TRUCK FLEET ANALYSIS

Case Scenario

Az National Trucking (ANT) - is a fictional national trucking corporation headquartered in California.

It is a 10-year-old organization with 400 employees, most of whom are long-distance truck drivers. ANT’s primary business is to provide long distance trucking services of all general and

Problem Statement

As the fleet managers of Az National Trucking Corporation, we are in charge of ensuring that all the drivers in the fleet are in compliance with the rules and regulations of the corporation and do not present an insurance risk due to

a) Speeding
b) Unsafe driving
c) Lane departure
d) Other unsafe driving practices

Objective

To reduce the risk of accidents among drivers by :
Analyzing and identify dangerous truck drivers
Highlighting drivers with risk factor greater than or equal to 7.0

Process Flow

Download the data

Performing Exploratory Data Analysis

Local disk

Move files from local system to HDFS using SCP command

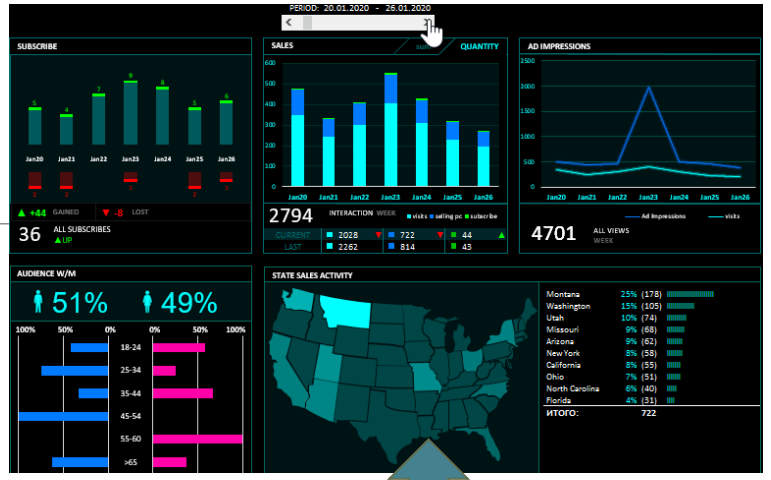
Cloudera Server

Create tables in Hive writing Hive queries

Import data in Impala using LOAD DATA INPATH command

Trucks and Geolocation CSV files loaded into HDFS as tables

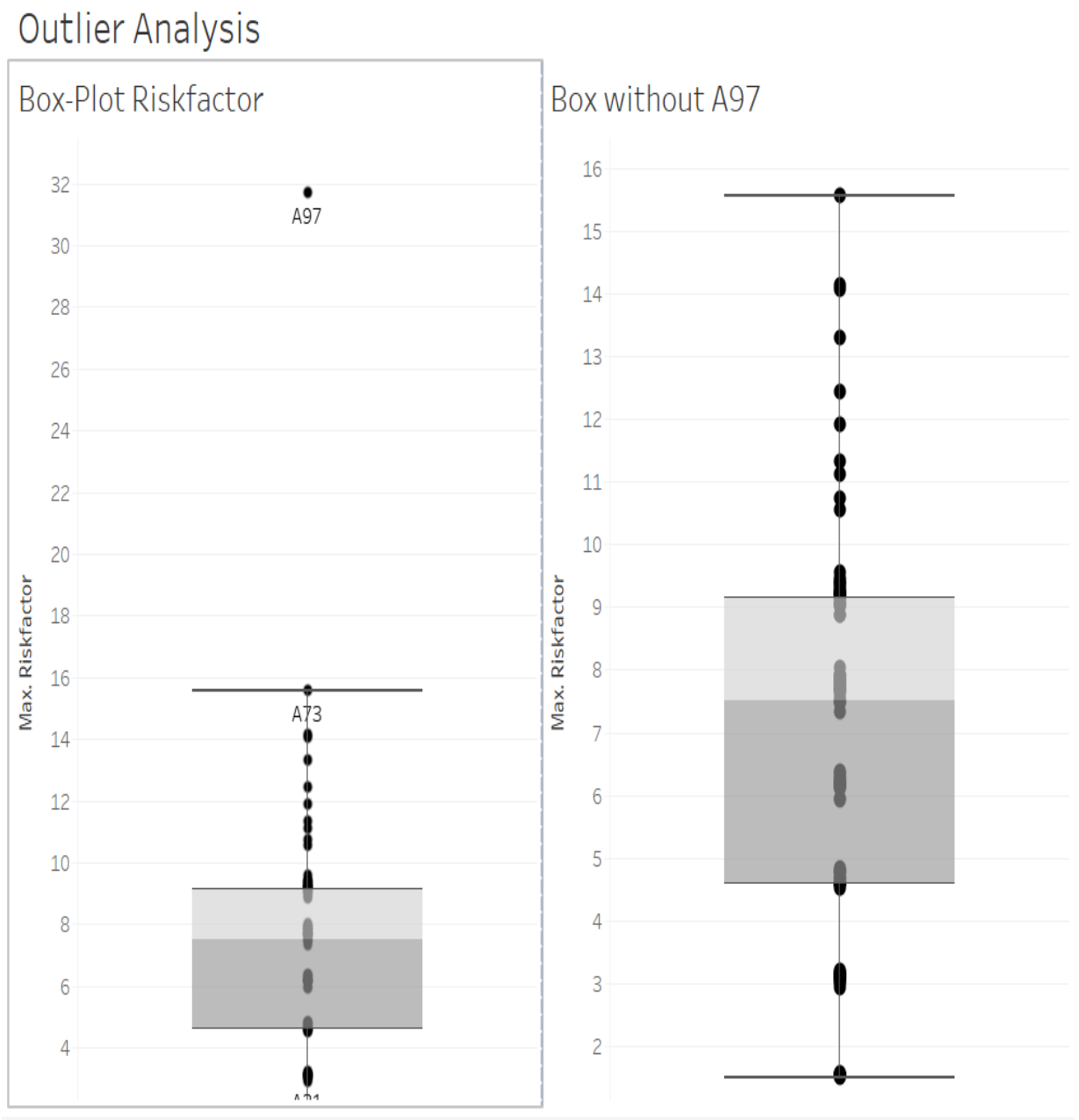
Import data into Tableau using Impala JDBC driver



Create an interactive visualization dashboards in Tableau

The boxplot in the Image 1 shows that DriverID A97 has a risk factor of 31.69 which is an outlier

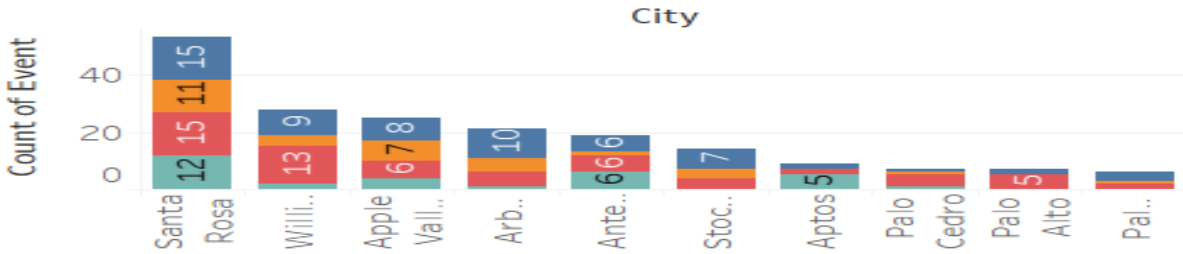
- In order to avoid skewness and to bring the Risk Factor values between 1 and 10, the outlier is removed, and scaling is done(Image 2)
- However, though driver A97 is an outlier, independent in-depth analysis has been done to further understand why he had a very high-risk factor when compared to other drivers



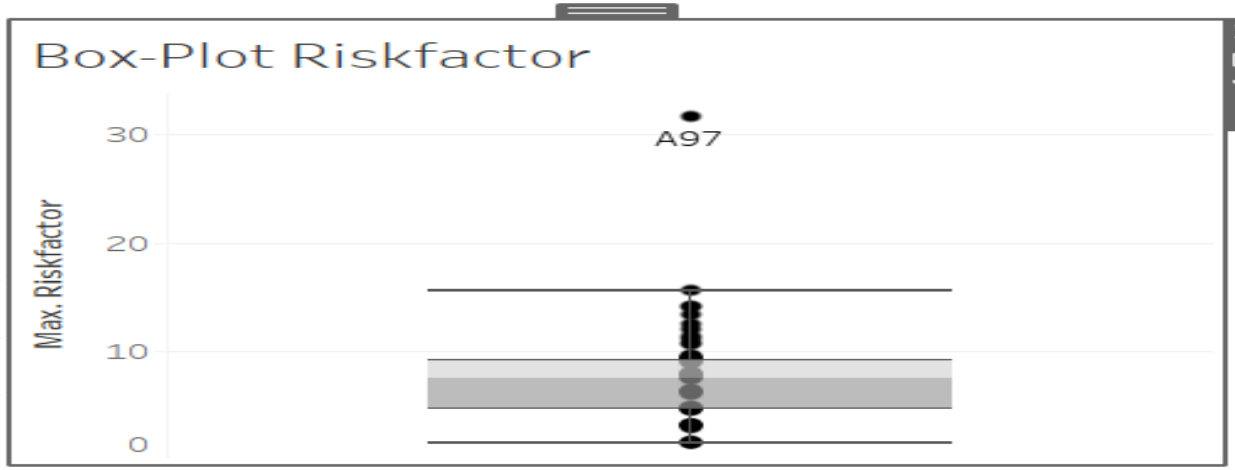
Driver A97 Analysis

Abnormal Events in Top 10 Cities

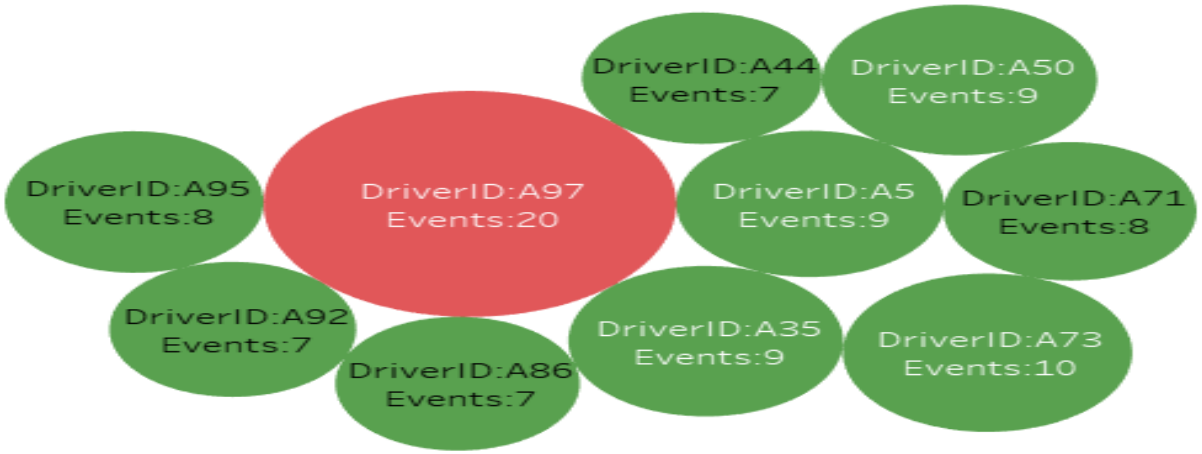
Event distribution per city



Box-Plot Riskfactor

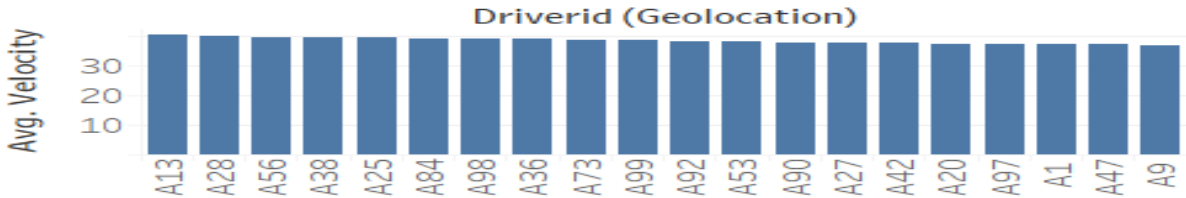


Top 10 Risky Drivers



Total abnormal Events

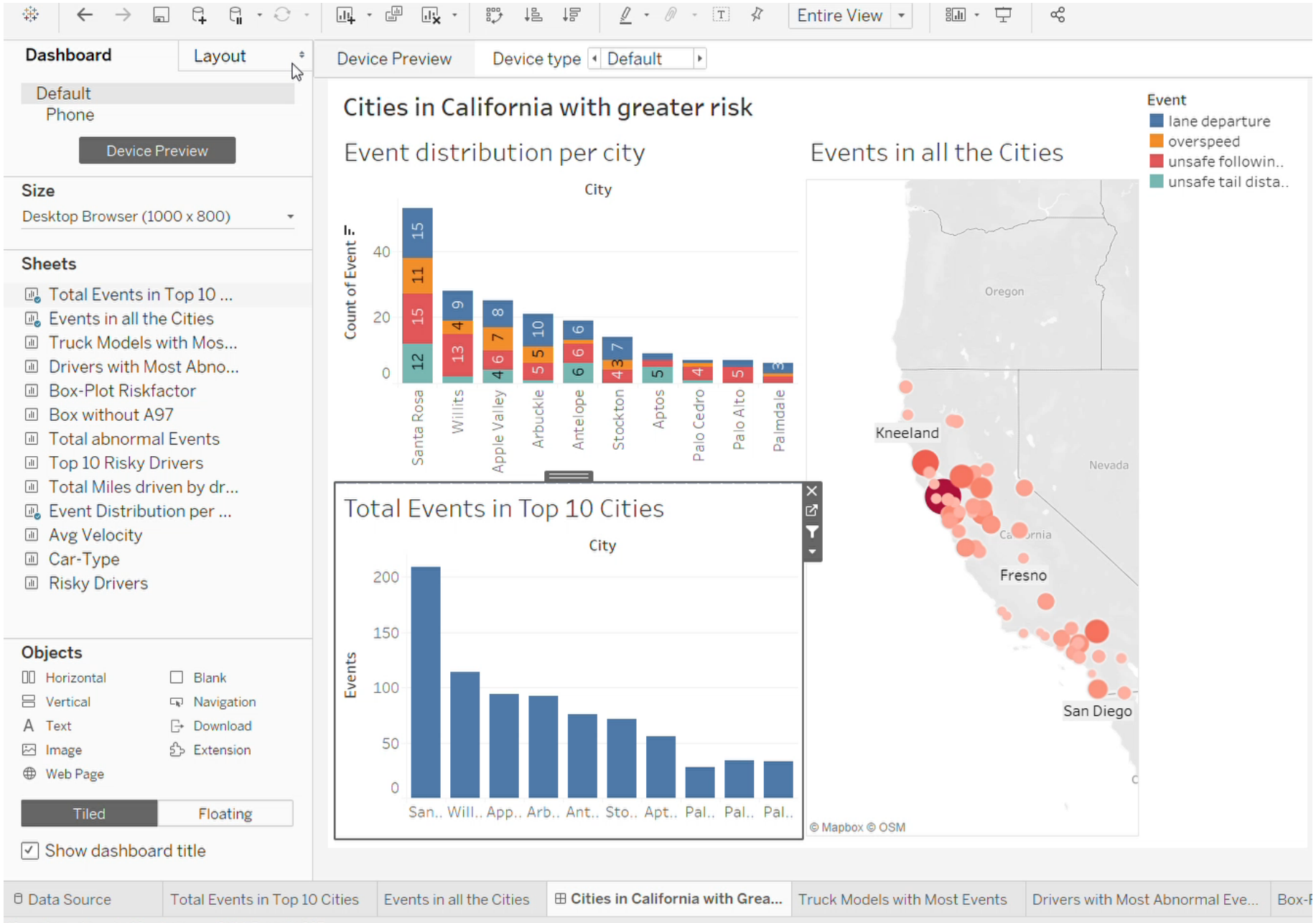
Avg Velocity



- Driver A-97 has been driving with an avg Velocity which is greater than the most risky drivers
- A97 did not comply with the rules and has posed risk with all the factors
- Also driver A97 can not be claimed as new or inexperienced driver since he had already driven 632K miles

Takeaway:

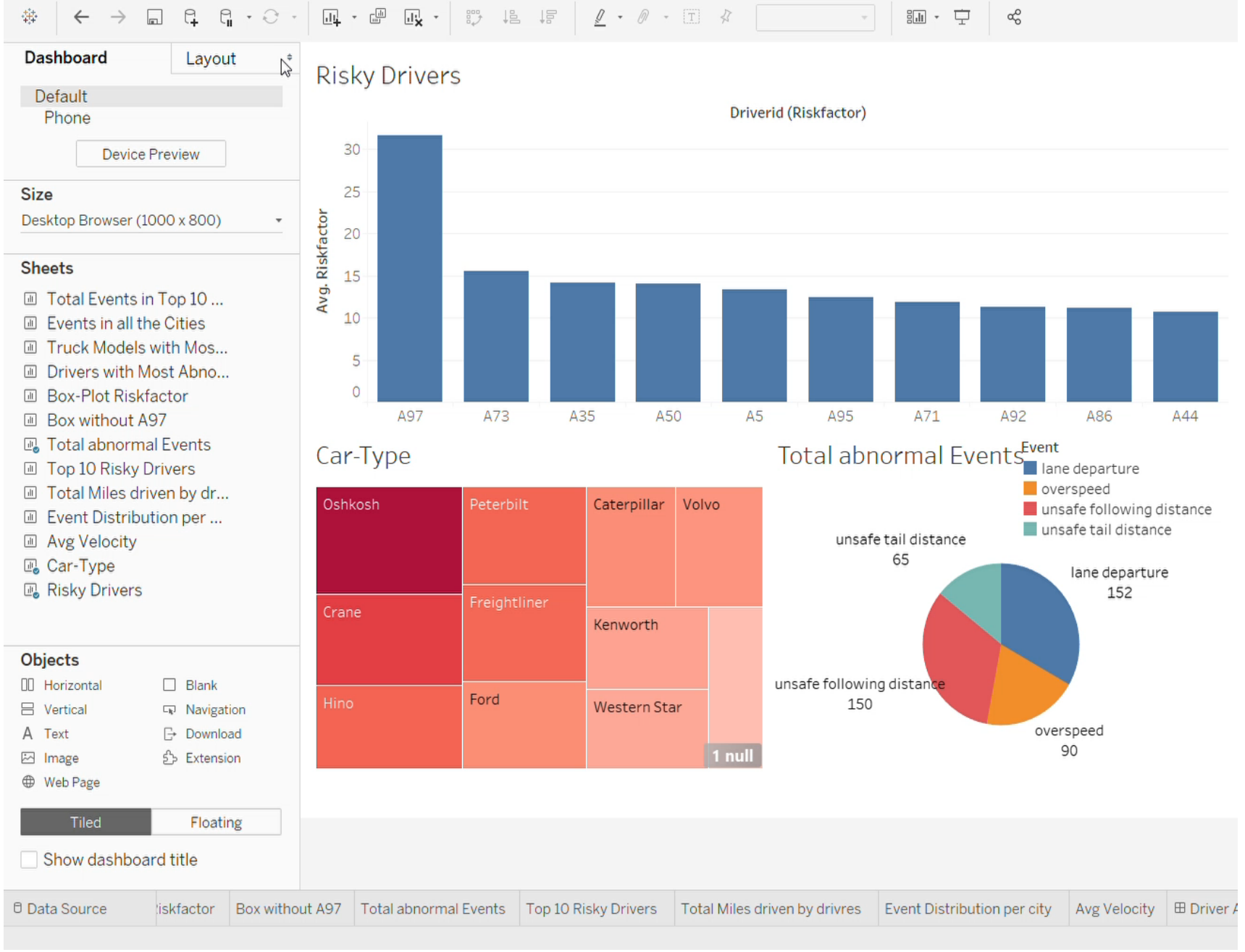
- This plot displays top 10 cities in California
- The City with the highest number of abnormal event is “Santa Rosa”
- To answer this question further analysis should be done like condition of roads in that city or weather condition, and Traffic condition



Analysis of Drivers with Risk Factor >7

Insights:

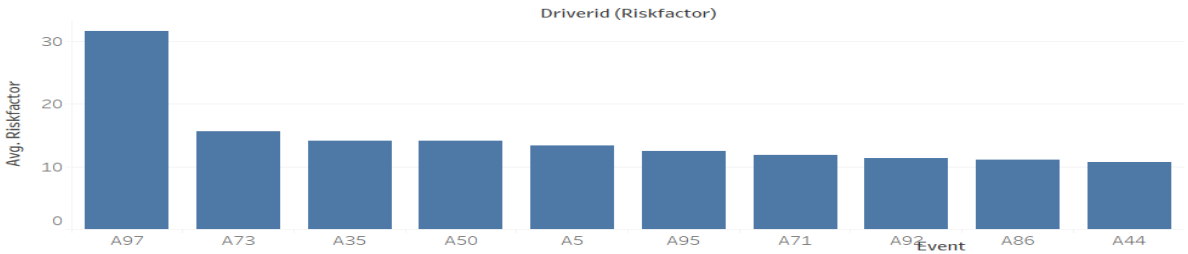
- There are 8 drivers with risk factor>7 who are A73, A35, A50, A5, A95, A71, A92 and A96
- These drivers should be considered as “RISKY” by the organization and should be monitored
- In the plot By clicking on the bar of each risky driver, their corresponding abnormal events and vehicle model can be seen



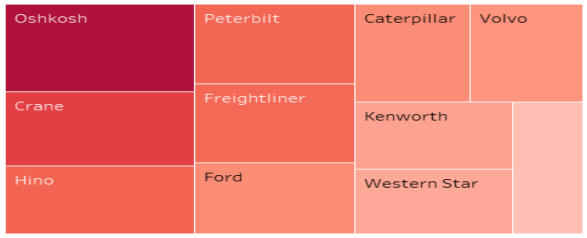
App for Manager

Drives with Risk Factor > 7

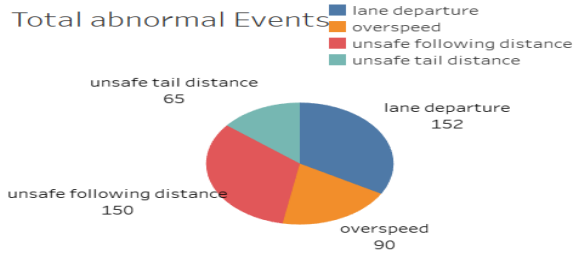
Risky Drivers



Car-Type

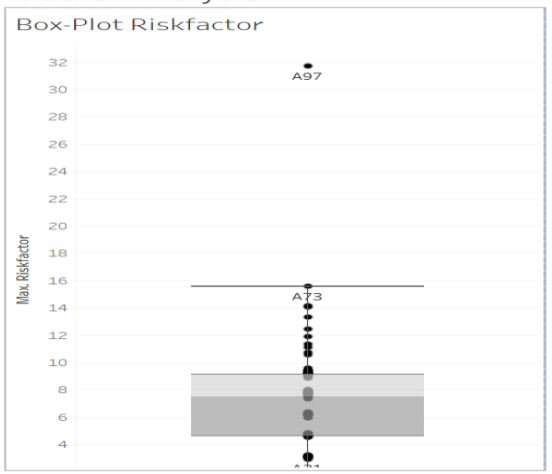


Total abnormal Events

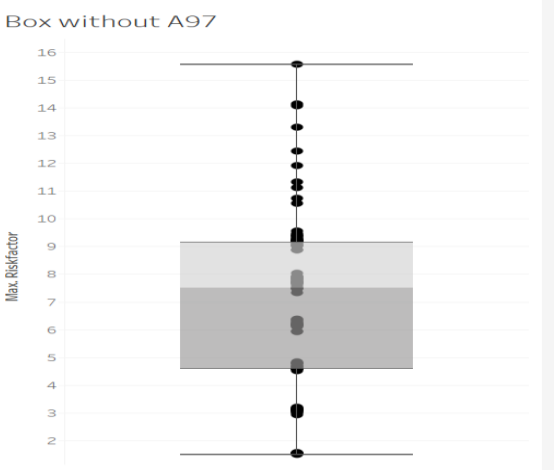


Outlier Analysis

Box-Plot Riskfactor



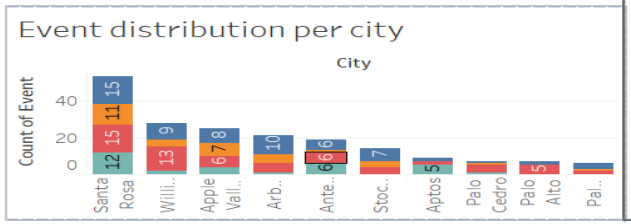
Box without A97



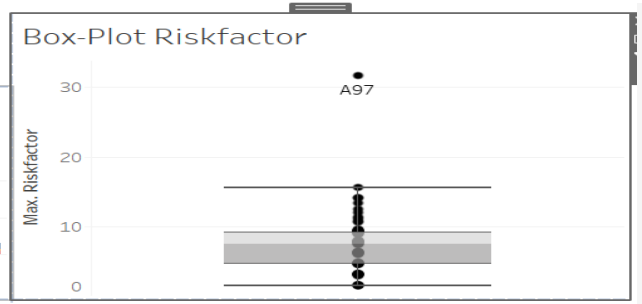
Driver A97 Analysis

Abnormal Events in Top 10 Cities

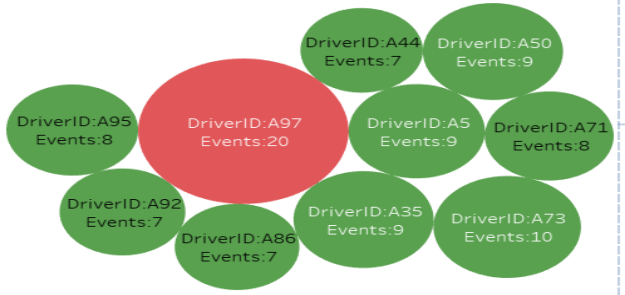
Event distribution per city



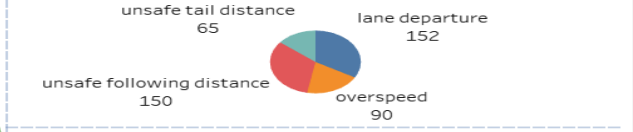
Box-Plot Riskfactor



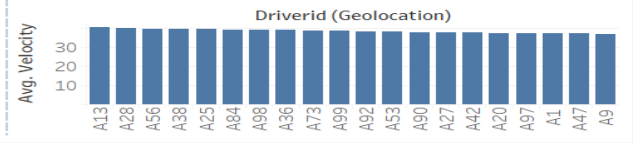
Top 10 Risky Drivers



Total abnormal Events

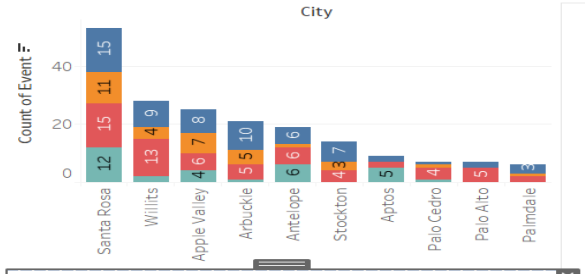


Avg Velocity

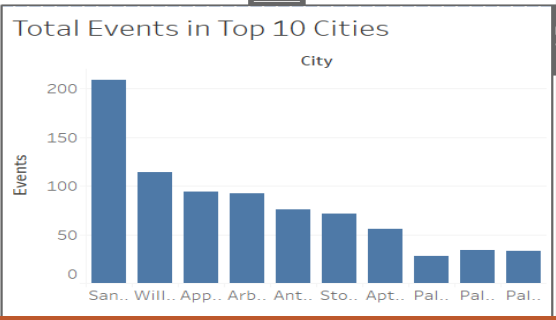


Cities in California with greater risk

Event distribution per city



Total Events in Top 10 Cities



Events in all the Cities