1 - EC2 Dashboard (Account Created)

2 - S3 Bucket (Data Lake)

3 - Glue (Crawler Created)

AWS Glue

Data catalog
Databases
  Tables
  Connections
Crawlers
  Classifiers
Schema registries
  Schemas
Settings

ETL
AWS Glue Studio ↗
Jobs ↗ *New*
Jobs (legacy)
  ML Transforms
Blueprints
Workflows
Triggers
Dev endpoints
  Notebooks ↗ *New*
  Notebooks (legacy)

Crawlers > S3toGlue

[Run crawler] [Edit]

Name	S3toGlue
Description	
Create a single schema for each S3 path	false
Security configuration	
Tags	-
State	Ready
Schedule	
Last updated	Wed Aug 03 16:33:51 GMT+100 2022
Date created	Wed Aug 03 16:33:51 GMT+100 2022
Database	carlist
Table level	
Service role	service-role/AWSGlueServiceRole-carlist
Selected classifiers	
Data store	S3
Include path	s3://mountasserbucket
Connection	
Exclude patterns	

Configuration options

Schema updates in the data store	Update the table definition in the data catalog.
Object deletion in the data store	Mark the table as deprecated in the data catalog.

4 - Glue (Database Built)

AWS Glue

Data catalog
Databases
  Tables
  Connections
Crawlers
  Classifiers
Schema registries
  Schemas
Settings

ETL
AWS Glue Studio ↗
Jobs ↗ *New*
Jobs (legacy)
  ML Transforms
Blueprints
Workflows
Triggers
Dev endpoints
  Notebooks ↗ *New*
  Notebooks (legacy)
Interactive Session *New*

Security
Security configurations

Tutorials
Add crawler

[Edit table] [Delete table]   [View properties] [Compare versions] [Edit schema]

Name	mountasserbucket
Description	
Database	carlist
Classification	csv
Location	s3://mountasserbucket/
Connection	
Deprecated	No
Last updated	Wed Aug 03 20:14:16 GMT+100 2022
Input format	org.apache.hadoop.mapred.TextInputFormat
Output format	org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Serde serialization lib	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
Serde parameters	field.delim ,

Table properties

skip.header.line.count	1	sizeKey	17598653	objectCount	1	UPDATED_BY_CRAWLER	S3toGlue	CrawlerSchemaSerializerVersion	1.0
recordCount	409271	averageRecordSize	43	CrawlerSchemaDeserializerVersion	1.0	compressionType	none	columnsOrdered	true
areColumnsQuoted	false	delimiter	,	typeOfData	file				

Schema
Showing: 1 - 7 of 7

	Column name	Data type	Partition key	Comment
1	price	bigint		
2	year	bigint		
3	mileage	bigint		
4	city	string		
5	state	string		
6	make	string		

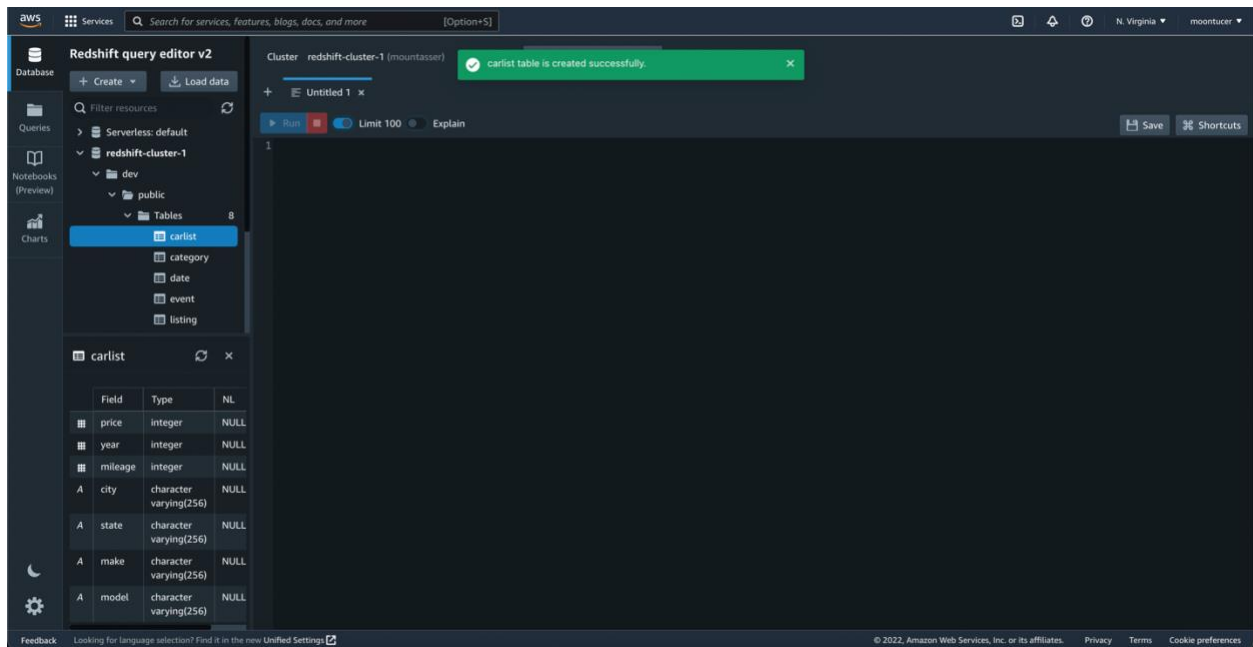## 5 - IAM (Role Created)

6 - Redshift (Data Warehouse)

7 - Redshift (IAM Role Associated to Cluster)
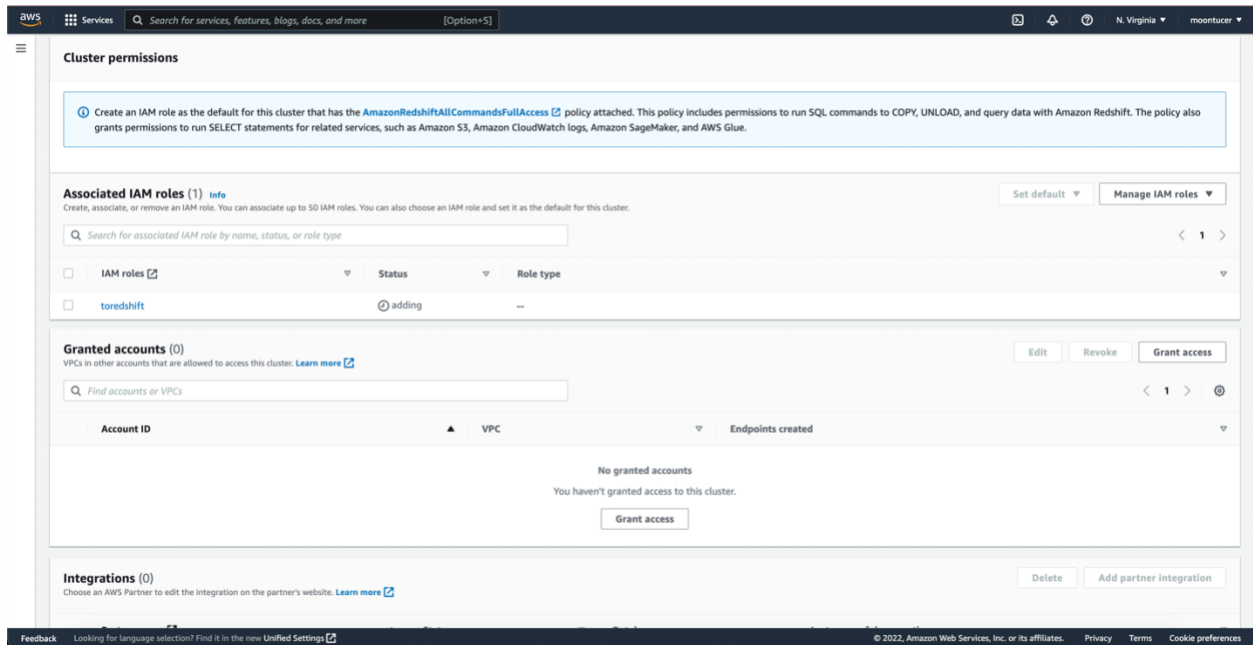
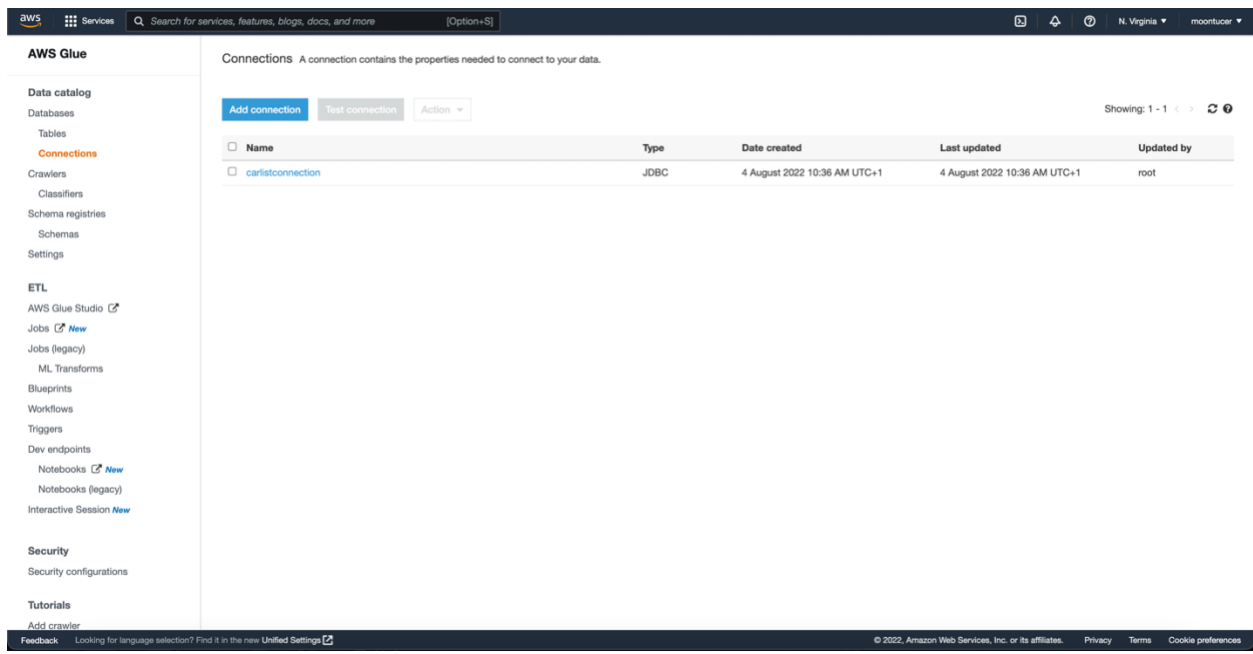**8 - Redshift (Database connected w query editor)**
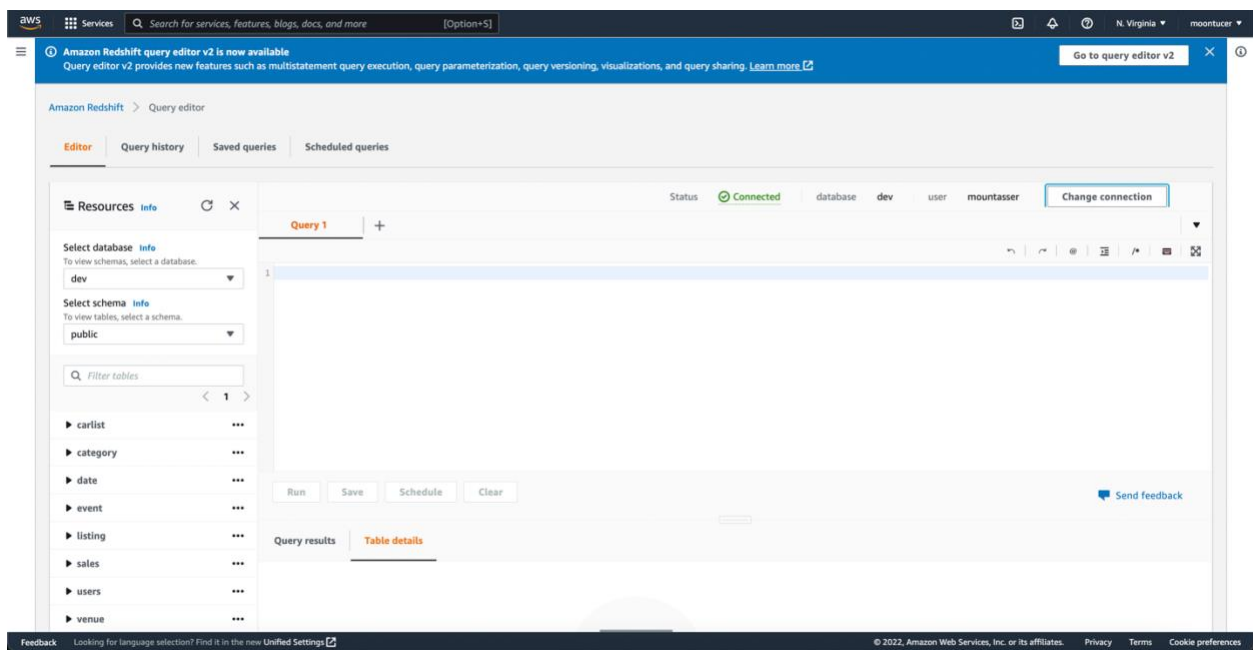
9 - Redshift (Table created w query editor v2)

10 - Glue (Connection added to cluster)

11 - VPC (S3 Endpoint Created)

12 - VPC (Security Group Creation w TCP inbound)

13 - VPC (NAT Gateway Created)

14 - Updating NAT Routes

15 - Glue (Crawler to ingest in cluster)
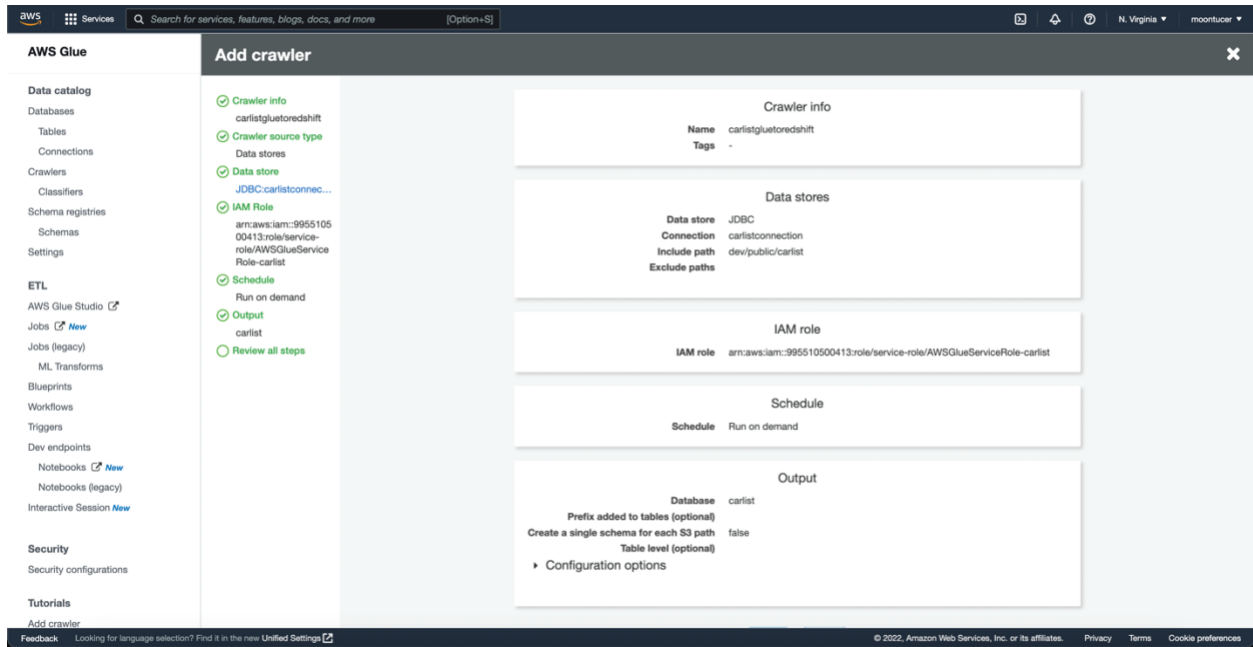
16 - Running ETL Job b4 Analytics
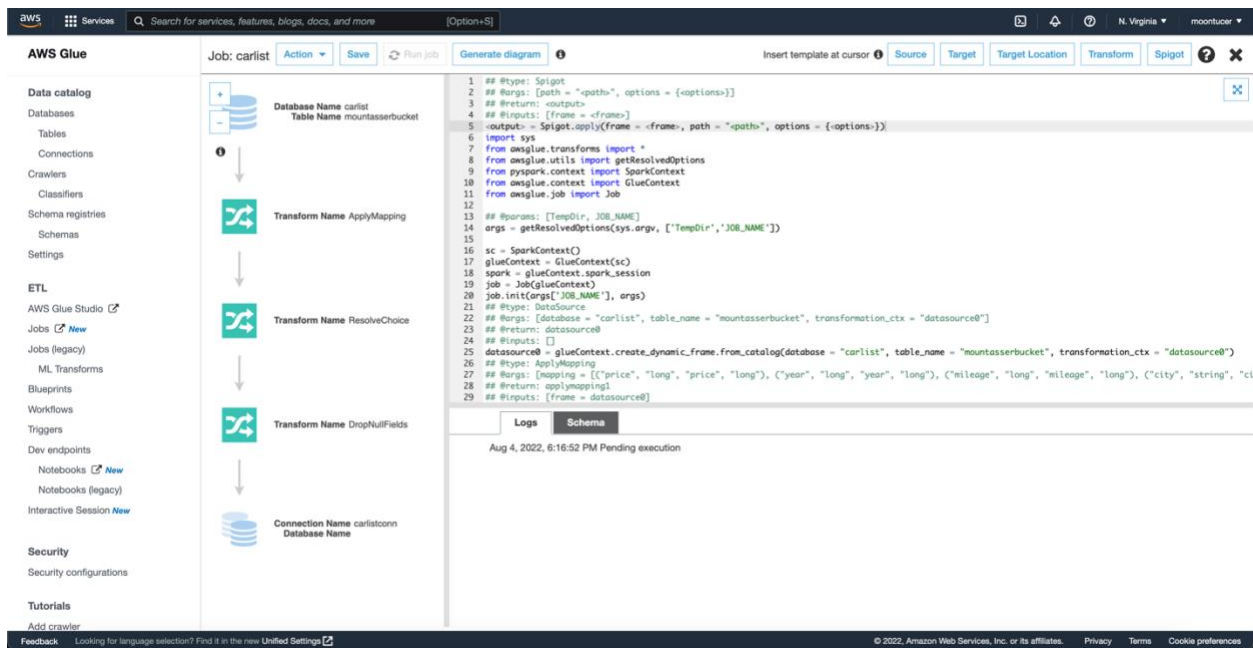
17 - Redshift Analytics 1

18 - Redshift Analytics 2

The most popular car models

Legend:
- SILVERADO
- TAHOE4WD
- SUBURBAN4WD
- TAHOEPREMIER
- TAHOELT
- CORVETTECONVERTIBLE
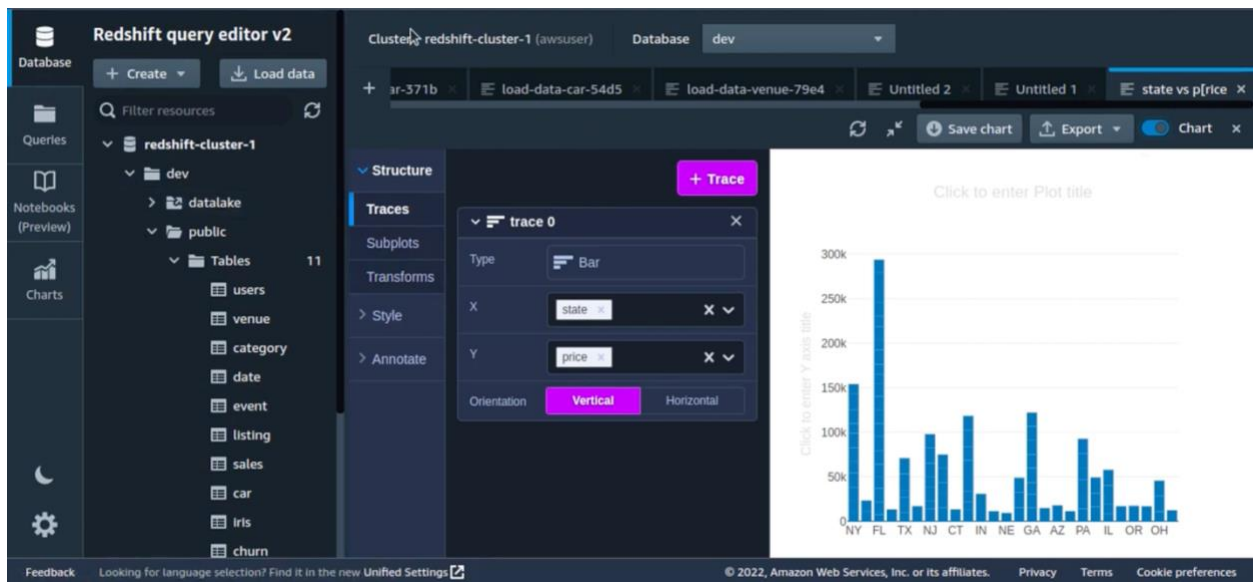- CORVETTE2DR
- CORVETTESTINGRAY
- CORVETTEGRAND
- SUBURBAN2WD
- TAHOE2WD
- CAMAROCONVERTIBLE
- CAMAROCOUPE
- CORVETTE
- TAHOE4DR
- CORVETTEZ06
- CORVETTECOUPE
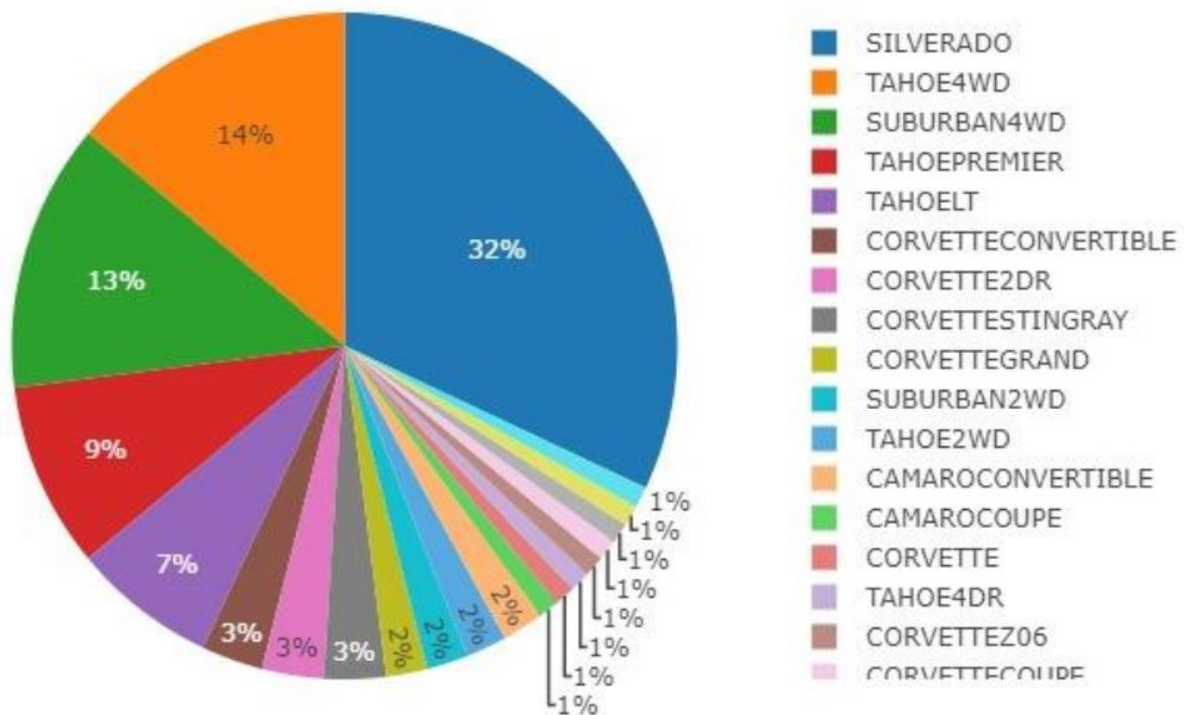
19 - Redshift Analytics 3

The Prices of the cars according to the year of manufacturing