



# ANALYSIS OF ROAD ACCIDENTS IN CALIFORNIA



**GROUP - 8**

**Hamsalakshmi Ramachandran 017423666**

**Soumya Challuru Sreenivas 017518618**

**Sugandha Chauhan 017506190**

**Prathyusha Pingili 017545970**

# INTRODUCTION

- In California, traffic accidents are not merely statistics; they represent a pressing public safety issue that affects thousands of lives each year.
- As one of the most populous states in the U.S., California's extensive network of highways, interstates, and rural roads sees a diverse range of traffic incidents.
- The complexity of managing such a vast fluctuating, dynamic traffic patterns necessitate a detailed understanding of accident dynamics to develop effective mitigation strategies.

# MOTIVATION



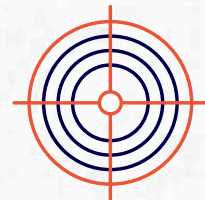
## **Raise awareness**

Helps raise awareness of safe driving practices by highlighting common causes and consequences of accidents



## **Identifying gaps**

Analyzing real-world accident scenarios identifies gaps in current safety measures, paving the way for innovative solutions to prevent accidents and reduce their impacts.



## **Significance for the technology world**

Analyzing accident scenarios can highlight technological gaps or opportunities for innovation in vehicle safety features, infrastructure, and traffic management systems.

# TECHNOLOGY STACK

---

DATABASE - POSTGRESQL, MONGODB

DATAWAREHOUSE - AMAZON REDSHIFT

ETL TOOLS - AWS GLUE

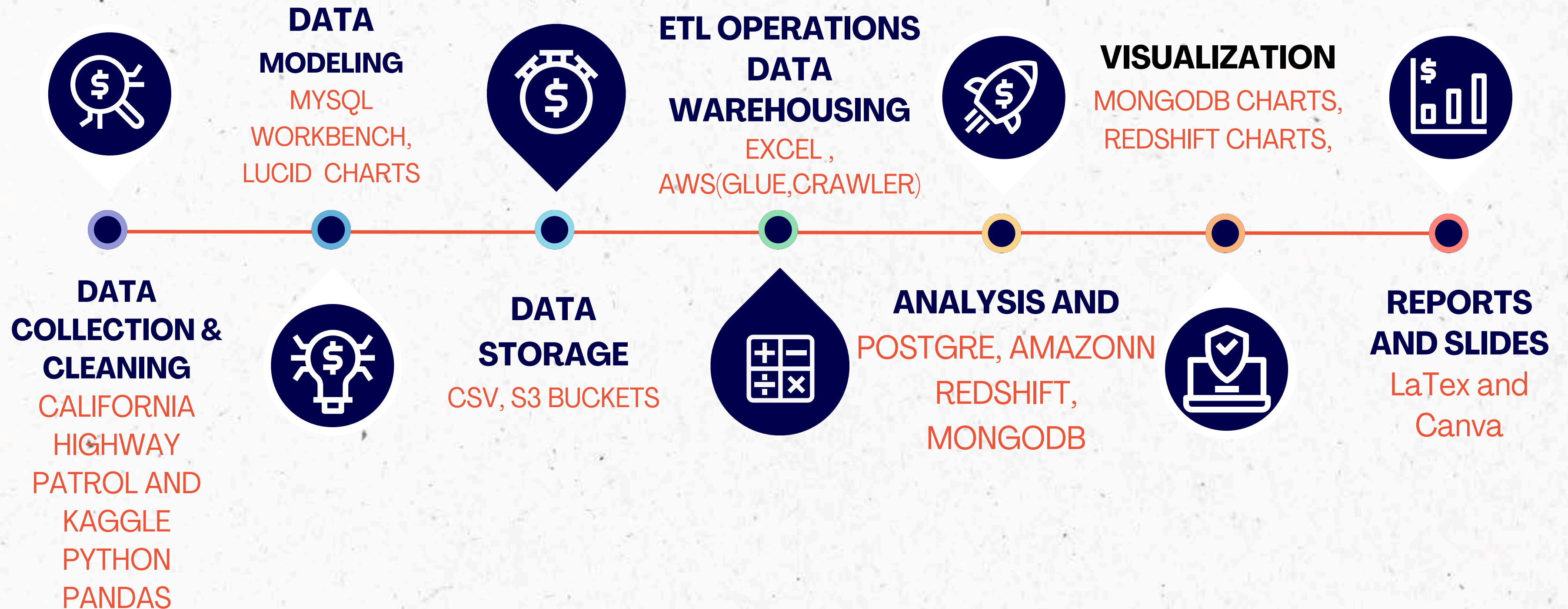
STORAGE - AWS S3 BUCKET

PROGRAMMING - PYTHON AND SQL

VISUALIZATION - REDSHIFT& MONGO CHARTS

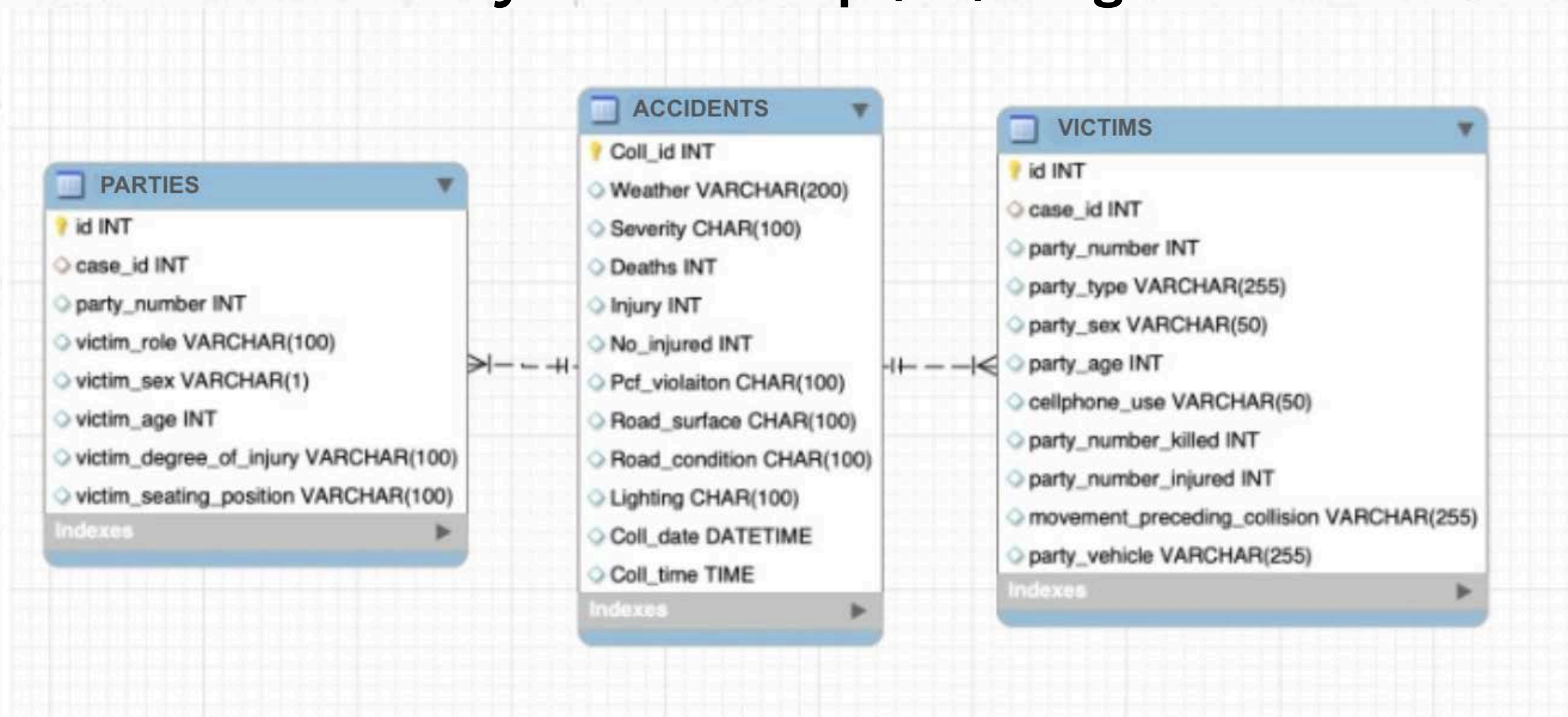


# DESIGN STEPS

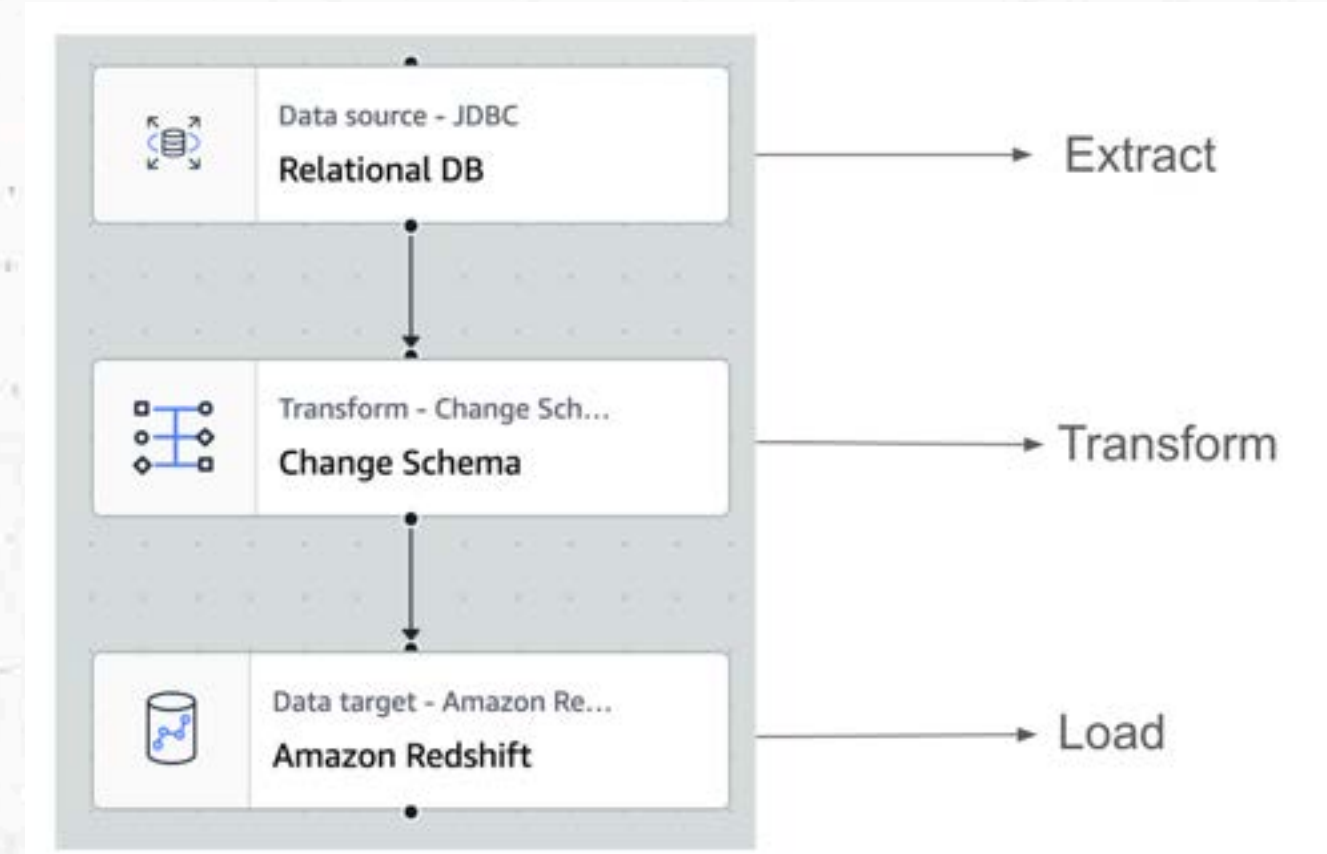
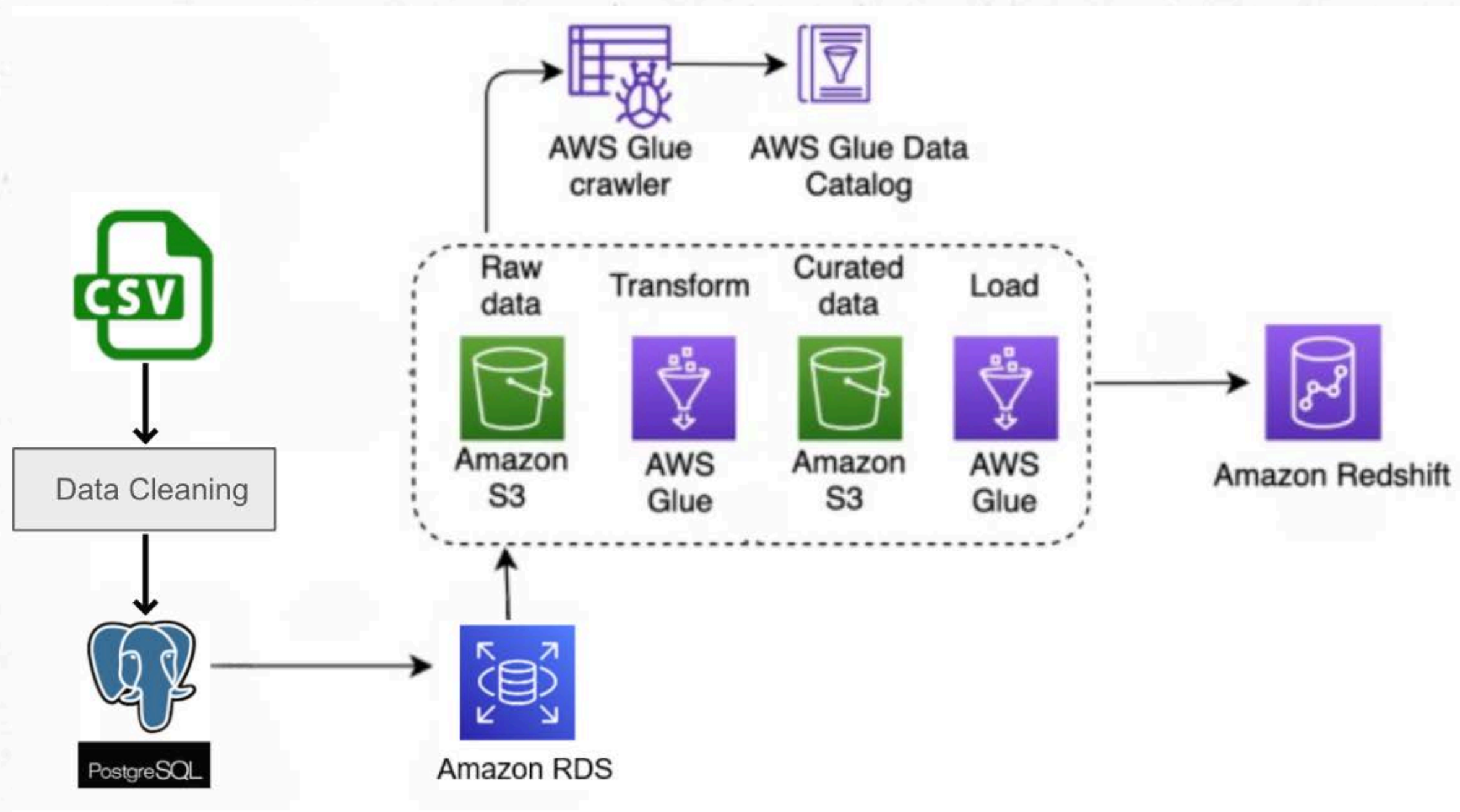


# DATA MODEL

## Entity Relationship (ER) Diagram



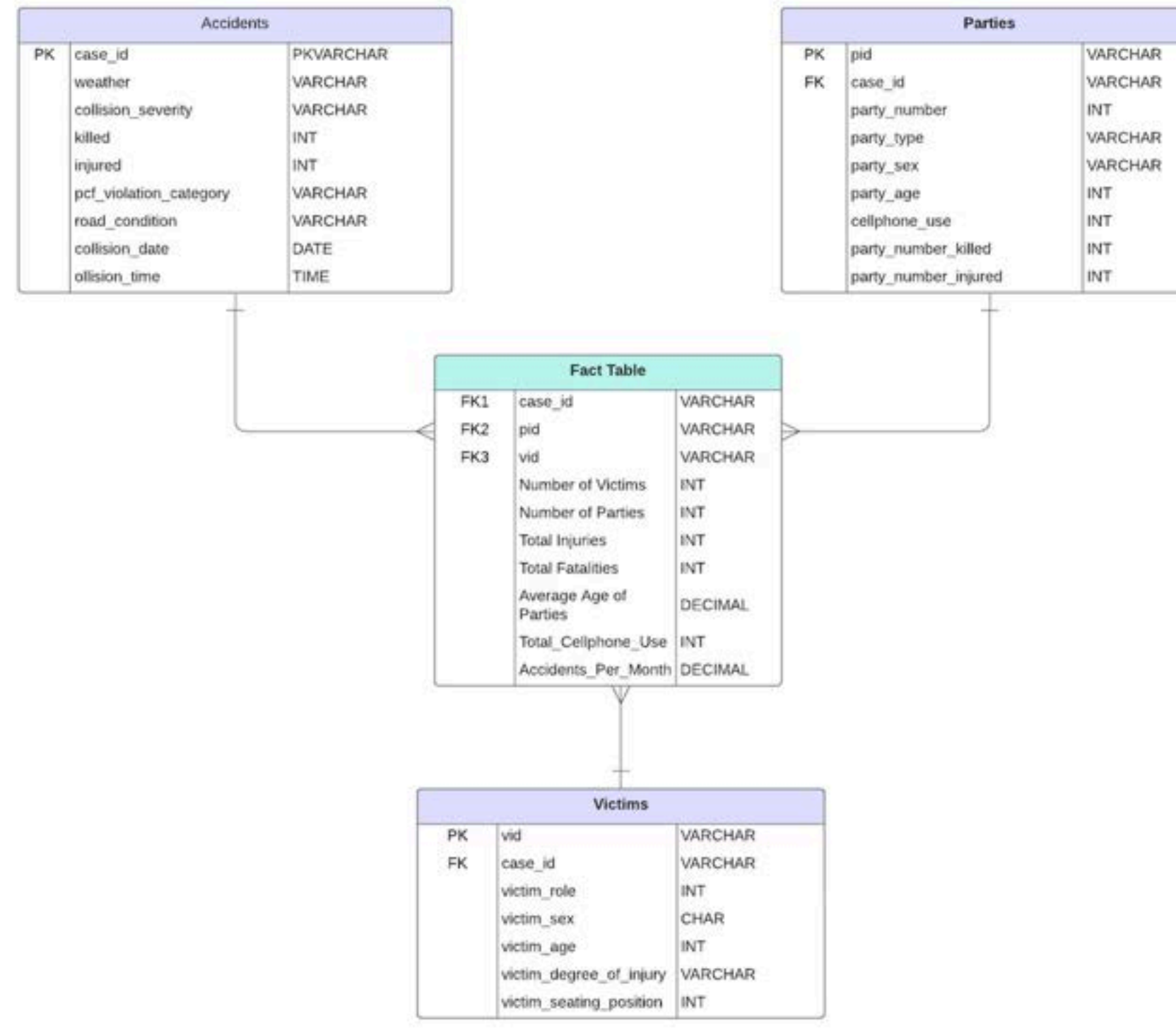
# WORKFLOW 1





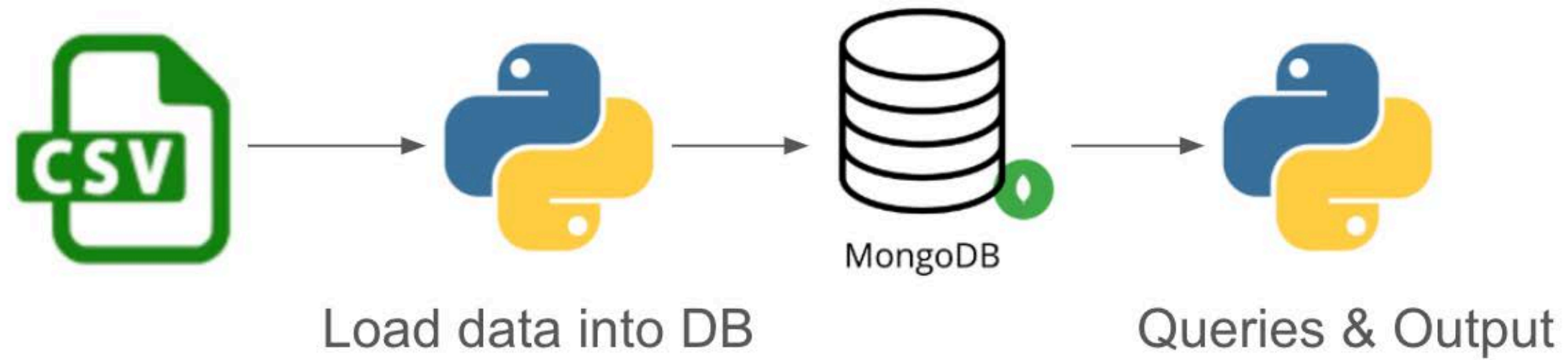
# DATA MODEL

## Star Schema





# WORKFLOW 2



# RESULTS FROM REDSHIFT

The screenshot displays the AWS Redshift query editor v2 interface. The left sidebar contains navigation options: Editor, Queries, Notebooks, Charts, History, and Scheduled queries. The main editor area shows a SQL query in a file named 'Untitled 2'. The query is as follows:

```
1 SELECT
2     EXTRACT(DOW FROM a.collision_date) AS day_of_week,
3     COUNT(*) AS accident_count
4 FROM
5     accidents AS a
6 GROUP BY
7     day_of_week
8 ORDER BY
9     accident_count DESC;
```

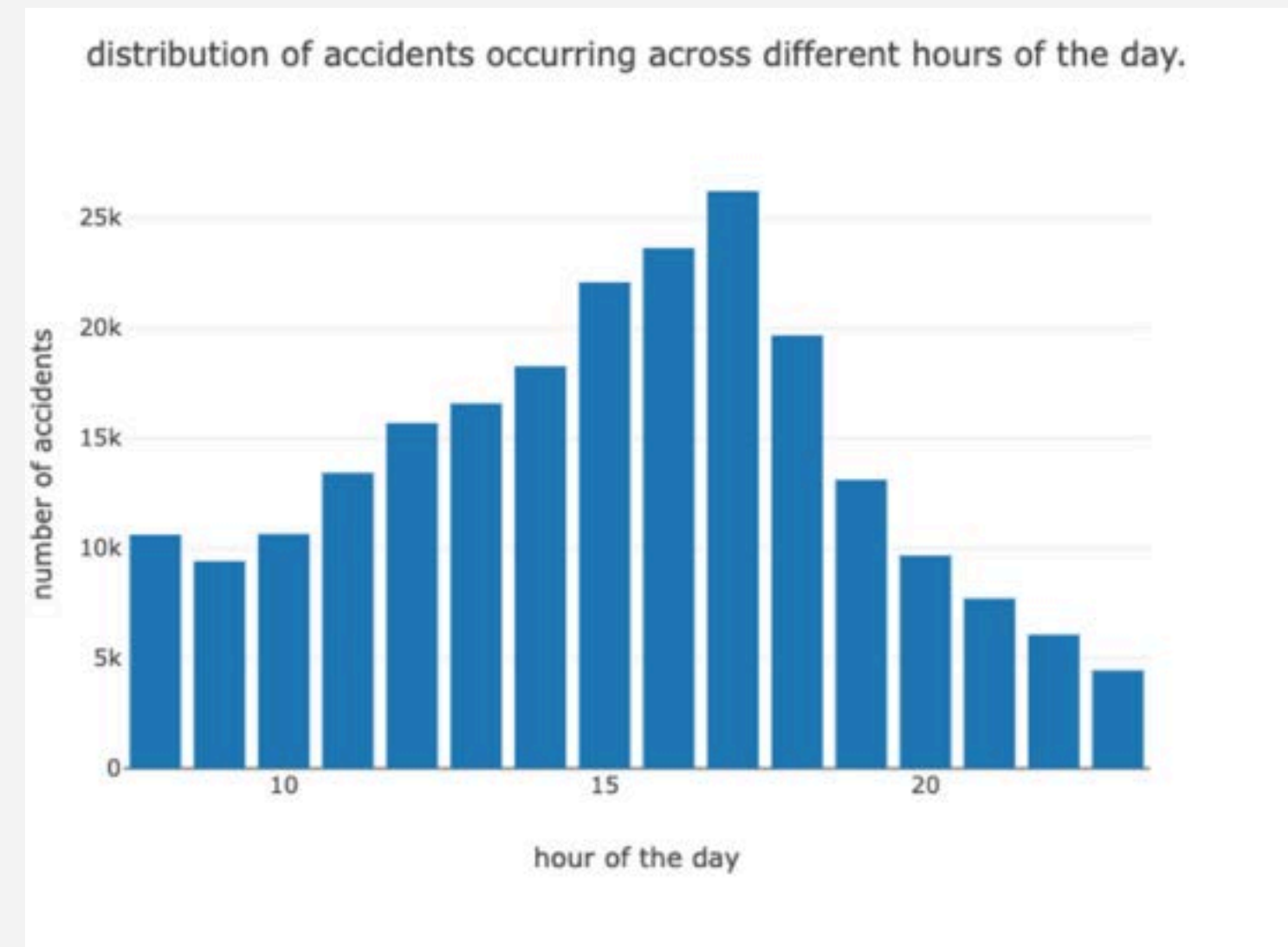
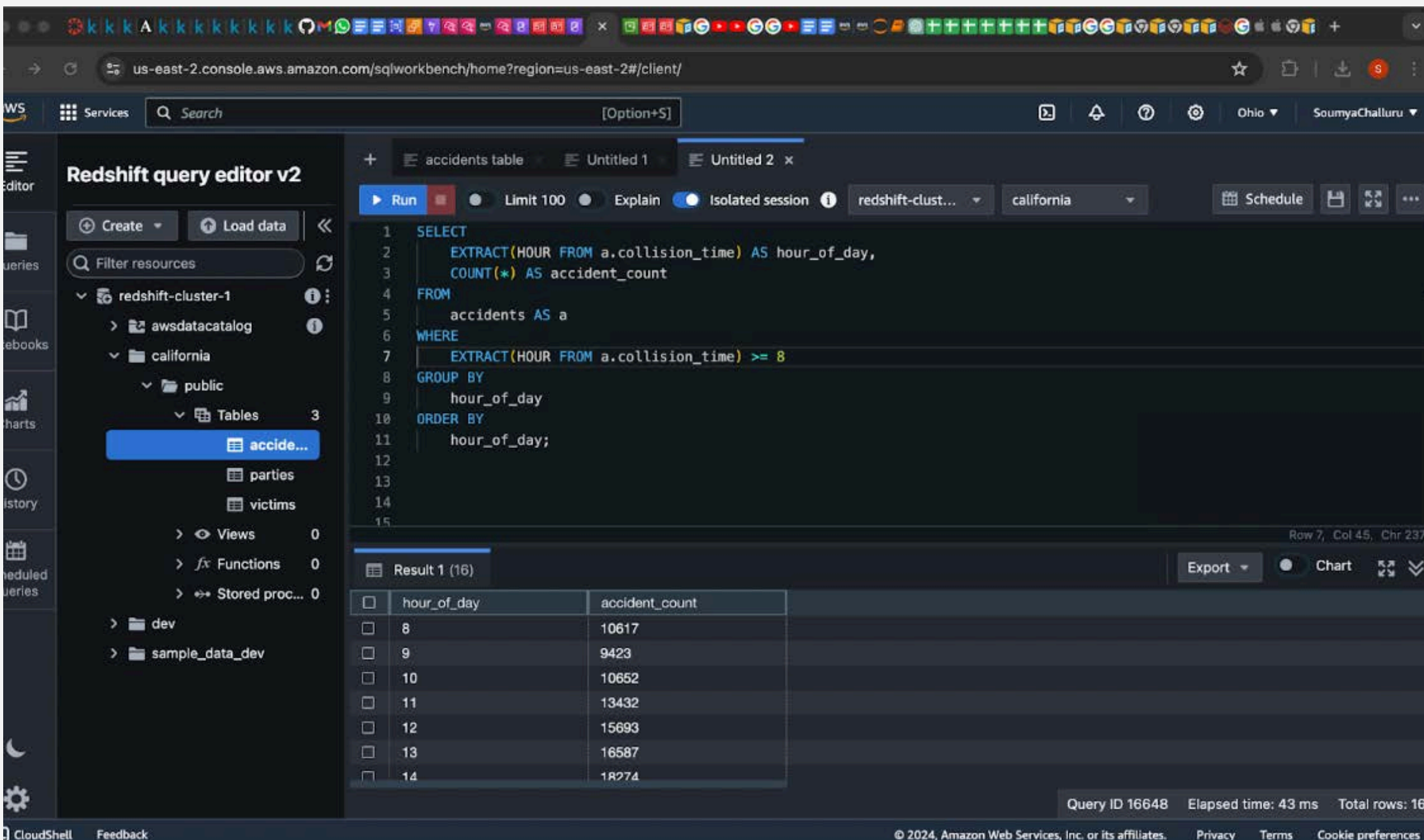
Below the query editor, the results are displayed in a table titled 'Result 1 (7)'. The table has two columns: 'day\_of\_week' and 'accident\_count'. The data is as follows:

day_of_week	accident_count
2	38807
5	37432
3	37410
4	37088
6	36999
0	36796
1	36140

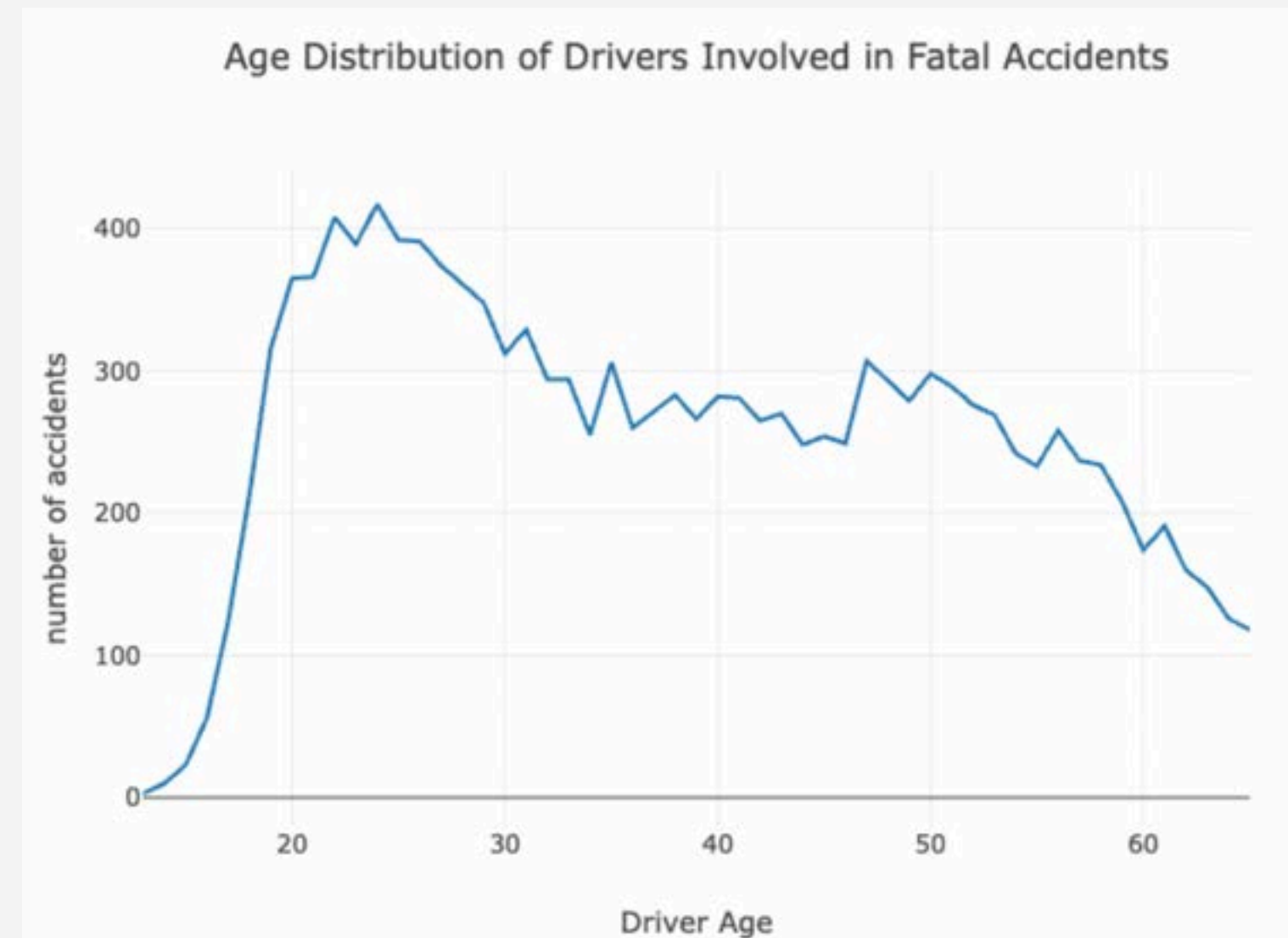
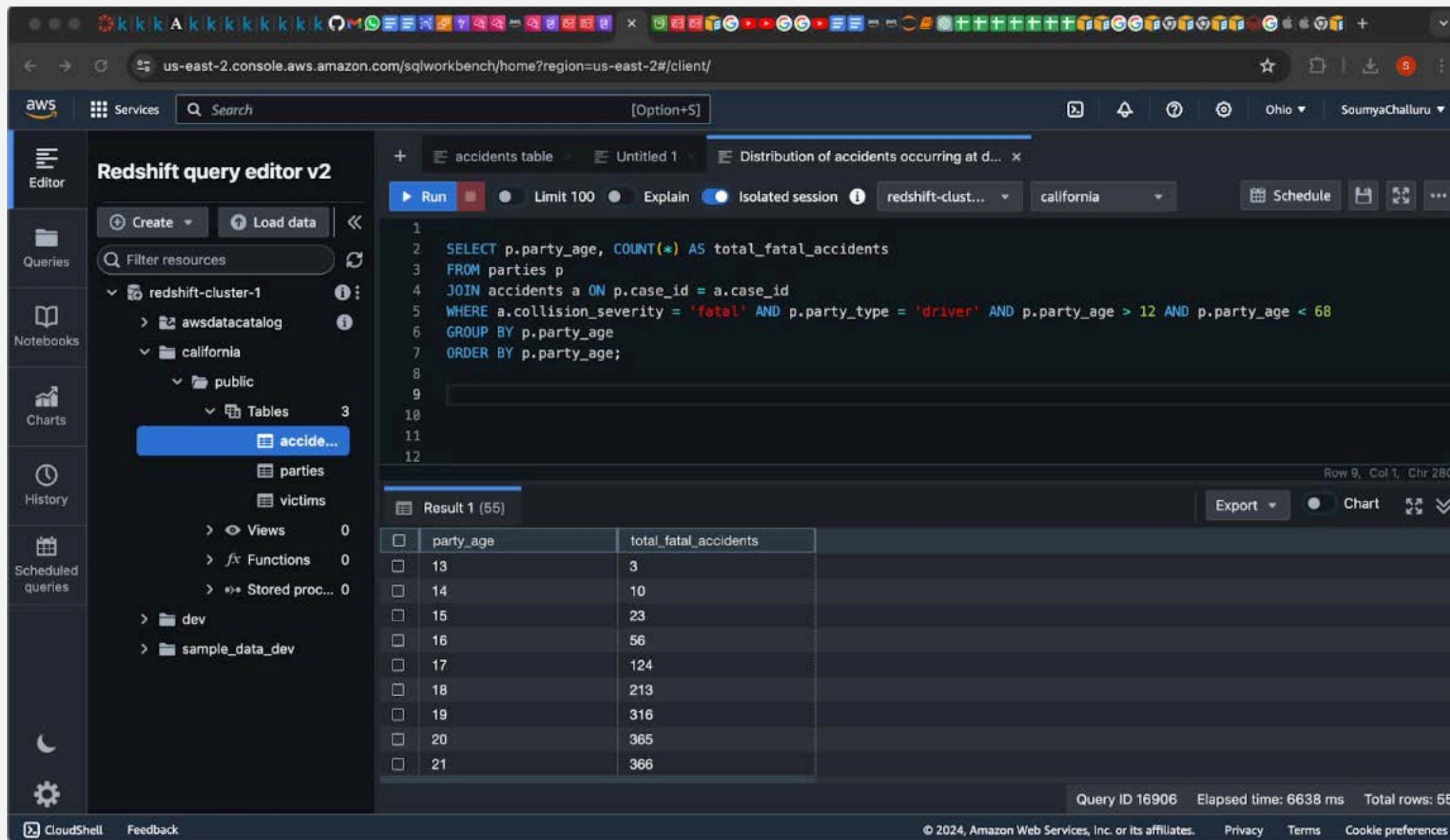
The bottom status bar indicates 'Query ID 15269', 'Elapsed time: 7 ms', and 'Total rows: 7'. The footer includes 'CloudShell', 'Feedback', and copyright information for Amazon Web Services, Inc. or its affiliates, along with links for Privacy, Terms, and Cookie preferences.



# RESULTS FROM REDSHIFT



# RESULTS FROM REDSHIFT





# RESULTS FROM REDSHIFT

us-east-2.console.aws.amazon.com/sqlworkbench/home?region=us-east-2#/client/

aws Services Search [Option+S]

Redshift query editor v2

Create Load data

Filter resources

redshift-cluster-1

- awsdatacatalog
- california
  - public
    - Tables 3
      - accide...
- parties
- victims
- Views 0
- Functions 0
- Stored proc... 0
- dev
- sample\_data\_dev

Run Limit 100 Explain Isolated session redshift-clust... california Schedule

```
1 SELECT movement_preceding_collision, COUNT(*) AS total_fatal_accidents
2 FROM parties p
3 JOIN accidents a ON p.case_id = a.case_id
4 WHERE a.collison_severity = 'fatal'
5 GROUP BY movement_preceding_collision
6 ORDER BY total_fatal_accidents DESC;
```

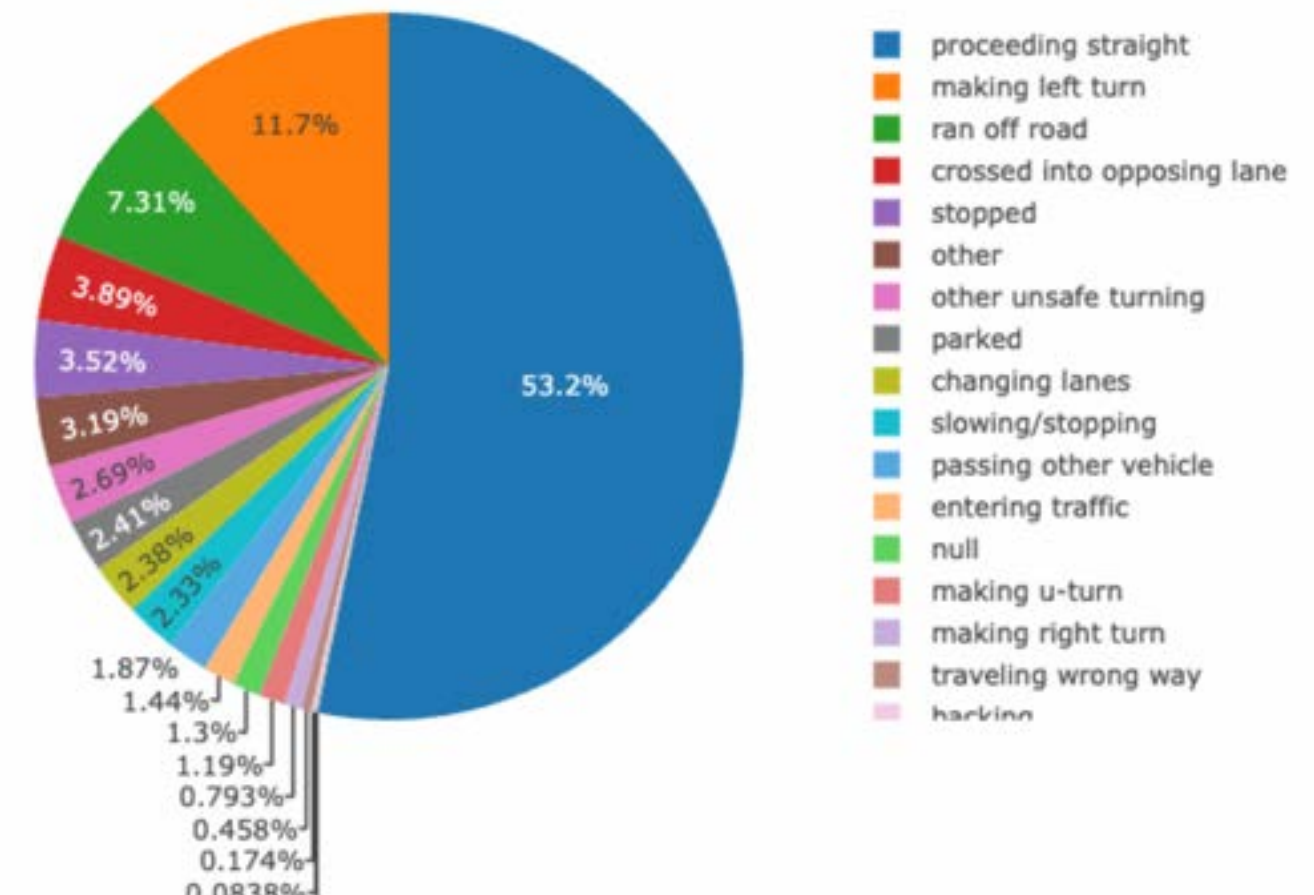
Result 1\* (19)

movement_preceding_...	total_fatal_accidents
proceeding straight	8260
making left turn	1815
ran off road	1134
crossed into opposing lane	604
stopped	546
other	495
other unsafe turning	418
parked	374
changing lanes	370
slowing/stopping	362
passing other vehicle	290
entering traffic	224

Query ID 17109 Elapsed time: 88 ms Total rows: 19

© 2024 Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Most Common Movement Preceding Collision in Fatal Accidents



# RESULTS FROM REDSHIFT

us-east-2.console.aws.amazon.com/sqlworkbench/home?region=us-east-2#/client/

aws Services Search [Option+S]

Redshift query editor v2

Create Load data

Filter resources

redshift-cluster-1

awsdatacatalog

california

public

Tables 3

accide...

parties

victims

Views 0

Functions 0

Stored proc... 0

dev

sample\_data\_dev

Run Limit 100 Explain Isolated session redshift-clust... california Schedule

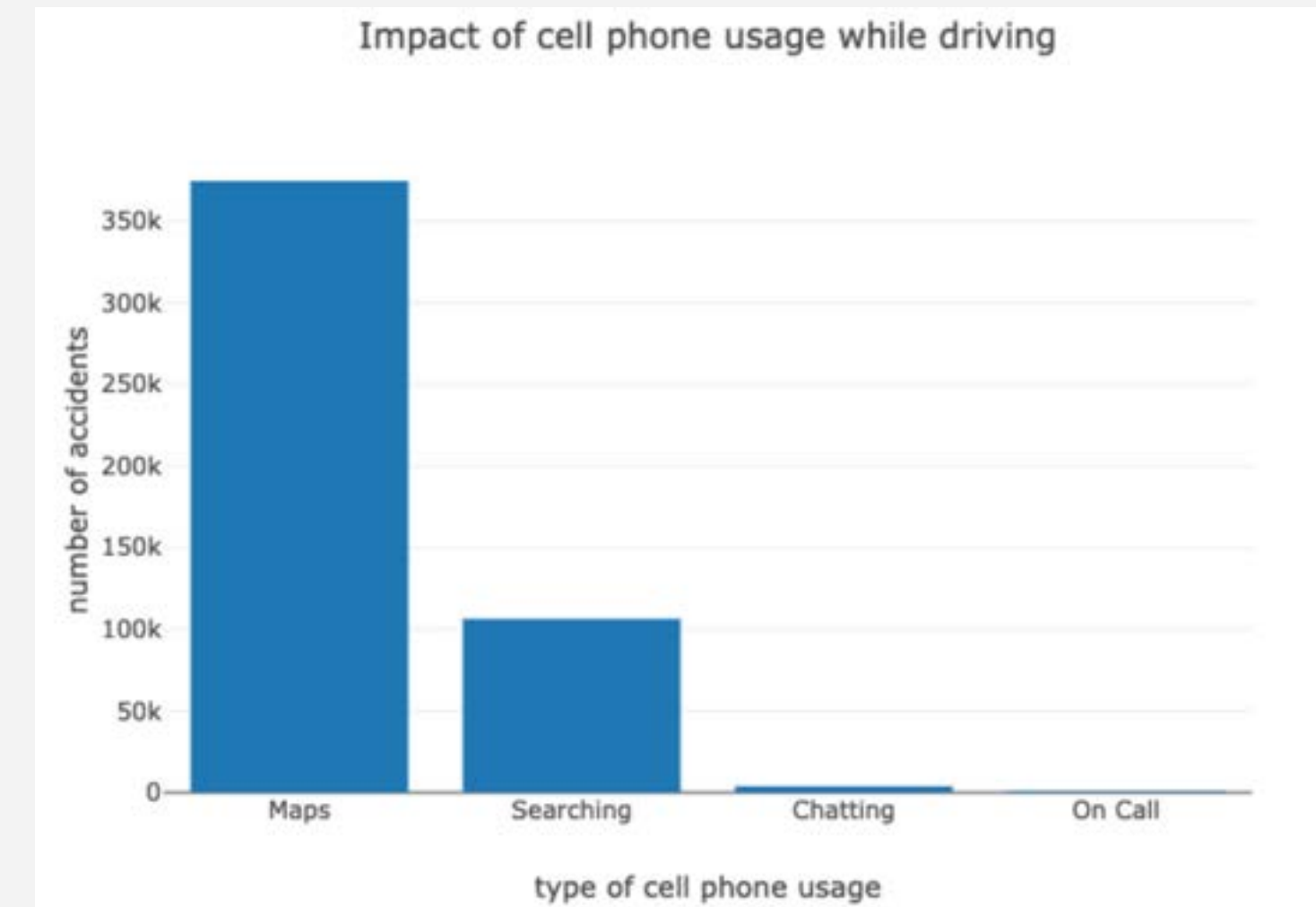
```
1 SELECT
2 CASE cellphone_use
3 WHEN 1 THEN 'On Call'
4 WHEN 2 THEN 'Chatting'
5 WHEN 3 THEN 'Maps'
6 WHEN 4 THEN 'Searching'
7 ELSE 'Other Use'
8 END AS cellphone_activity,
9 COUNT(*) AS total_accidents
10 FROM parties
11 GROUP BY cellphone_use
12 ORDER BY total_accidents DESC;
```

Result 1 (4)

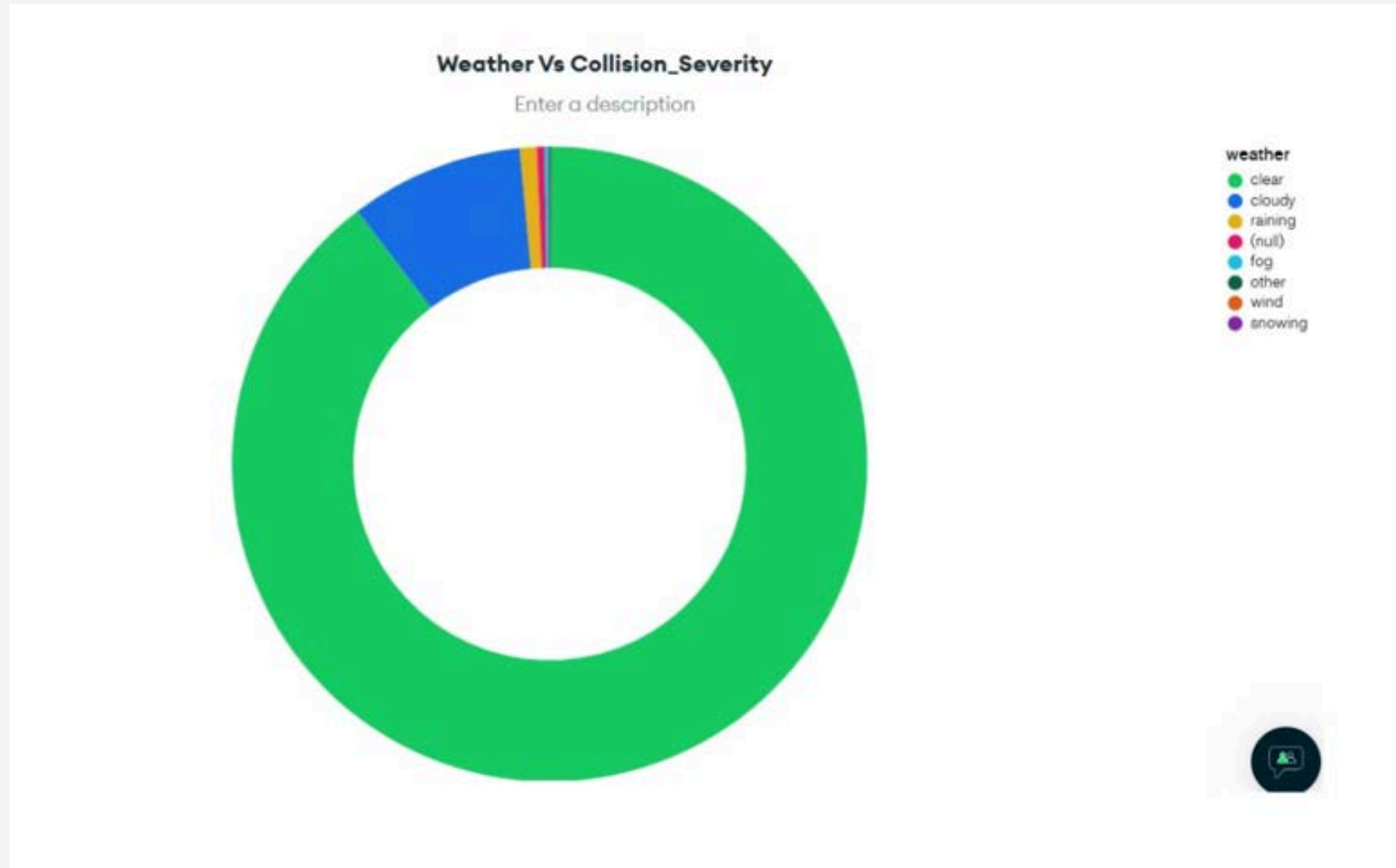
cellphone_activity	total_accidents
Maps	374445
Searching	106382
Chatting	3848
On Call	1010

Query ID 17686 Elapsed time: 5901 ms Total rows: 4

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

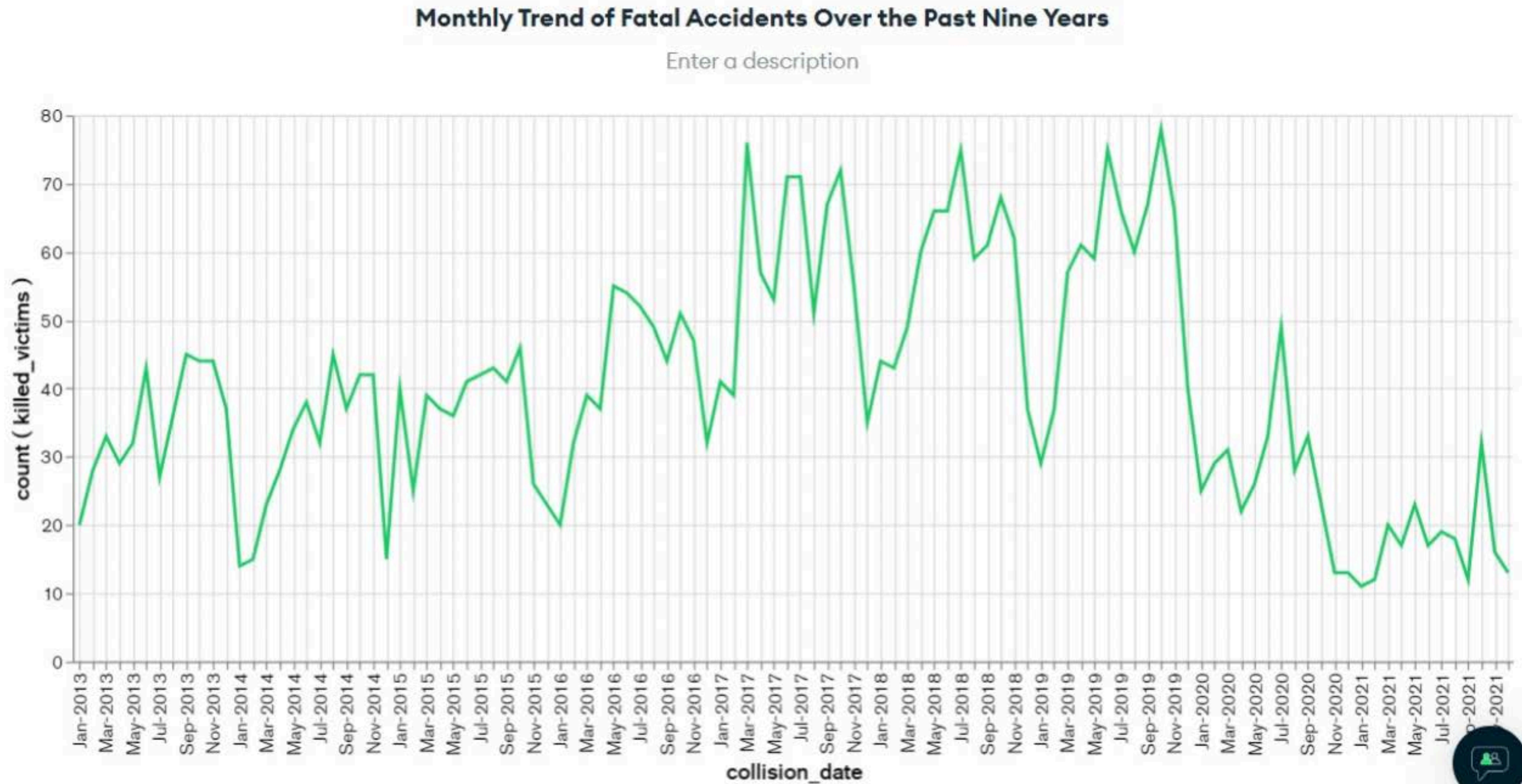


# RESULTS FROM MONGODB





# RESULTS FROM MONGODB





- We learned how to use AWS Solution Stack to build end-to-end Data Engineering/ Data Warehousing Solutions.
- We uploaded the data to the PostgreSQL database from csv files. Before uploading data, we made sure to clean the files according to the PostgreSQL datatype format.
- Data cleaning with an ETL tool like AWS Glue and Excel is much easier than python pandas.

# TECHNICAL DIFFICULTIES

- In Redshift, the default datatype for time was timestamp but the requirement was to have time without timezone. So we had to default setting of the Redshift datatype.
- While making connections from Glue to Redshift, we faced issues with endpoints, when making connections from S3 to AWS Glue to Redshift. We reconfigured the entire Glue pipeline from scratch.
- Learning to use AWS was time consuming as its terminologies were difficult to understand.

# CONCLUSION

In conclusion, the analysis of road accident data in California reveals challenges that require extensive solutions. The data emphasizes the urgency of addressing key risk factors such as distracted driving behavior. By making use of evidence-based strategies, and raising awareness for responsible driving, we can strive towards a future where road accidents are minimized, and the safety of all road users is safeguarded.



# THANK YOU!

GOT QUESTIONS?