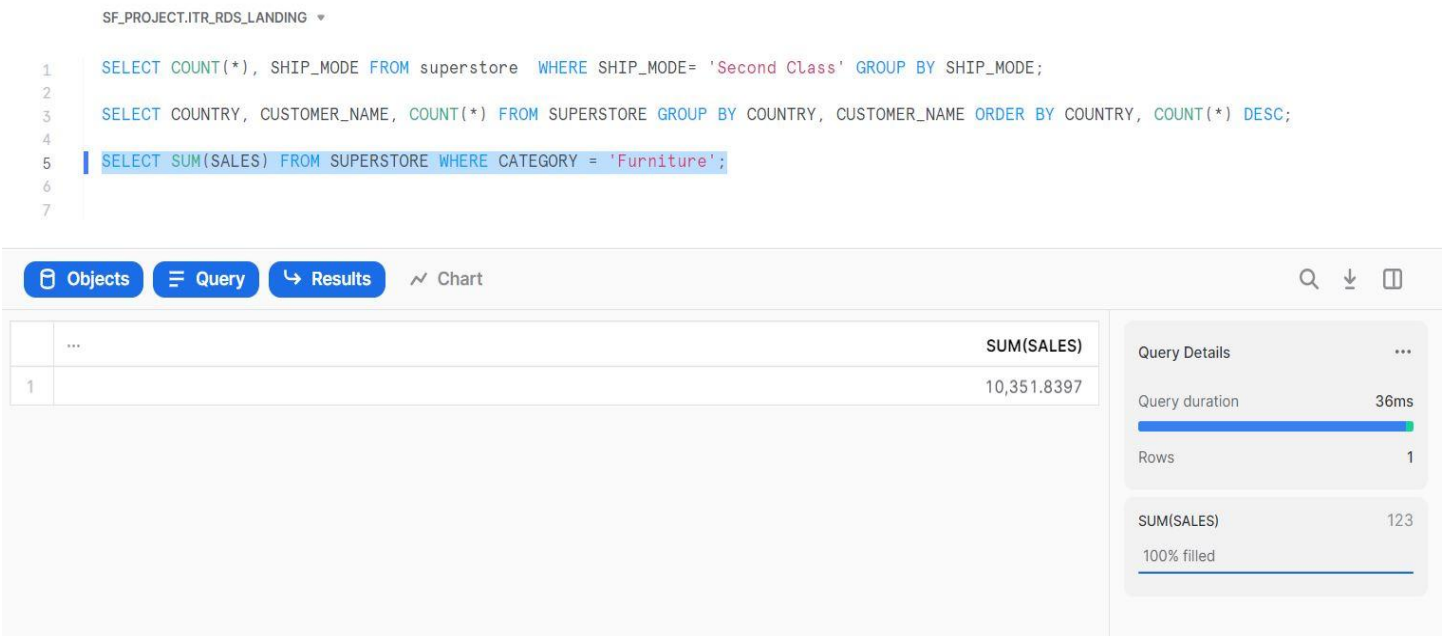
--3. Total Sales for Category 'Furniture'

SELECT SUM(SALES) FROM SUPERSTORE WHERE CATEGORY = 'Furniture';



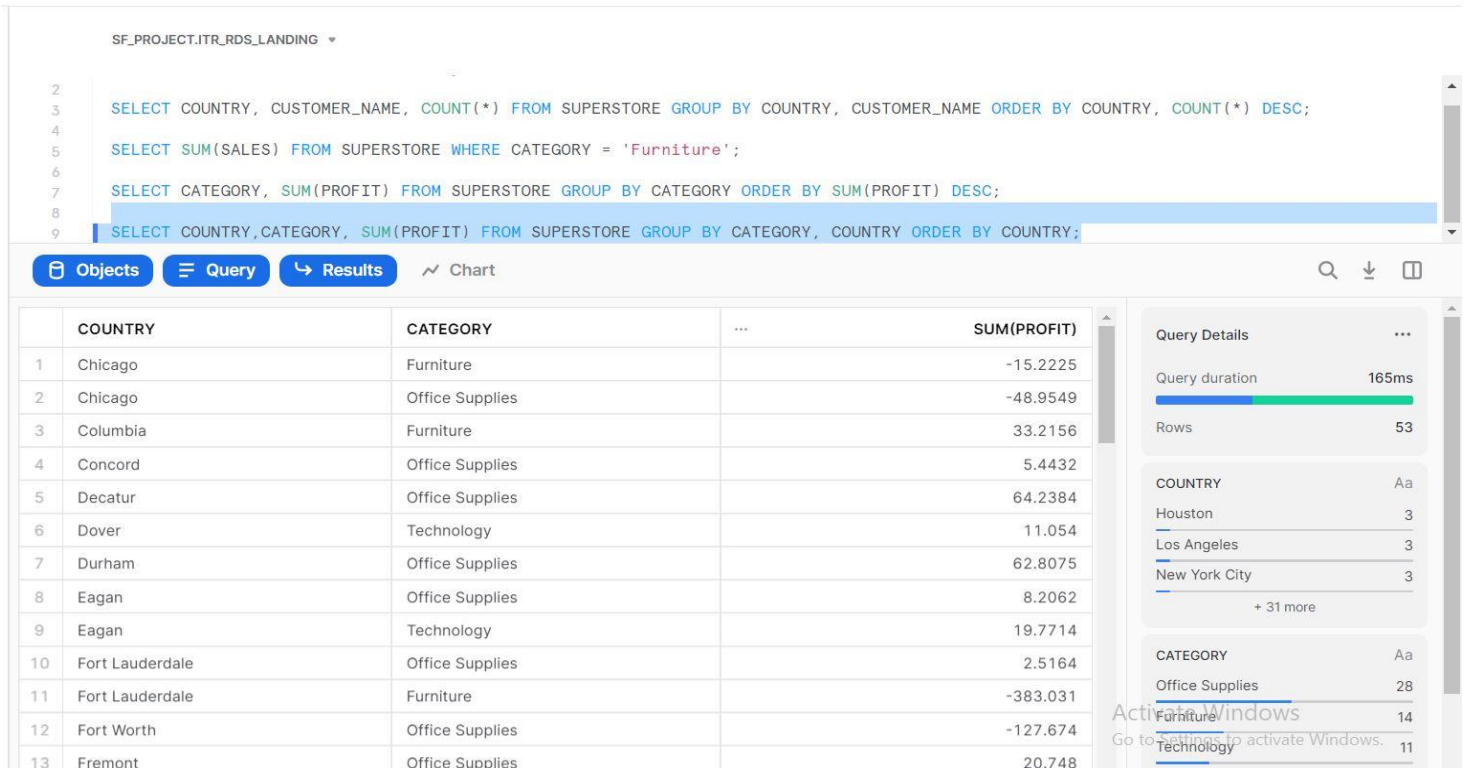
--4. Which Product provides the maximum profit

SELECT CATEGORY, SUM(PROFIT) FROM SUPERSTORE GROUP BY CATEGORY ORDER BY SUM(PROFIT) DESC;



--5. Calculate the total profit made by each product category country wise

SELECT COUNTRY,CATEGORY, SUM(PROFIT) FROM SUPERSTORE GROUP BY CATEGORY, COUNTRY ORDER BY COUNTRY;



--6. Which region of United States have majority loss?

SELECT REGION, MIN(PROFIT) FROM SUPERSTORE GROUP BY REGION;