



TECH
CONSULTING

VEHICLE SALES

By Juan Granillo



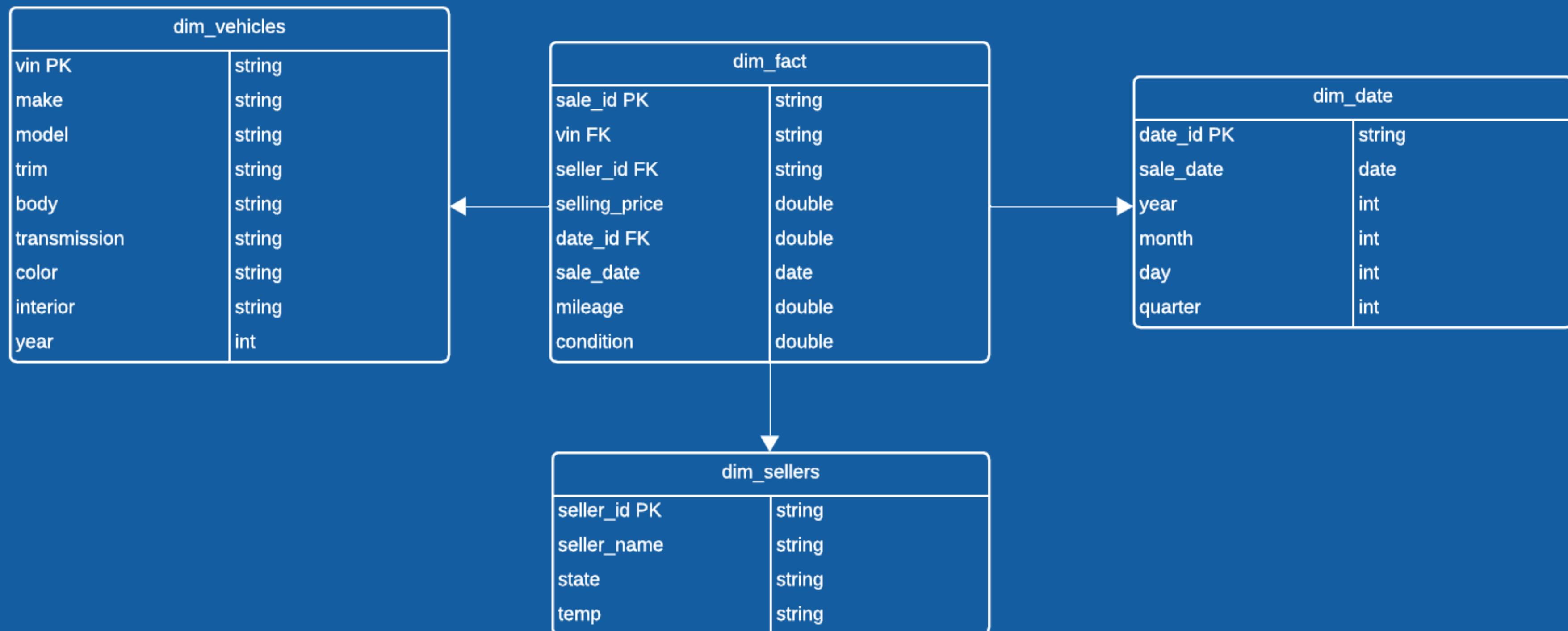


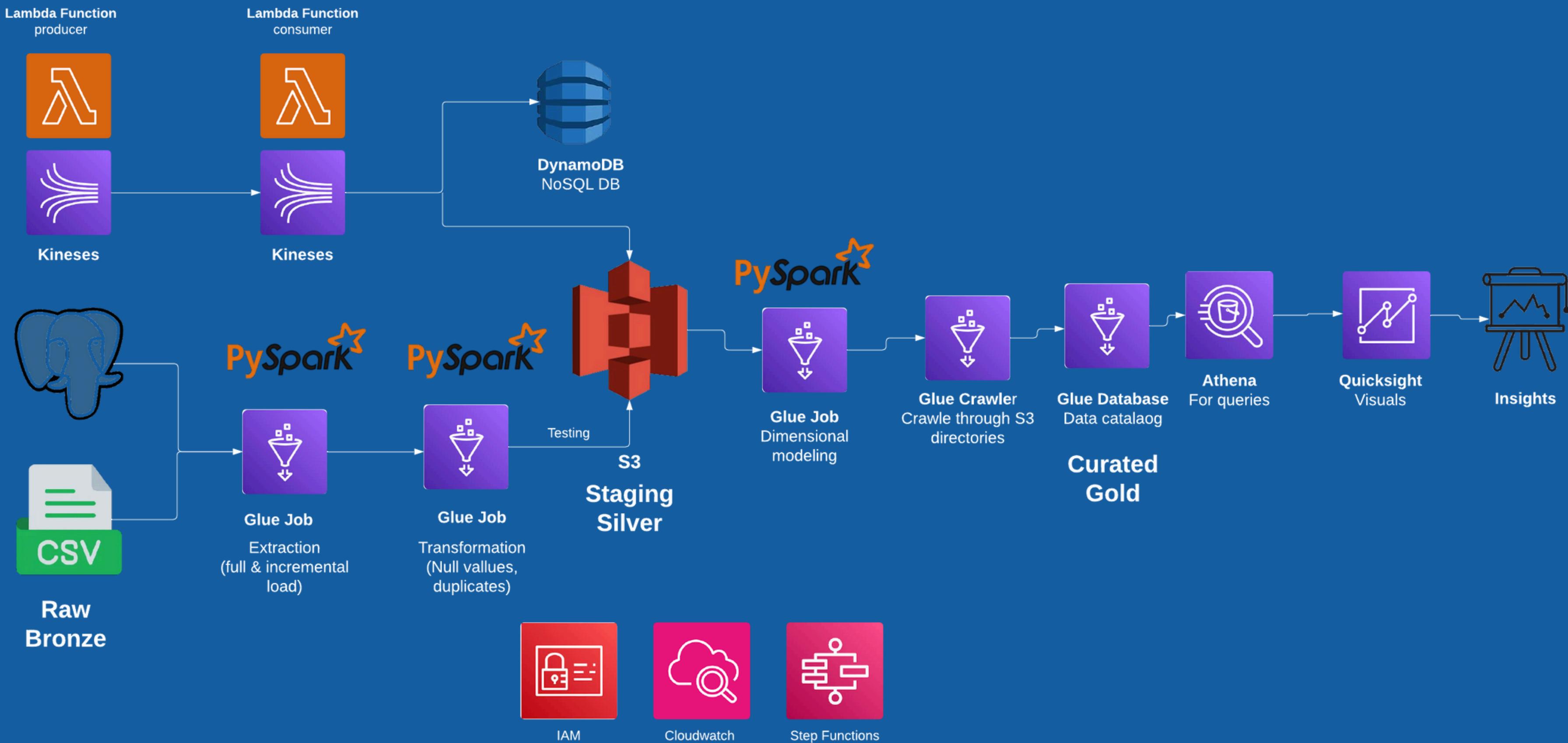
PROBLEM STATEMENT

The challenge is to analyze these factors together and gain insights into how climate and vehicle characteristics influences vehicle sales patterns in different states

BUSINESS USE CASE

Improved decision-making for stock management, targeted promotions, and efficient resource allocation based on climate trends.





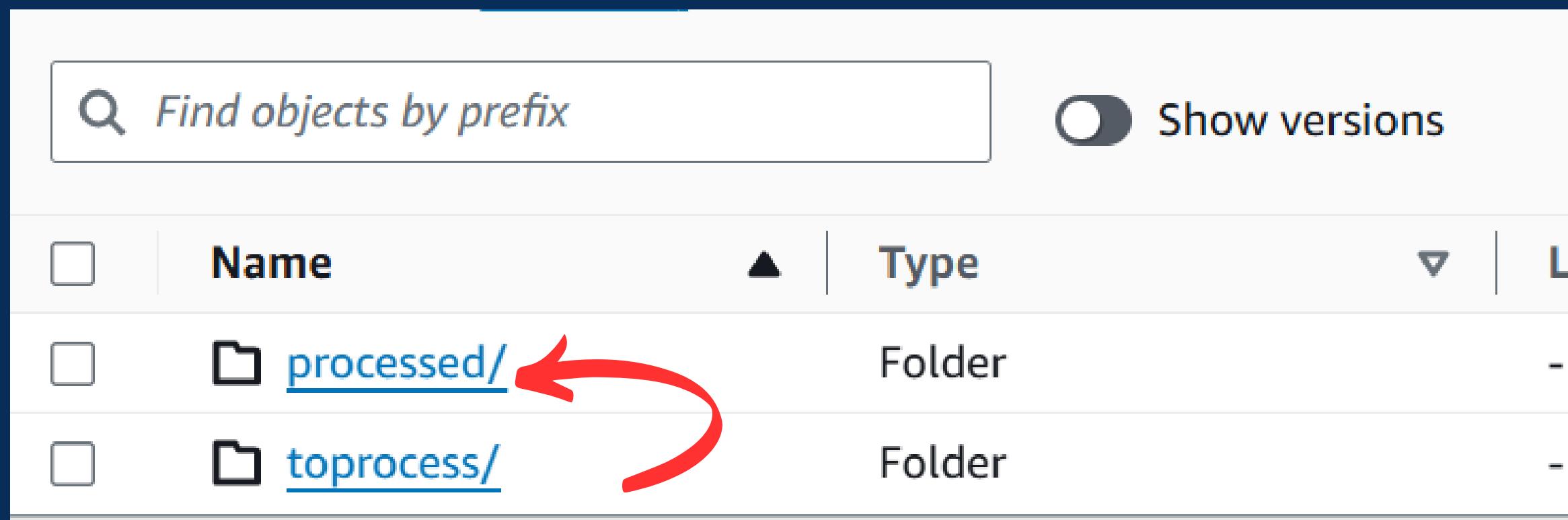
Implemented CDC

- If timestamp on data > timestamp on table then ingest

```
testdb=> select * from timestamp_table;
 id |          created_at
---+-----
  1 | 2004-05-01 12:00:00
 24 | 2024-10-16 03:15:04.141387
 25 | 2024-10-19 21:41:24.797246
(3 rows)
```

ETL

- Raw data is ingested from Postgres server, transformed and loaded into stagingsilver/toprocess



<input type="checkbox"/>	Name	Type	<input type="checkbox"/> L
<input type="checkbox"/>	<u>processed/</u>	Folder	-
<input type="checkbox"/>	<u>toprocess/</u>	Folder	-

Testing data quality

- Data is checked for any nulls, duplicates, or negative values
- Using Lambda function
- If status code returns anything but 200, step function is halted

Function Logs:

```
START RequestId: b34f97ac-d8c2-4629-b55d-4798853191ba Version: 1
Total Duplicates: 0
Total Nulls: {}
Total Negative Values: 0
END RequestId: b34f97ac-d8c2-4629-b55d-4798853191ba
```

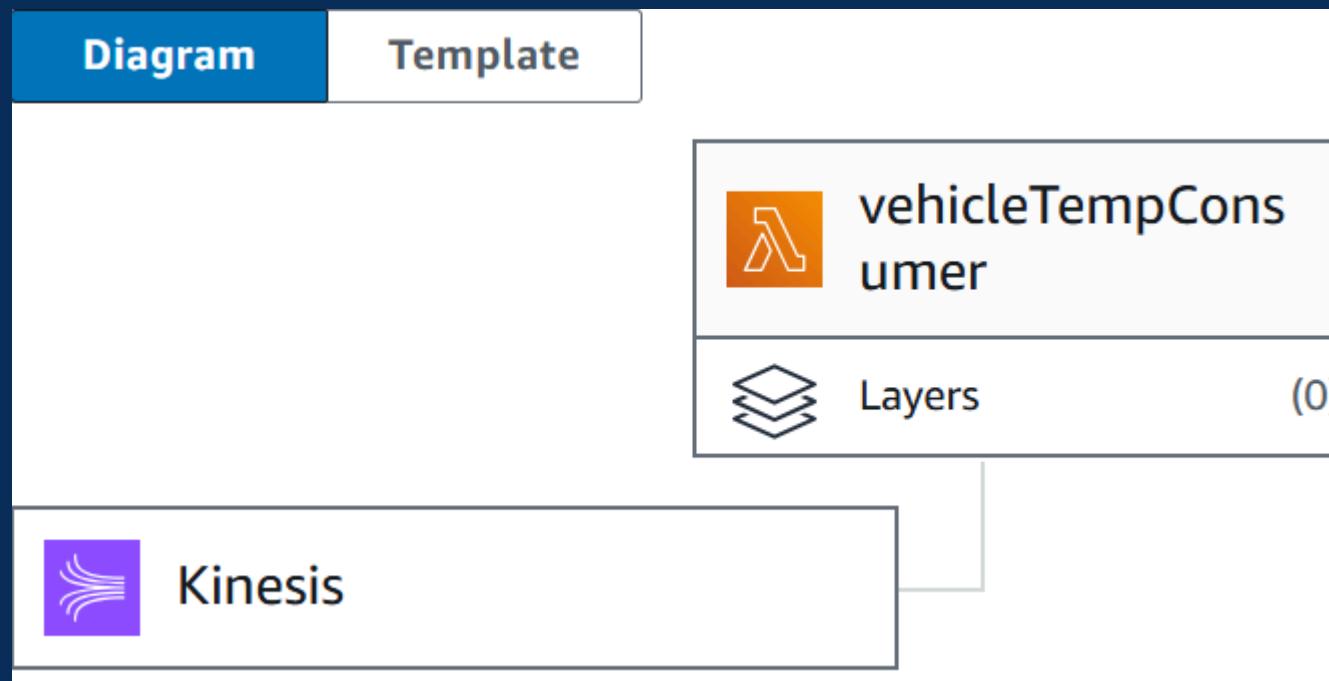
Dimensional Modelling

- Data is moved from toprocess/ to processed/
- Data is split into appropriate tables with appropriate PK and FK
- Moved into separate folders within goldcurated to be able to be crawled

	Name	Type	Last modified
	 dim_date/	Folder	-
	 dim_sellers/	Folder	-
	 dim_temp/	Folder	-
	 dim_vehicles/	Folder	-
	 fact_table/	Folder	-

Real Time Data

- Consumer triggered by Kinesis
- Writes to DynamoDB
- DynamoDB exported to S3 when step function is ran



Items returned (50)

C Actions ▾ Create if

	state (String)	average_rainfall	average
<input type="checkbox"/>	or	45	60
<input type="checkbox"/>	ar	50	65
<input type="checkbox"/>	pa	39	58

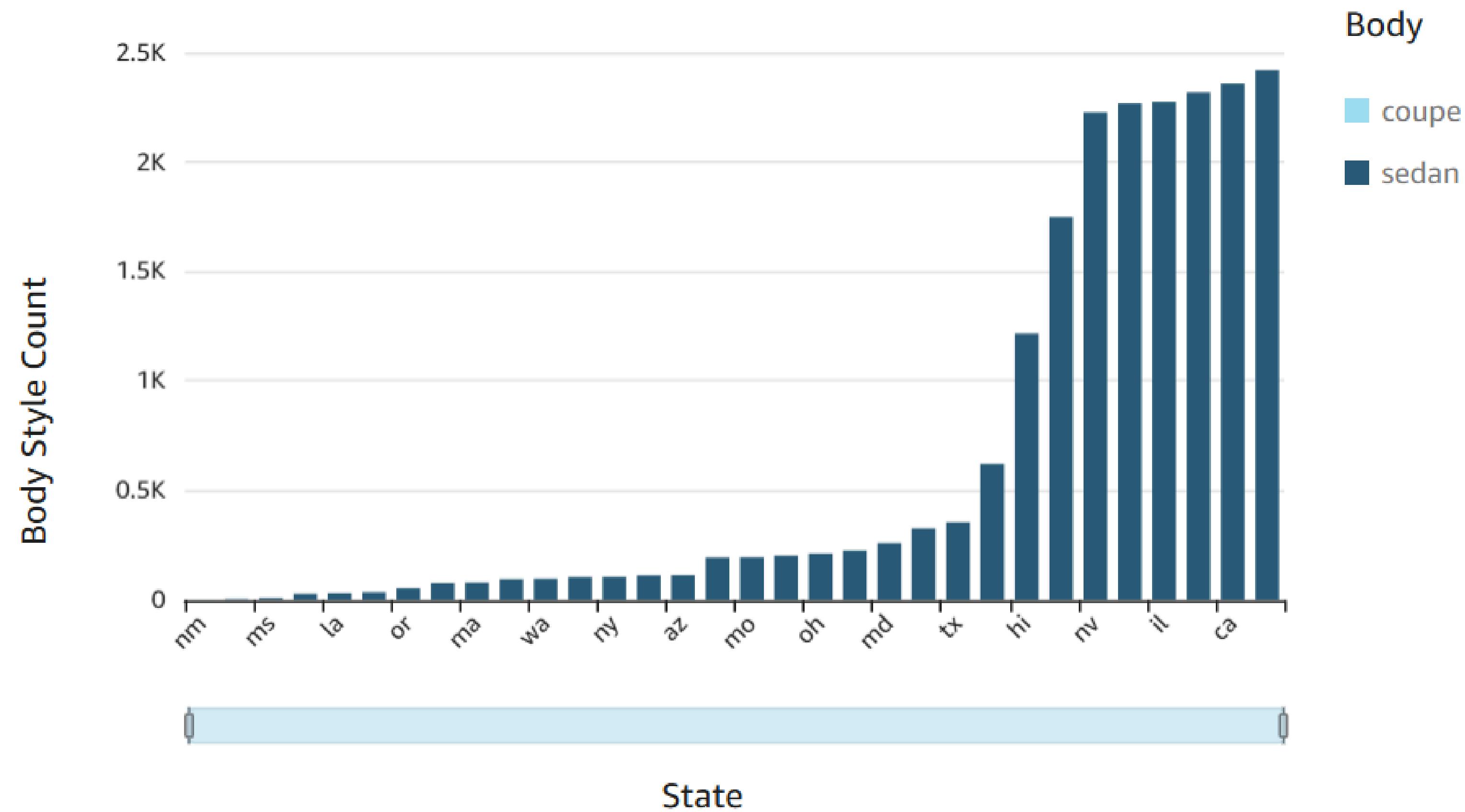
Data querying through Athena

- Glue crawlers crawl through tables in S3
- Crawler is linked to database within Glue data catalog
- Data is then queried through Athena

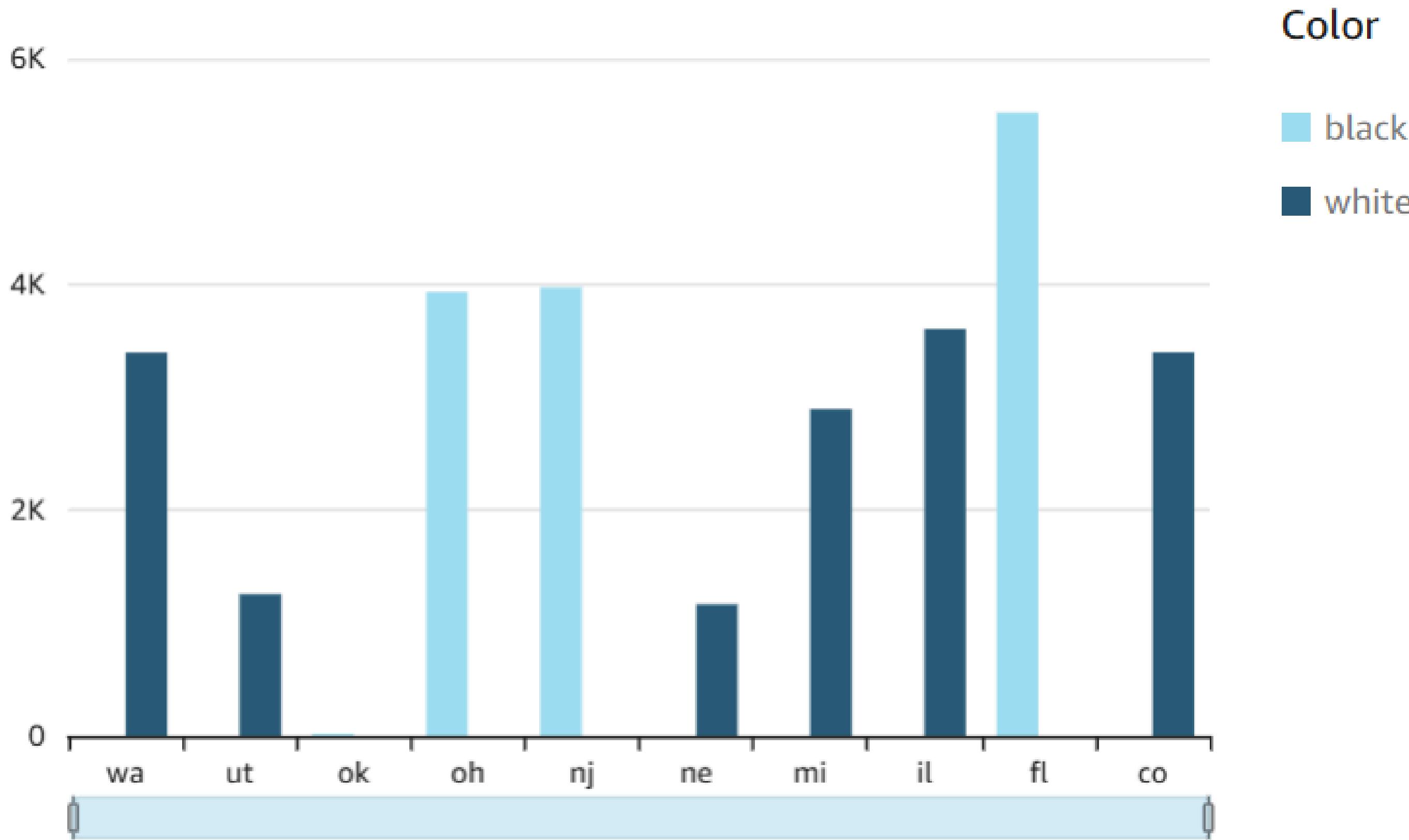
The screenshot shows the results of an Athena query. At the top, it indicates the query is 'Completed' with a green checkmark. Below that, it shows 'Time in queue: 67 ms', 'Run time: 522 ms', and 'Data scanned: 8.1 GB'. The main area is titled 'Results (10)' and includes a search bar labeled 'Search rows'. There are two buttons: 'Copy' and 'Download results'. The data is presented in a table with columns: '#', 'vin', 'seller_id', 'sellingprice', and 'saledate'. The table contains three rows of data.

#	vin	seller_id	sellingprice	saledate
1	5YJXCAE20MF123456	1	110000.0	2024-04-05 14:45:00.000
2	wp1aa2a2xkla12345	1	85000.0	2024-04-02 09:30:00.000
3	19uuua66245a069478	138	6500.0	2014-12-16 12:30:00.000

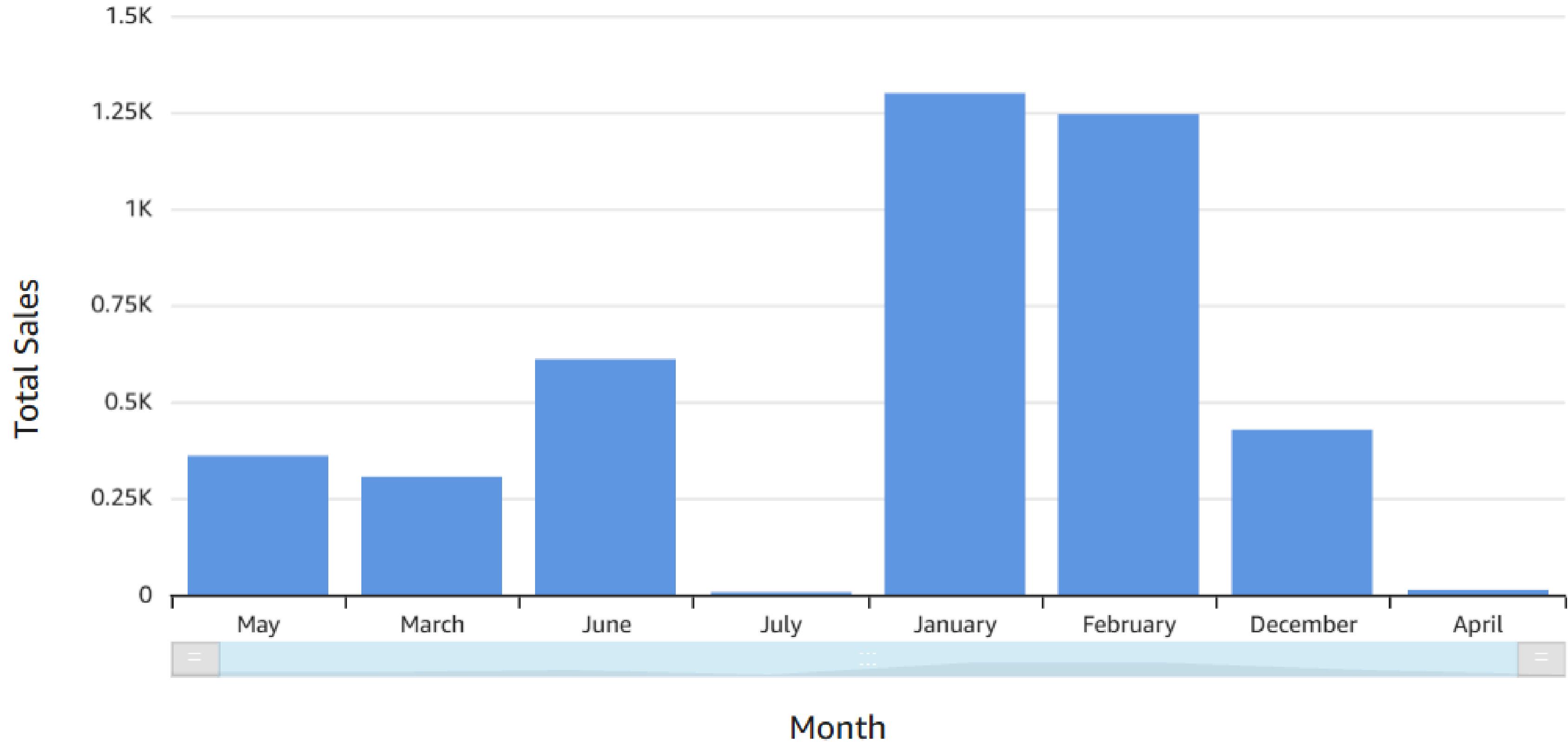
Sum of Body Style per State



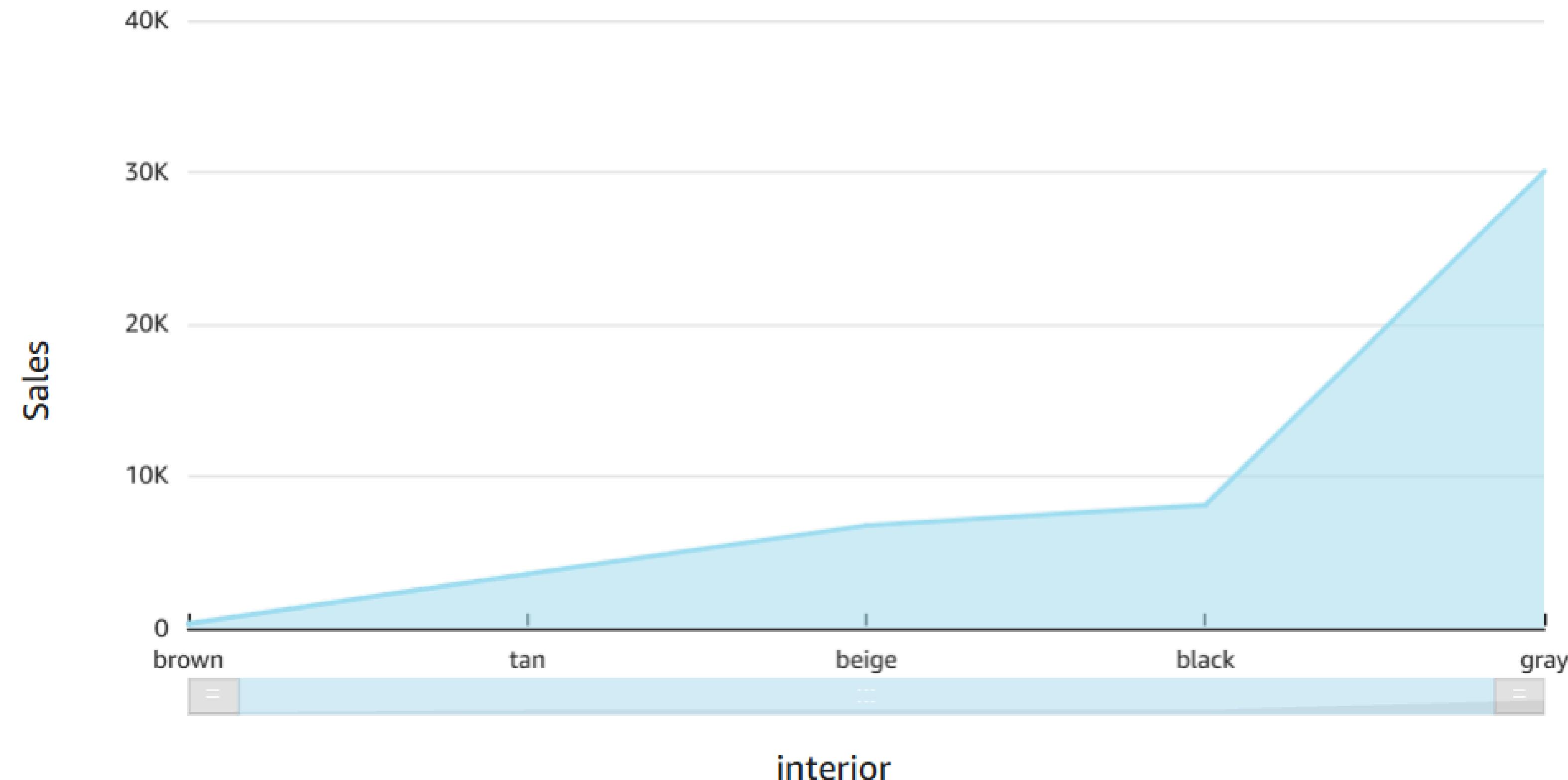
Color Count By State



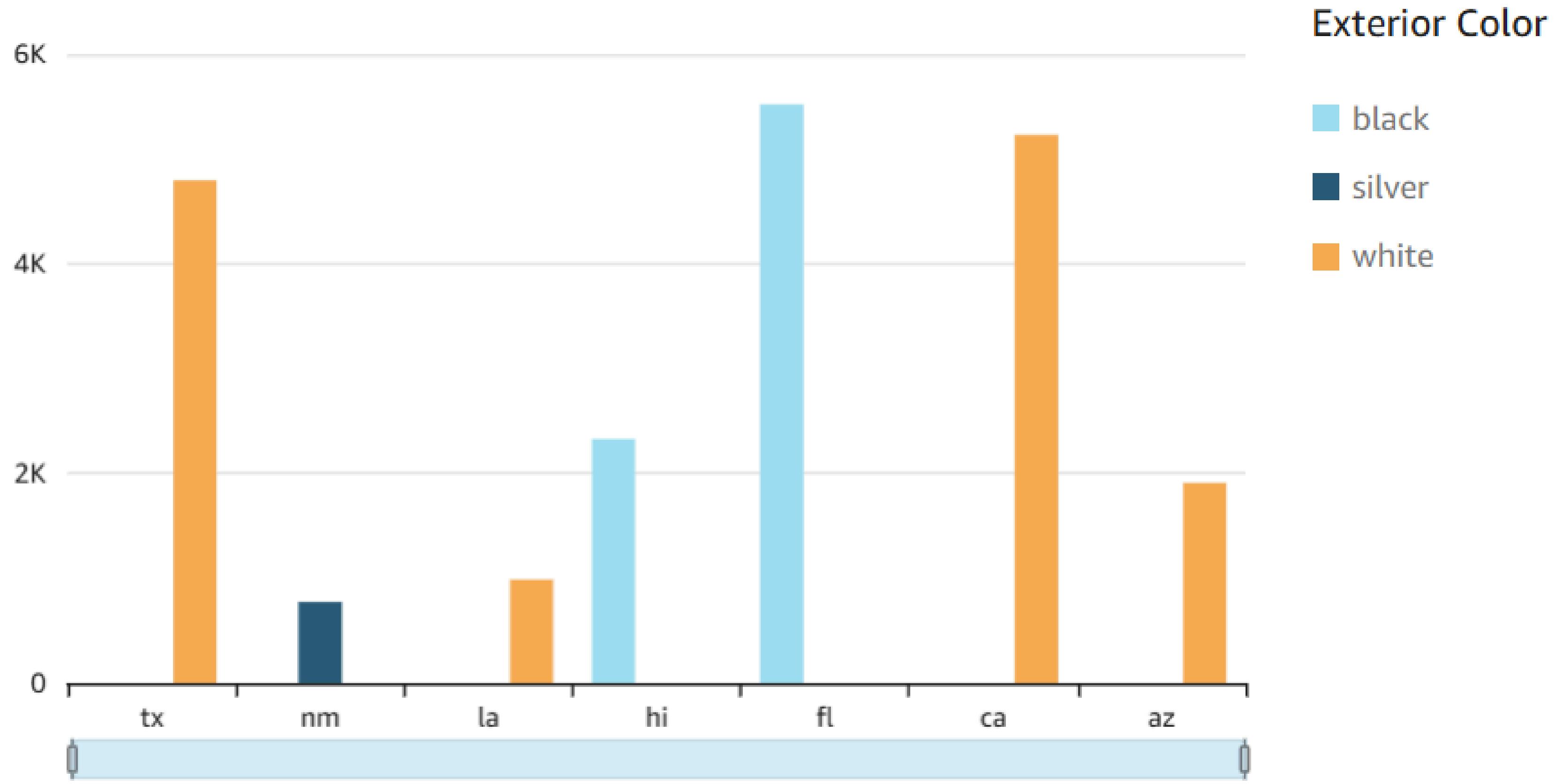
Sales Per Month



Total Sales per Avg Rainfall



Top Exterior Colors where State Avg Temp > 72



Challenges faced

- Ensuring integrity of schema in data
- Cost management
- Adding layers to Lambda Functions
- Managing IAM role permissions

Future Enhancements

- Be able to check for updated and deleted entries
- Using Redshift
- Using EMR



TECH
CONSULTING

Q&A

Thank you

