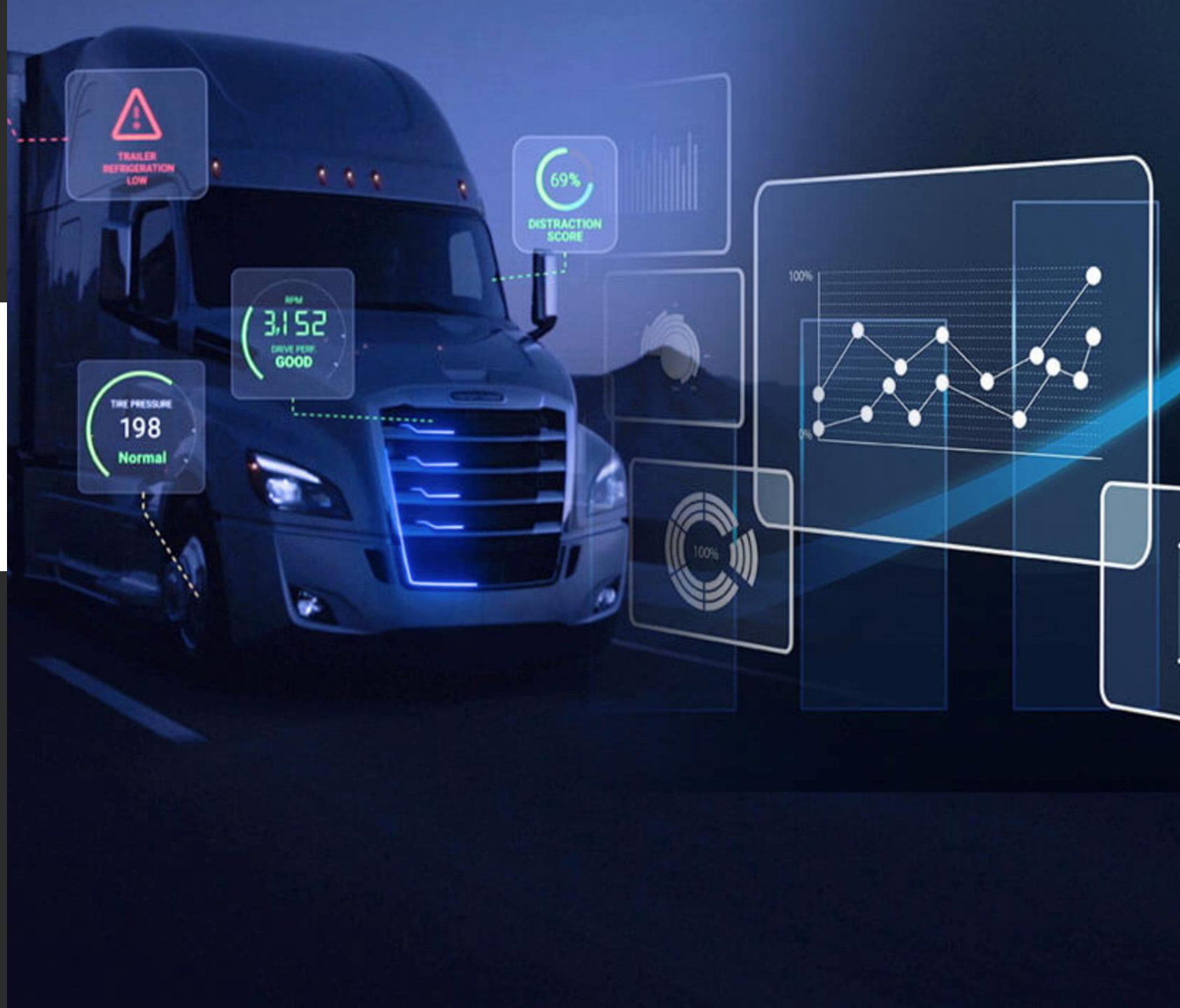
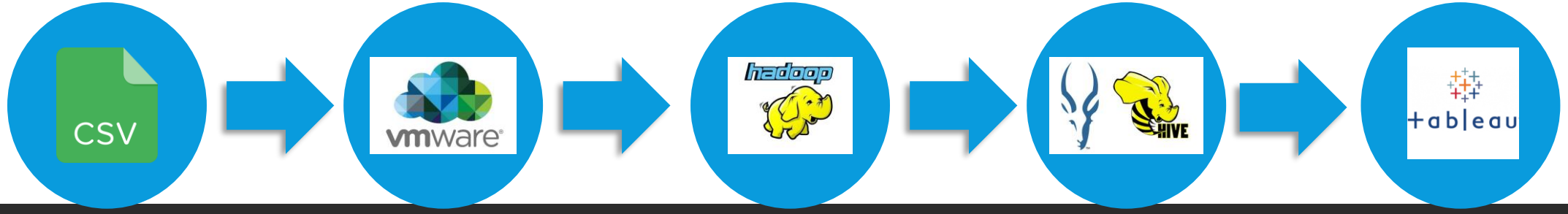
DATA-DRIVEN ANALYSIS FOR ANT TRUCK FLEET OPTIMIZATION



GROUP 4
- Apurwa Chandrakant Bidaye
- Sanjana Annu
- Ishika Singh Chandel
- Chandrika Bhagat

PROCESS FLOW



CSV - Copy the data from local system to VMware local.

VMWare – Transferred data to VMWare by employing drag and drop functionality.

Hadoop – Utilize VMware tools or Hadoop commands to transfer the data from the VMware local storage to the default location in HDFS.

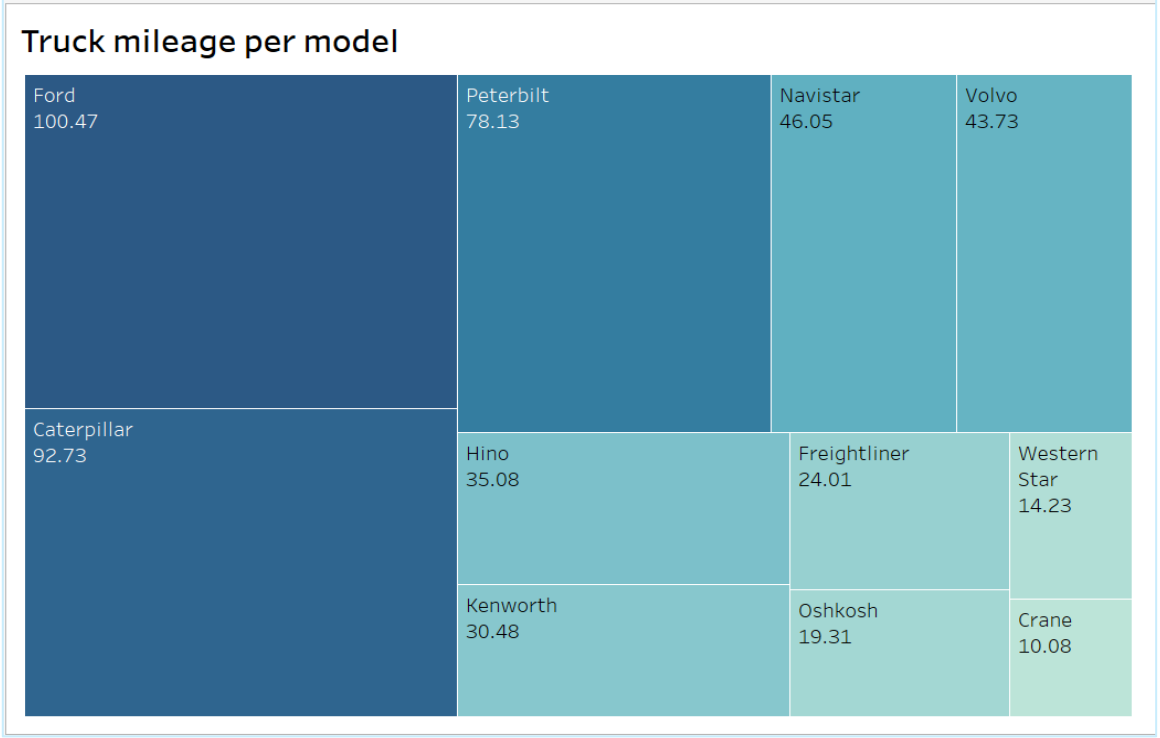
Hive/Impala – Create new tables with appropriate schemas to accommodate the data that has been imported.

Tableau – Use Tableau to connect to Cloudera Hadoop, specifying the necessary connection details.

BUSINESS OBJECTIVES

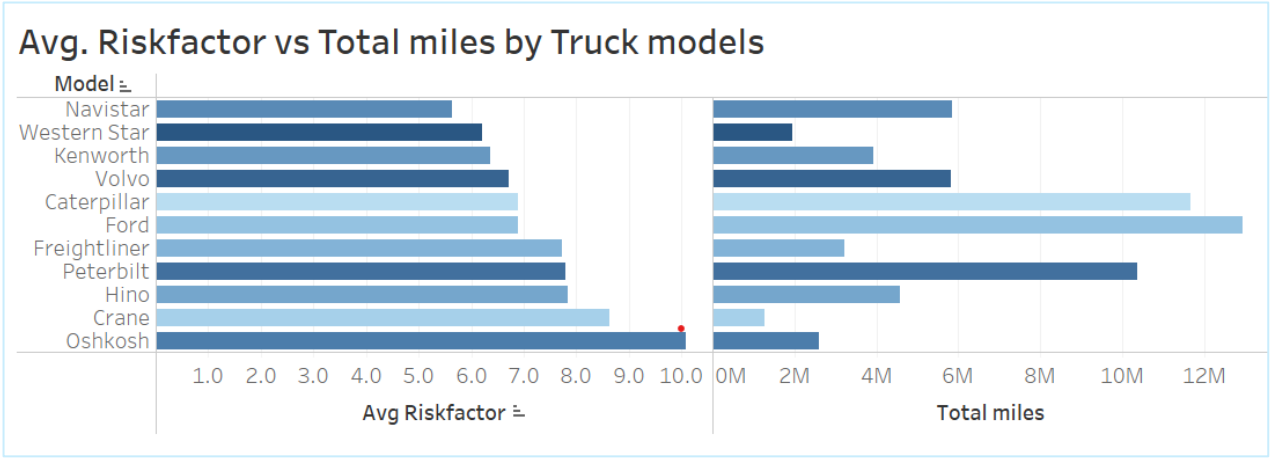
- Identify and select truck models that demonstrate the reduce operational cost by considering factors such as truck mileage, risk factors associated with events, engine performance and overall fuel consumption.
- Conducting risk analysis to identify high risk commercial truck drivers on a national scale.
- Offering insights and suggestions for truck managers to effectively identify and manage drivers with potential risk factors, enhancing overall safety measures.
- Developing risk mitigation strategies by carefully analyzing a spectrum of risk factors.

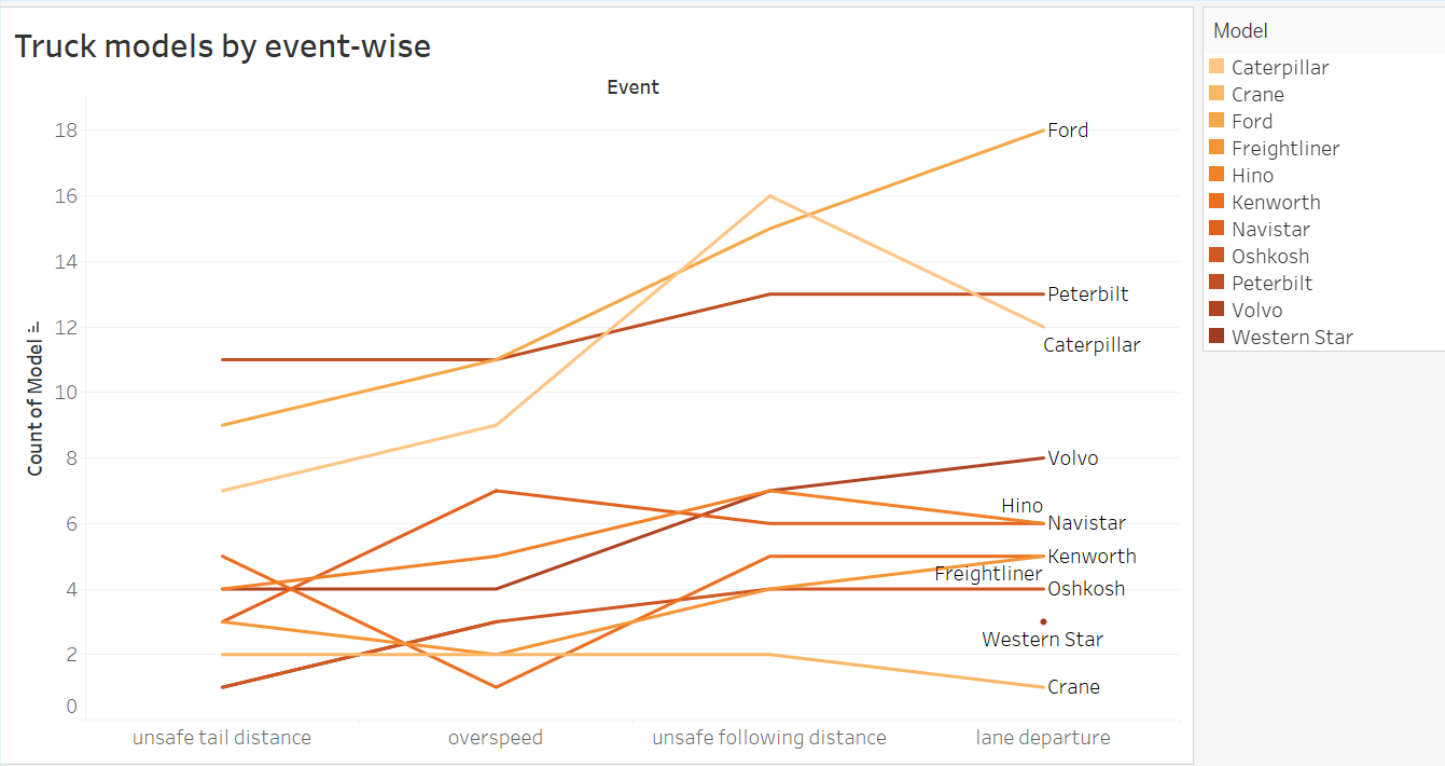
TRUCK MODEL ANALYSIS



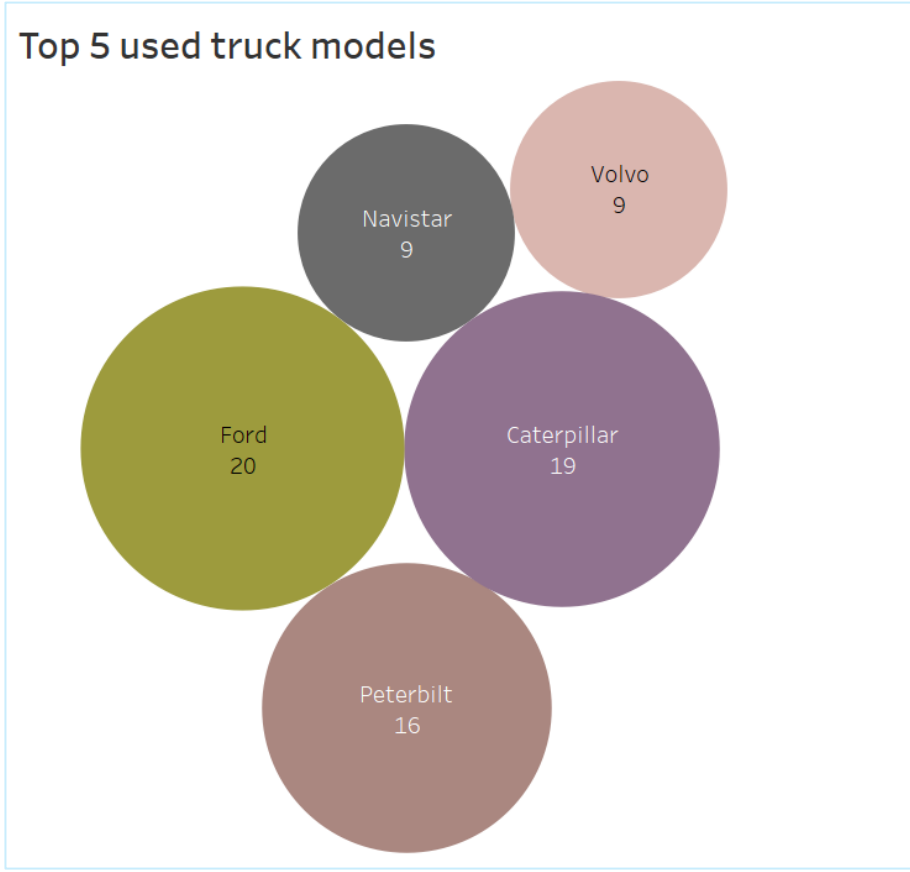
- Tree map shows different categories of the truck model where darkest color and largest rectangle identifies the highest truck mileage.
- It is observed that, organization can consider choosing truck models from top 3 such as Ford, Caterpillar and Peterbilt.

- Considering risk factor is the main concern while deciding a truck model, it is observed that Ford, caterpillar shows risk factor below 7 even though they have the highest total miles driven.



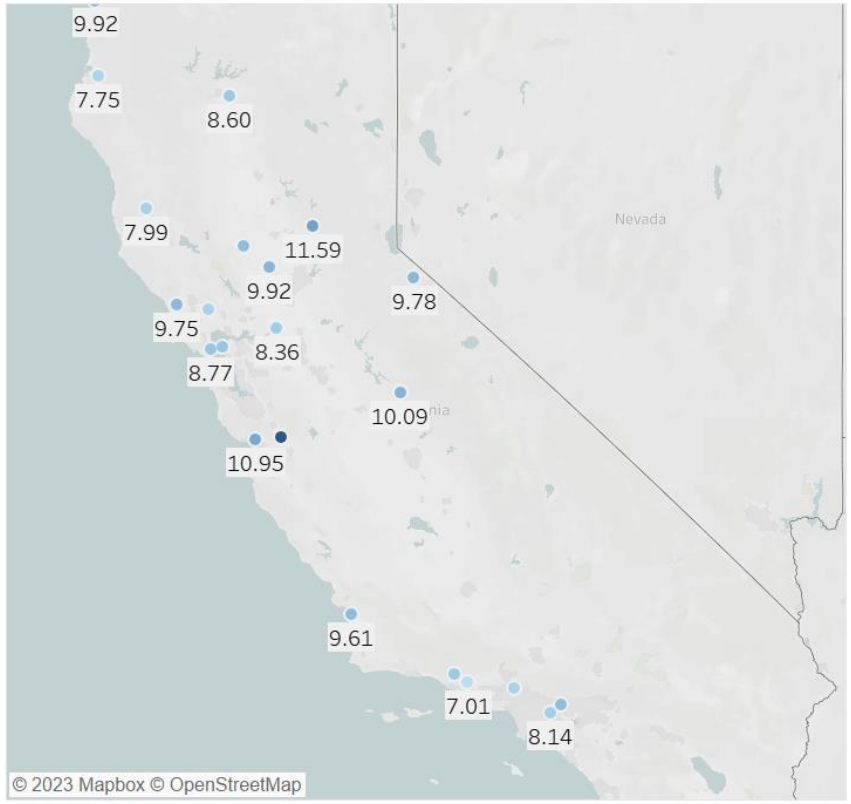


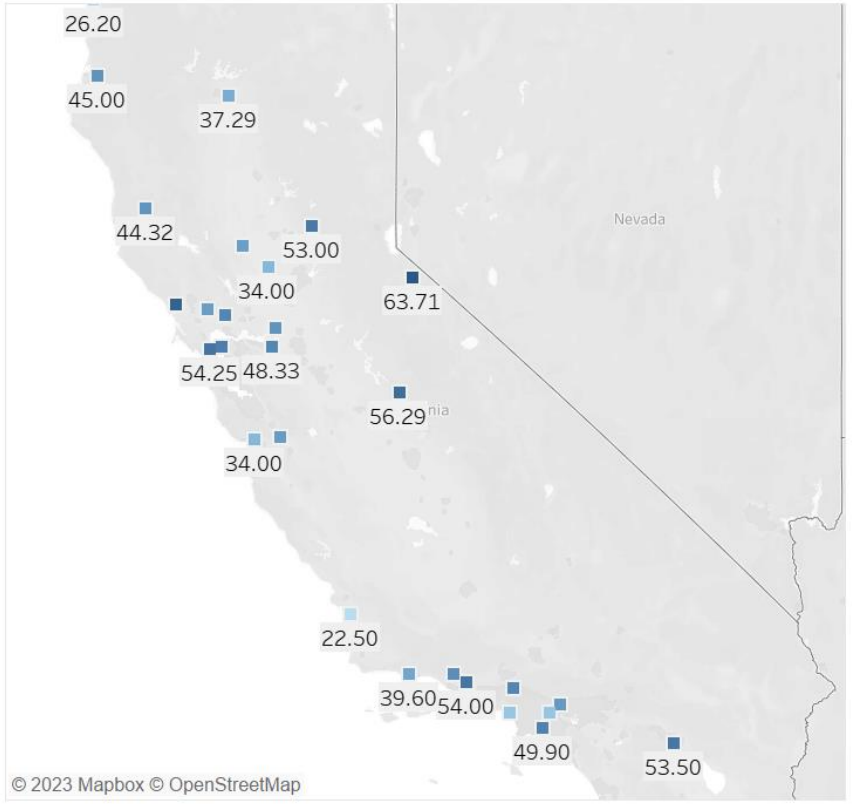
- Here, x-axis represents the events and y-axis represents the count of truck models per event.
- Ford has highest number of events in lane departure whereas Caterpillar shows most events in unsafe following distance. Also, Peterbilt is top in unsafe tail distance.



- Bubble chart visualization shows the usage frequency of each model through the size of bubbles.
- Ford, Caterpillar and Peterbilt is mostly used in the fleet management.

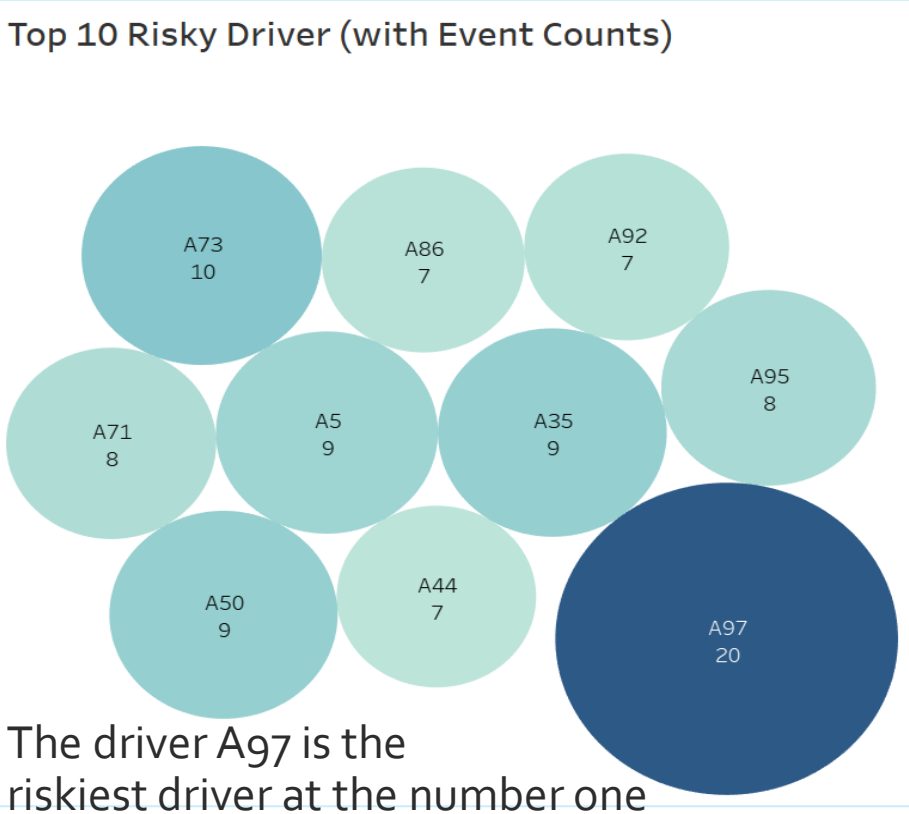
CITIES HAVE A HIGH- RISK FACTOR IN CALIFORNIA





- The cities shown are the cities with an avg risk factor greater than or equal to 7.
- The other chart shows the avg velocity of across these cities.
- The cities of Occidental, Hollister, and Oceano have the highest risk factors.
- Identifying cities with high risk factor allows the company to strategically allocate resources, optimize routes and tailor safety measures.

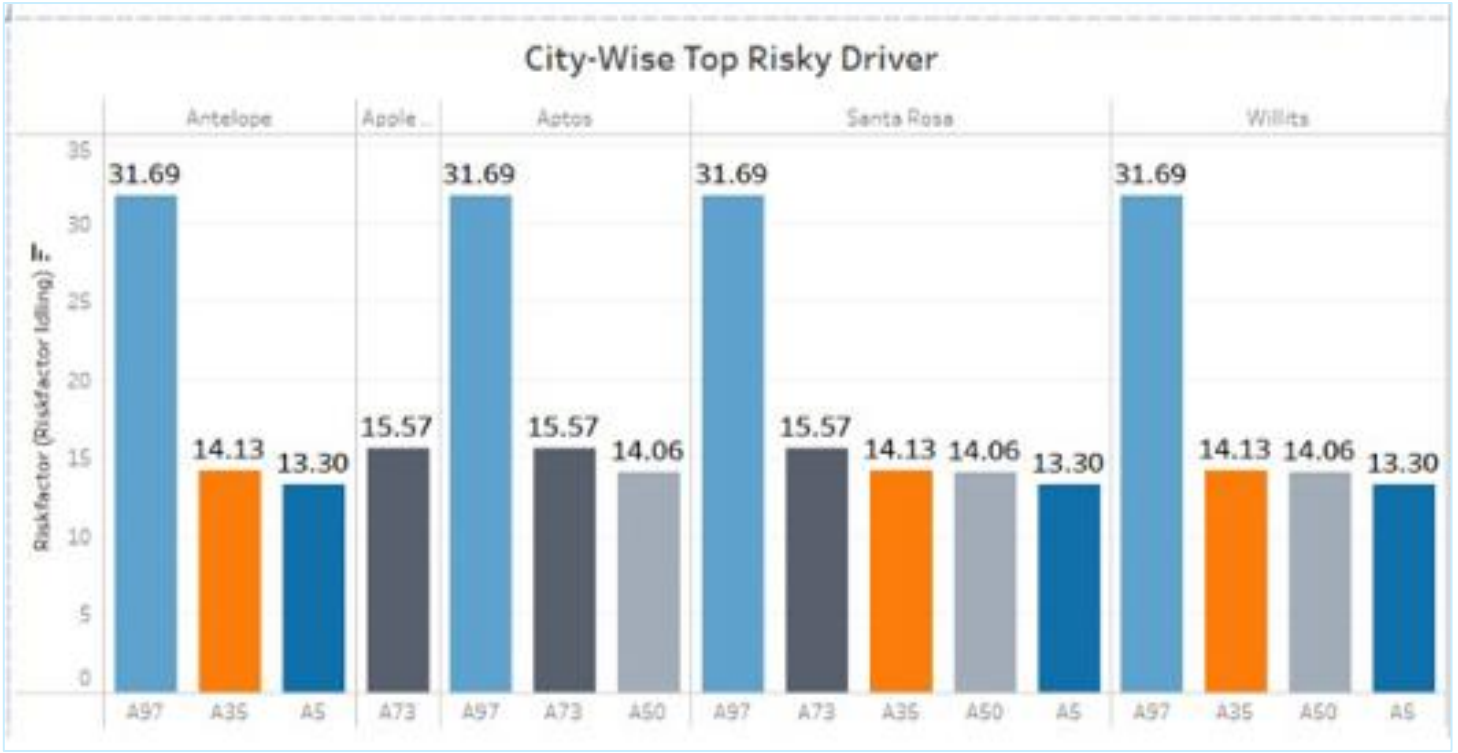
DRIVER ANALYSIS



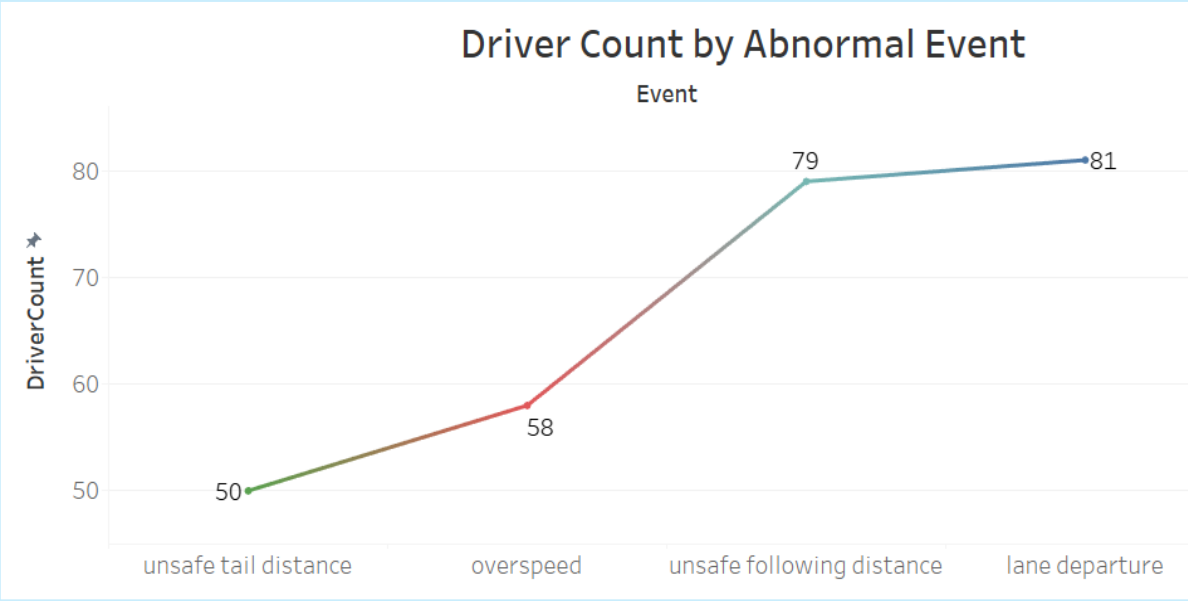
The driver A97 is the riskiest driver at the number one position, and the fleet manager should avoid this driver.

Identifying the riskiest driver allows the company to investigate the driver further and take appropriate actions.

- This chart shows the top risky driver across all the cities of California
- This insight can help the company decide on allocating resources more efficiently, directing efforts toward those drivers who pose the greatest safety concerns.

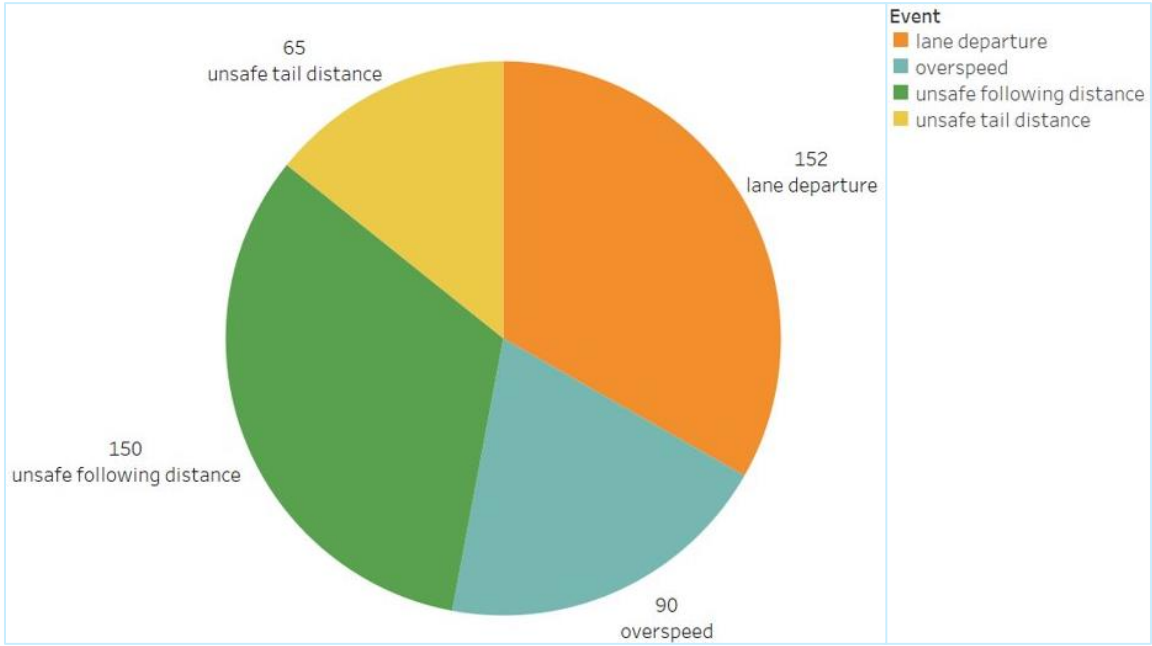


EVENT ANALYSIS



- The line graph above can help identify the most prevalent safety issues among drivers.
- Analyzing event data enables the company to pinpoint drivers who frequently engage in risky behaviors.
- This can be used to implement corrective measures, provide individualized coaching, and encourage safer driving habits..

- The pie chart shows the scale of each event with the other events, with the total of each event shown on the graph.
- Lane departures and unsafe following distances are the most common.
- Identifying the most common events can guide targeted interventions and driver training to enhance risk management across the fleet.



CONCLUSION

- Ford, Caterpillar and Peterbilt truck model are used most often are the safest model that can be used considering the risk factor and total miles covered by them.
- Based on our study Driver id with A97 is exceedingly risky in every risk factor category.
- Driver id with A73 is most likely to cause accidents following A97.
- Almost the same driver ids appeared frequently across the various risk factors considered, suggesting that these drivers are indeed associated with higher levels of risk.
- Lane departure is the highest risk causing event, thus using features like lane assist will help significant mitigation of risk.

THANK YOU