DATA 226- DATAWAREHOUSE

Homework 1

Name : Vimalanandhan Sivanandham

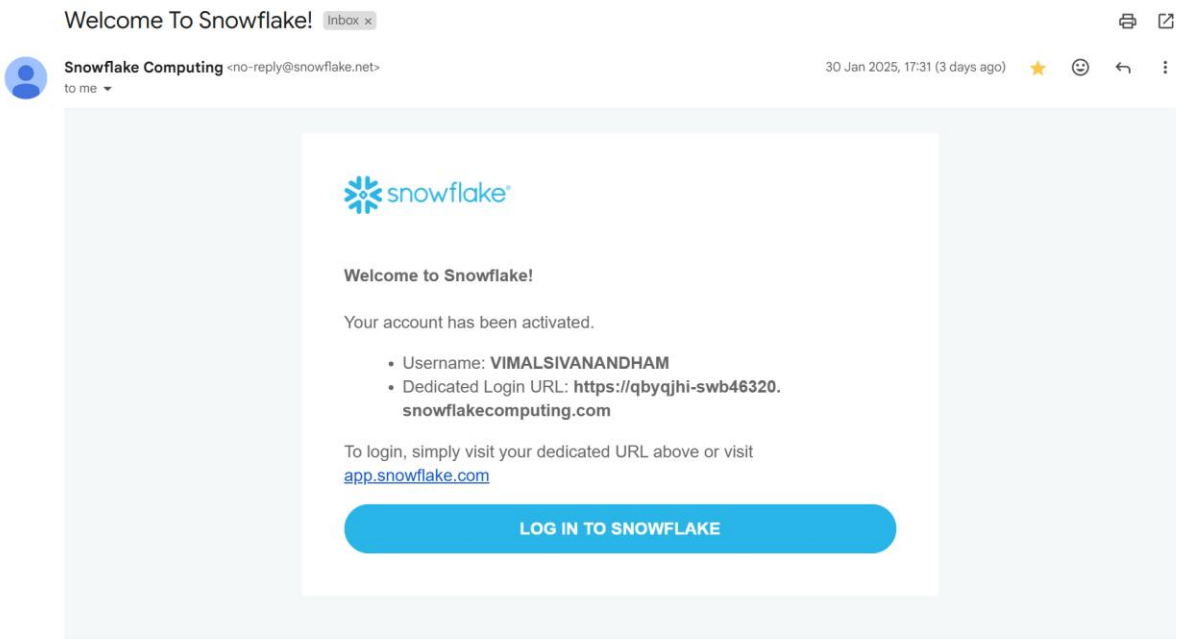
SJSU ID: 017596436

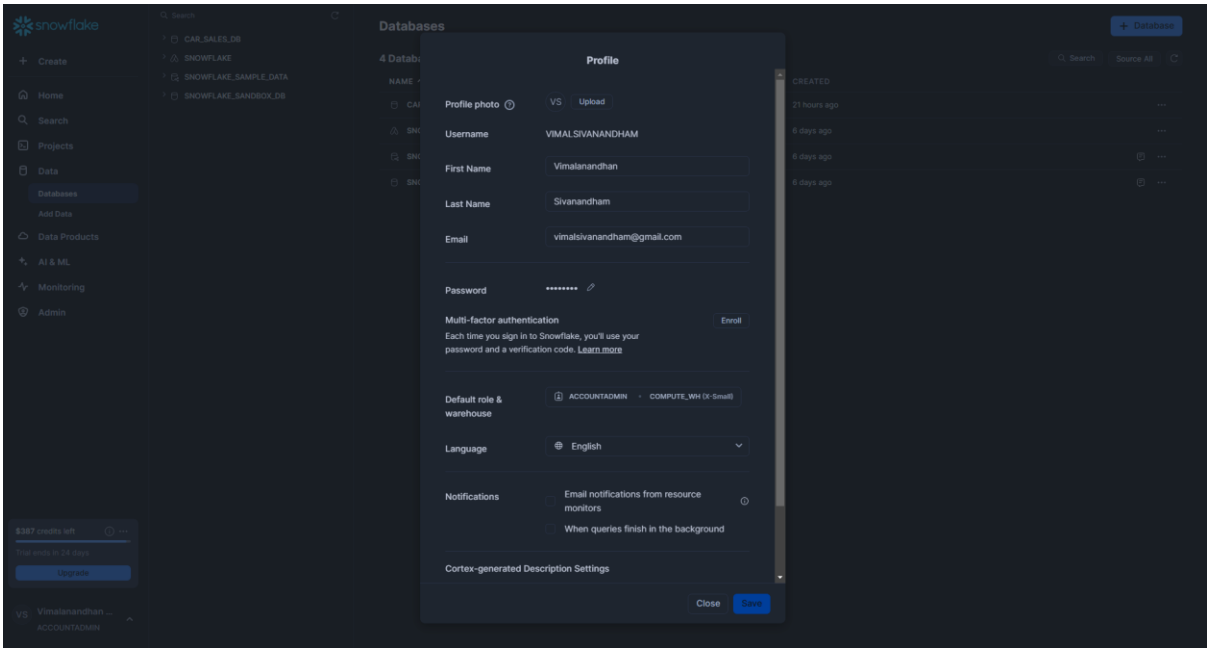
Assignment - Snowflake Environment Setup

Perform the following in Snowflake:

Car Sales csv(synthetic) dataset is under the Files/Homeworks/Car_Sales_Data.csv

1. (+1) **Create Snowflake Account:**





Account Created.

2. (+2) Create Database 'car_sales_db', Schema 'car_sales_schema', Table 'car_sales_tbl' (understand the dataset structure). Use the warehouse of size XSMALL.

   Created **Database** named as **'car_sales_db'**

   Created **Schema** named as **'car_sales_schema'**

   Created **Table** named as **'car_sales_tbl'**

   While creating the warehouse of Size **XSMALL**

   CODE:

```
-- Create the database

CREATE DATABASE car_sales_db;


CREATE WAREHOUSE compute_wh

WITH WAREHOUSE_SIZE = 'XSMALL';


-- Create the schema inside the database


CREATE SCHEMA car_sales_schema;
```

-- Use the schema

USE SCHEMA car_sales_schema;

--- Creation Table

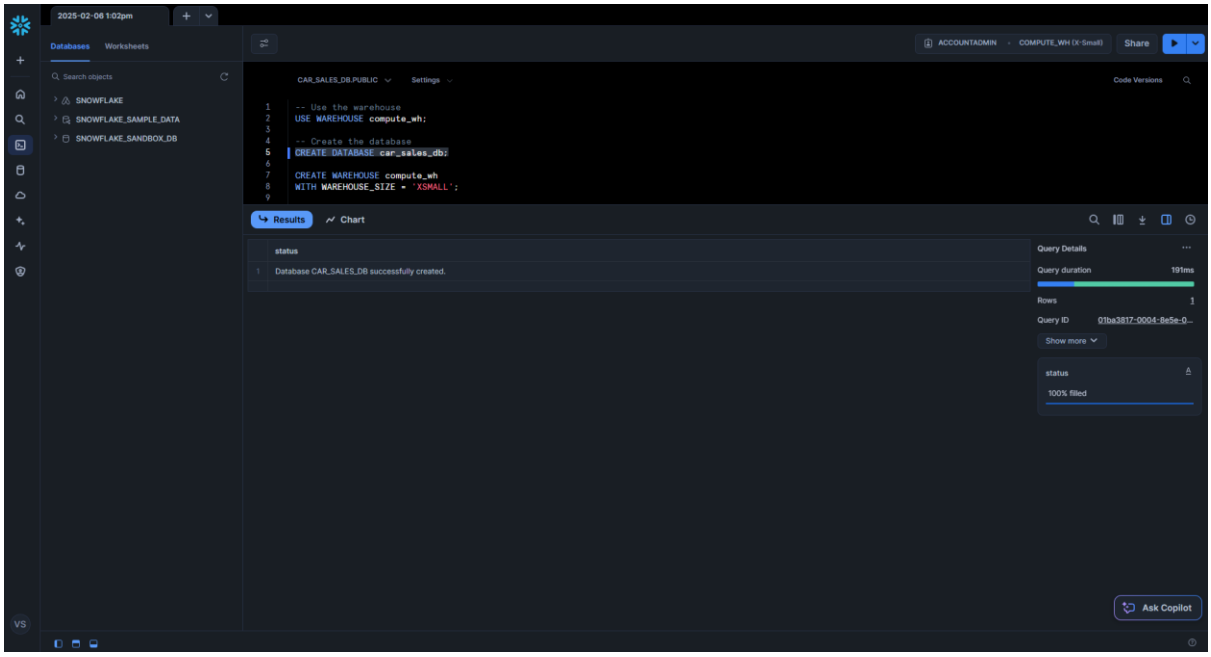
CREATE OR REPLACE TABLE car_sales_tbl (

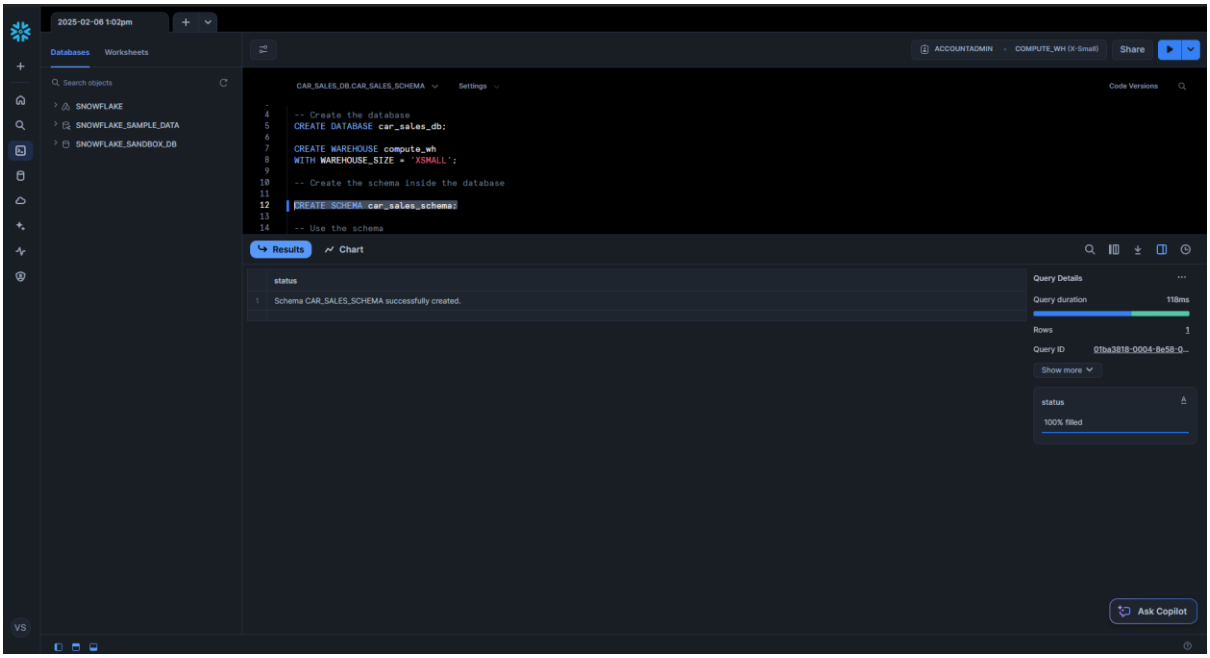
  car_name STRING,

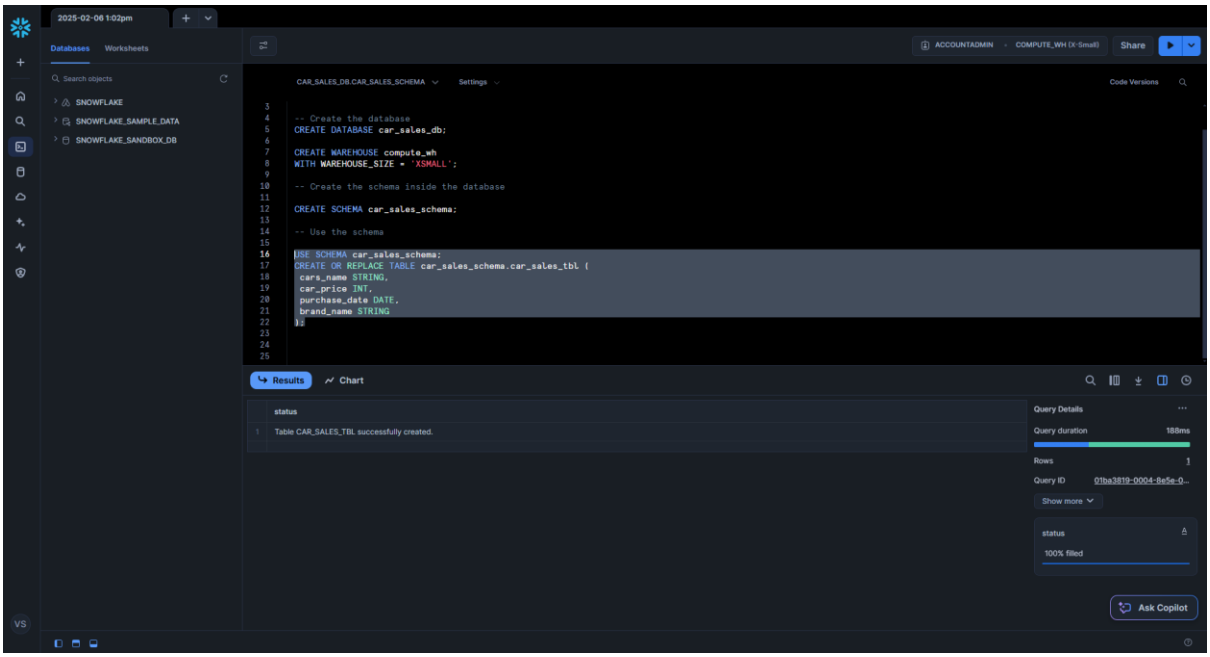
  sale_date DATE,

  units_sold INT

);



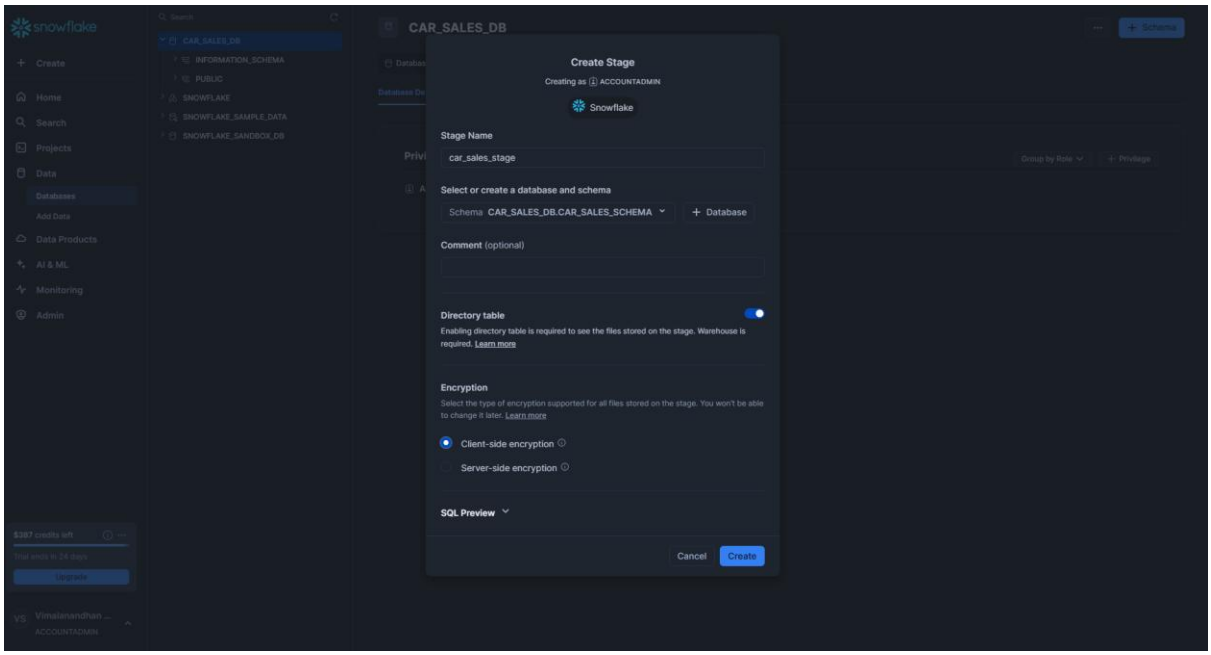

2025-02-06 1:02pm
Databases
Worksheets
Search objects
SNOWFLAKE
SNOWFLAKE_SAMPLE_DATA
SNOWFLAKE_SANDBOX_DB
ACCOUNTADMIN
COMPUTE_WH (X-Small)
Share
CAR_SALES_DB.CAR_SALES_SCHEMA
Settings
Code Versions
-- Create the database
CREATE DATABASE car_sales_db;
CREATE WAREHOUSE compute_wh
WITH WAREHOUSE_SIZE = 'XSMALL';
-- Create the schema inside the database
CREATE SCHEMA car_sales_schema;
-- Use the schema
Results
Chart
status
Schema CAR_SALES_SCHEMA successfully created.
Query Details
Query duration
118ms
Rows
1
Query ID
01ba3818-0004-8e58-0...
Show more
status
100% filled
Ask Copilot

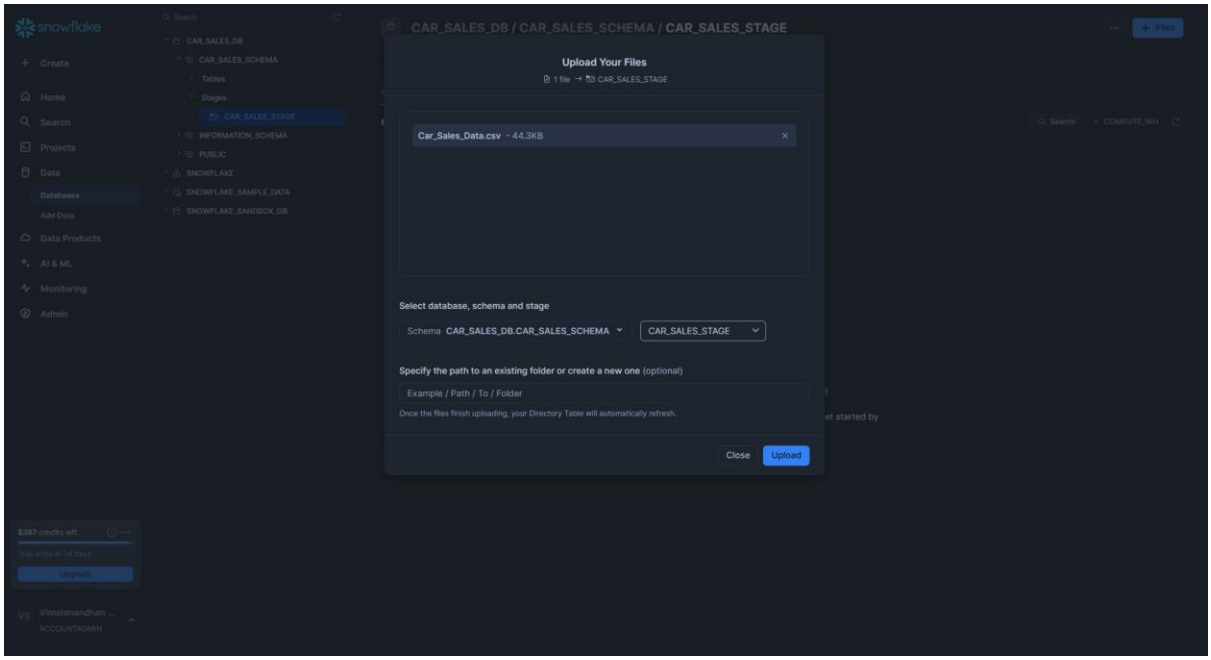

2025-02-06 1:02pm
Databases
Worksheets
Search objects
SNOWFLAKE
SNOWFLAKE_SAMPLE_DATA
SNOWFLAKE_SANDBOX_DB
ACCOUNTADMIN
COMPUTE_WH (X-Small)
Share
CAR_SALES_DB.CAR_SALES_SCHEMA
Settings
Code Versions
-- Create the database
CREATE DATABASE car_sales_db;
CREATE WAREHOUSE compute_wh
WITH WAREHOUSE_SIZE = 'XSMALL';
-- Create the schema inside the database
CREATE SCHEMA car_sales_schema;
-- Use the schema
USE SCHEMA car_sales_schema;
CREATE OR REPLACE TABLE car_sales_schema.car_sales_tbl (
cars_name STRING,
car_price INT,
purchase_date DATE,
brand_name STRING
);
Results
Chart
status
Table CAR_SALES_TBL successfully created.
Query Details
Query duration
188ms
Rows
1
Query ID
01ba3819-0004-8e5e-0...
Show more
status
100% filled
Ask Copilot

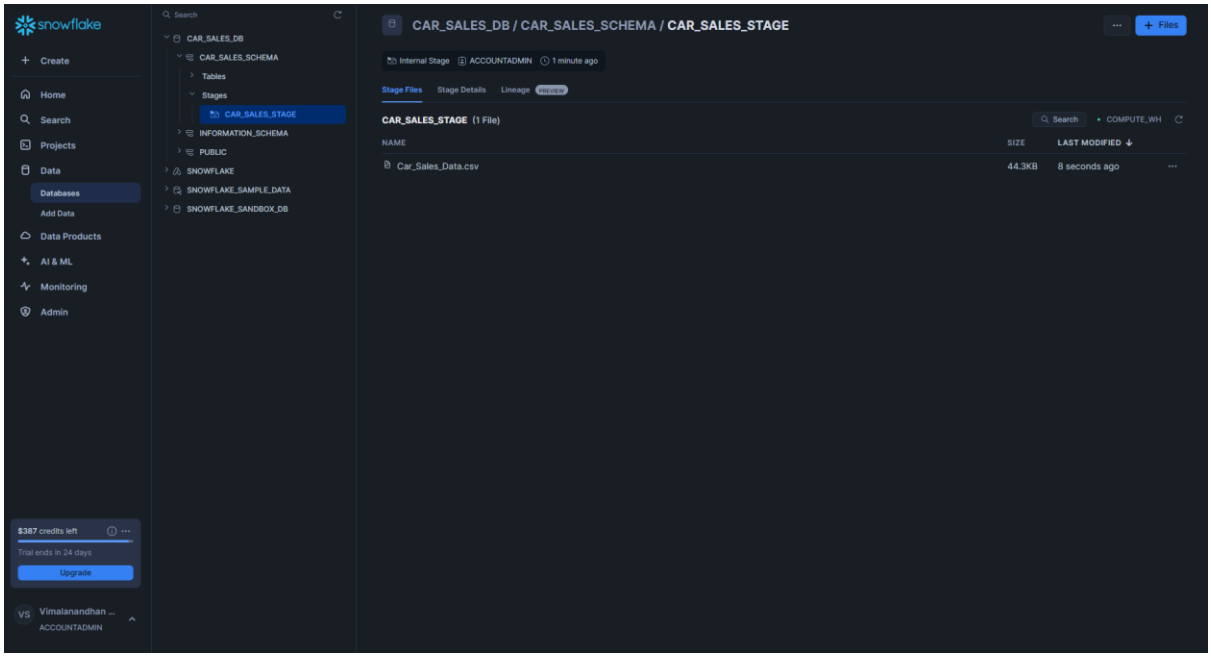
3. (+1) Create Stage 'car_sales_stage' in Snowflake and load the csv.

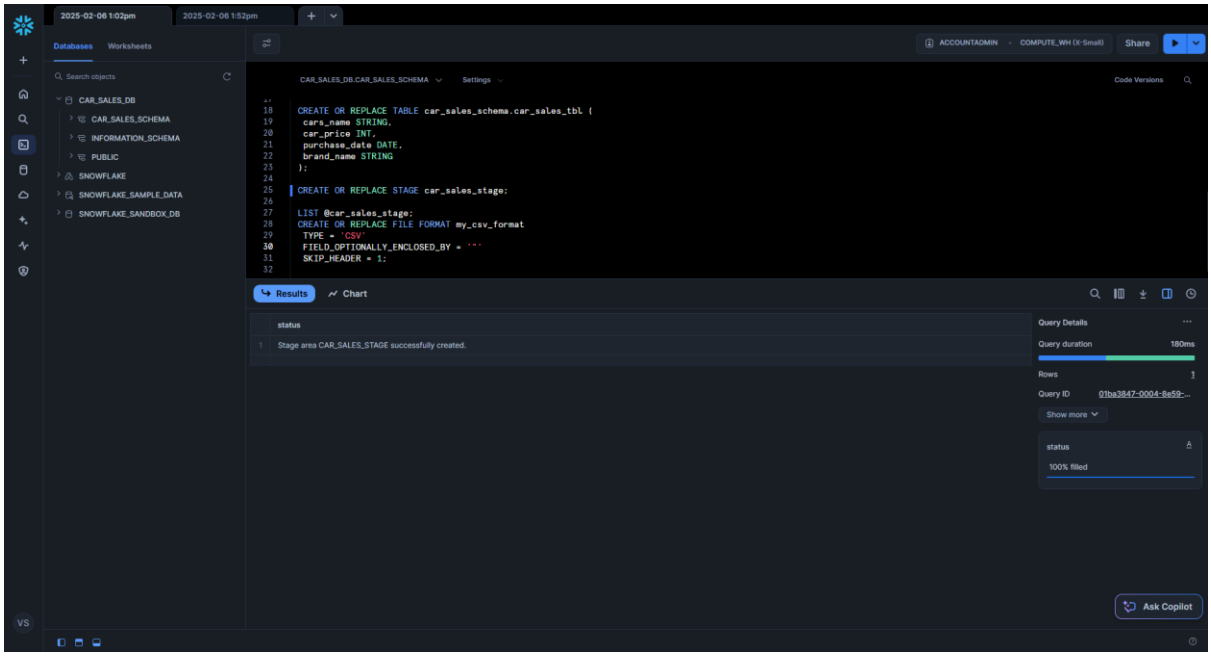
CODE:

CREATE OR REPLACE STAGE car_sales_stage; LIST @car_sales_stage; CREATE OR REPLACE FILE FORMAT my_csv_format TYPE = 'CSV' FIELD_OPTIONALLY_ENCLOSED_BY = '"' SKIP_HEADER = 1;









4. (+2) Create file format and Load Data into the 'car_sales_data' table (COPY INTO).

CODE:

-----Creation of File

```
CREATE OR REPLACE FILE FORMAT car_sales_file_format
TYPE = 'CSV'
FIELD_OPTIONALLY_ENCLOSED_BY = '"'
SKIP_HEADER = 1;
```

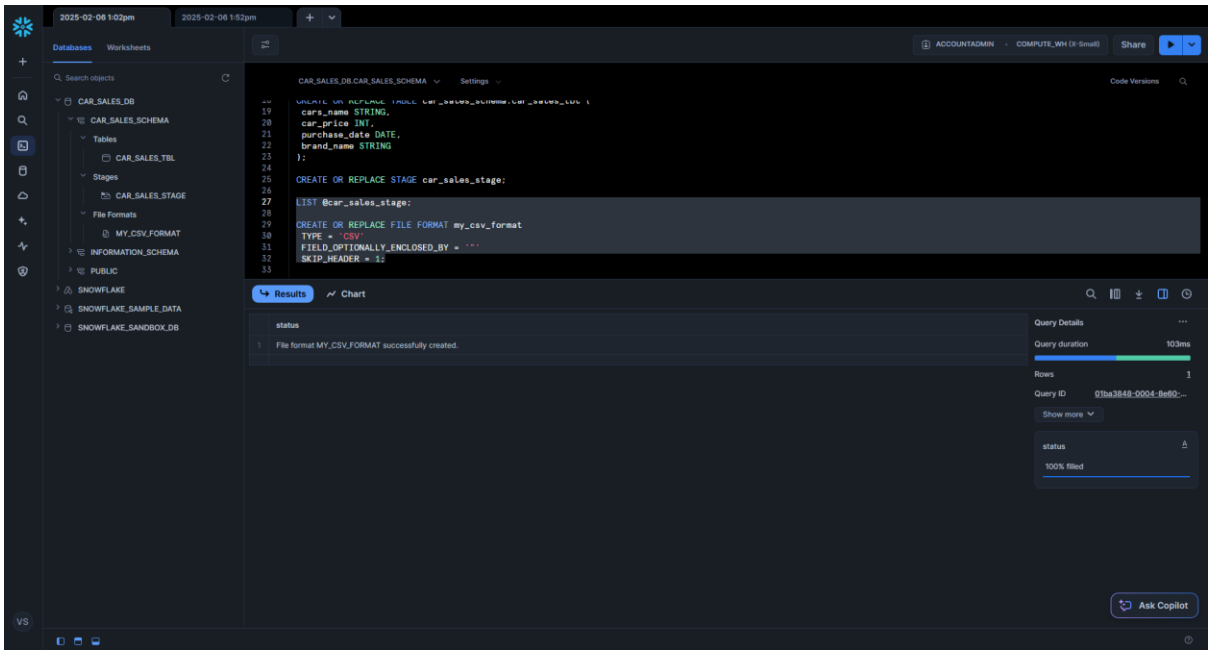
-----Load Data into Table

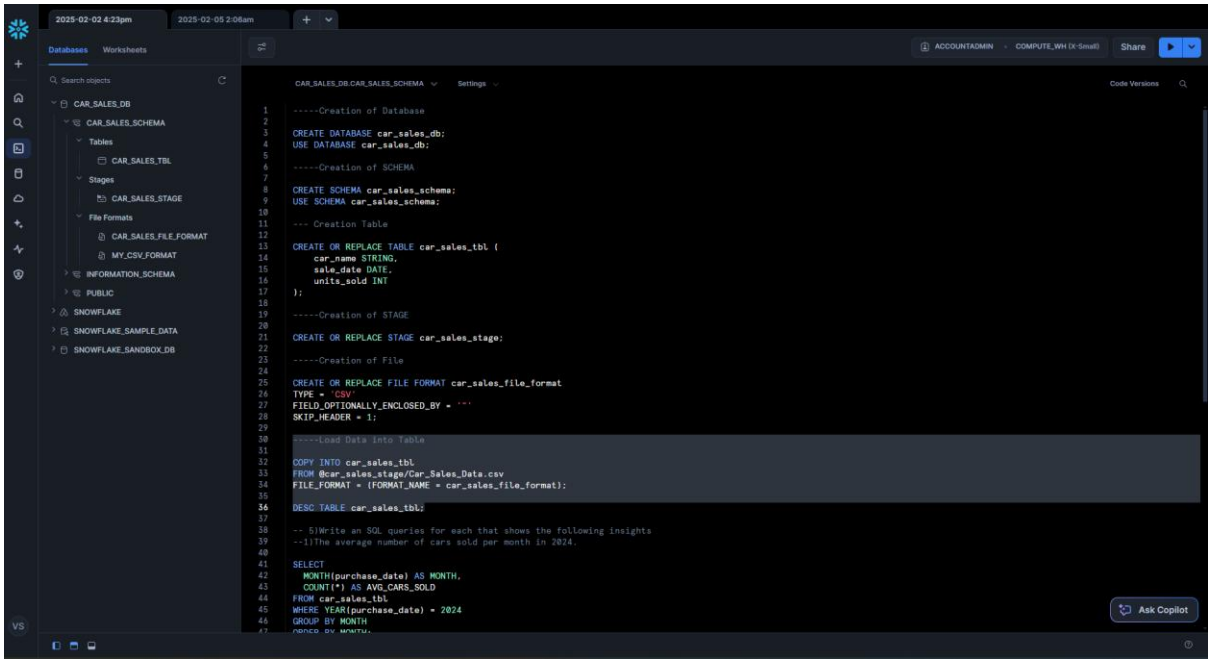
COPY INTO car_sales_tbl

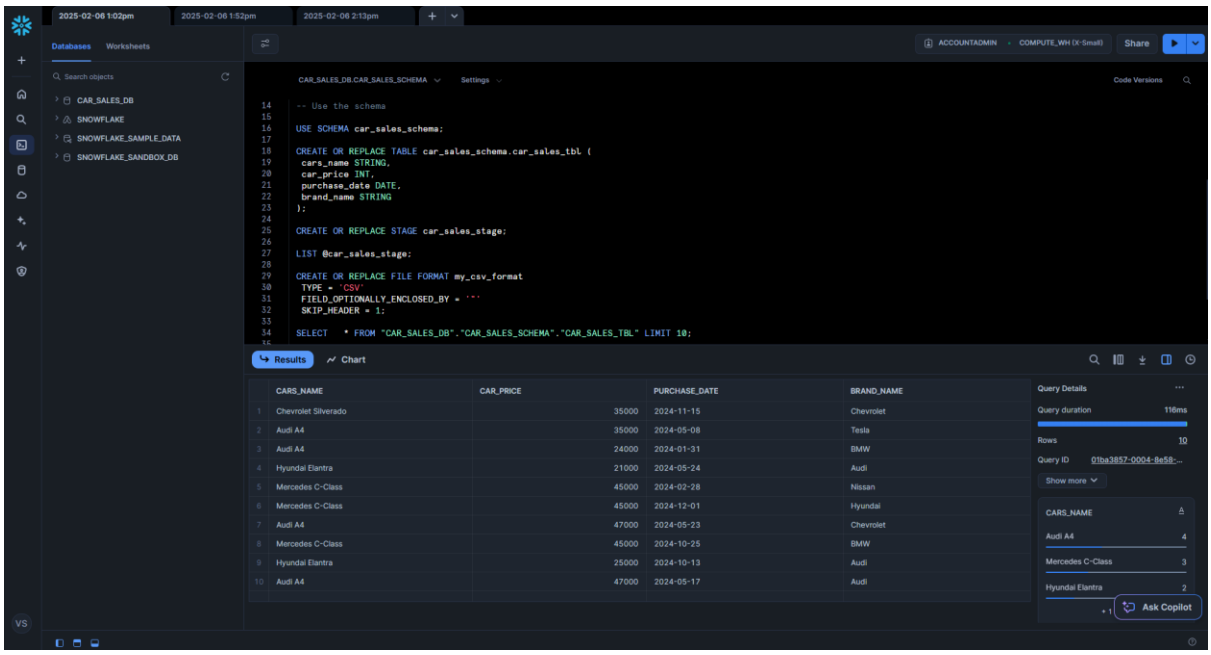
FROM @car_sales_stage/Car_Sales_Data.csv

FILE_FORMAT = (FORMAT_NAME = car_sales_file_format);

DESC TABLE car_sales_tbl;







5. (+6) Write an SQL queries for each that shows the following insights

    1. The average number of cars sold per month in 2024.

With regard to step 5, the expected output format is mentioned below:

A. Expected Output

MONTH	AVG_CARS_SOLD
1	4
2	116

3	94
4	106
5	89
6	99

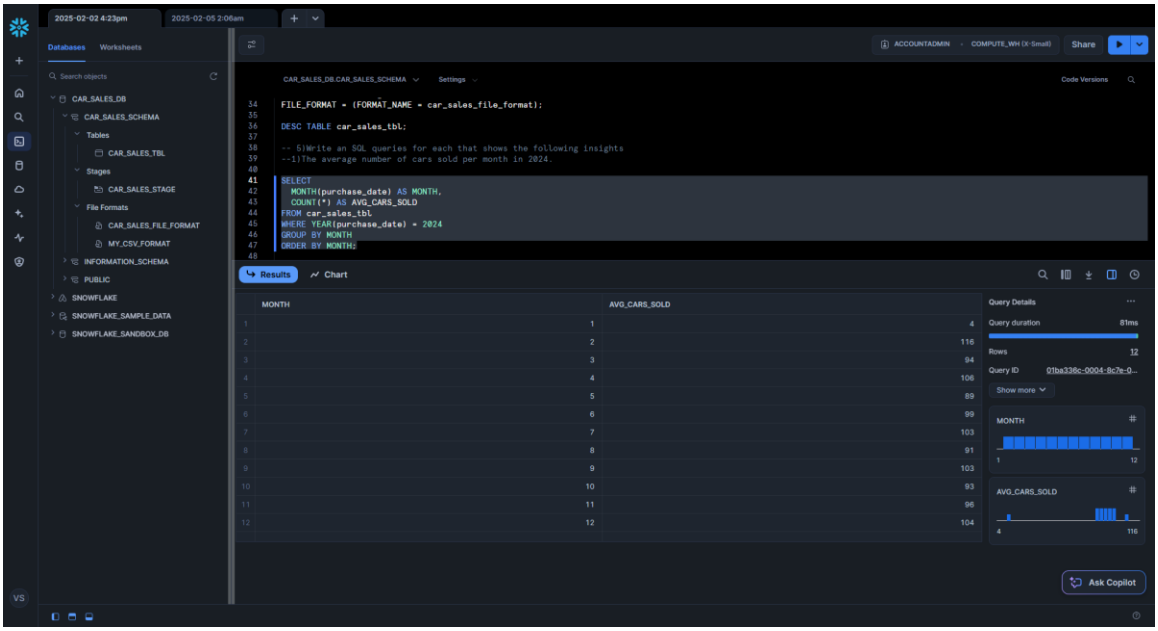
....Till Month 12

CODE:

5)Write an SQL queries for each that shows the following insights

1)The average number of cars sold per month in 2024.

```
SELECT
 MONTH(purchase_date) AS MONTH,
 COUNT(*) AS AVG_CARS_SOLD
FROM car_sales_tbl
WHERE YEAR(purchase_date) = 2024
GROUP BY MONTH
ORDER BY MONTH;
```



2. The top 5 best-selling cars along with the count in 2024.

B. Expected Output

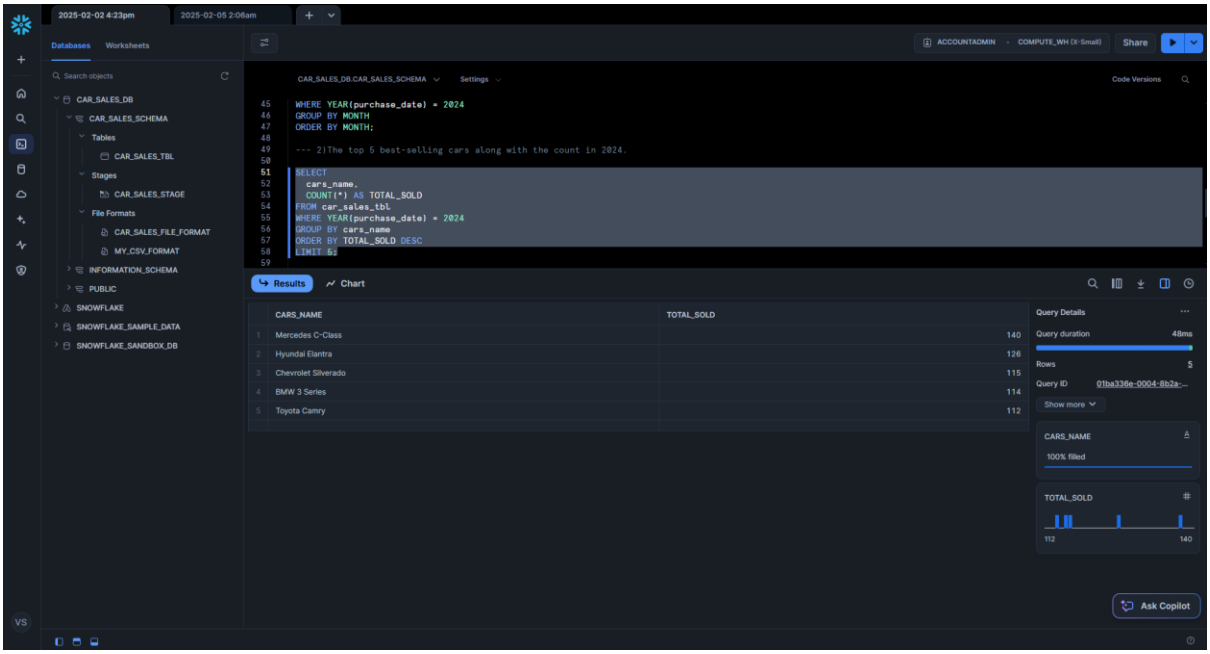
CARS_NAME	TOTAL_SOLD
Mercedes C-Class	140
Hyundai Elantra	126

....till top 5

CODE:

2)The top 5 best-selling cars along with the count in 2024.

```
SELECT
 cars_name,
 COUNT(*) AS TOTAL_SOLD
FROM car_sales_tbl
WHERE YEAR(purchase_date) = 2024
GROUP BY cars_name
ORDER BY TOTAL_SOLD DESC
LIMIT 5;
```



3. The top 3 least-selling cars along with the count in 2024.

C. Expected Output

CARS_NAME	TOTAL_SOLD
Audi A4	79

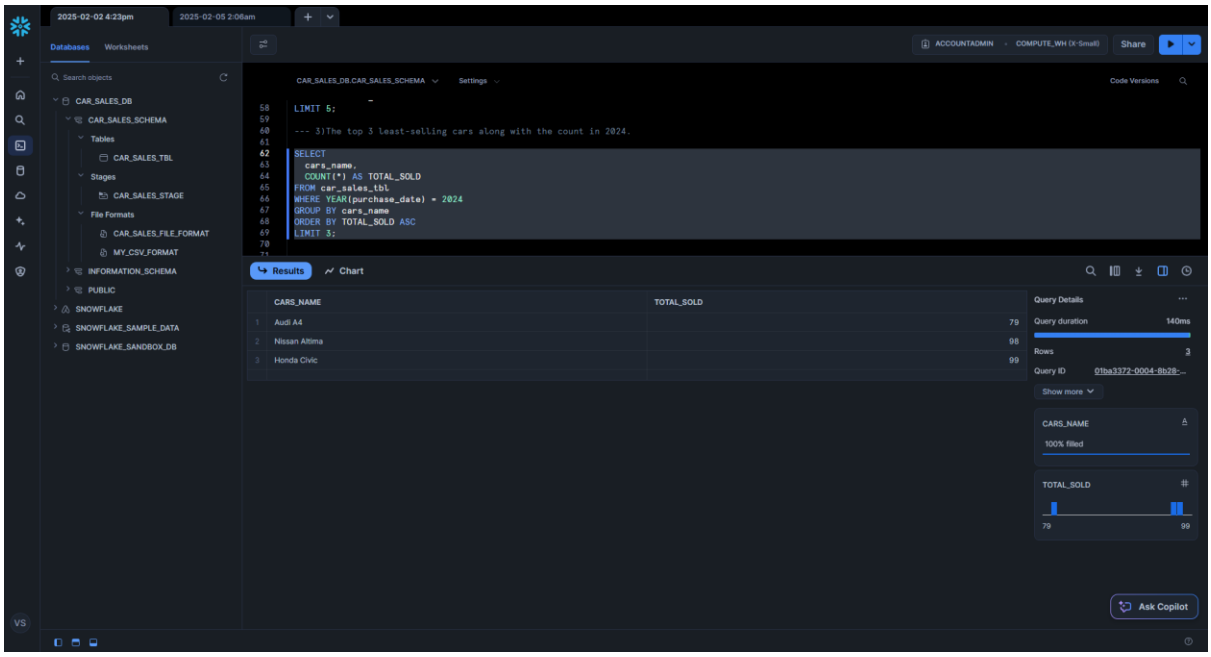
..till least 3

CODE:

3)The top 3 least-selling cars along with the count in 2024.

```
SELECT
 cars_name,
 COUNT(*) AS TOTAL_SOLD
FROM car_sales_tbl
WHERE YEAR(purchase_date) = 2024
GROUP BY cars_name
ORDER BY TOTAL_SOLD ASC
```

LIMIT 3;


2025-02-02 4:23pm
2025-02-05 2:06am
Databases
Worksheets
Search objects
CAR_SALES_DB
CAR_SALES_SCHEMA
Tables
CAR_SALES_TBL
Stages
CAR_SALES_STAGE
File Formats
CAR_SALES_FILE_FORMAT
MY_CSV_FORMAT
INFORMATION_SCHEMA
PUBLIC
SNOWFLAKE
SNOWFLAKE_SAMPLE_DATA
SNOWFLAKE_SANDBOX_DB
ACCOUNTADMIN
COMPUTE_WH (X-Small)
Share
CAR_SALES_DB.CAR_SALES_SCHEMA
Settings
Code Versions
58 LIMIT 5;
59
60 --- 3)The top 3 least-selling cars along with the count in 2024.
61
62 SELECT
63 cars_name,
64 COUNT(*) AS TOTAL_SOLD
65 FROM car_sales_tbl
66 WHERE YEAR(purchase_date) = 2024
67 GROUP BY cars_name
68 ORDER BY TOTAL_SOLD ASC
69 LIMIT 3;
70
Results
Chart
CARS_NAME
TOTAL_SOLD
1 Audi A4 79
2 Nissan Altima 98
3 Honda Civic 99
Query Details
Query duration 140ms
Rows 3
Query ID 01ba3372-0004-8b28-...
Show more
CARS_NAME
100% filled
TOTAL_SOLD
79
99
Ask Copilot