ANT Truck Data Analysis

GROUP 7

- Chia-Yao Chiu
- Chitvan Bhadoria
- Shilpa Nidhi Kirubanidhi
- Shubhrata Gupta
- Tanmayee Bhavsar
- Shivani Datar
- Sai Kiran Jammula

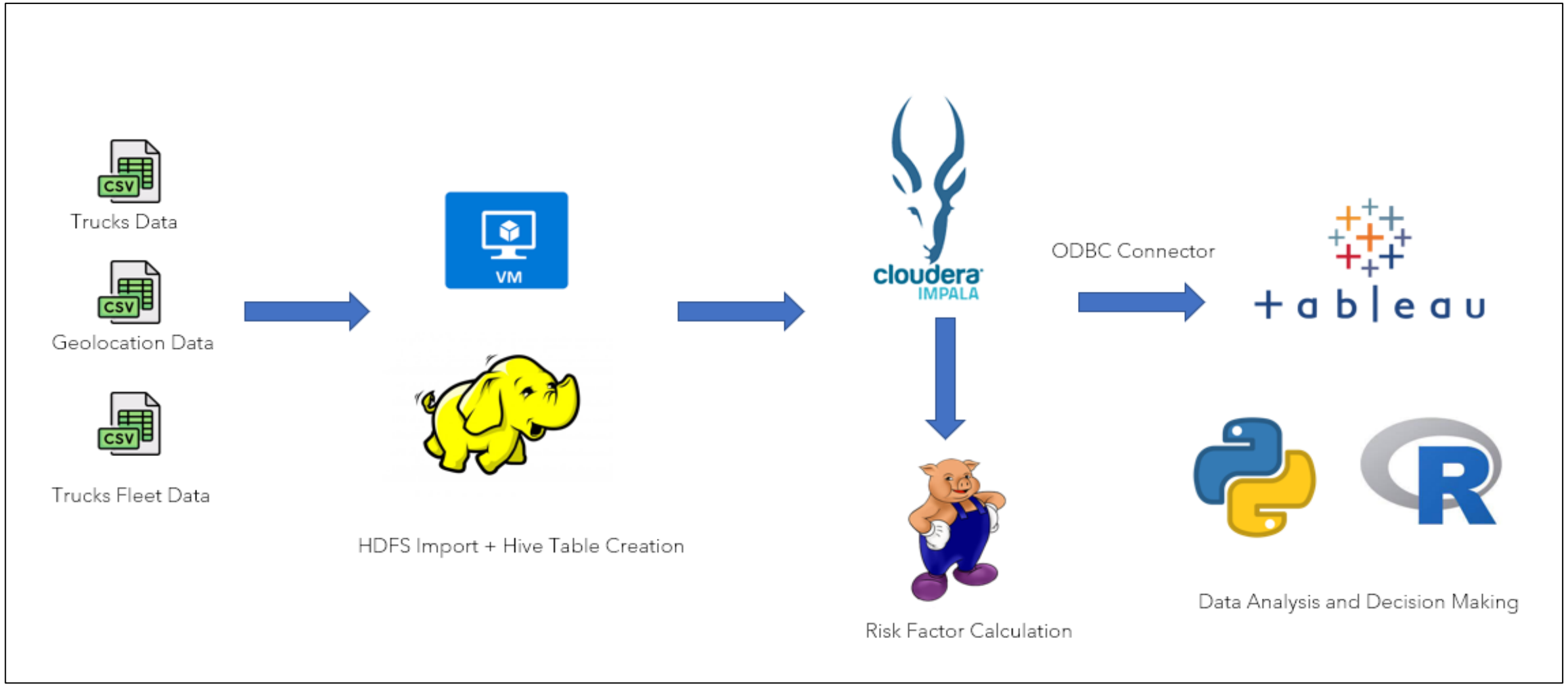
Business Objective

- Our goal is to identify high-risk commercial truck drivers and develop solutions to mitigate the issue, by analyzing key factors such as driver violations, risk factors, and other relevant data.

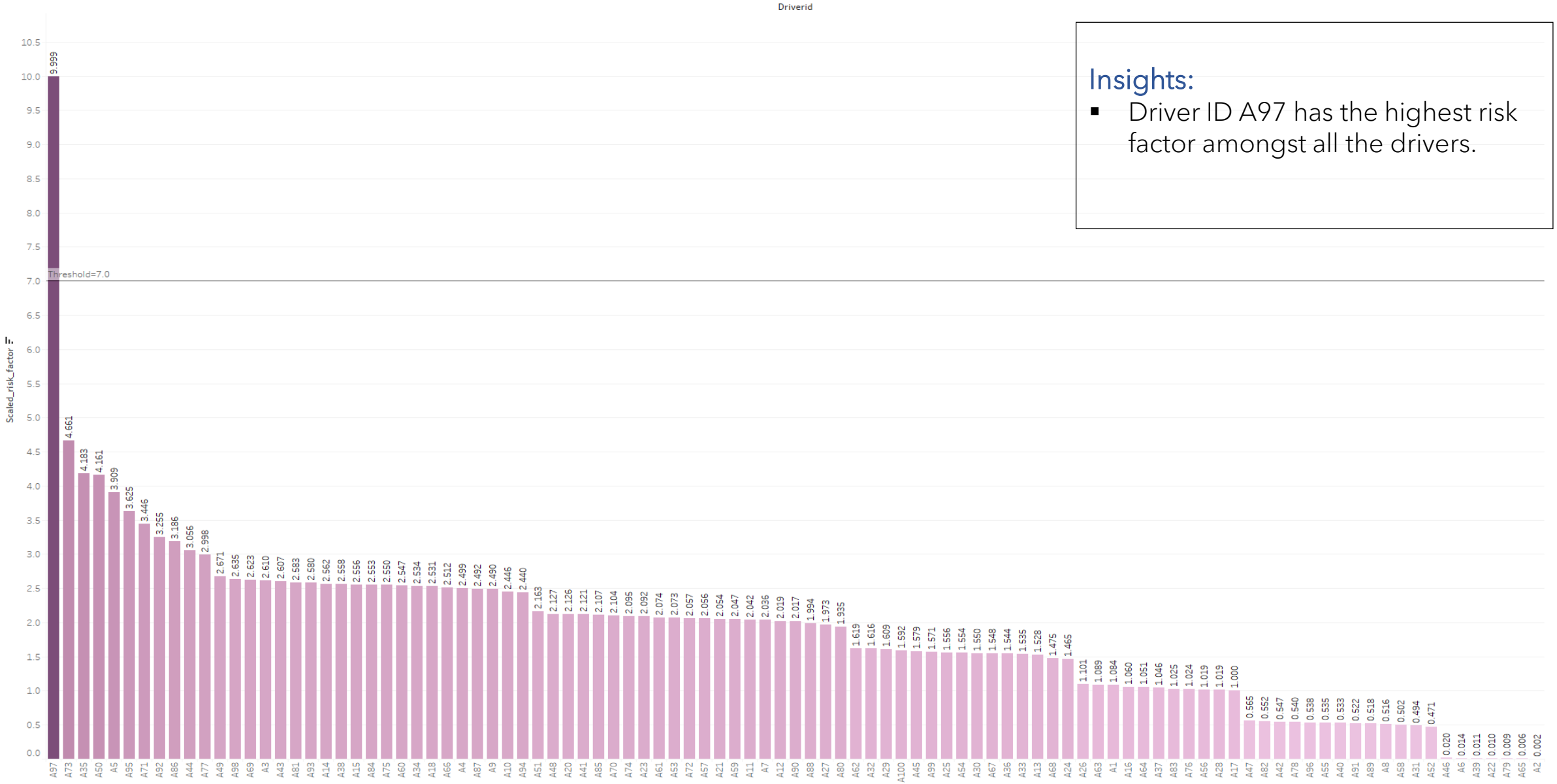
- By analyzing data on geolocation, truck model, and driver behavior, we aim to identify key performance indicators such as average speeds, mileage, risk factors, and gas consumption, to inform decision-making and improve outcomes.

- Through data analysis, we seek to minimize risks associated with each driver and reduce the number of accidents caused by large commercial trucks, with the ultimate goal of decreasing injuries and fatalities in California.

Data to Decision Model



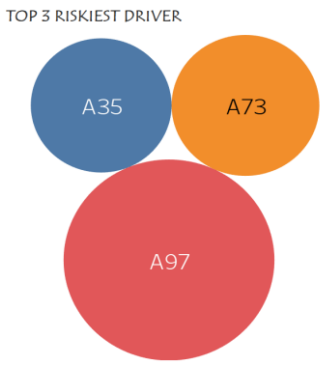
Driver Vs Scaled Risk Factor

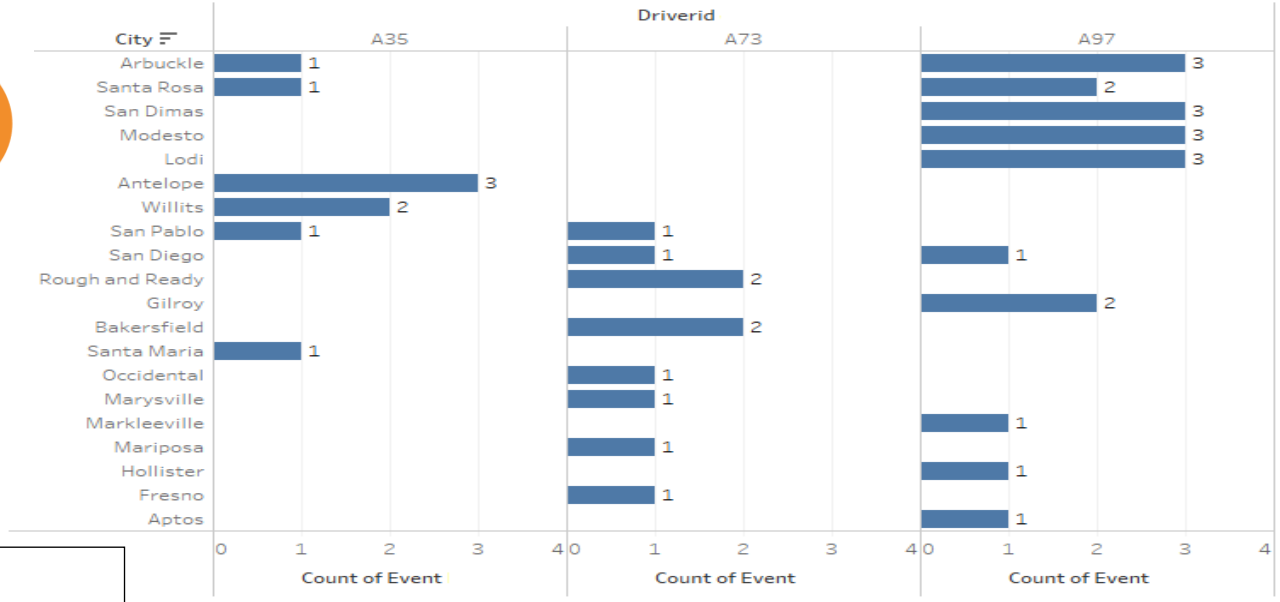


Insights:

- Driver ID A97 has the highest risk factor amongst all the drivers.

Top Risky Drivers



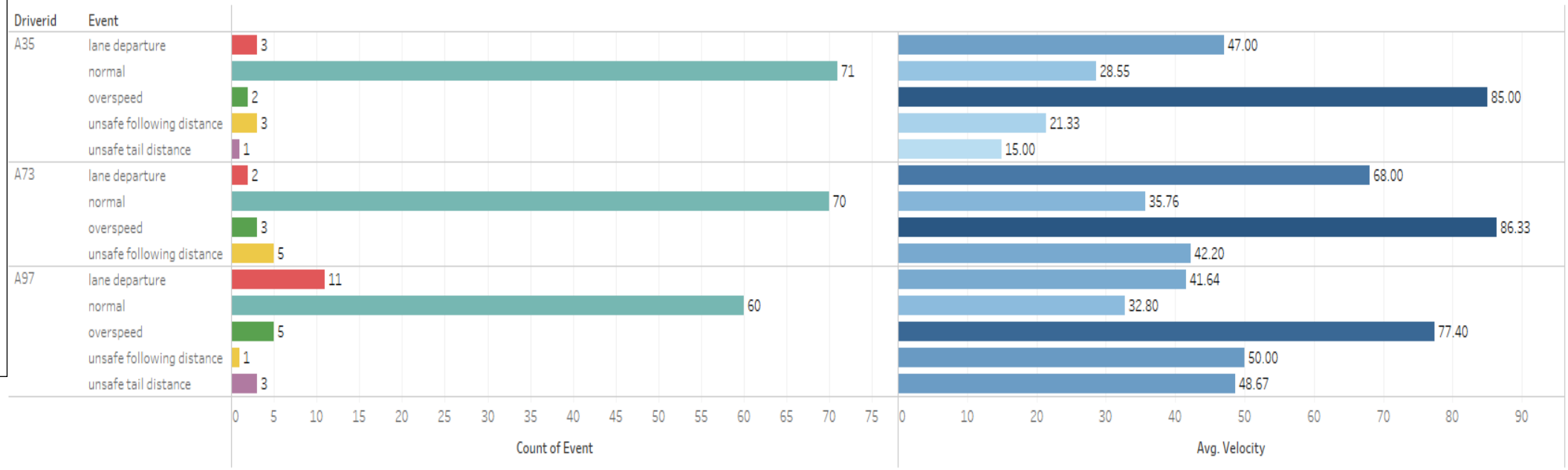


Insights:

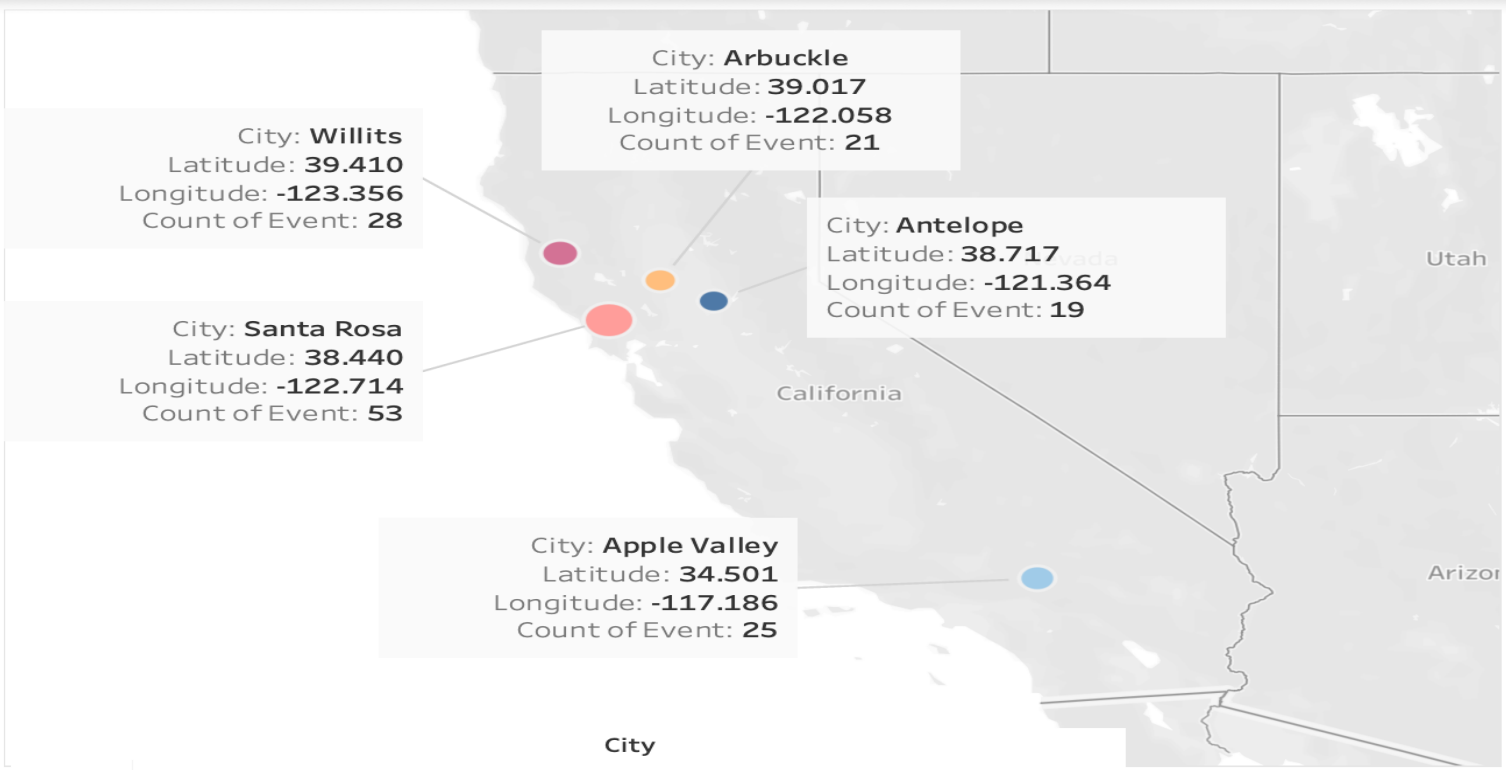
- Arbuckle, San Dimas, Modesto, and Lodi are the riskiest cities for driver ID A97 based on past incidents.
- Driver A97 has had fewer normal events and more incidents related to other factors, contributing to its high-risk level.

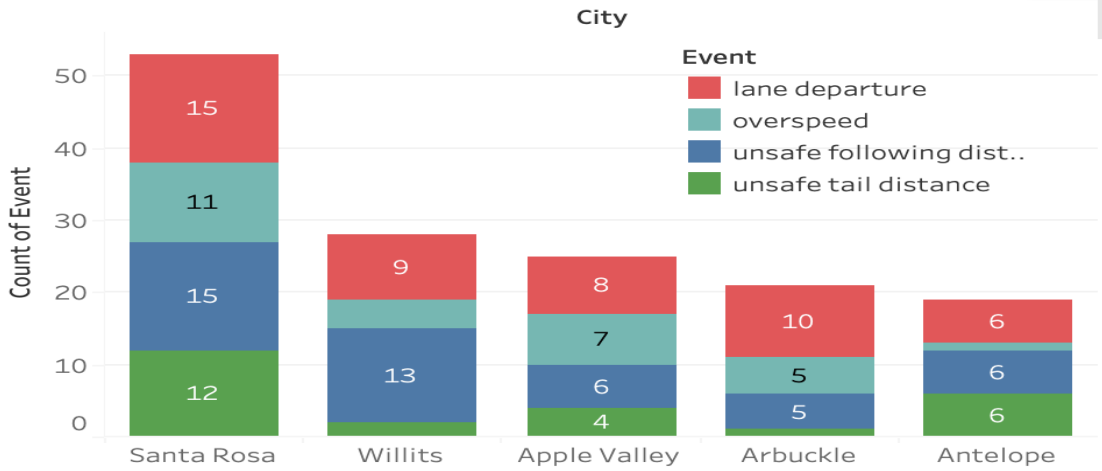
Recommendations:

- Install lane departure warning systems and collision avoidance systems on trucks to alert drivers when they are drifting out of their lane or following too closely.
- Install speed monitoring devices on trucks to track driver speed and ensure that they are adhering to speed limits.
- Enforce speeding violations by ticketing drivers and imposing fines.
- Use telematics technology to alert drivers when they are approaching the speed limit and remind them to slow down.

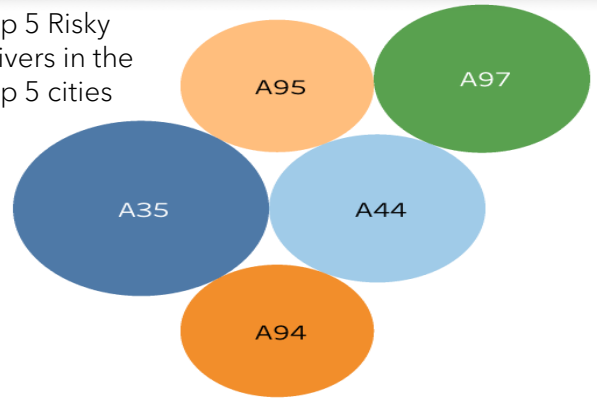


Geographical Analysis





Top 5 Risky Drivers in the Top 5 cities



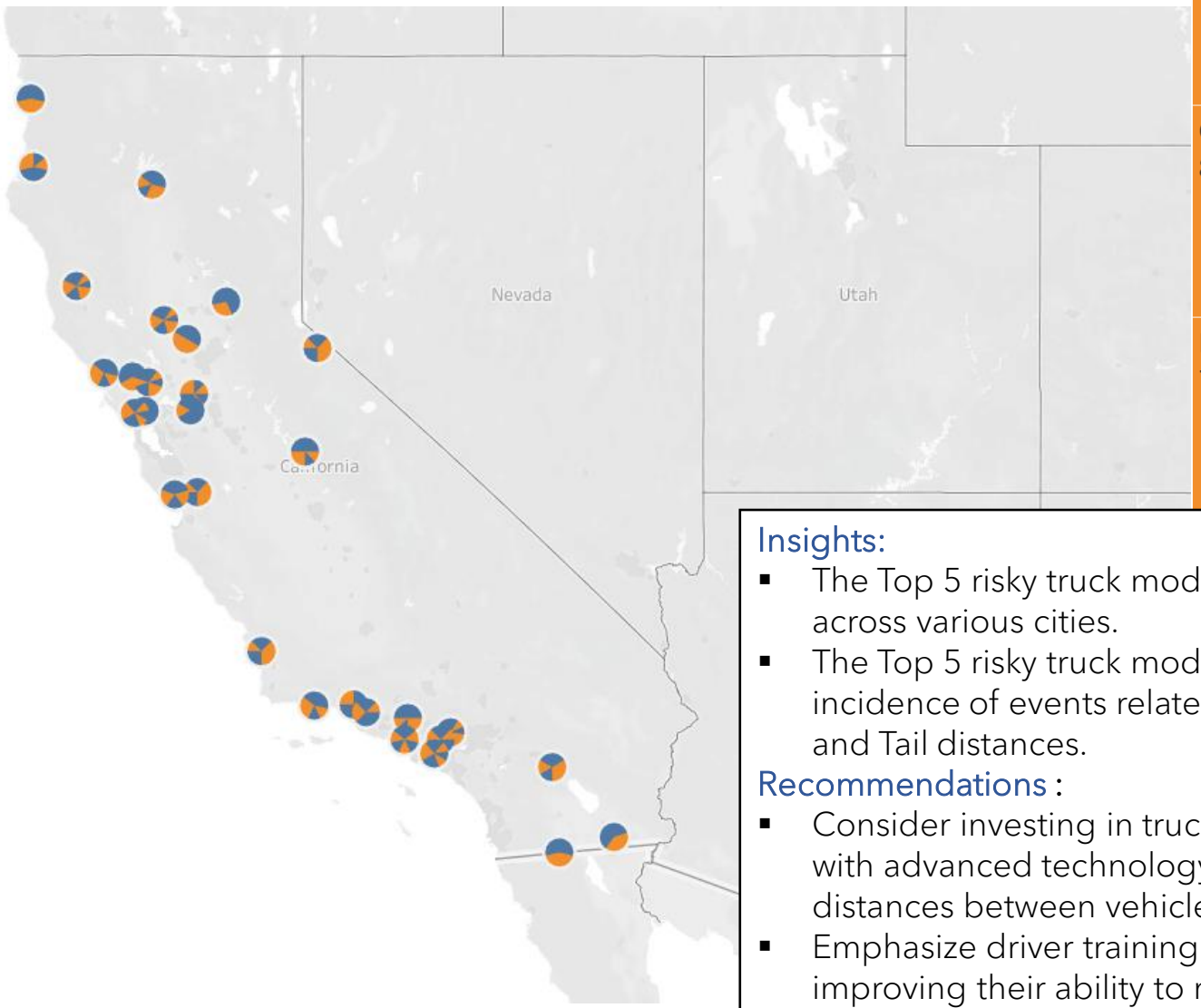
Insights:

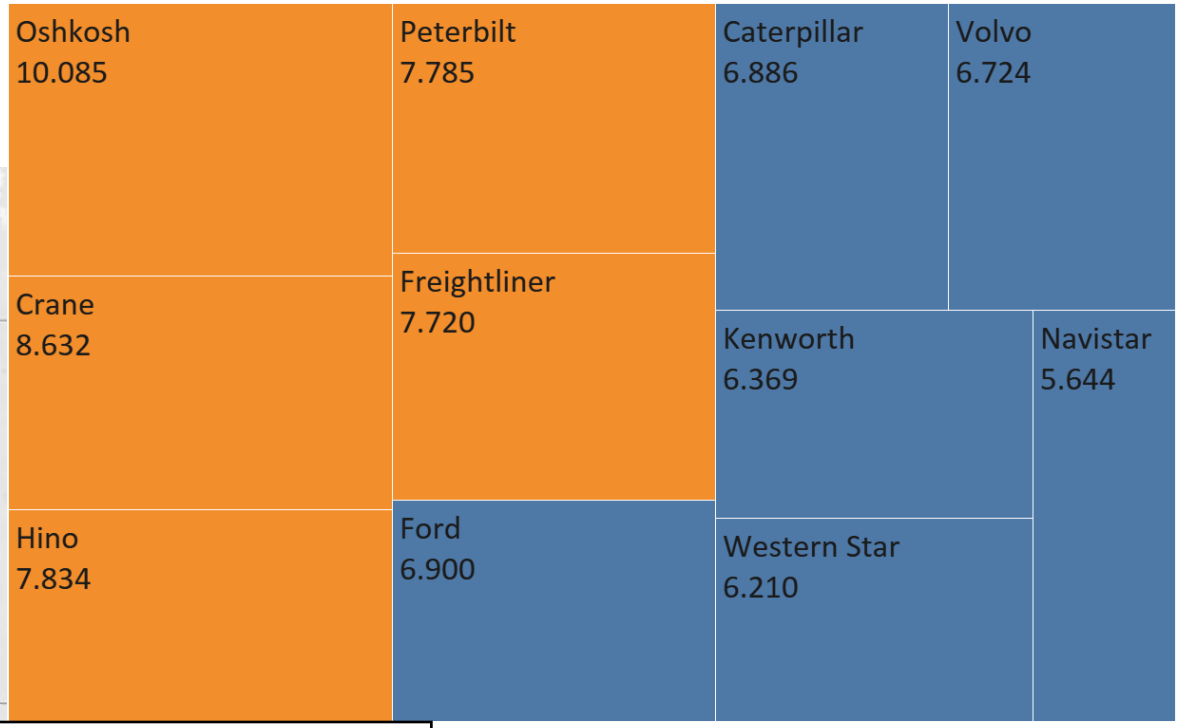
- The top 4 cities with most events appear clustered with highest events occurring in Santa Rosa.
- The most occurring events in these cities is lane departure and unsafe following distance.
- The most prominent drivers with high risk factors are indicated.

Recommendations:

- We can target the drivers with high risk factor in these cities for better training.
- The training for new recruits and existing risky drivers could be more focused on avoiding lane departure and unsafe following distance issues

Risky Truck Models



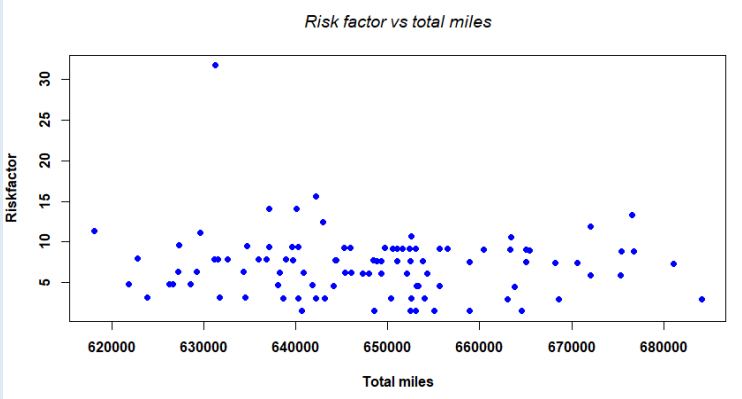


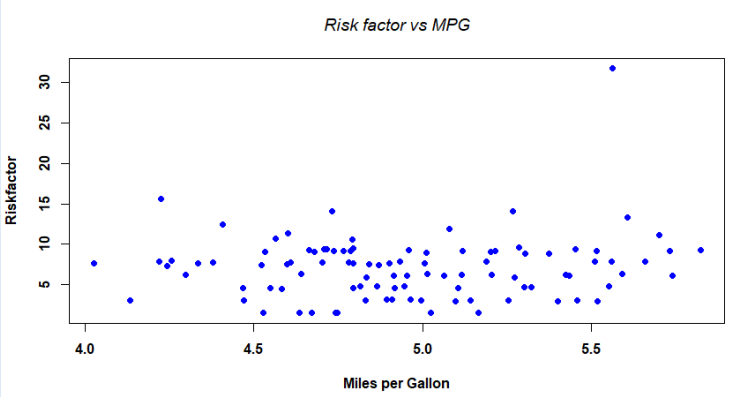
Insights:

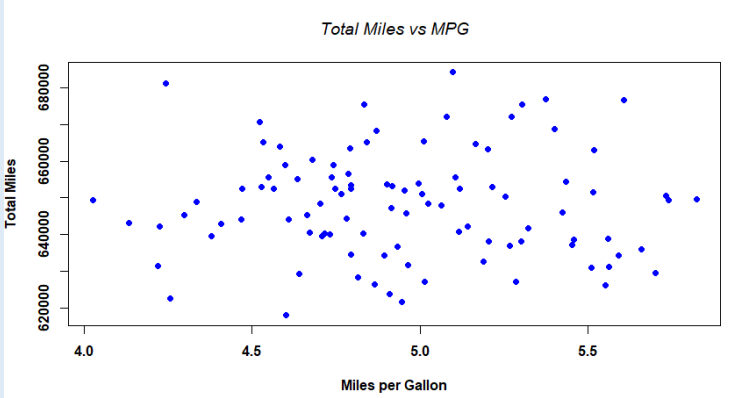
- The Top 5 risky truck models are used equally across various cities.
- The Top 5 risky truck models have a higher incidence of events related to Unsafe Following and Tail distances.

Recommendations :

- Consider investing in truck models equipped with advanced technology for detecting distances between vehicles.
- Emphasize driver training that focus on improving their ability to maintain safe following and tail distances.







Exploratory Analysis

- Interpreting correlation between data
  - ✓ -1 implies strongly negative correlation
  - ✓ 0 implies no correlation
  - ✓ +1 implies strongly positive correlation
- Results shows there is no much correlation between any 2 fields
- This shows that we need:
  - ✓ More Historical data
  - ✓ More precise data

	MeanDecreaseGini		
normal	10.95408558	may12_gas	0.69885234
unsafefollowingdistance	1.93712934	may12_miles	0.66566295
lanedeparture_vel	1.61514747	jun11_miles	0.64935741
lanedeparture	1.18147085	overspeed_vel	0.64614040
unsafetaildistance_vel	1.01335473	jul11_gas	0.56247871
unsafetaildistance	0.86211101	normal_vel	0.55079296

Random Forest Classifier

Geo Location Table

	truckid	driverid	event	city	velocity
truckid	1.000000	1.000000	0.016464	0.012769	0.003179
driverid	1.000000	1.000000	0.016464	0.012769	0.003179
event	0.016464	0.016464	1.000000	-0.004142	0.029318
city	0.012769	0.012769	-0.004142	1.000000	0.002736
velocity	0.003179	0.003179	0.029318	0.002736	1.000000

Trucks_mg table

	model	Miles	Gas
model	1.000000	0.082476	0.094211
Miles	0.082476	1.000000	0.874419
Gas	0.094211	0.874419	1.000000

Risk Factor Table

	driverid	events	totmiles	riskfactor
driverid	1.000000	0.188600	-0.005567	0.189634
events	0.188600	1.000000	-0.050809	0.998911
totmiles	-0.005567	-0.050809	1.000000	-0.092068
riskfactor	0.189634	0.998911	-0.092068	1.000000

Conclusion

- Driver A97 is the riskiest of all and we recommend using advanced speed monitoring equipment and lane departure alerting systems.
- Raising awareness about safe driving practices and violations could help mitigate risk for drivers who frequently drive on routes in northwest California.
- To reduce risk for drivers, the company should consider providing training specifically for those with the highest risk factors and increasing awareness about safe driving practices and violations for those who regularly drive in northwest California.
- Invest in trucks equipped with advanced distance-detecting technology to enhance their proficiency in maintaining safe following and tail distances.