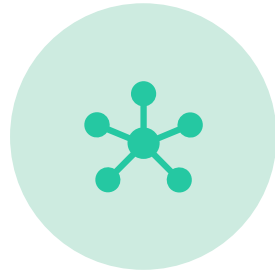


# BIG DATA-ENABLED DRIVER RISK ANALYTICS FOR FLEET SAFETY



MIS 6346.001 :  
BIG DATA



GROUP NO – 9



TEAM  
MEMBERS:

Shibani Kumar

Jeffrey Li

Akshil Makwana

Juhee Makwana

Jenil Mandavia

Aakanksha Shekhar Khare

# PROBLEM STATEMENT



Accidents caused by large trucks are a major concern in the transportation industry, often leading to severe injuries and fatalities.



The objective of this project is to identify and flag high-risk truck drivers by analyzing telematics data, including abnormal driving events and total mileage.



By calculating a Risk Factor for each driver using big data technologies, this project aims to assist fleet managers in improving road safety and minimizing insurance risks.



We are using Hadoop Ecosystems to analyze truck fleet data for Az National Trucking (ANT) to improve safety and compliance.

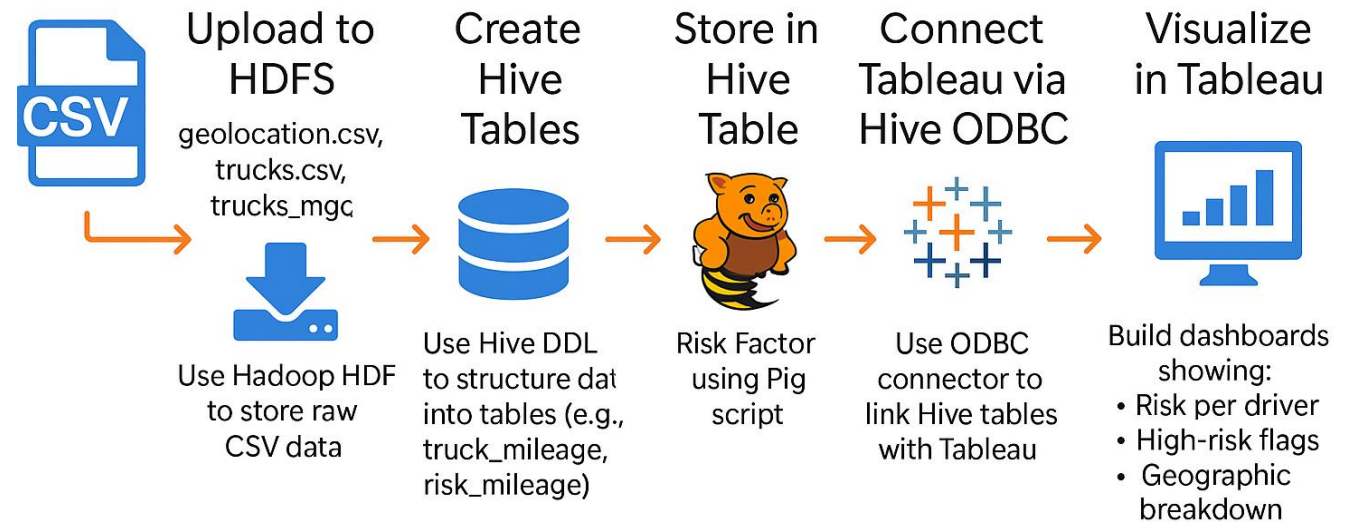
# DATA PROCESSING

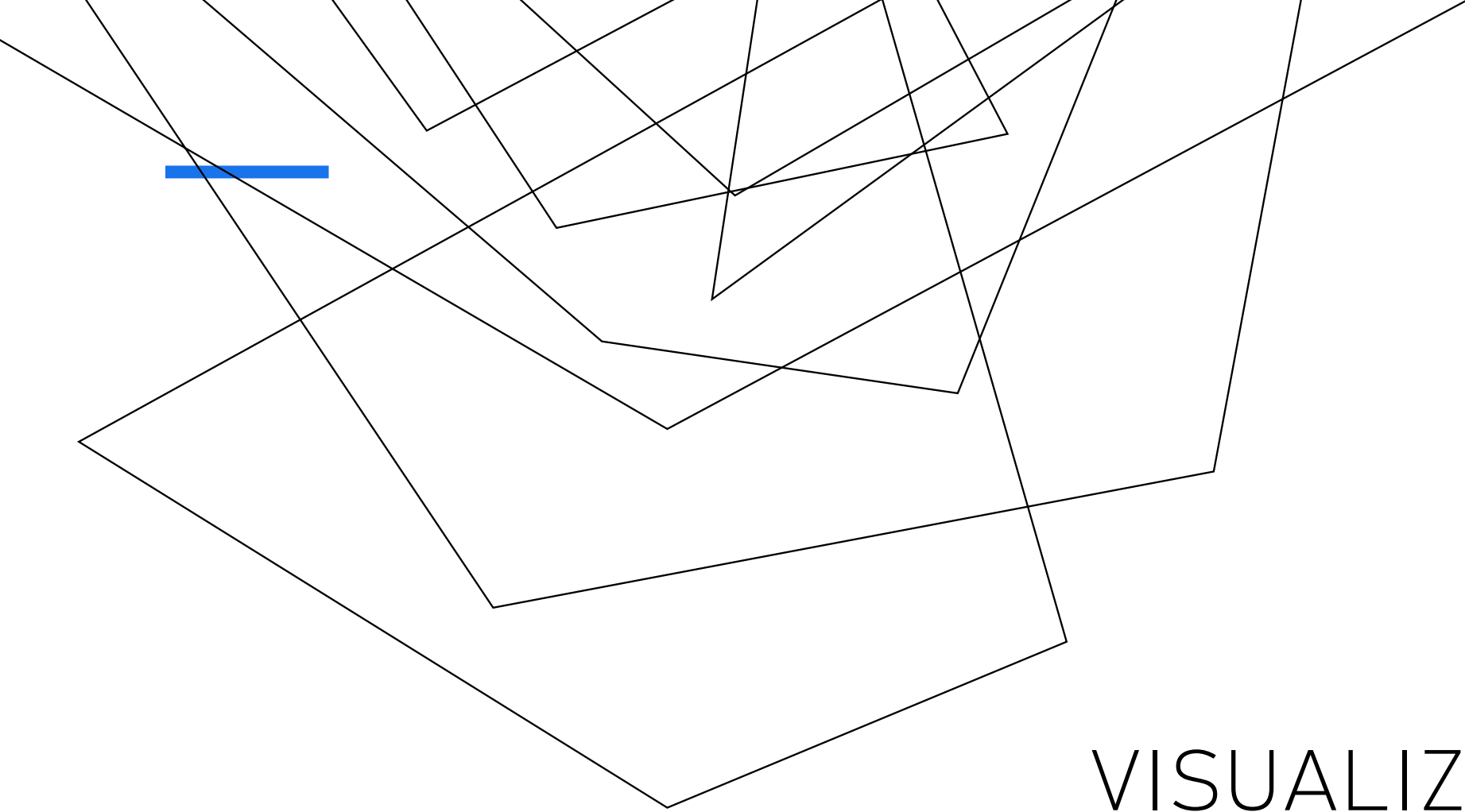
**Data Load:** CSV files (geolocation, trucks, trucks\_mg) are uploaded to Cloudera HDFS.

**Hive Integration:** Data is structured into Hive tables for querying through Hive DDL.

**Risk Calculation:** Apache Pig processes geolocation and driver mileage to compute a risk score based on event frequency and mileage.

**Visualization:** Final risk data is accessed via Hive tables and visualized in Tableau using ODBC connector.





VISUALIZATIONS

# VISUALIZATION 1

## Highest Risk Driver Identified

Driver A97 tops the list with a risk factor of 31.69, significantly higher than other drivers.

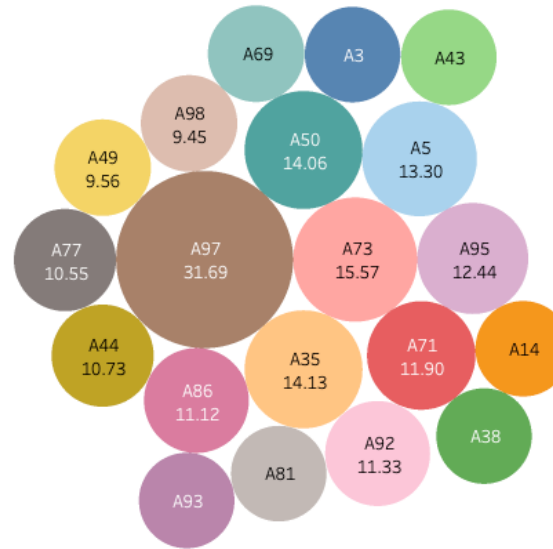
## Event Hotspot Locations

Risk events for driver A97 are concentrated in California, more specifically San Diego and around the Bay Area.

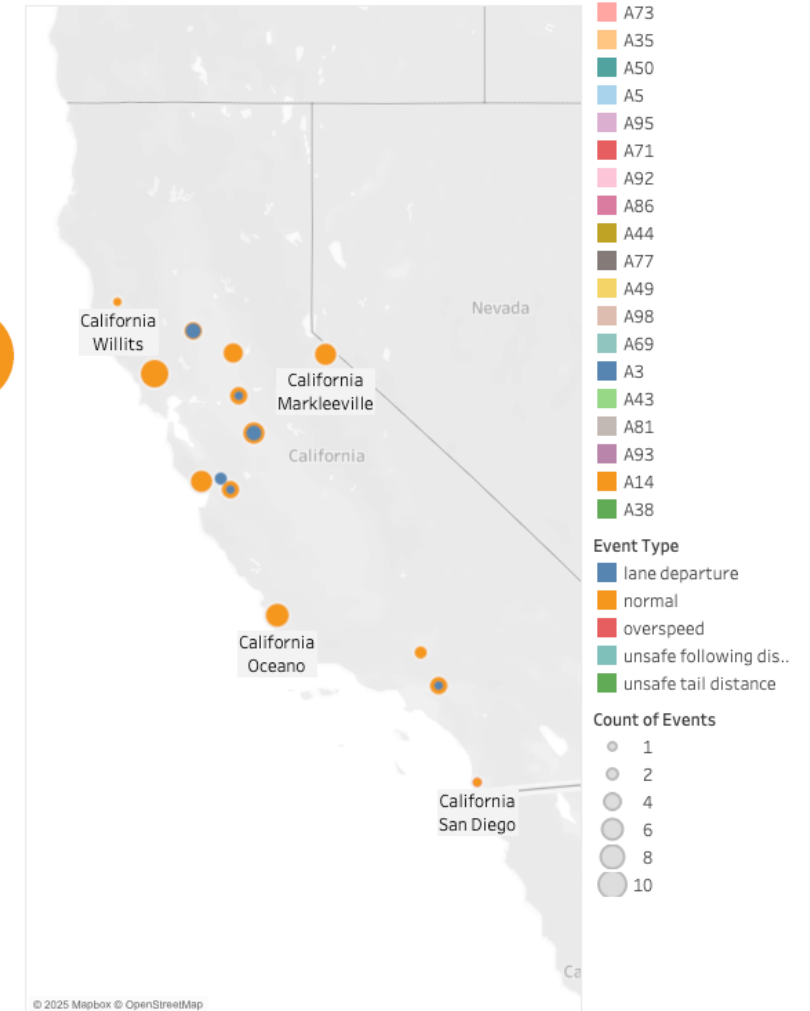
## Types of Risk Events

Most frequent events: Lane Departure.

Top 20 Riskiest Drivers



Map – Event Distribution for Driver A97



## VISUALIZATION 2

### Highest Risk Truck Model Identified

Oshkosh tops the list with a risk factor of 10.085, followed by Crane and Hino.

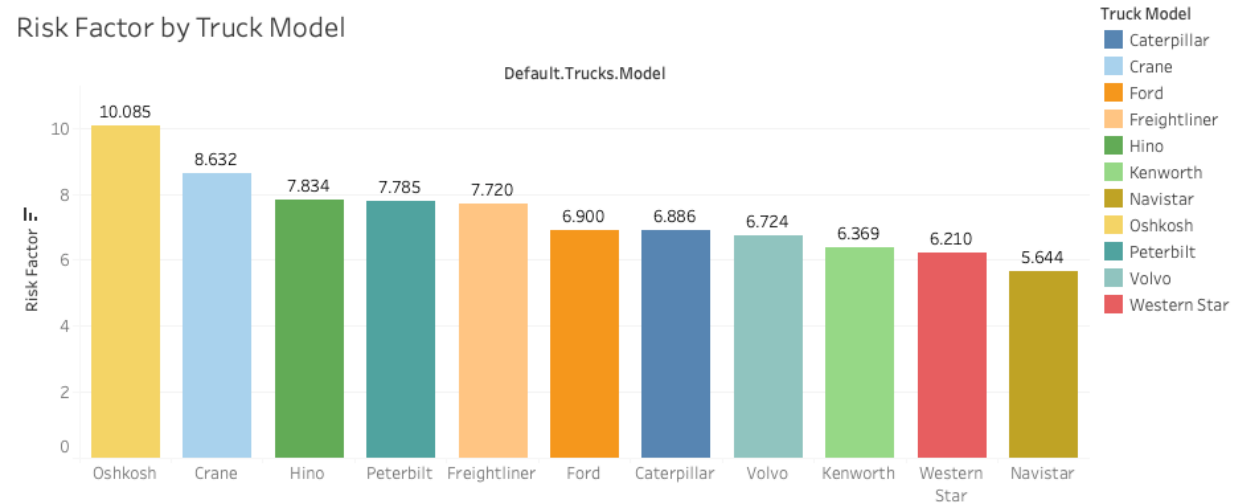
### Most Miles Driven

Ford trucks have the highest mileage at nearly 13 million miles, followed by Caterpillar and Peterbilt.

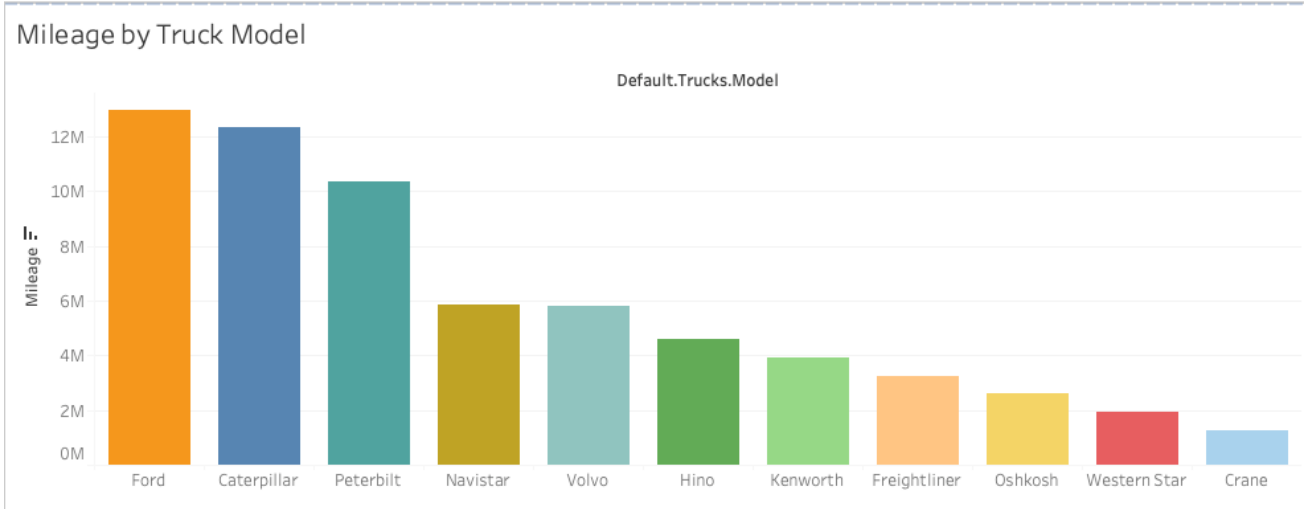
### Most and Least Balanced Truck Model

Navistar has the fourth highest mileage and lowest risk factor out of all truck models, while Crane has the lowest mileage and second highest risk factor.

Risk Factor by Truck Model



Mileage by Truck Model



# VISUALIZATION 3

## High-Risk Cities Identified

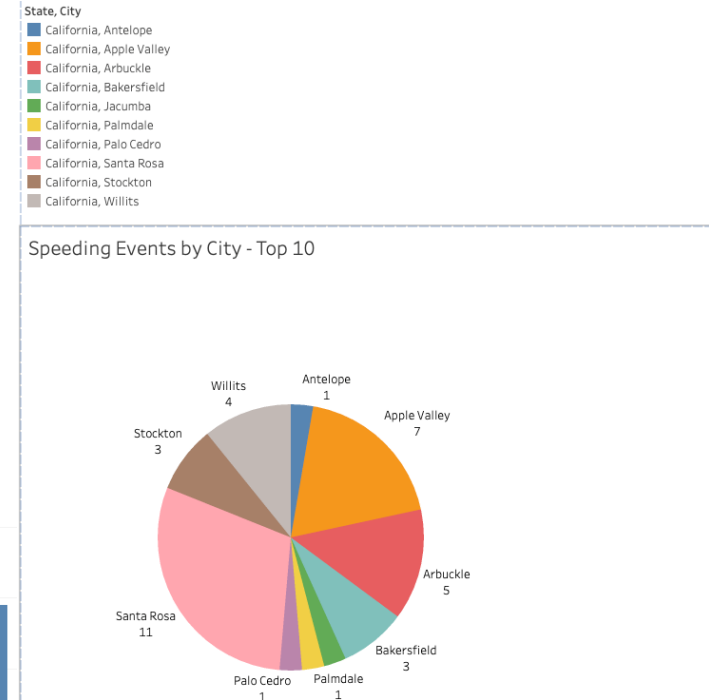
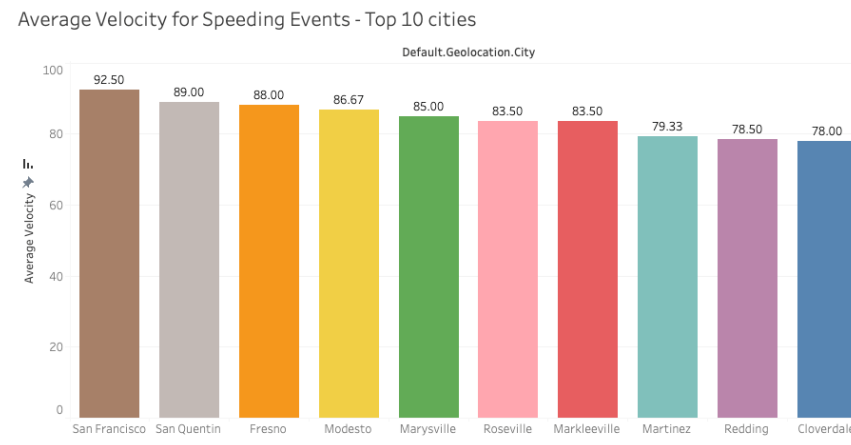
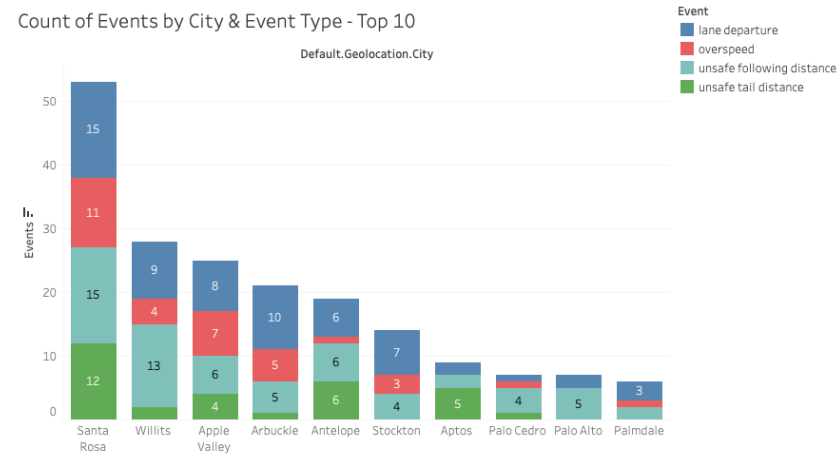
Santa Rosa, Willits, and Apple Valley show the highest number of risk events, especially lane departures and unsafe following distances.

## Top Contributors to Speeding

Santa Rosa (11 events), Apple Valley (7), and Arbuckle (5) are the top three cities with the most speeding incidents.

## Speeding Velocity Trends

Cities like San Francisco (92.5 mph) and San Quentin (89.0 mph) record the highest average speeds during speeding events.



# VISUALIZATION 4

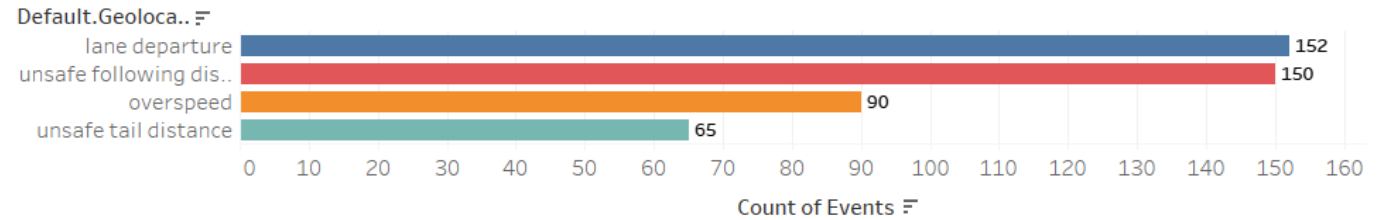
## Most Common Violation Event Type

Lane departures, at a total of 152, make the most common violation event type, followed by unsafe following distances at 150. These make 33.26% and 32.82% of violations, respectively.

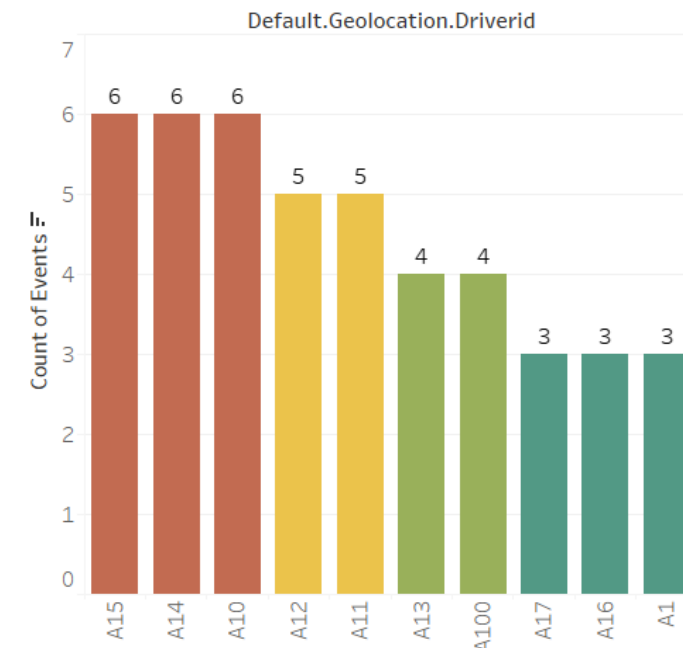
## Most Violation Events by Driver

Drivers A15, A14, and A10 were involved in the most violation events at six.

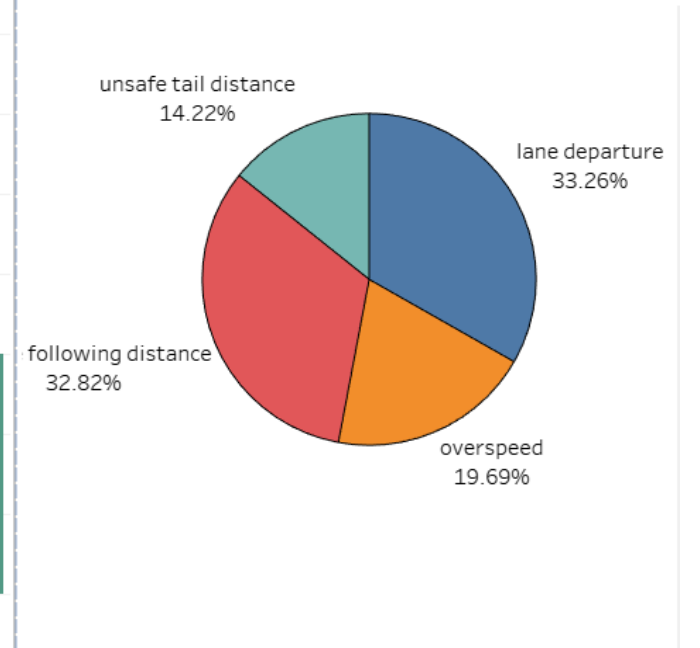
Violation Event Type Counts



Top 10 Drivers by Total Violation Events



Violation Type Distribution







# CONCLUSION

## 1. High-Risk Driver Intervention

Driver A97 shows the highest risk score (31.69) due to frequent lane departure events.

- Recommend driver-specific investigation and safety training.

## 2. Location-Based Risk Management

Event clusters observed in San Dimas and around the Bay Area.

- Conduct regional audits to address environmental or routing issues.

## 3. Fleet Optimization Strategy

Navistar demonstrates higher mileage and lower risk.

- Increase utilization of this model.

Crane exhibits low mileage with higher risk.

- Consider phasing out or re-evaluating deployment.

## 4. City-Level Speeding Control

Santa Rosa, Apple Valley, and Arbuckle contribute to 70%+ of speeding cases.

- Recommend installing speed limiters or real-time driver alerts in these zones.



THANK YOU!