



Visualizing Factors Associated with Risky Driving

Group 1: Katie Dyo, Regina George Suresh, Connor Lowe, Varmi Sanghani, Palak Sharma

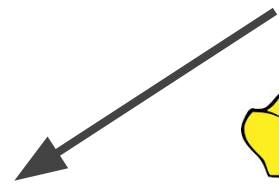
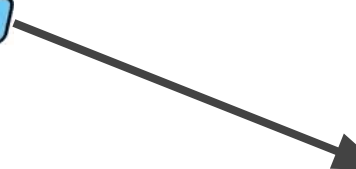


Problem Statements and Objectives

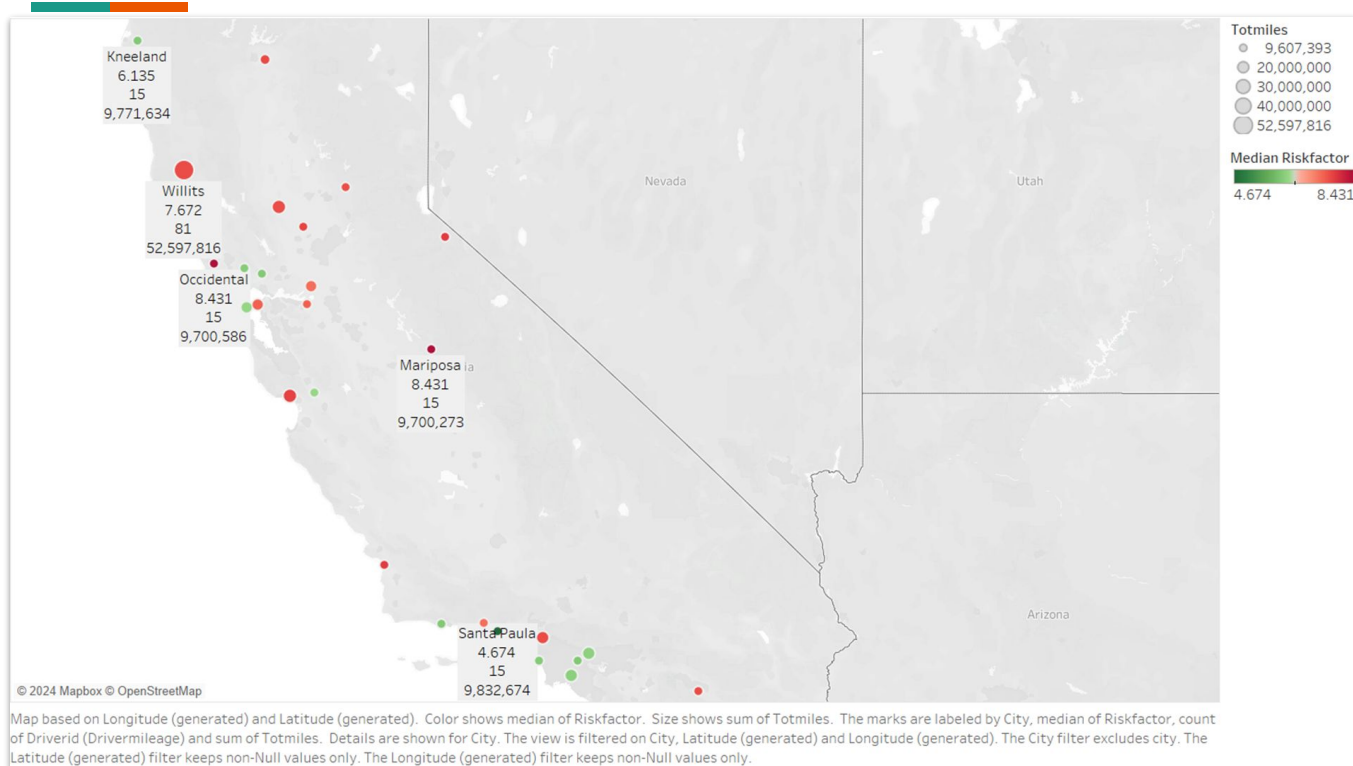
What are some patterns that can be identified in association with high driver risk? Are these patterns indicative of risk and can they be mitigated?

- 1.) Acquire driver data
- 2.) Leverage HDFS to move data in HIVE server that can connect to Tableau
- 3.) Generate additional tables before doing final Tableau connection
- 4.) Define a set of questions to guide us through our analysis
- 5.) Leverage Tableau and it's visualizations to experiment with the data
- 6.) Create visualizations that communicate an answer to the generated guiding questions
- 7.) Compile all final visuals into a single, easy to use dashboard

Workflow



Q1.) What cities are the riskiest/least risky?

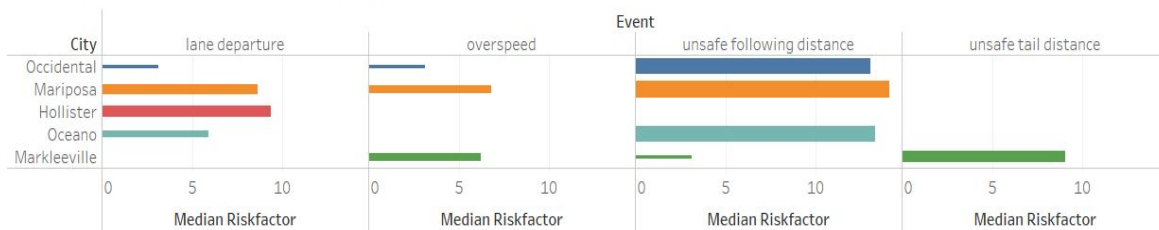


Highlighted Cities are some of the riskiest and least risky in terms of median Risk Factor

Areas with a smaller median risk factor tend to be more grouped together (as can be seen in Southern California)

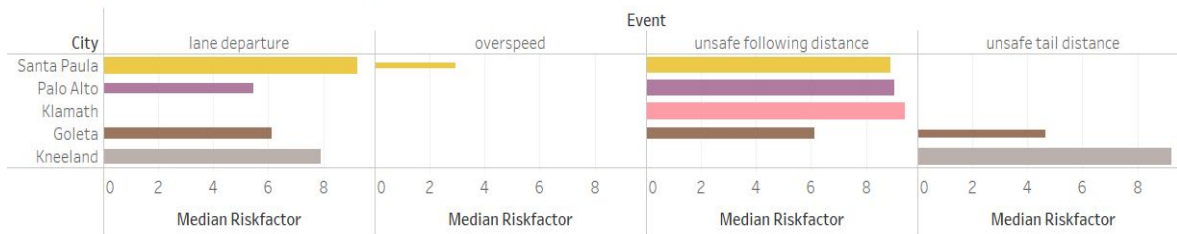
Q2.) What are the event distributions of the riskiest cities and least risky cities?

Event Distribution in the riskiest cities



- In the riskiest cities, such as Occidental and Mariposa, **unsafe following distance** is the most prevalent event contributing to higher risk factors. Other events, including **lane departure** and **unsafe tail distance**, also play significant roles, highlighting a pattern of dangerous driving behavior in these locations.

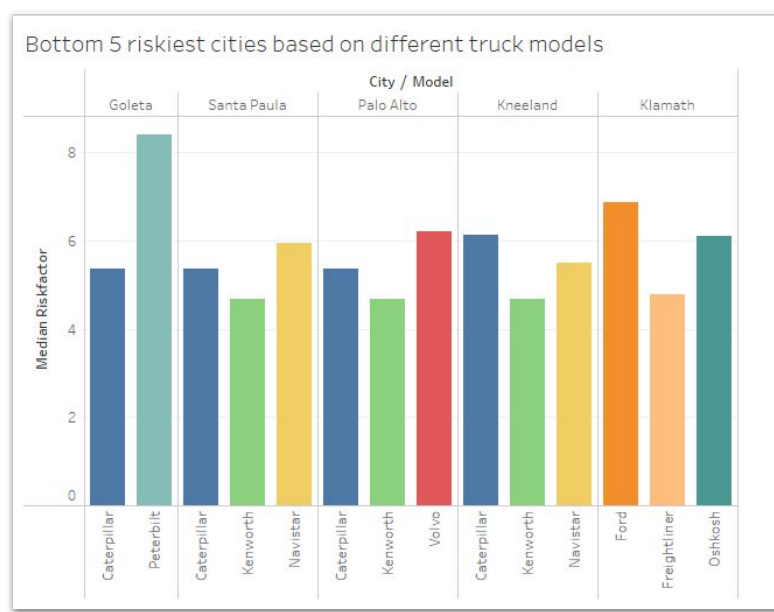
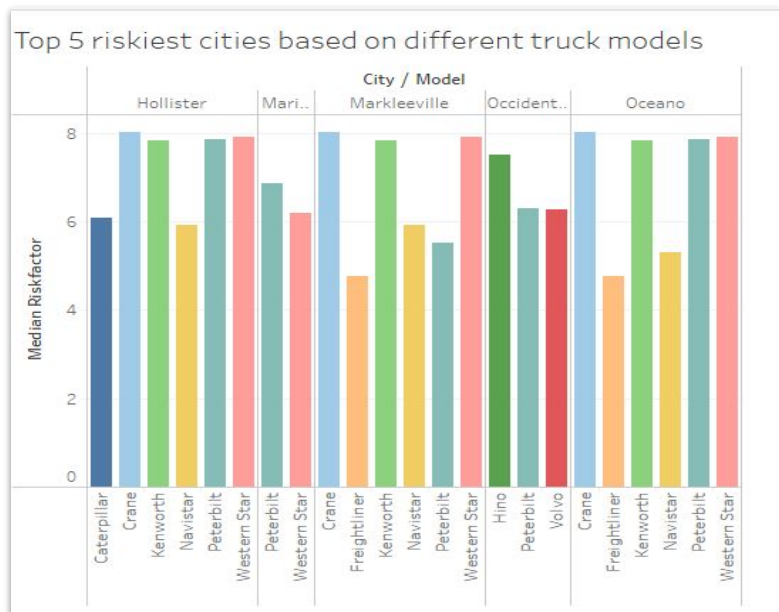
Event distribution in least risky cities



- In the least risky cities, such as Santa Paula and Palo Alto, both **unsafe following distance** and **lane departure** are notable contributors to risk factors. However, both are observed at much lower levels compared to the riskiest cities, reflecting relatively safer driving behavior in these areas.

Q3.) Is there a correlation between risky cities and trucks models?

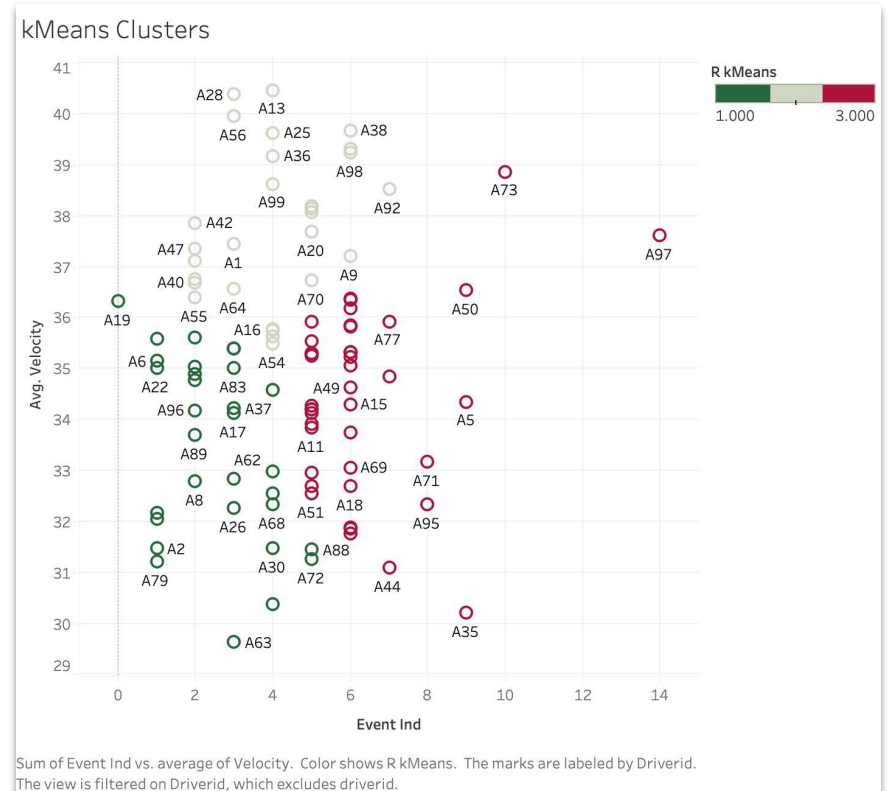
The key purpose of this visualization is to identify the highest risk combinations of cities and truck models, based on some underlying "risk factor" metric.



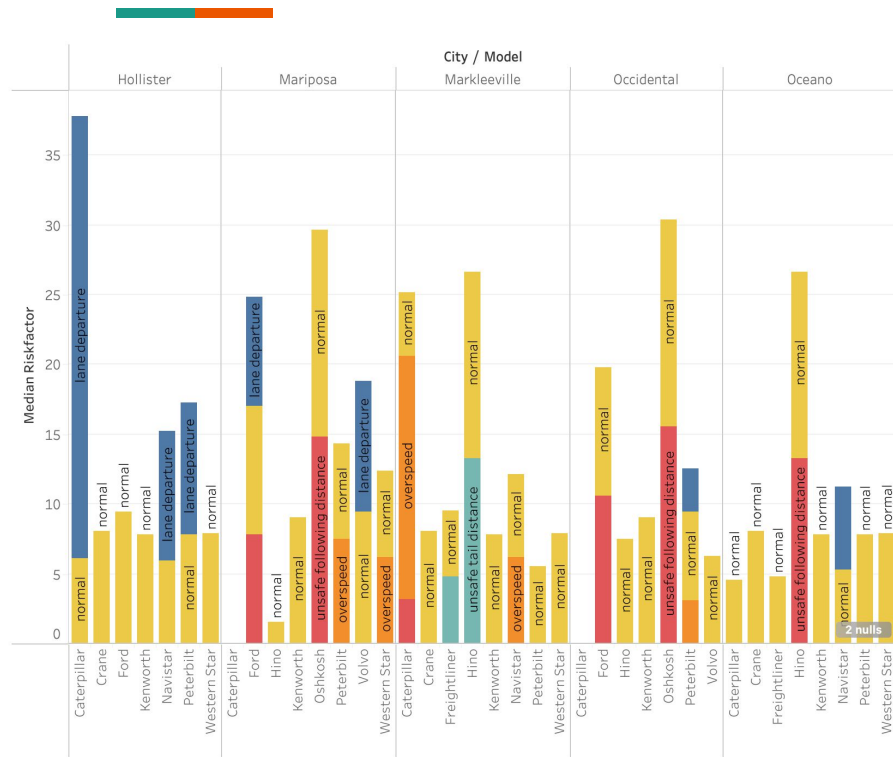
Q4.) Is average velocity a good predictor of riskiness?

Cluster 1 are the safest drivers

Cluster 3 are the riskiest drivers



Q5.) What event types are associated with risky cities and trucks?



Model Performance:

Caterpillar and Kenworth models often appear in cities with higher median risk factors, suggesting a potential link between model type and risk exposure.

City-Specific Risk Factors:

Cities like Hollister and Occidental show higher median risk factors across various truck models, indicating potentially more challenging driving conditions or stricter enforcement of safety regulations.

Conclusion



There was a (weak) correlation between the number of drivers and the median risk factor, and Northern California had more cities with a higher median risk factor.

Unsafe following distance and lane departure are the 2 events that have the biggest influence in identifying most and least risky cities.

The drivers of Crane, Peterbilt and Western Star trucks appear to be some of the riskiest in the high risk cities, while the drivers of Caterpillar and Kenworth trucks are some of the least risky drivers in the low risk cities.

Average velocity is not the best indicator of riskiness/an event occurring.

Of two of the riskiest cities, Hollister had a very low variety of event types (norm/lane dept), while Mariposa had a greater variety of event types (norm/lane dept/overspeed/unsafe dist)