

**BUAN 6346.002**  
**BIG DATA PROJECT**  
**Group 4**



# **Truck Driver Risk Analysis**

- Anirudh Madhavan
- Srinidhi Ls
- Vishal Kanna
- Pavan Katady
- Hyungi Kim
- Parveen Kumar
- Rithwik Reddy

# Contents

- Problem Statement
- Objectives
- Data Flow
- Tableau Table relationships
- Dashboard
- ML
- Conclusions

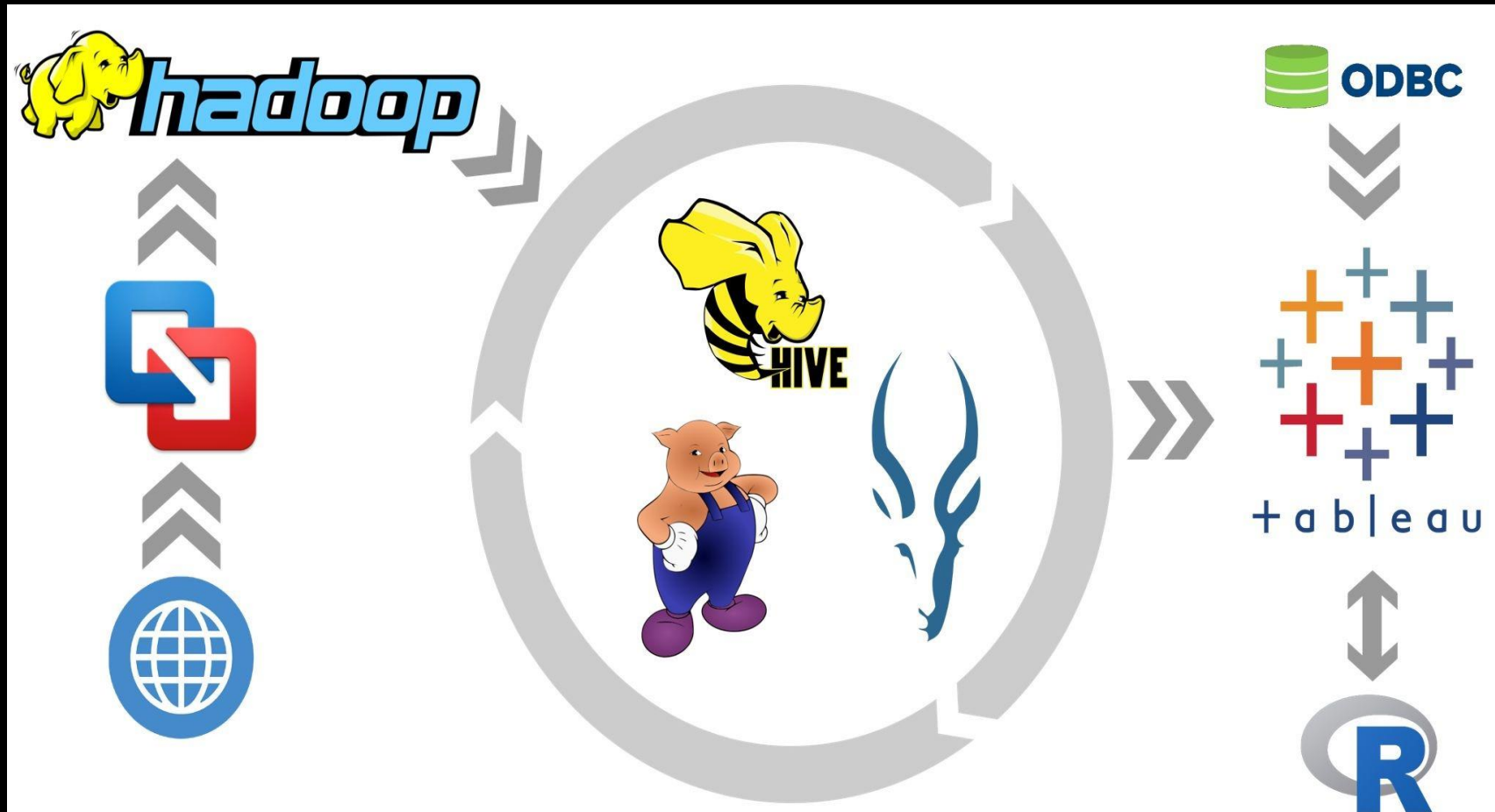
# Problem Statement

- Large truck accidents remain a major concern in the United States
- Injuries and fatalities caused by such accidents demand urgent attention
- Trucks equipped with devices to log location and event data
- Data streamed to a data center for processing and revising truck movements
- Goal: Utilize data-driven insights to improve safety and prevent accidents

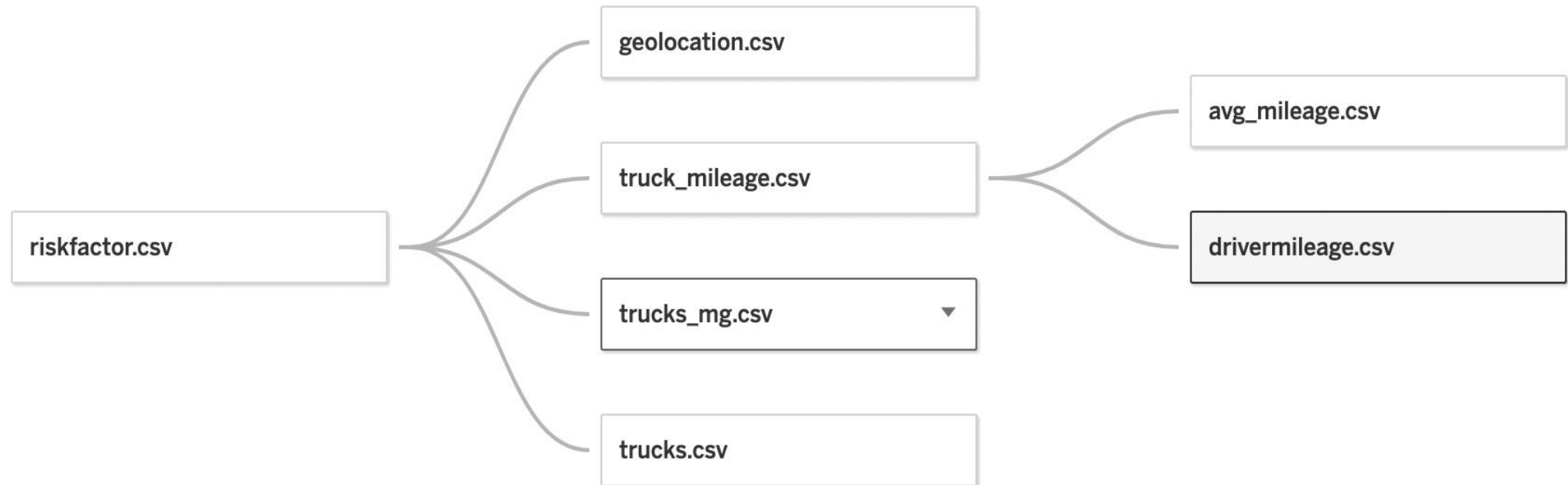
# Objectives

1. Identify the most hazardous commercial truck drivers
2. Calculate and predict risk factors for driver identification based on:
  - a. Events
  - b. Total distance travelled
  - c. Average speeds
  - d. Mileage
3. Create a dashboard for in-depth analysis
4. Minimize risks associated with each driver by analyzing their data, contributing to reduced accidents involving large commercial trucks
5. Provide key insights and visual illustrations to support decision-making

# Data Flow Process

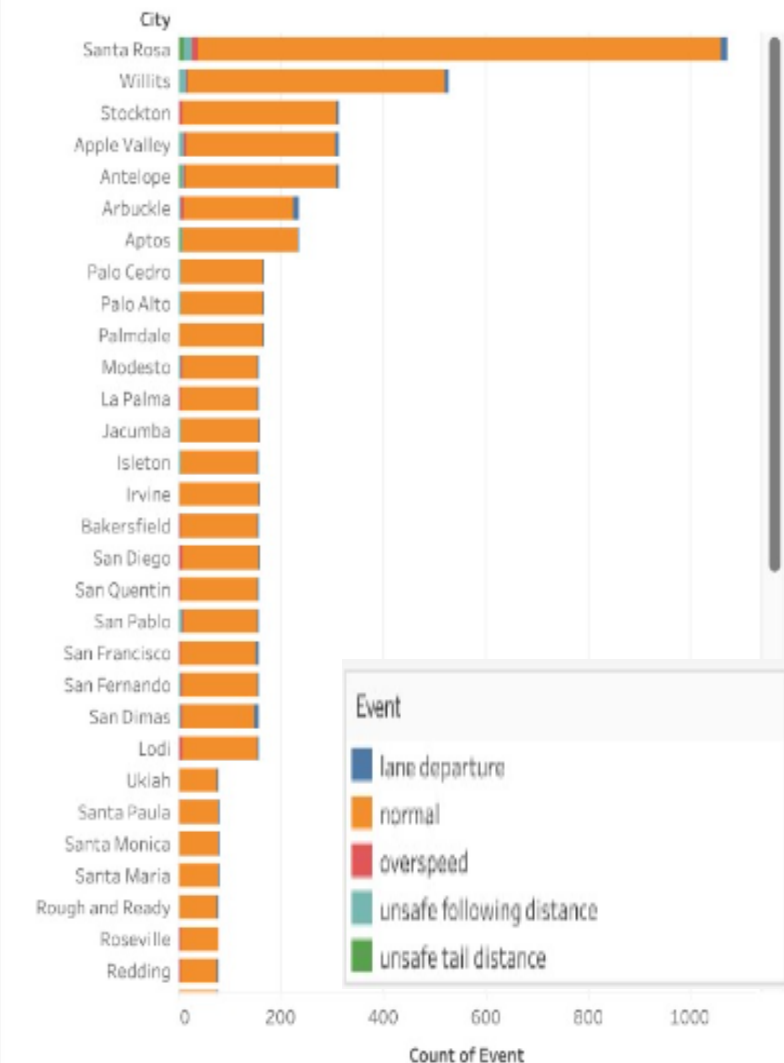


# TABLEAU TABLE RELATIONSHIPS

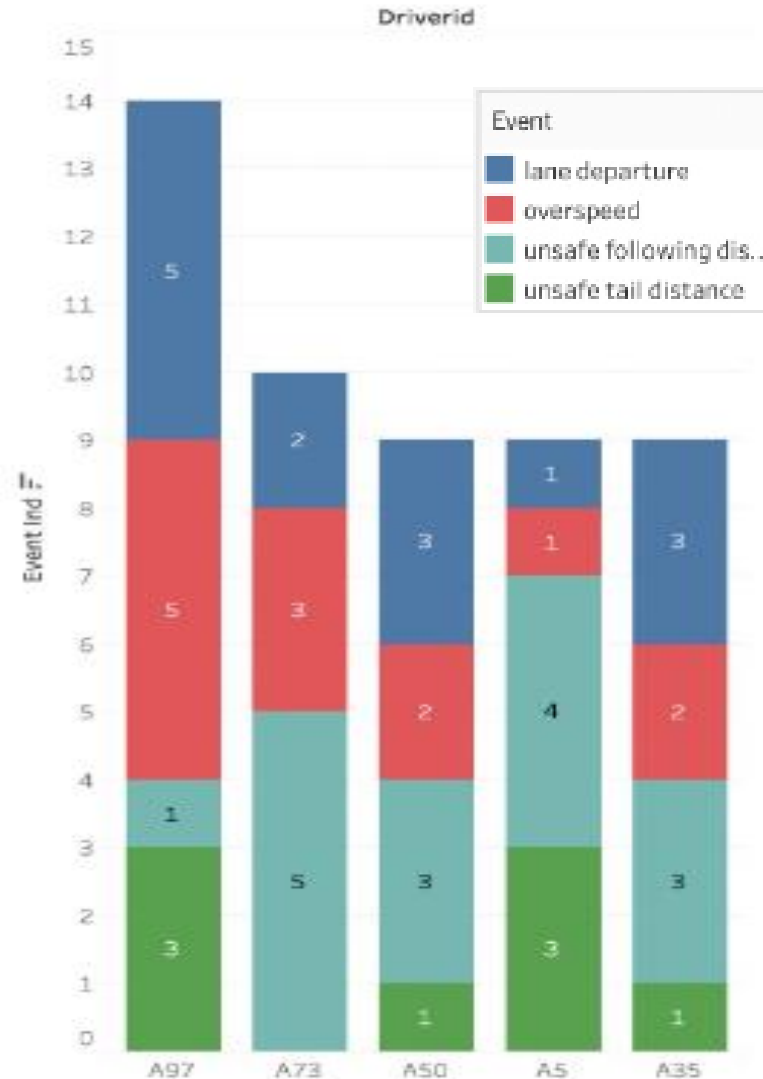


# Analysis 1 – Key Insights on Violations

Number of events by city to show which cities have the most



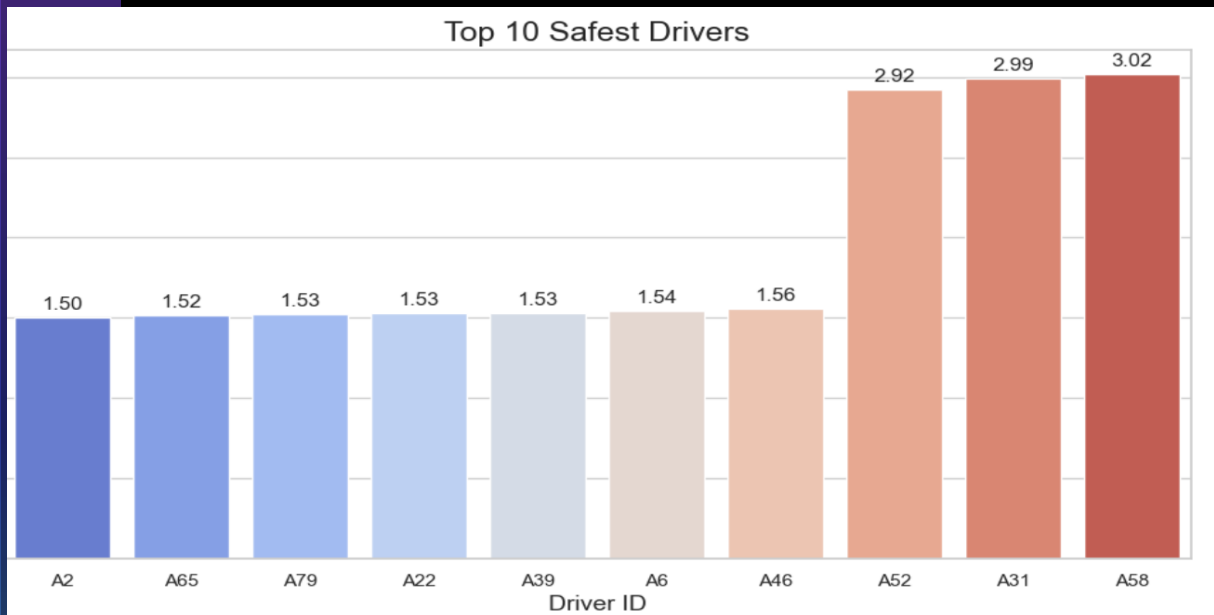
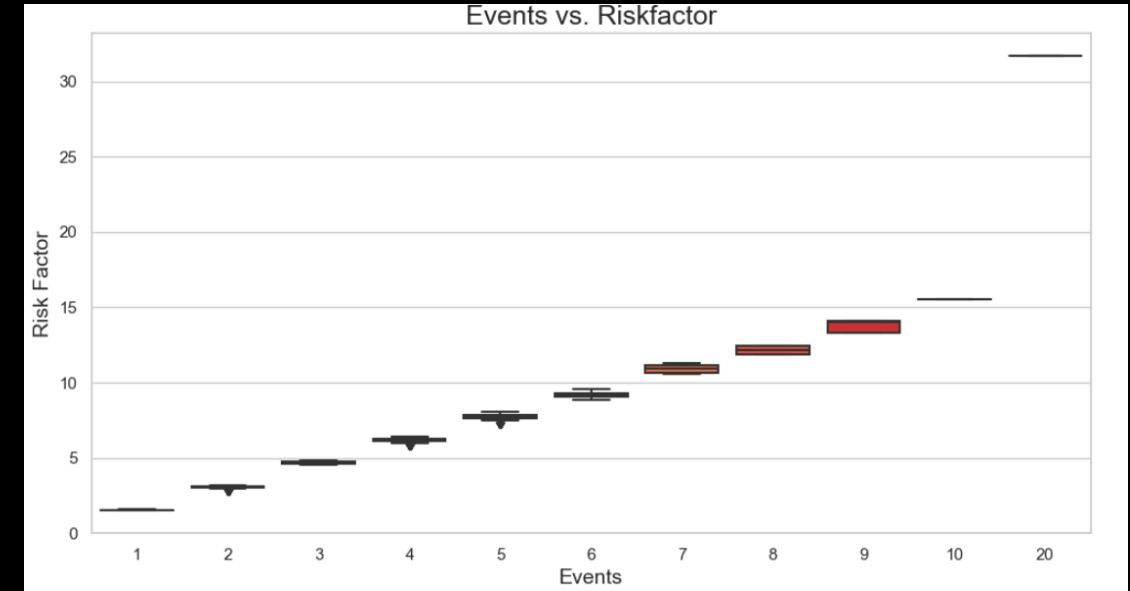
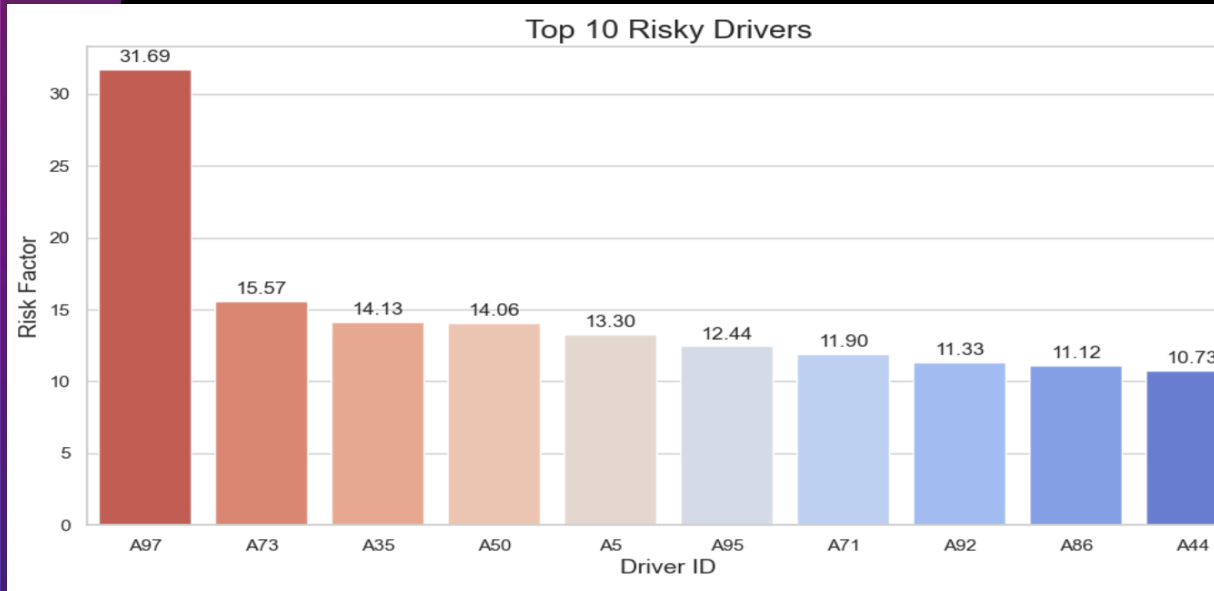
Event types for Top 5 risky drivers



1.Santa Rosa: The most common city in the dataset with the highest number of violations, so maximum events not necessarily mean anything here

2.Driver A97: Committed the most violations (14 times), followed by Driver A73

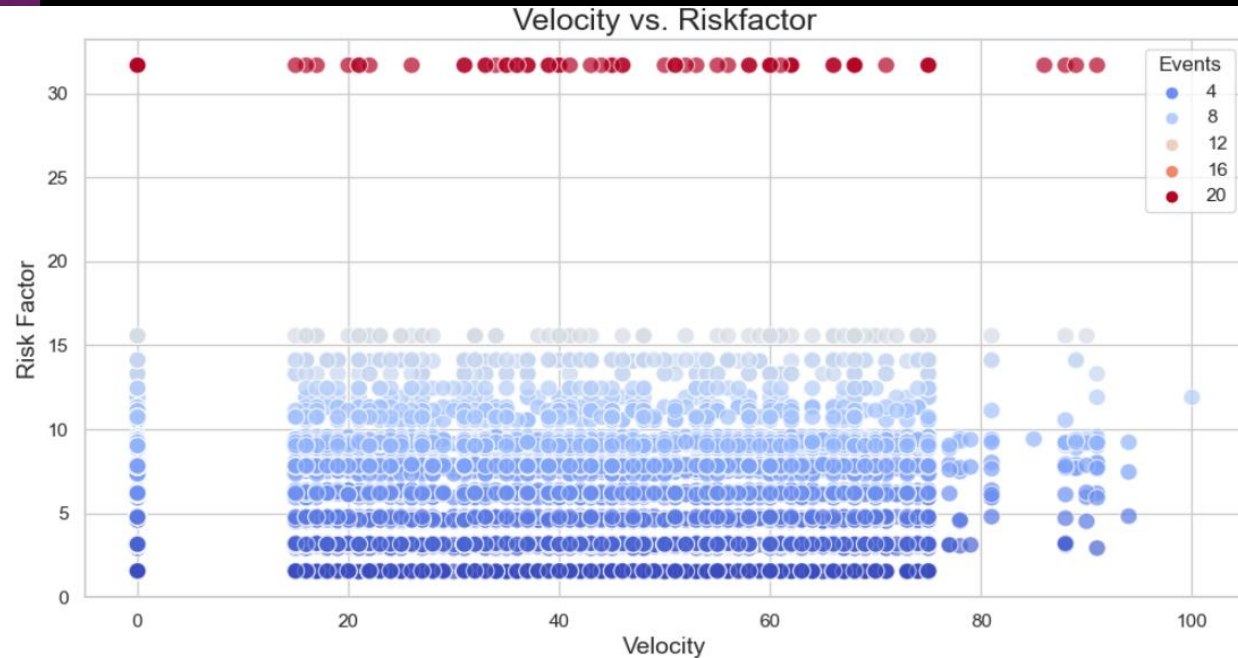
# Analysis 2– Risk Factor Analysis



- Risk factor is used as a parameter to categorize safe and risky drivers.
- A97 has the highest risk factor- no surprise
- A2 is the safest driver with risk factor- 1.50
- (Considering drivers with abnormal events only !)



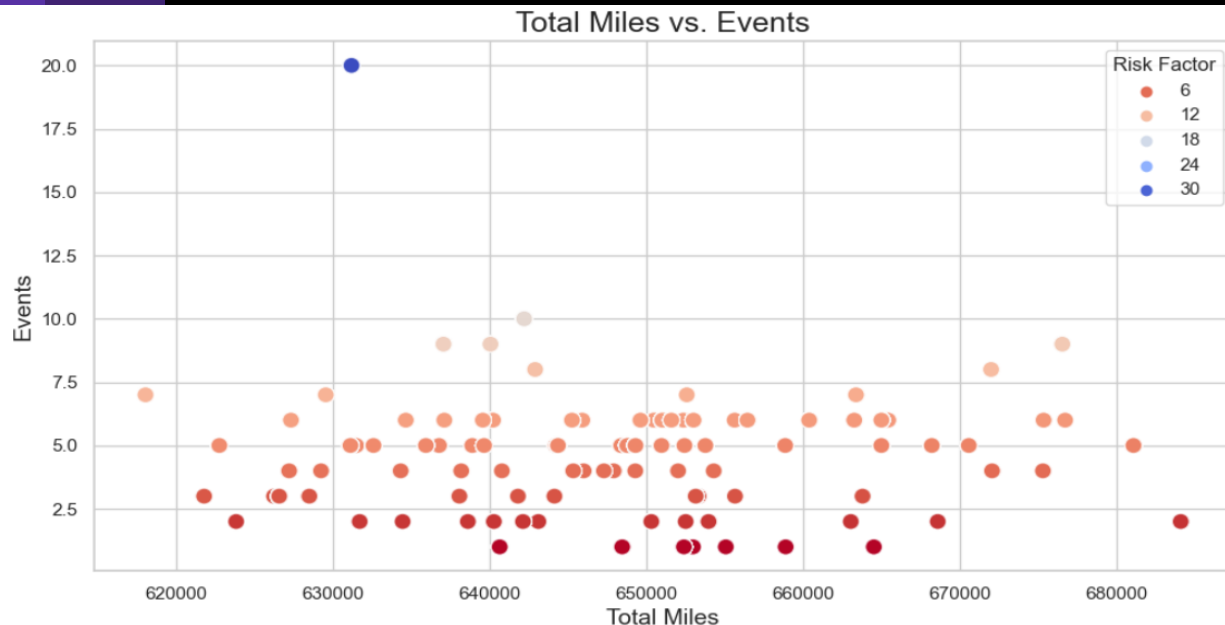
## Analysis 3– Velocity vs Risk Factor Analysis



1. Increase in velocity **does not** have a significant impact on the Risk Factor.

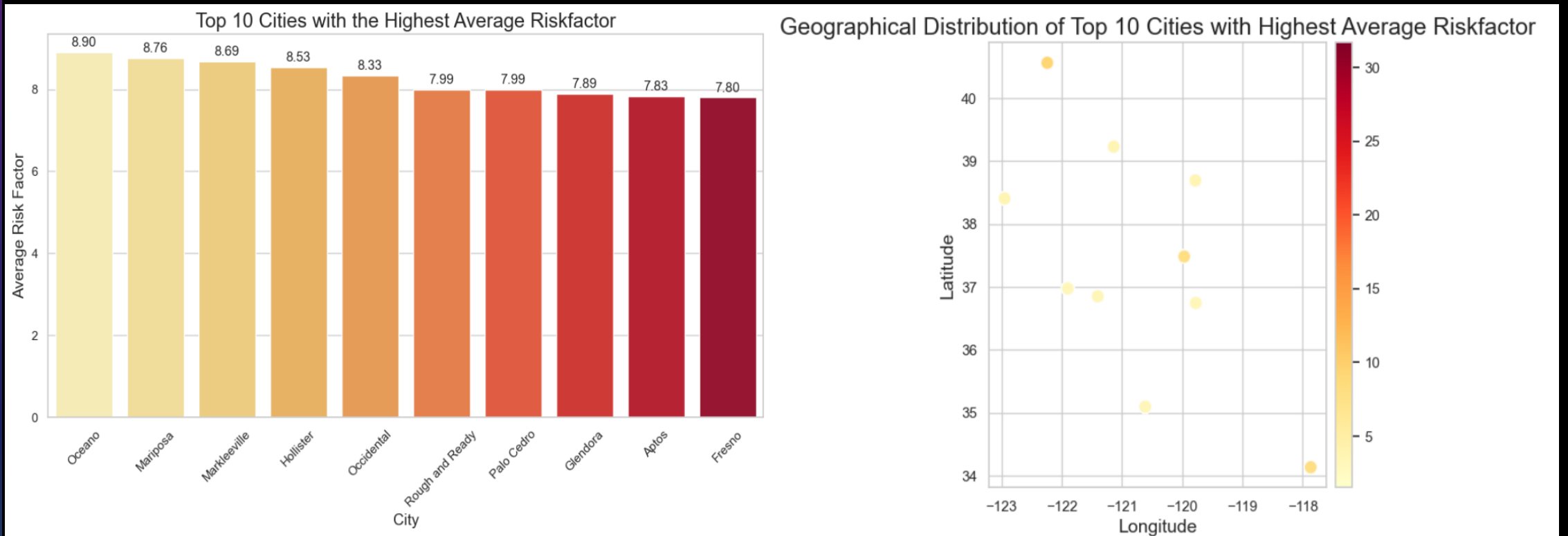
2. Risk factor mostly ranges between 6 and 12 for total miles between 620,000 and 680,000 miles.

3. Notable **outlier** observed at around 630,000 total miles.



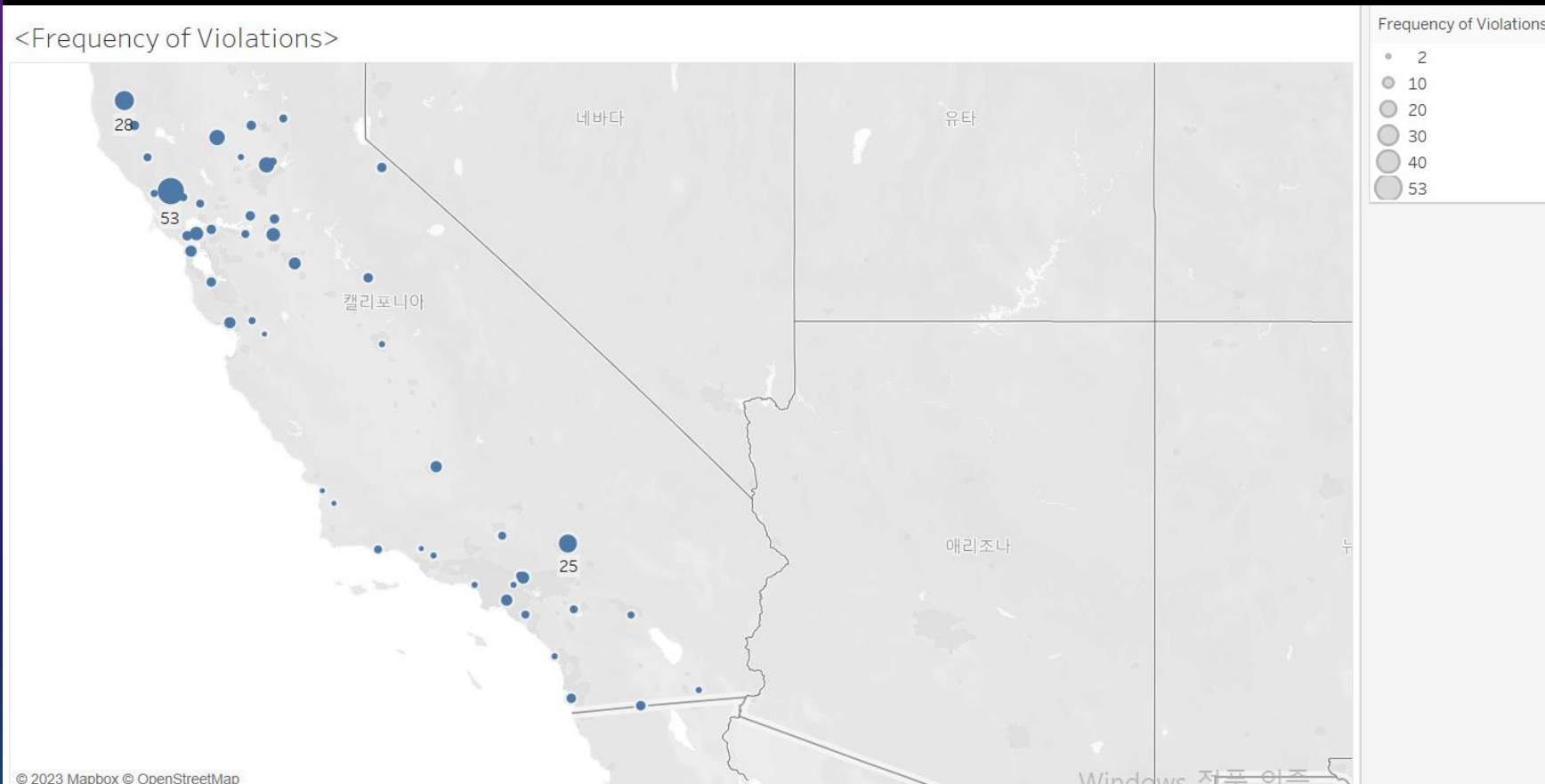
# Analysis 4

- Geographical distribution of top 10 cities in California with the highest average risk factor has been represented.
- Oceano is the city with the highest risk factor

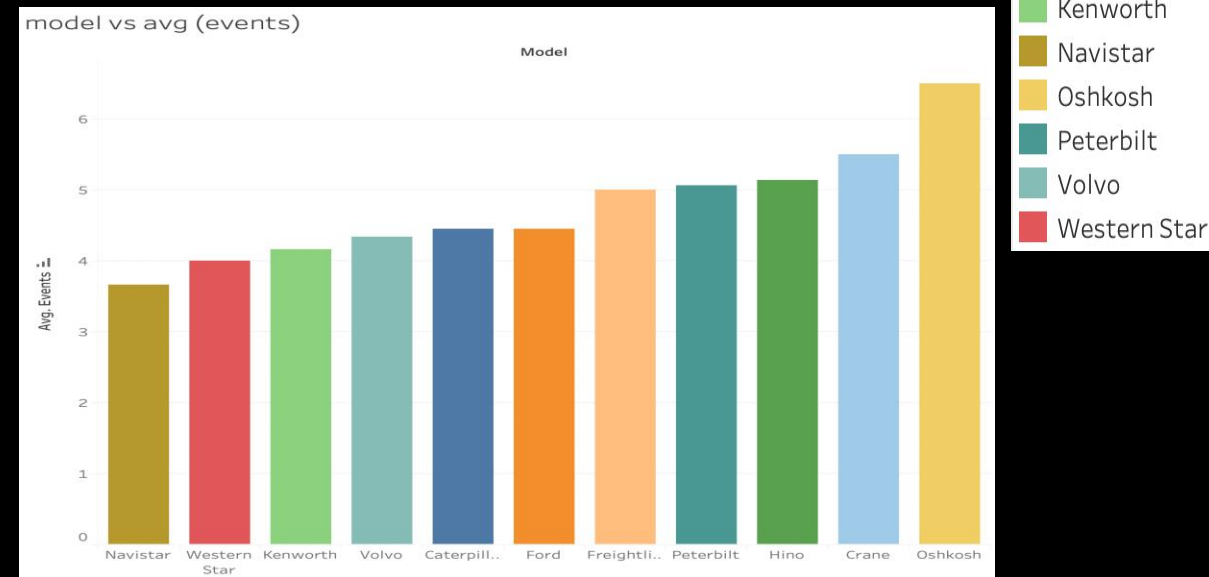
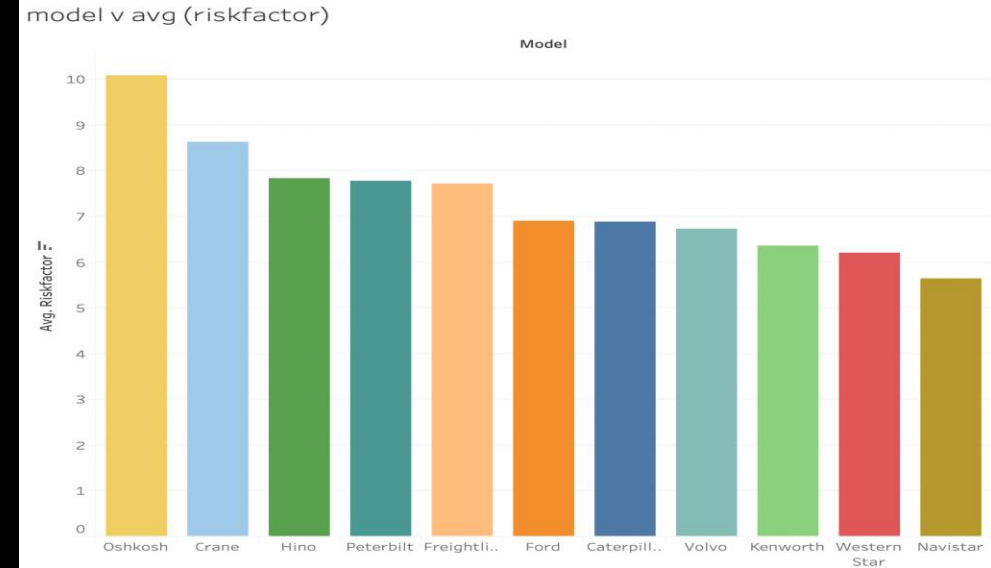
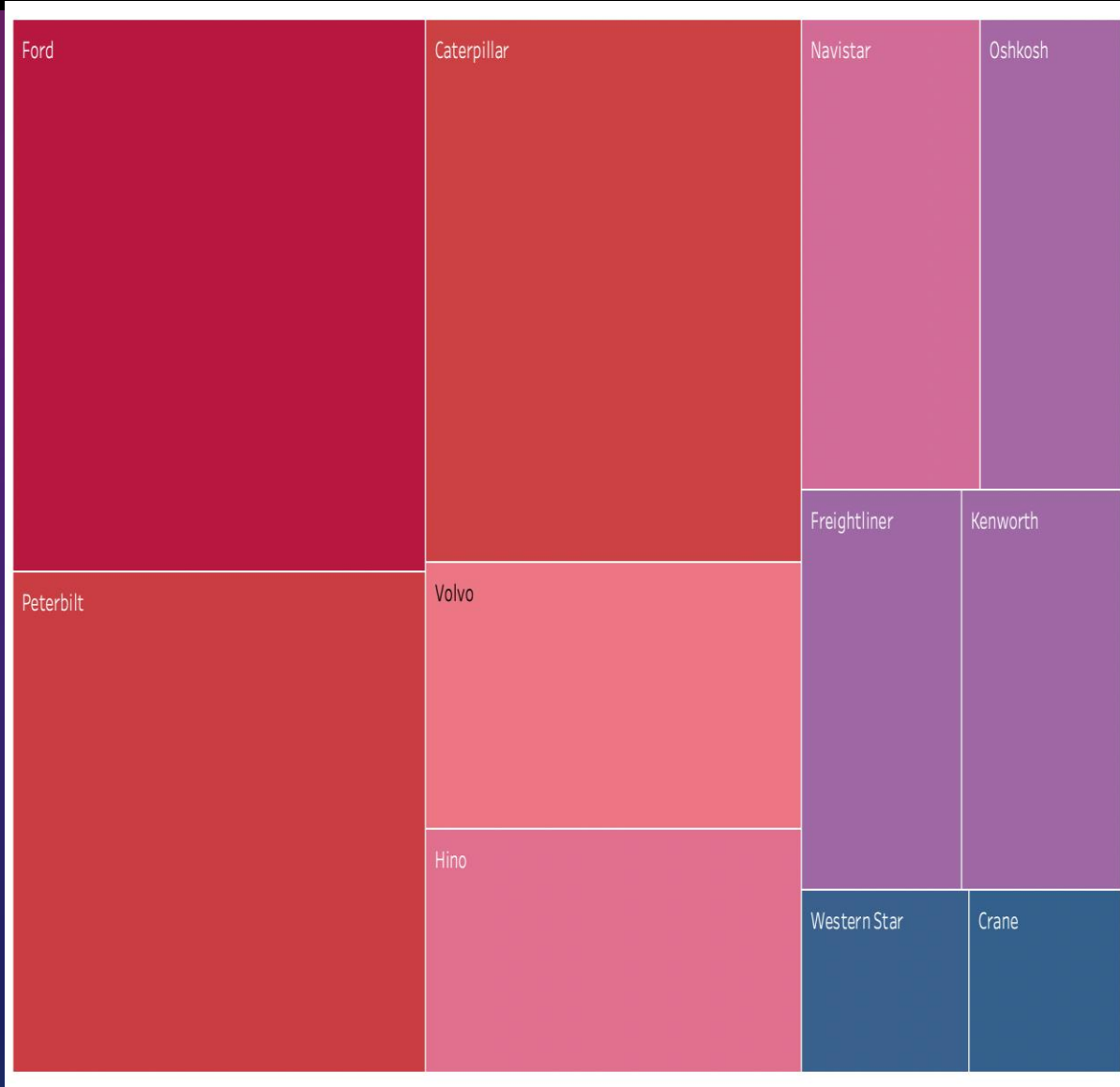


# Analysis 5

- This is the Frequency of Violation distribution in California. The riskiest city is Santa Rosa where 53 violations occurred in here.
- And it is followed by Willits with 28 incidents and Apple Valley as 19 violations.



# Analysis 6 – Truck model & risk factor



Different truck models have varying frequencies of abnormal events, which results in various degrees of risk factor.

# Challenges

- Limited knowledge of Tableau
- Wanted to leverage ChatGPT's capabilities
- Connecting Hadoop with Jupyter Notebook

# Turning Challenges into Opportunities

- Developed a machine learning model with the help of ChatGPT which in turn calls ChatGPT.
- ChatGPT integration for seamless interaction.
- Easy-to-use GUI for risk analysis.
- Real-time insights and recommendations.

# Introducing "RiskWise": *Your Interactive AI Risk Consultant*

RiskWise: Your Interactive AI Risk Consultant

Feature Importances:

Top 8 Feature Importances:  
1. Avgmpg with an importance of 0.3290  
2. Gas with an importance of 0.1769  
3. Avg\_gas\_used\_yearly with an importance of 0.1148  
4. Miles with an importance of 0.0883  
5. Date with an importance of 0.0801  
6. avg\_miles\_driven\_yearly with an importance of 0.0773  
7. model with an importance of 0.0695  
8. Totmiles with an importance of 0.0641

Ask a question:

Hello! Find a way to reduce accidents(Reduce the riskfactors)

RiskWise: Your Interactive AI Risk Consultant

Feature Importances:

Top 8 Feature Importances:  
1. Avgmpg with an importance of 0.3290  
2. Gas with an importance of 0.1769  
3. Avg\_gas\_used\_yearly with an importance of 0.1148  
4. Miles with an importance of 0.0883  
5. Date with an importance of 0.0801  
6. avg\_miles\_driven\_yearly with an importance of 0.0773  
7. model with an importance of 0.0695  
8. Totmiles with an importance of 0.0641

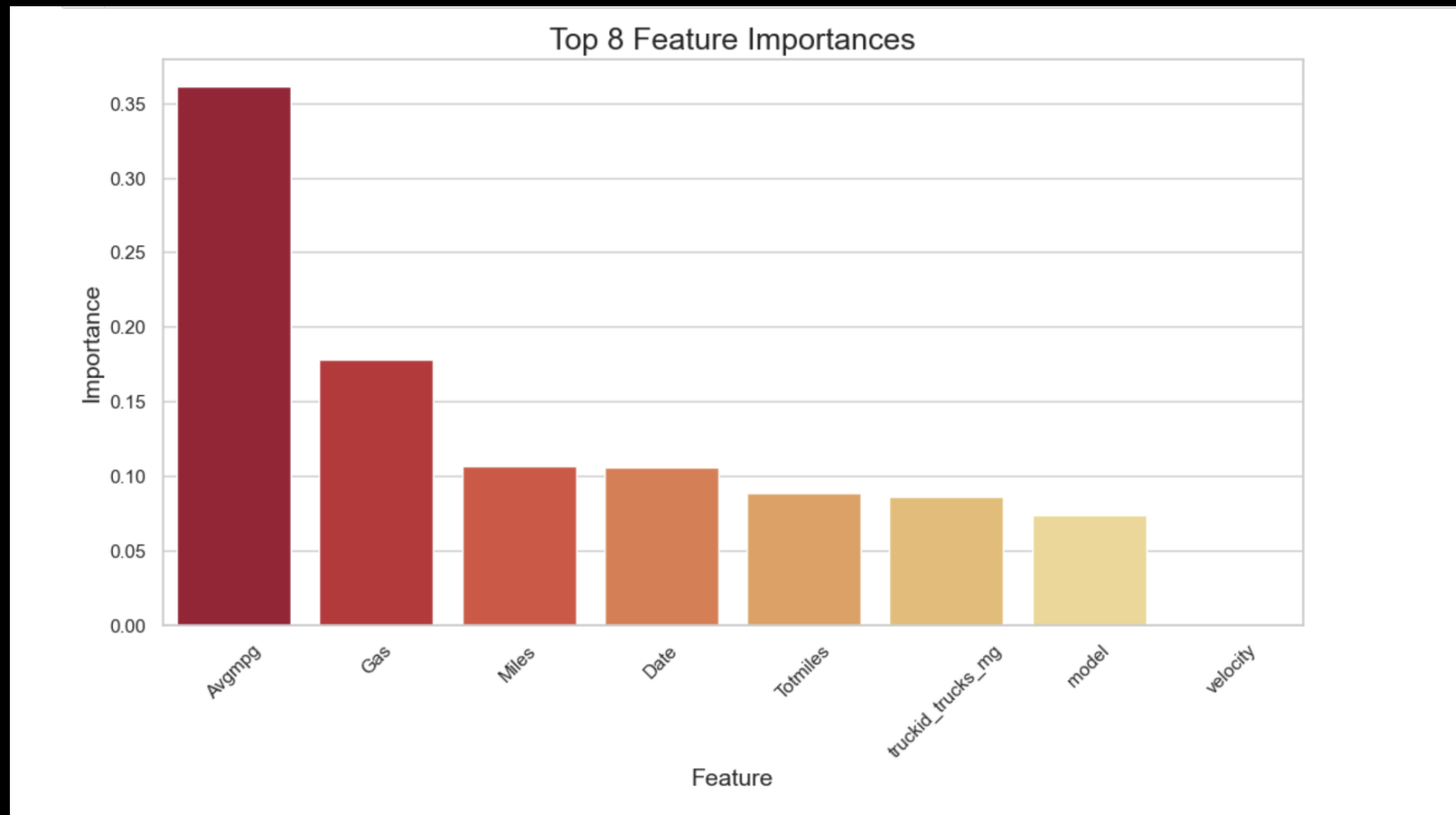
Ask a question:

Hello! Find a way to reduce accidents(Reduce the riskfactors)

Get Answer

Based on the feature importances above, it appears that the most important factor in reducing accidents is the average mpg of the vehicle. This suggests that the user should focus on increasing the fuel efficiency of their vehicles in order to reduce the risk of accidents. Additionally, it appears that other important factors are gas, the average gas used yearly, the miles driven, the date of purchase, the average miles driven yearly, the model of the vehicle, and the total miles driven. All of these factors should be considered when attempting to reduce the risk of accidents.

# Key Insights - Top 8 Features Impacting Risk Factor from Regression Model





# CONCLUSIONS

1. The top 8 features impacting risk factors are identified: a. Avgmpg (0.3290) b. Gas (0.1769) c. Avg\_gas\_used\_yearly (0.1148) d. Miles (0.0883) e. Date (0.0801) f. avg\_miles\_driven\_yearly (0.0773) g. model (0.0695) h. Totmiles (0.0641)
2. These features can be used to predict and assess risk factors, ultimately helping to improve driver safety and reduce accidents caused by large commercial trucks.
3. By focusing on these critical factors, companies can implement targeted interventions to minimize risk and enhance the overall safety of their fleet operations.



# THANK YOU!

We kindly ask you to take a moment to rate our

**"RiskWise":** *Your Interactive AI Risk Consultant*