

Used Redshift:

The screenshot shows the AWS Redshift Query Editor v2 interface. On the left, the sidebar includes 'Editor', 'Queries' (selected), 'Notebooks', 'Charts', 'History', and 'Scheduled queries'. The main area displays a query in the 'accidents table' tab:

```
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;
```

The results table shows the count of accidents for each day of the week:

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

At the bottom, it says 'Query ID 15269 Elapsed time: 7 ms Total rows: 7'.

The screenshot shows the AWS Redshift Query Editor v2 interface. On the left, the sidebar includes 'Editor', 'Queries' (selected), 'Notebooks', 'Charts', 'History', and 'Scheduled queries'. The main area displays a query in the 'accidents table' tab:

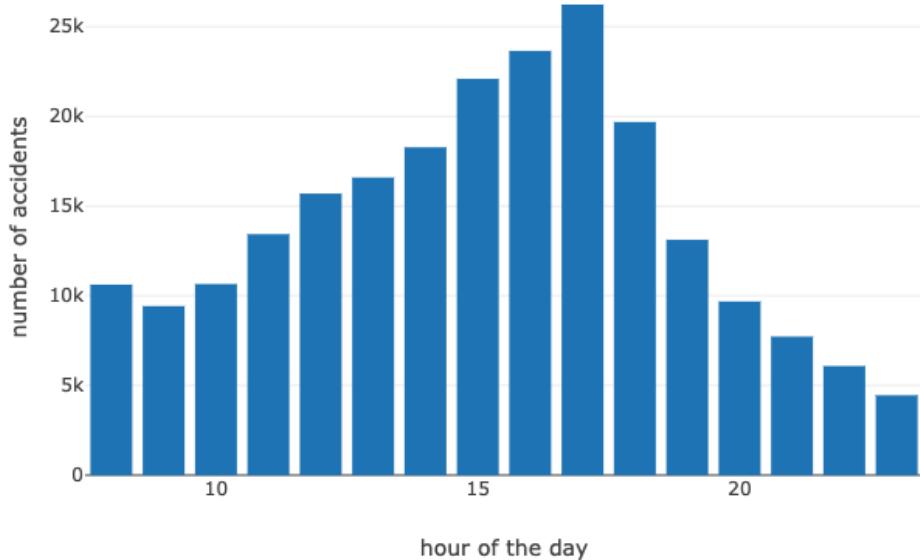
```
1 SELECT
2     EXTRACT(HOUR FROM a.collision_time) AS hour_of_day,
3     COUNT(*) AS accident_count
4 FROM
5     accidents AS a
6 WHERE
7     EXTRACT(HOUR FROM a.collision_time) >= 8
8 GROUP BY
9     hour_of_day
10 ORDER BY
11     hour_of_day;
```

The results table shows the count of accidents for each hour of the day from 8 to 14:

hour_of_day	accident_count
8	10617
9	9423
10	10652
11	13432
12	15693
13	16587
14	18274

At the bottom, it says 'Query ID 16648 Elapsed time: 43 ms Total rows: 16'.

distribution of accidents occurring across different hours of the day.



Screenshot of the AWS Redshift Query Editor v2. The left sidebar shows the schema structure under "redshift-cluster-1" and "california" database, including tables like "accidents", "parties", and "victims". The main area displays a SQL query and its results.

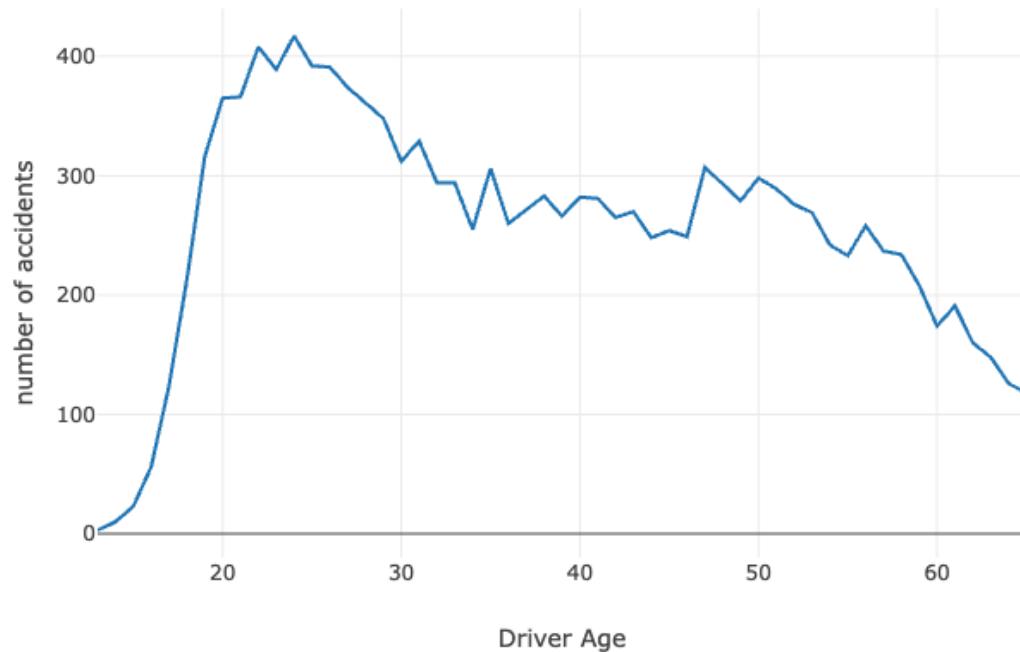
```
1 SELECT p.party_age, COUNT(*) AS total_fatal_accidents
2   FROM parties p
3   JOIN accidents a ON p.case_id = a.case_id
4   WHERE a.collision_severity = 'fatal' AND p.party_type = 'driver' AND p.party_age > 12 AND p.party_age < 68
5   GROUP BY p.party_age
6   ORDER BY p.party_age;
```

The results table shows the count of fatal accidents for each age group:

party_age	total_fatal_accidents
13	3
14	10
15	23
16	56
17	124
18	213
19	316
20	365
21	366

Query statistics at the bottom: Query ID 16906, Elapsed time: 6638 ms, Total rows: 55.

Age Distribution of Drivers Involved in Fatal Accidents



AWS Services Search [Option+S] Distribution of accidents occurring at d... □ Run □ Limit 100 Explain Isolated session redshift-clust... California Schedule ...

Redshift query editor v2

Queries Notebooks Charts History Scheduled queries

Editor

redshift-cluster-1 awsdatacatalog california public dev sample_data_dev

accide... parties victims Views Functions Stored proc... dev sample_data_dev

```

1
2 SELECT movement_preceding_collision, COUNT(*) AS total_fatal_accidents
3 FROM parties p
4 JOIN accidents a ON p.case_id = a.case_id
5 WHERE a.collision_severity = 'fatal'
6 GROUP BY movement_preceding_collision
7 ORDER BY total_fatal_accidents DESC;
8
9

```

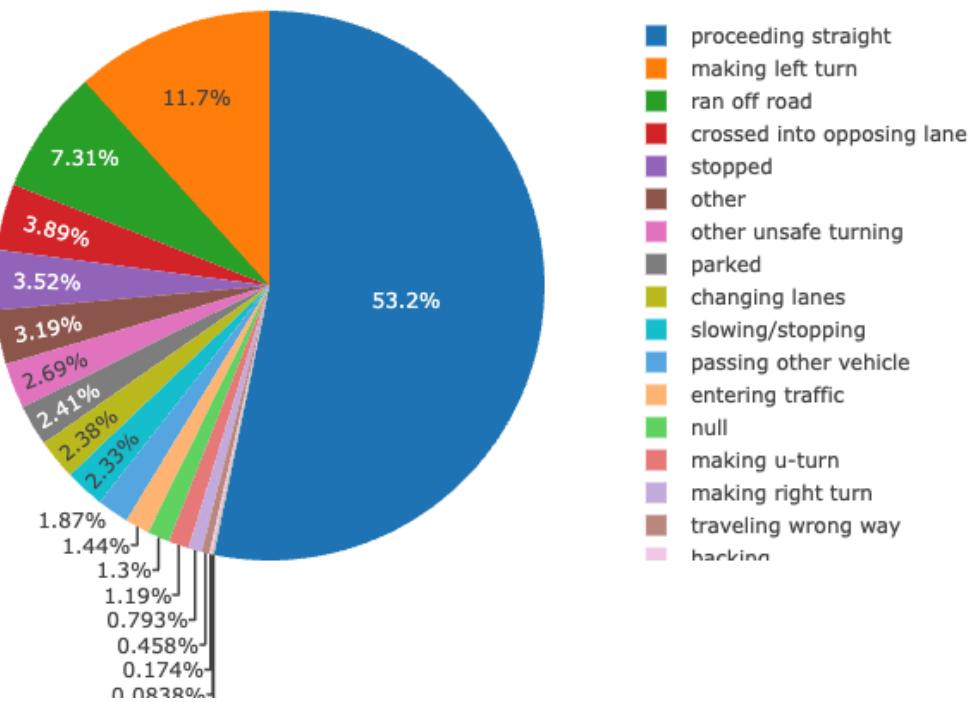
Result 1* (19)

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



Screenshot of the AWS Redshift Query Editor v2 interface showing a query results table.

Query:

```

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;

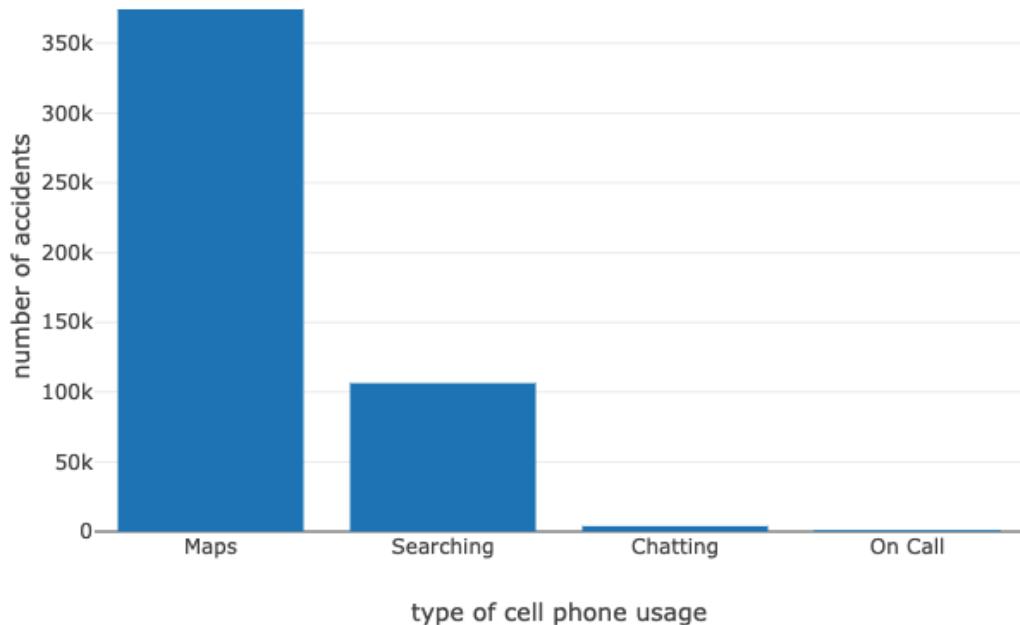
```

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

Impact of cell phone usage while driving



Used LaTeX:

The screenshot shows the Overleaf LaTeX editor interface. The left sidebar displays a file outline with sections like 'images', 'main.tex', 'File outline', 'ABSTRACT', and 'Subsection test'. The main code editor area contains the LaTeX source code for the document. The right side shows the compiled PDF output, which includes the title page with author details and a section titled '1 ABSTRACT'.

```
1 \documentclass[10pt,a4paper,onecolumn]{article}
2 \usepackage{graphicx, float}
3 \graphicspath{{images/}}% Required for inserting images
4 \usepackage[T1]{fontenc}
5 %usepackage{babel}
6 \usepackage{authblk}
7
8 \title{ANALYSIS OF ROAD ACCIDENTS IN CALIFORNIA}
9 \author{\texttt{(hamsalakshmi.ramachandran)}}
10 \affil{\texttt{[Master of Science in Data Analytics]}}
11 \affil{\texttt{[San Jose State University]}}
12 \affil{\texttt{[Spring 2024]}}
13 \affil{\texttt{[hamsalakshmi.ramachandran@sjtu.edu]}}
14
15 \author{\texttt{(Sugandha Chauhan)}}
16 \affil{\texttt{[Master of Science in Data Analytics]}}
17 \affil{\texttt{[San Jose State University]}}
18 \affil{\texttt{[Spring 2024]}}
19 \affil{\texttt{[sugandha.chauhan@sjtu.edu]}}
20
21 \author{\texttt{(Prathyusha Pingali)}}
22 \affil{\texttt{[Master of Science in Data Analytics]}}
23 \affil{\texttt{[San Jose State University]}}
24 \affil{\texttt{[Spring 2024]}}
25 \affil{\texttt{[prathyusha.pingali@sjtu.edu]}}
26
27 \author{\texttt{(Soumya Challuru Sreenivas)}}
28 \affil{\texttt{[Master of Science in Data Analytics]}}
29 \affil{\texttt{[San Jose State University]}}
30 \affil{\texttt{[Spring 2024]}}
31 \affil{\texttt{[soumya.challurusreenivas@sjtu.edu]}}
32
33
34 \begin{document}
```

1 ABSTRACT

Millions of people are injured and lose their lives in traffic accidents every year, raising numerous health and safety concerns worldwide. The frequency and severity of these incidents in California pose significant challenges to the government, transportation authorities, and politicians. Understanding the mechanics of accidents is critical for developing effective policies to reduce their impact and increase road safety. This research underscores the importance of analyzing road accidents in California, highlighting the need for comprehensive data gathering to discern patterns, risk factors, and feasible actions to minimize accident severity. The goals include identifying relevant variables, determining their impact on accident severity, and developing evidence-based policy and intervention strategies. This study investigates how weather, road infrastructure, driving behavior, and geographic location might be used to improve road safety in California. The goal is to prevent accidents while improving the quality of life for inhabitants and visitors, resulting in a safer environment.

1.1 Subsection test

This is a subsection

MYSQL Workbench:

Loading data in MYSQL work bench

```
import mysql.connector

mydb = mysql.connector.connect(
    host = 'localhost',
    user = 'root',
    passwd = '#Wisdom555',
    database = 'CALIFORNIA_ACCIDENTS',
)

print(mydb)

<mysql.connector.connection_cext.CMySQLConnection object at 0x1278eb5c0>

mycursor = mydb.cursor()

mycursor.execute("CREATE DATABASE CALIFORNIA_ACCIDENTS")
```

Creating and populating our database using CSV files

```
mycursor.execute("SHOW DATABASES")
for db in mycursor:
    print(db)

('Cal_accidents',)
('CALIFORNIA_ACCIDENTS',)
('information_schema',)
('mysql',)
('performance_schema',)
('sys',)
```

MongoDB

Loading of data:

The screenshot shows the MongoDB Compass interface with three collections listed:

- collisions**:
 - Storage size: 18.24 MB
 - Documents: 378 K
 - Avg. document size: 269.00 B
 - Indexes: 1
 - Total index size: 4.46 MB
- parties**:
 - Storage size: 23.63 MB
 - Documents: 490 K
 - Avg. document size: 270.00 B
 - Indexes: 1
 - Total index size: 5.89 MB
- victims**:
 - Storage size: 11.08 MB
 - Documents: 323 K
 - Avg. document size: 188.00 B
 - Indexes: 1
 - Total index size: 4.12 MB

The screenshot shows the MongoDB Compass interface focusing on the **collisions** collection.

Collection Details:

- Documents: 270.4K
- Aggregations
- Schema
- Indexes: 1
- Validation

Document Preview:

Type a query: { field: 'value' } or [Generate query](#) [+](#)

[Explain](#) [Reset](#) [Find](#) [Options](#)

[ADD DATA](#) [EXPORT DATA](#) [UPDATE](#) [DELETE](#)

1 - 20 of 270376

```
_id: ObjectId('662c3ab4f2139bbd24d8d532')
case_id: 351283
weather: "clear"
collision_severity: "other injury"
killed_victims: 0
injured_victims: 1
party_count: 2
psf_violation_category: "trafficsignalsandsigns"
road_surface: "dry"
road_condition: "normal"
lighting: "dark with street lights"
collision_date: "2017-01-01T00:00:00.000+00:00"
collision_time: "06:00:00"
```

```
_id: ObjectId('662c3ab4f2139bbd24d8d533')
case_id: 351526
weather: "clear"
collision_severity: "pain"
killed_victims: 0
injured_victims: 1
party_count: 2
psf_violation_category: "improperturning"
road_surface: "dry"
road_condition: "normal"
lighting: "dark with street lights"
collision_date: "2017-01-01T00:00:00.000+00:00"
collision_time: "07:00:00"
```

The screenshot shows the MongoDB Compass interface. The left sidebar displays the database structure with 'parties' selected under 'Databases'. The main area shows the 'parties' collection with 487.9K documents. A query builder is present, and the results pane displays two documents from the collection.

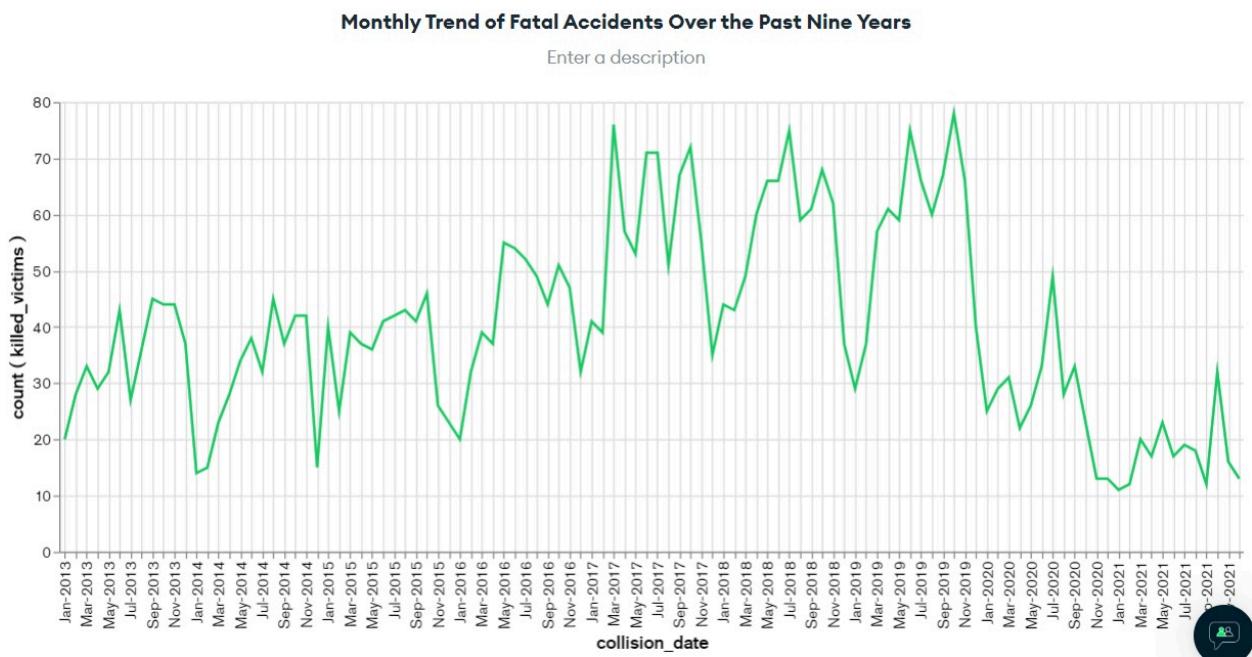
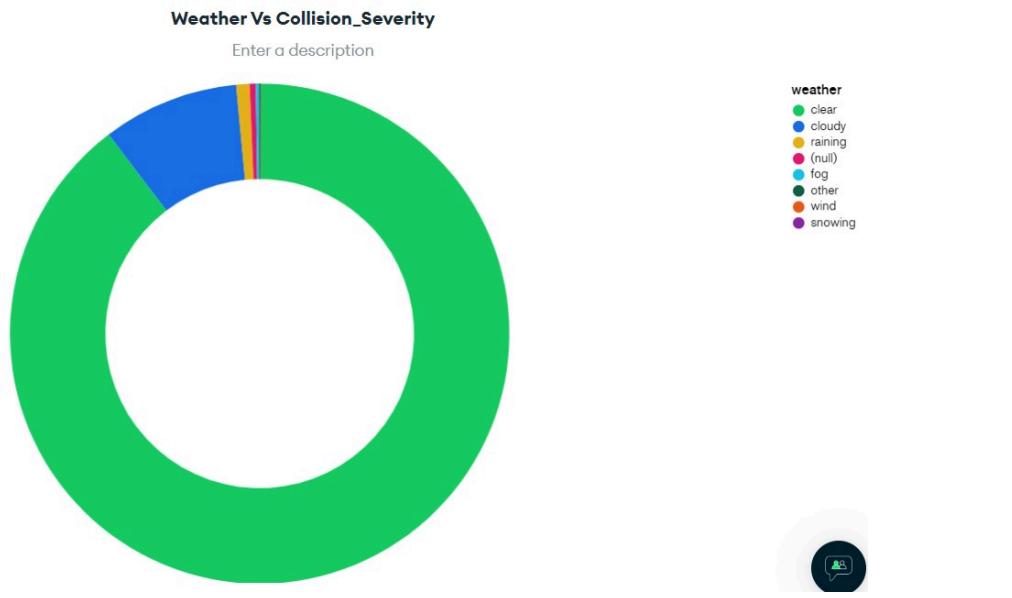
```
_id: ObjectId('662c3d6ef2139bbd24dcf55b')
id: 16
case_id: 3511283
party_number: 1
party_type: "driver"
party_sex: "female"
party_age: 43
cellphone_use: 3
party_number_killed: 0
party_number_injured: 1
movement_preceding_collision: "proceeding straight"
party_vehicle: "HYUNDAI"

_id: ObjectId('662c3d6ef2139bbd24dcf55c')
id: 17
case_id: 3511283
party_number: 2
party_type: "driver"
party_sex: "male"
party_age: 28
cellphone_use: 3
party_number_killed: 0
party_number_injured: 0
movement_preceding_collision: "proceeding straight"
party_vehicle: "INTERNATIONAL HARVESTER"
```

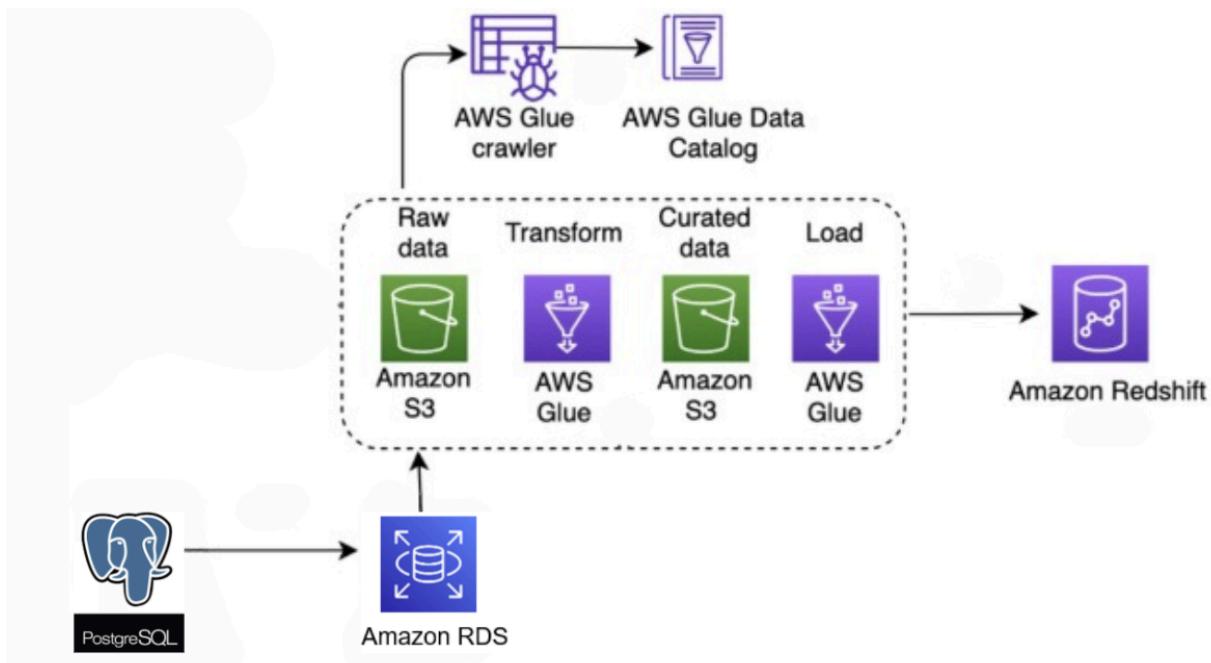
The screenshot shows the MongoDB Compass interface. The left sidebar lists databases and collections: cluster0.n0od6..., admin, config, dbproject (collisions, parties, victims), local, prathyu, sample_mflix. The 'victims' collection is selected in the dbproject database. The main pane displays the 'Documents' tab with 822.7K documents. A search bar at the top says 'Type a query: { field: 'value' } or Generate query'. Below it are buttons for Explain, Reset, Find, and Options. The document list shows three entries:

- `_id: ObjectId('662c410df2139bbd24e46ee2')`
id: 4
case_id: 3511283
party_number: 1
victim_role: 1
victim_sex: "female"
victim_age: 43
victim_degree_of_injury: "other visible injury"
victim_seating_position: 1
- `_id: ObjectId('662c410df2139bbd24e46ee3')`
id: 247
case_id: 3515256
party_number: 2
victim_role: 1
victim_sex: "male"
victim_age: 35
victim_degree_of_injury: "complaint of pain"
victim_seating_position: 1
- `_id: ObjectId('662c410df2139bbd24e46ee4')`
id: 265
case_id: 3515353
party_number: 2
victim_role: 1

Charts:



Workflow



Jira Kanban:

A screenshot of the Jira Software interface showing a Kanban board for the project "DB_project". The board is divided into three columns: "TO DO 1 OF 1", "IN PROGRESS 2 OF 2", and "DONE 5 OF 5".

- TO DO 1 OF 1:**
 - ETL Needs to be done for mongo db, need to use amazon glue for joining the tables
- IN PROGRESS 2 OF 2:**
 - PPT presentation (Status: DP-19)
 - Github (Status: DP-20)
- DONE 5 OF 5:**
 - Data modeling (ER diagram and schemas) (Status: DP-18)
 - Report writing on latex (Status: DP-17)
 - MongoDB (Status: NOSQL)
 - Amazon glue (Status: EXTRACT TRANSFORM & LOAD OPERATI...)

The left sidebar shows the project navigation and various management options like "Timeline", "Board", "Calendar", "Goals", "Issues", "Add view", "Code", "Project pages", "Add shortcut", and "Project settings".

GRAMMARLY

Untitled document

ABSTRACT: taken from the project proposal

Millions of people are injured and lose their lives in traffic accidents every year, which raises serious health and safety concerns worldwide. The frequency and severity of these incidents in California pose difficult problems for the public, law enforcement, transportation authorities, and politicians. Understanding the mechanics of accidents is critical for designing effective methods to reduce their impact and increase road safety. Our analysis underlines the importance of analyzing road accidents in California, highlighting the necessity for extensive data gathering to discover patterns, risk factors, and feasible actions to minimize accident severity. The goals include identifying relevant variables, determining their impact on accident severity, and developing evidence-based policy and intervention initiatives. This study investigates how weather, road infrastructure, driving behavior, and geographic location might be used to improve road safety in California. The goal is to prevent accidents while improving the quality of life for inhabitants and visitors, resulting in a safer environment.

INTRODUCTION:

In California, traffic accidents represent a significant and complex challenge in the state's transportation environment. California has an extensive network of highways, interstates, and country roads, and is home to a wide variety of accidents, from minor fender benders to catastrophic collisions. Factors such as population density, diverse

B I U | H1 H2 ⌂ | ≡ ≡ | ↴

16 readability score ▾

Hide assistant >>

88 Overall score See performance >

Goals Adjust goals >

Generative AI 🤖

Review suggestions 10

Correctness Clarity Engagement Delivery Style guide

effective

Use our best version It provides informed decision-making for...

Remove the comma year,

Remove the phrase types of

Add a hyphen speeding related

Delivery - Complete the sentence ...
...June alone, showed there were 2,061 deaths...

Accept Dismiss ...

Rephrase sentence Accidents in California present various challenges...

Change the wording a large number of

Change the wording records of traffic collisions

Choose a different word useful

Increase the impact of your text New!

Check for plagiarism

Get Expert Writing Help

Plagiarism