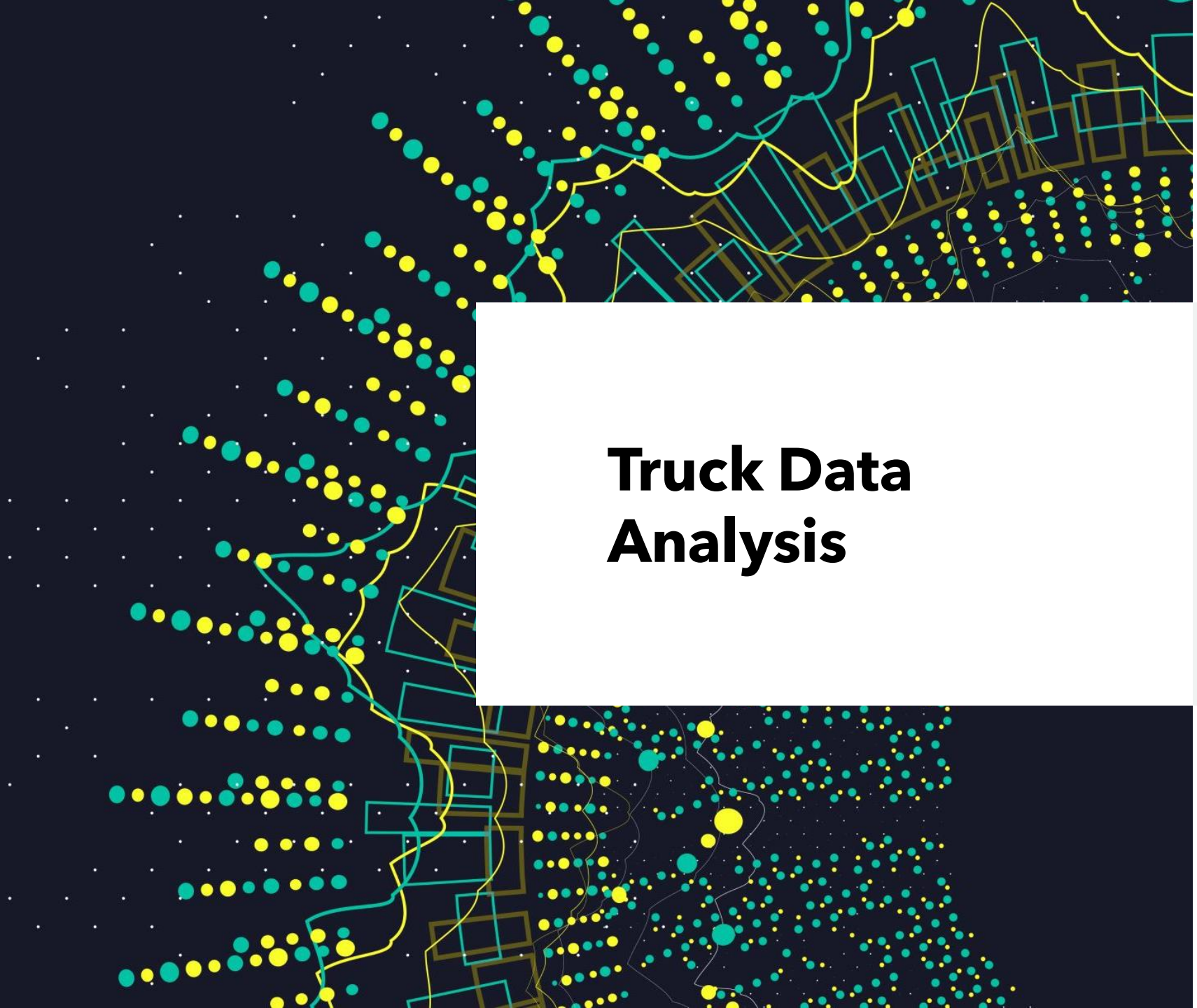


Truck Data Analysis

BUAN 6346.502

GROUP 5

Group Members

- Asmita Shetty
- Jaya Prakash Sathravada Gopi
- Pavan Kumar Siripuram
- Prayag Prasad Verma
- Salman Shaik
- Sibi Chakravarthy Vasudevan

Problem Statement

ANT (Az National Trucking) aims to guarantee that all of its drivers follow the laws and guidelines that control the trucking business with a particular emphasis on reducing risk factors that may result in mishaps. The company is worried about several driver behavior issues and operational characteristics, such as speed and mileage, in addition to other compliance requirements specified by state-specific legislation the FMCSA regulation, and the Department of Transportation.

Objective

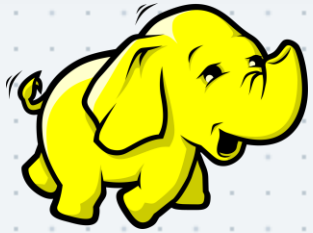
- Identify the most hazardous commercial truck drivers
- The ultimate goal is to uphold the highest standards of fleet operation safety and efficient.
- Thereby ensuring the well-being of drivers and reducing operational costs related to non-compliance and insurance

Workflow process

Download given datasets

Move to VMWare

Load to HDFS



Import to Impala

Create risk factors using Pig

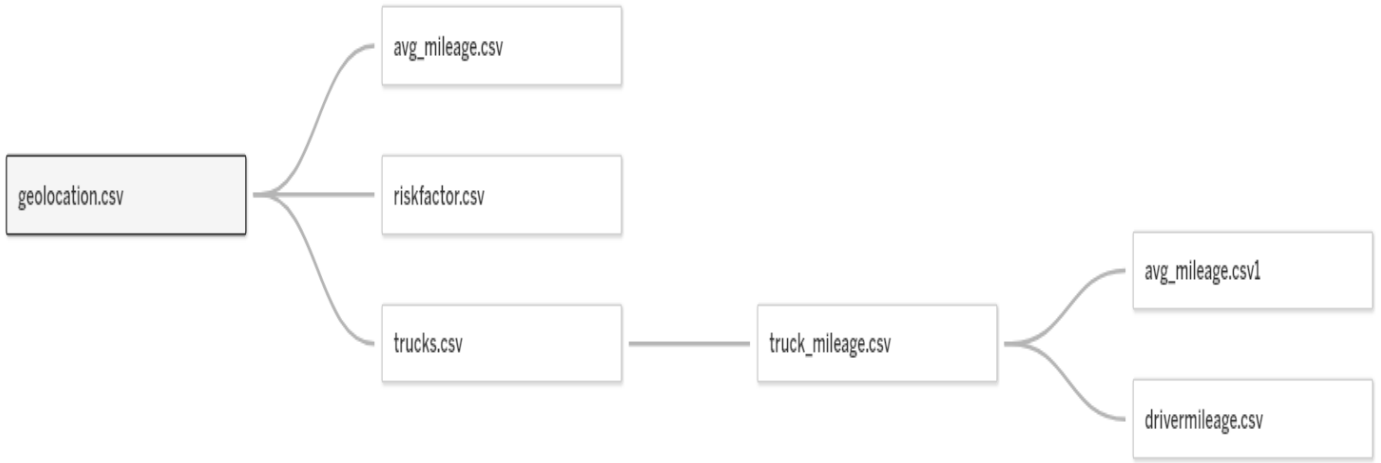
Connect the tables to tableau

About Data

Important Columns used from the tables that we considered for analysis in the datasets are:

- Driver ID
- Truck Id
- Events
- Risk Factor
- State, City
- Truck model
- Velocity

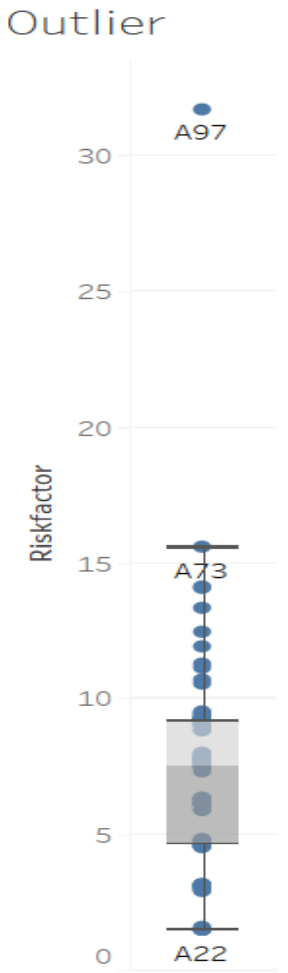
Relational Database



Outliers

INSIGHTS

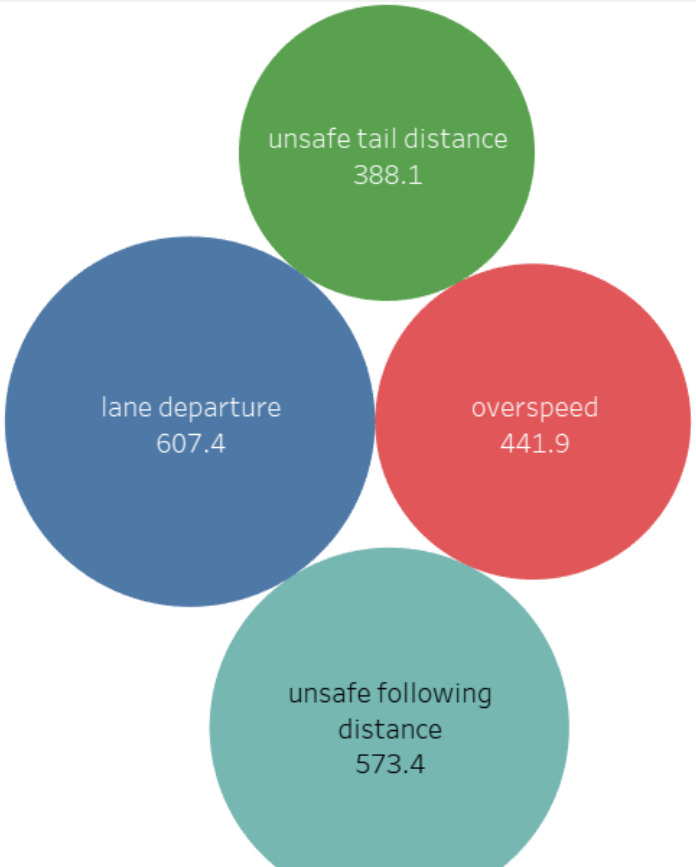
- From the box plot for Risk Factor, we can clearly see that Truck ID **A97** has the highest risk factor.
- This can be considered an outlier as inferred from the graph.
- Therefore, we propose to remove the outlier A97 , as it can alter our analysis.



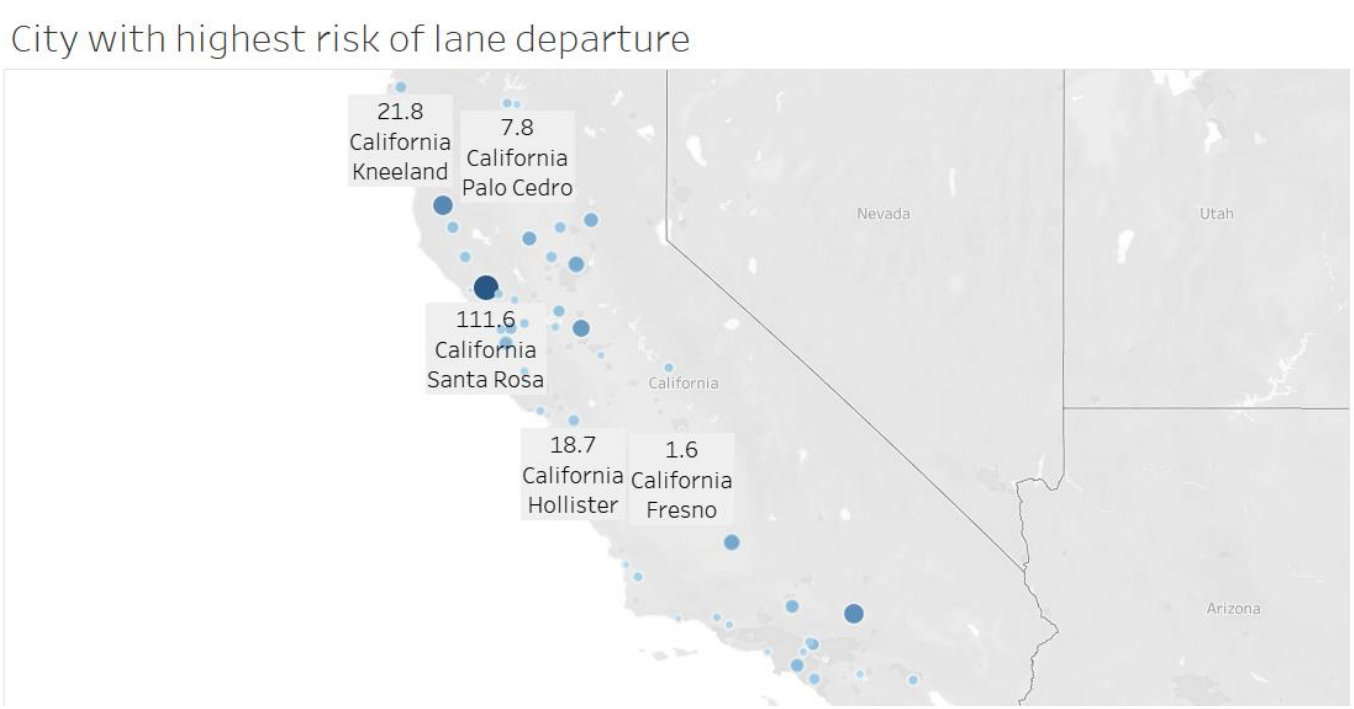
Events with Highest Risk

- Four events that pose the highest risk, in descending order:

1. Lane departure
2. Unsafe following distance
3. Overspeed
4. Unsafe tail distance

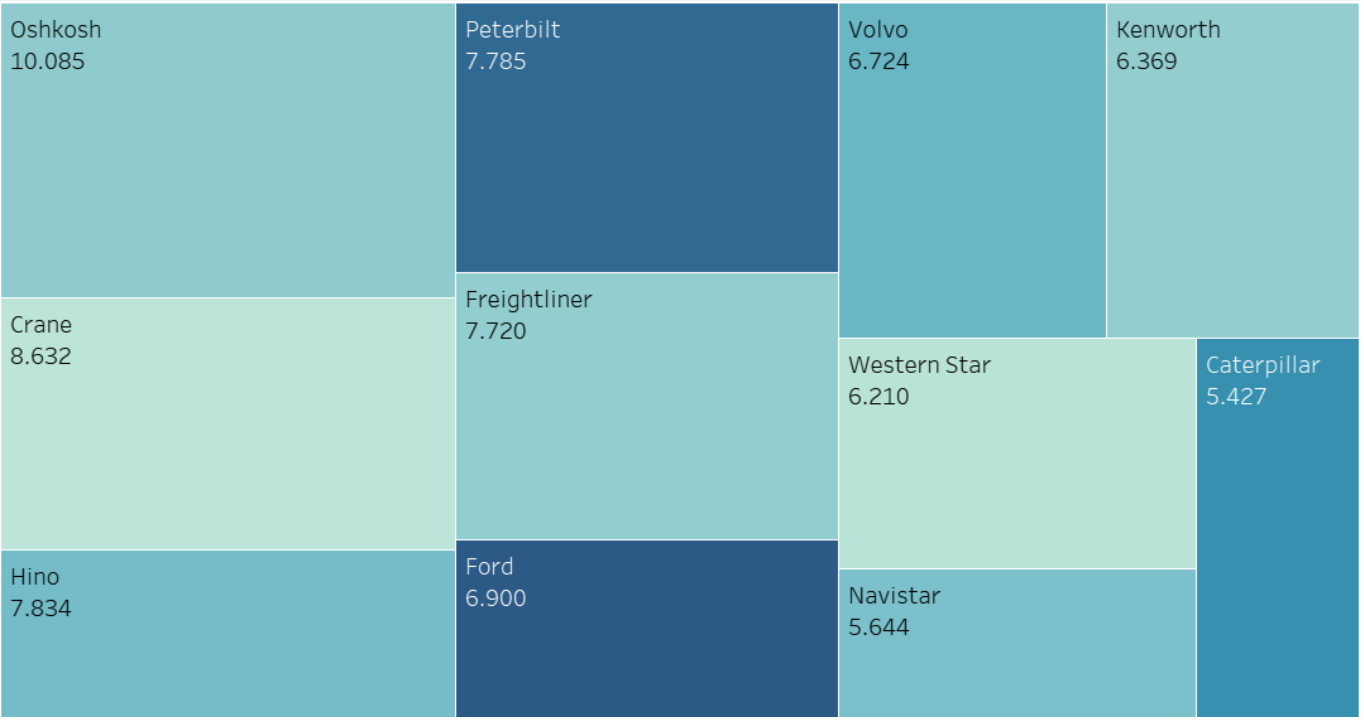


Cities with Highest Risk of Lane Departure



From the above plot, we can conclude that for the city "Santa Rosa" in California, there is a highest risk of Lane Departure with a sum of risk factor of "111.6"

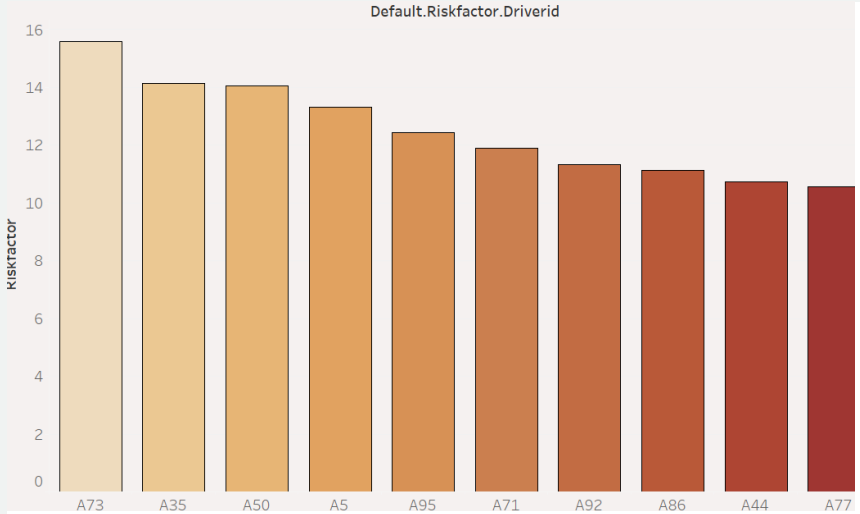
Truck Models with Highest Risk

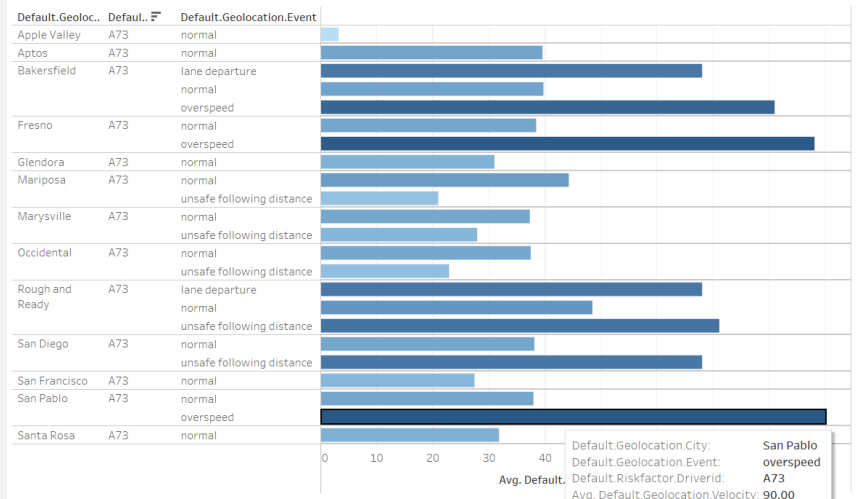


The top 3 truck models with highest risk are:

1. Oshkosh
2. Crane
3. Hino

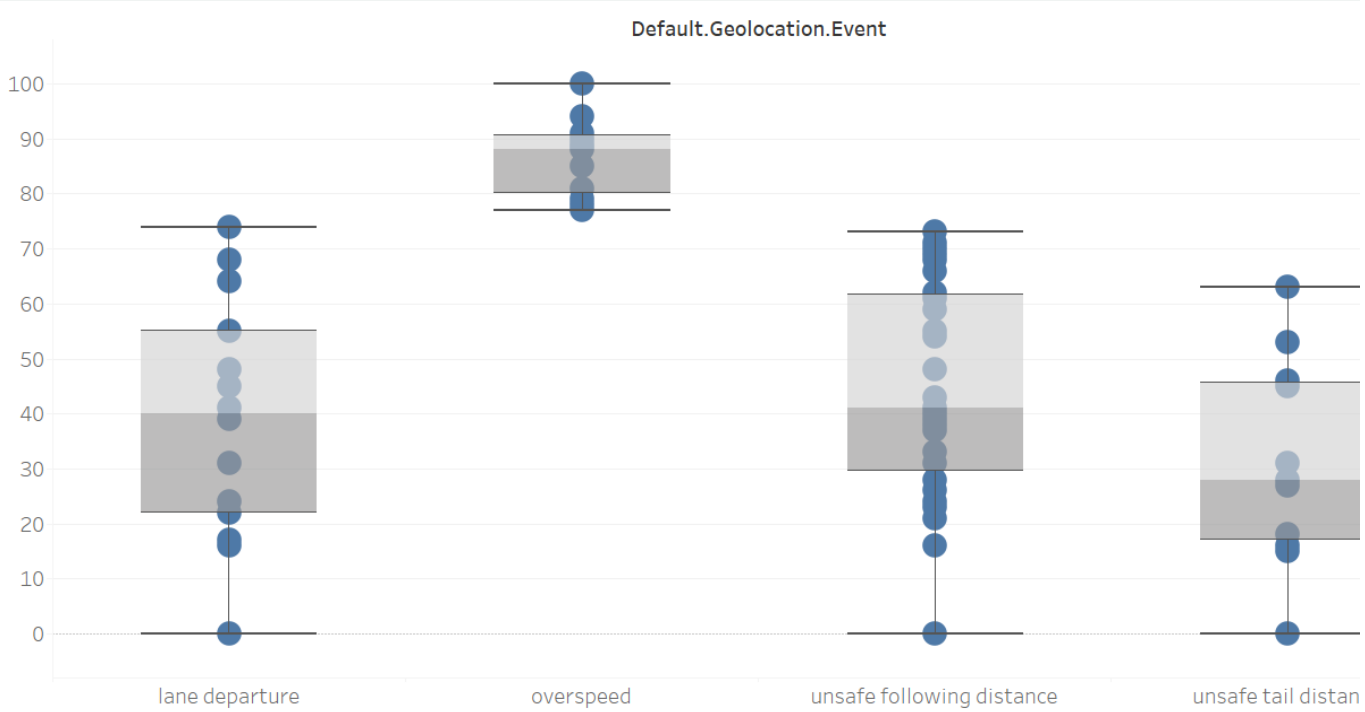
Top 10 Risky Drivers





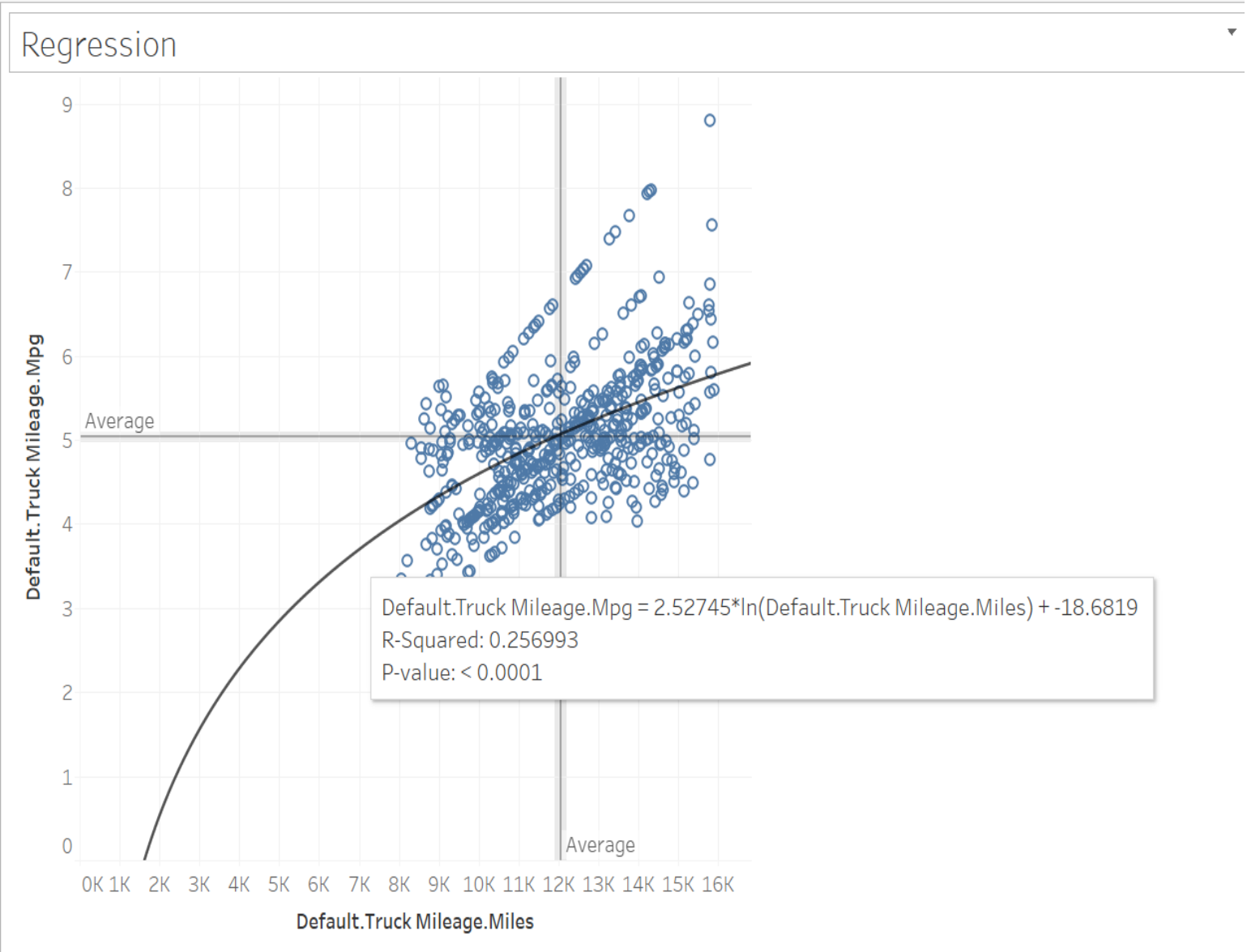
- From our plot, we can say that A73 is the riskiest driver, followed by A35 and A50.
- The factor causing most risk has generally been Over-speeding with an average velocity of 90.
- The highest risk has been shown in San Pablo

Truck Speeds for different Events



- We have plotted the velocity distributions for different types of events
- Clearly, for Over-speeding, the truck velocities are very high.
- Similarly, for unsafe tail distance, the velocities are very low this could be due to not maintaining safe distances in traffic.
- Lane departure and unsafe following distance have velocities in the middle (neither too high nor too low)

Business Insight



- Hypothesis : That the longer the truck travels the better mpg it gives
- We have seen an upward positive trend
- The average mileage is 5.037mpg with a confidence interval 95%

Conclusion

- Risk factor for every driver could depend on the truck model, events, and velocities.
- Truck Id A97 is an outlier.
- Type of abnormal event can depend on geographical location (city).
- We can cut the gas expenditure, by sending older trucks for longer distances and newer trucks for shorter trips.