

GROUP 9

BIG DATA – BUAN 6346.002

PROFESSOR GASAN ELKHODARI

Meet our Team:

Thriksha Giriraju

Ashwini Keshavamurthy

Aditya Patil

Nimitha Siddaraju

Rahul Talele

PROJECT AT A GLANCE

OBJECTIVE

Analyze truck geolocation data using Cloudera to identify high-risk driving and improve U.S. fleet safety.

SCOPE

Process & visualize large-scale fleet data.

Build a structured geolocation database.

Enable data-driven risk reduction.



Big data analytics empowers proactive risk management, enhancing safety and performance across trucking fleets.

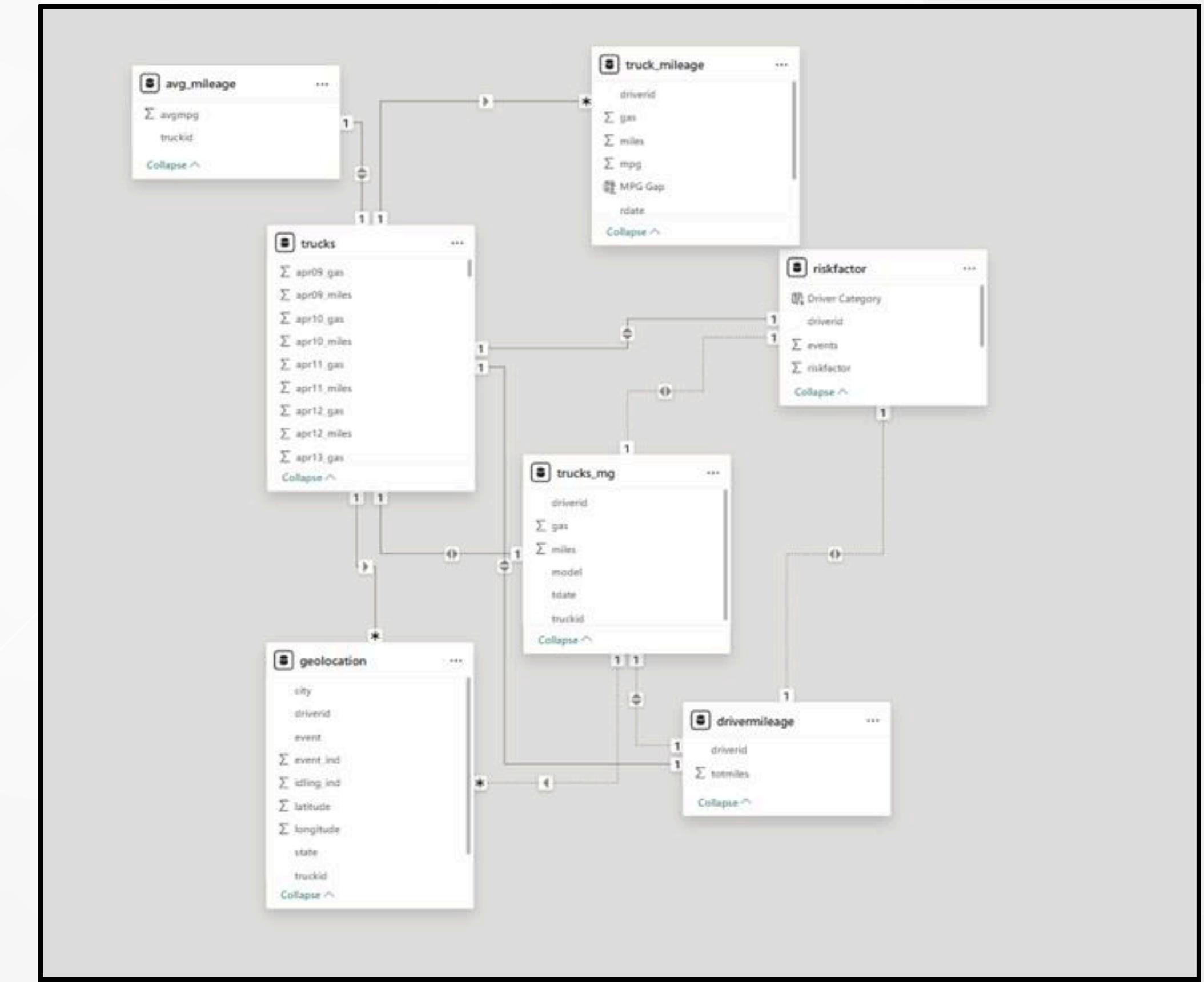
DATA MODEL: TRUCK PERFORMANCE & RISK ANALYSIS

trucks & trucks_mpg: Core performance data (mileage, MPG, model).

drivermileage & riskfactor: Driver behavior, risk scores.

avg_mileage & truck_mileage: Benchmarking vs actual MPG.

geolocation: Event locations (lat/long).



QUESTION - 1

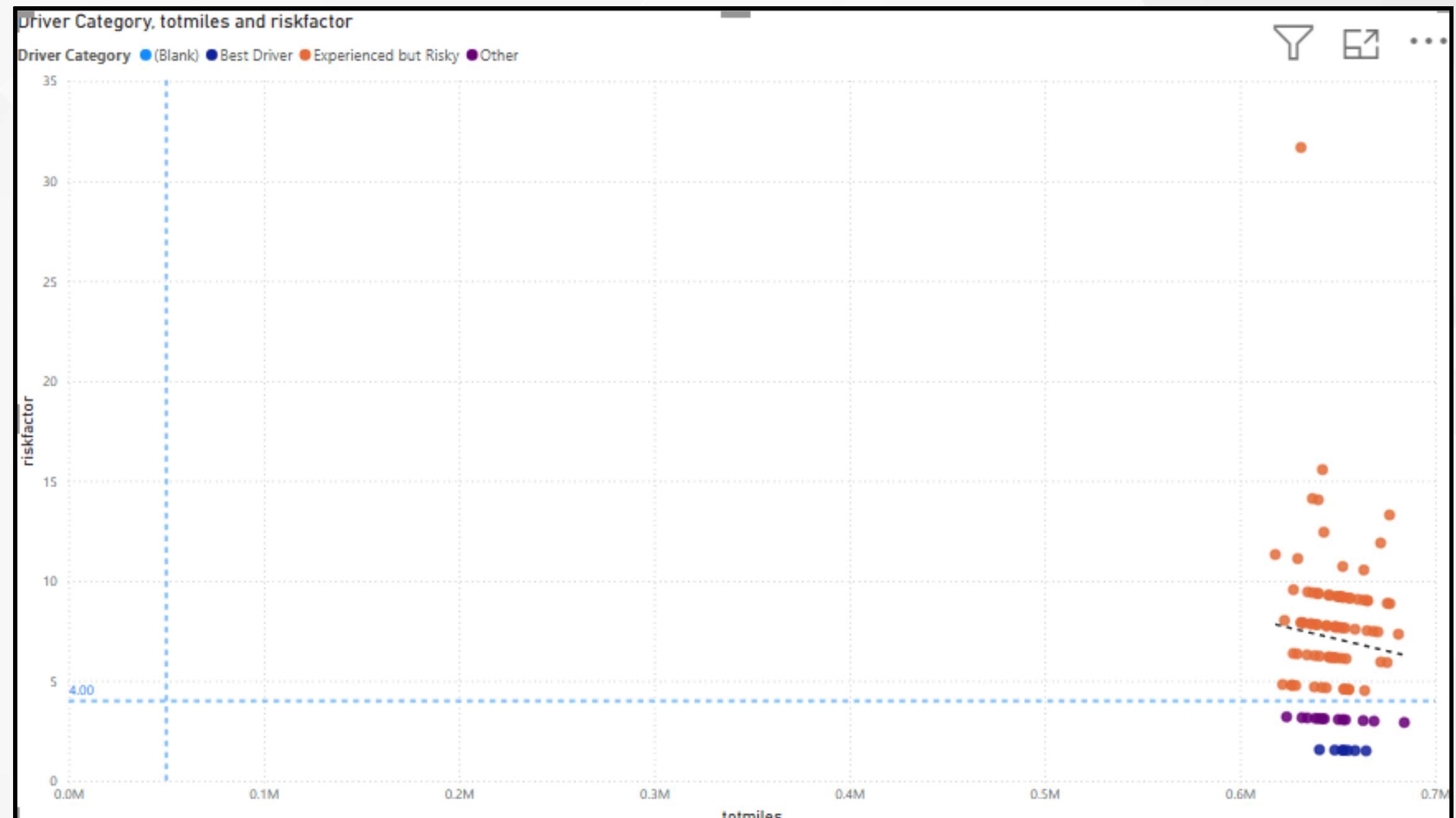
Which driver categories demonstrate high mileage with low risk, and which require closer safety monitoring?



DRIVER PERFORMANCE: RISK VS. TOTAL MILES

X-axis:
Total Miles Driven

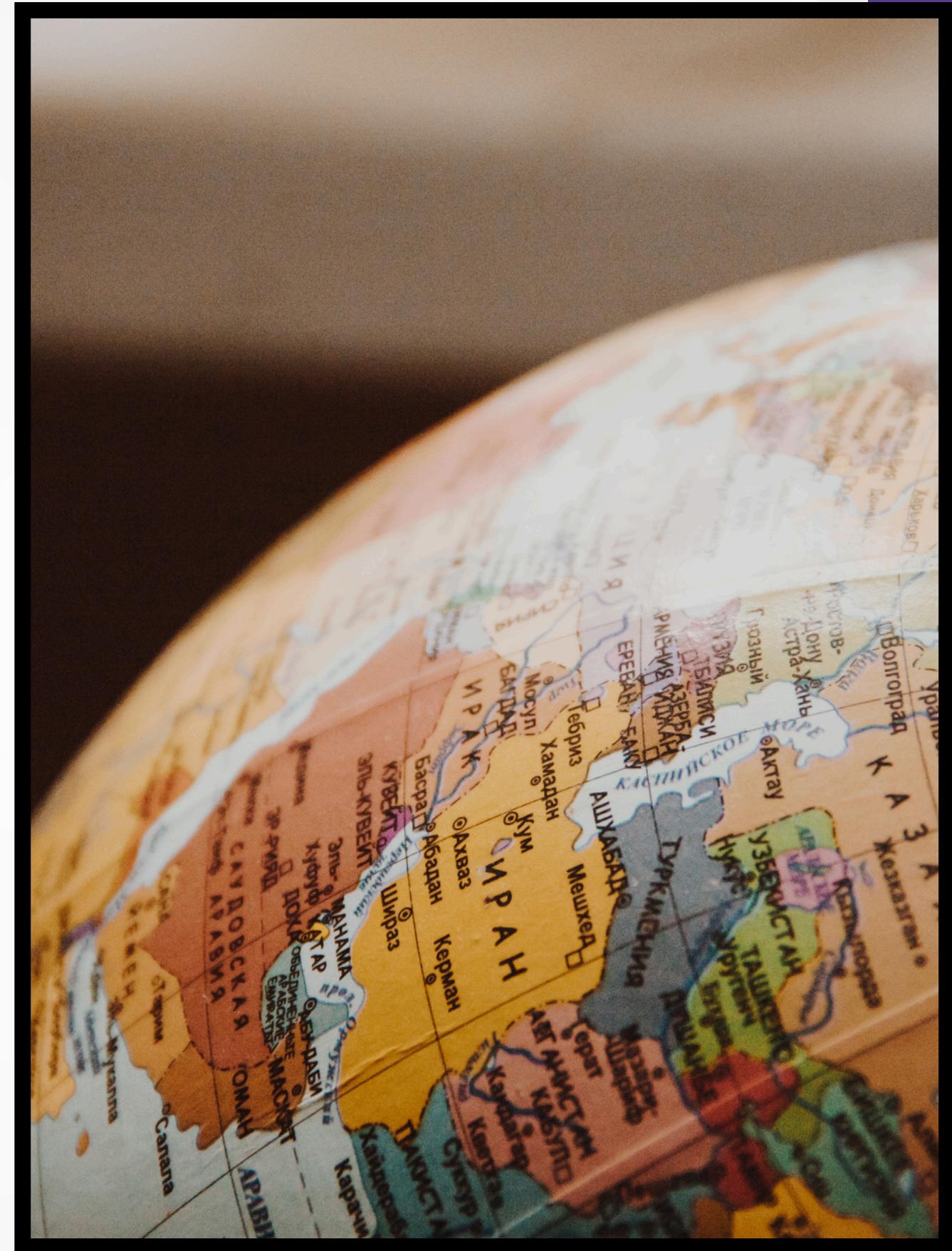
Y-axis:
Risk Factor



High-risk drivers (orange) cluster at high mileage with elevated risk. Best drivers (blue) maintain low risk even at high mileage.

QUESTION - 2

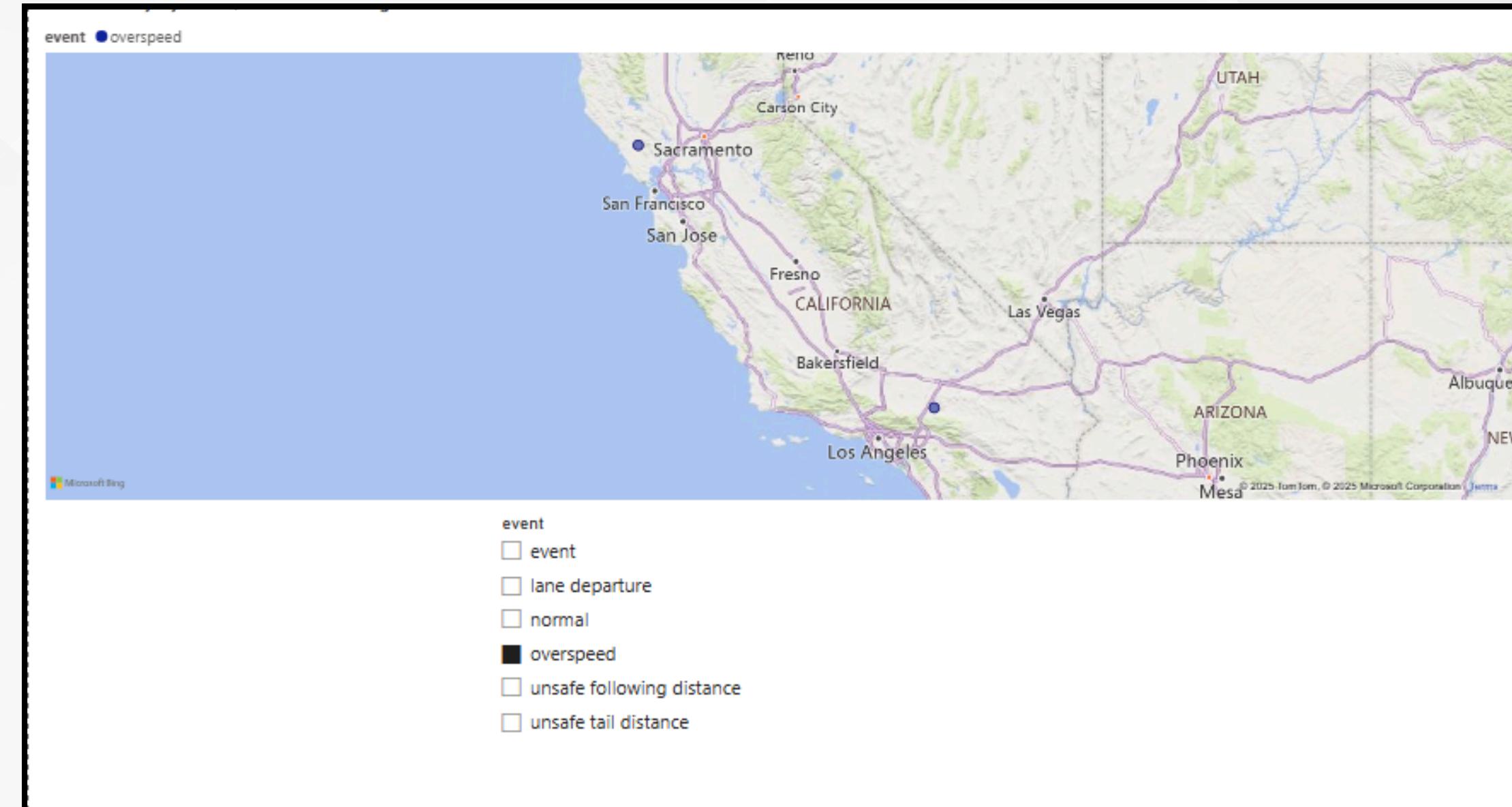
Where are unsafe driving behaviors—specifically unsafe following distances—most concentrated geographically?



GEOLOCATION OF EVENTS IN CALIFORNIA

Visualizes overspeed incidents across major California cities

Most overspeed events are clustered around urban centers and major highways, highlighting high-risk corridors.

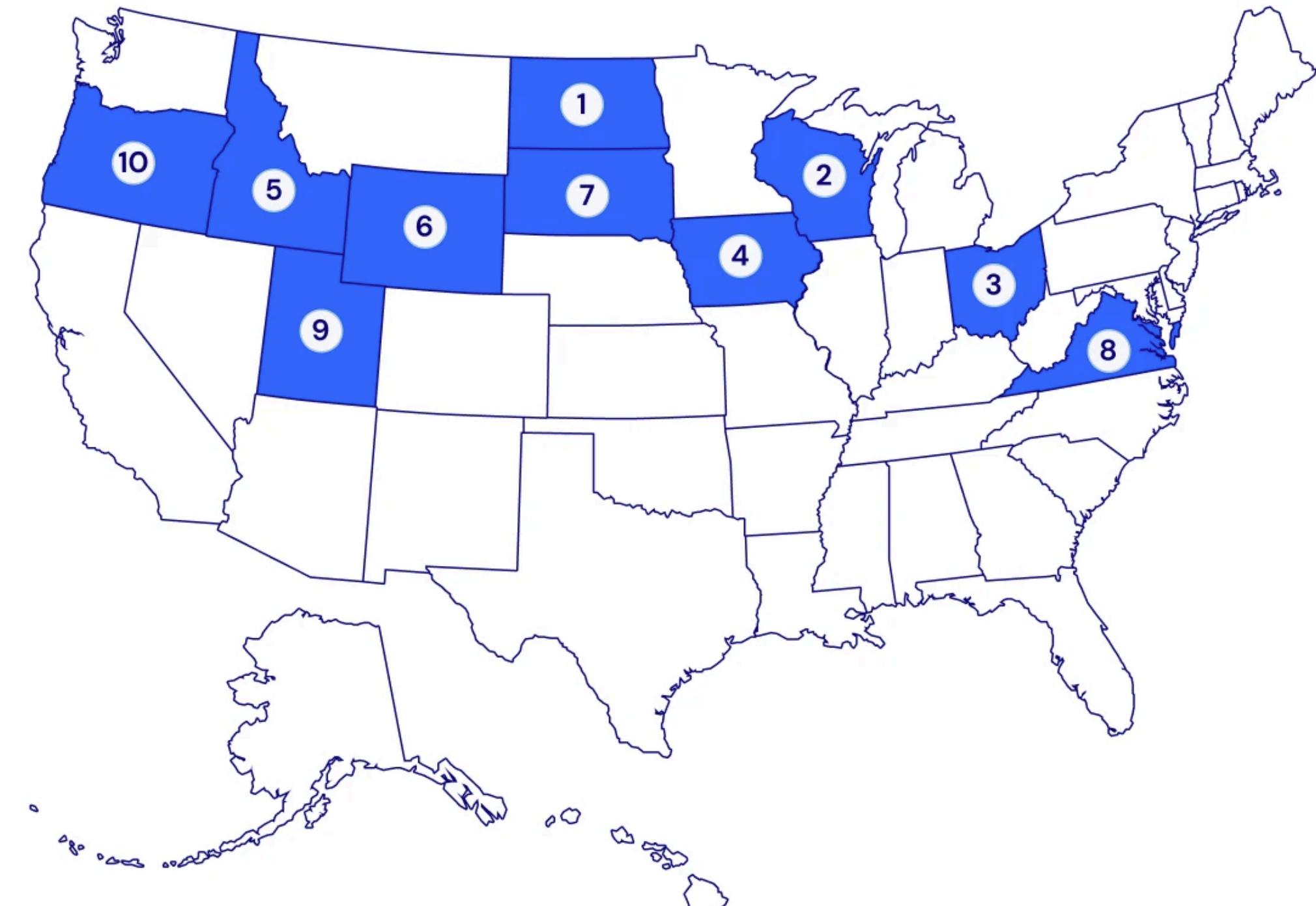


Concentrations near cities suggest a need for enhanced speed monitoring and potential driver training in congested areas.

QUESTION - 3

Which cities report the highest number of safety-related driving events (OverSpeeding)?

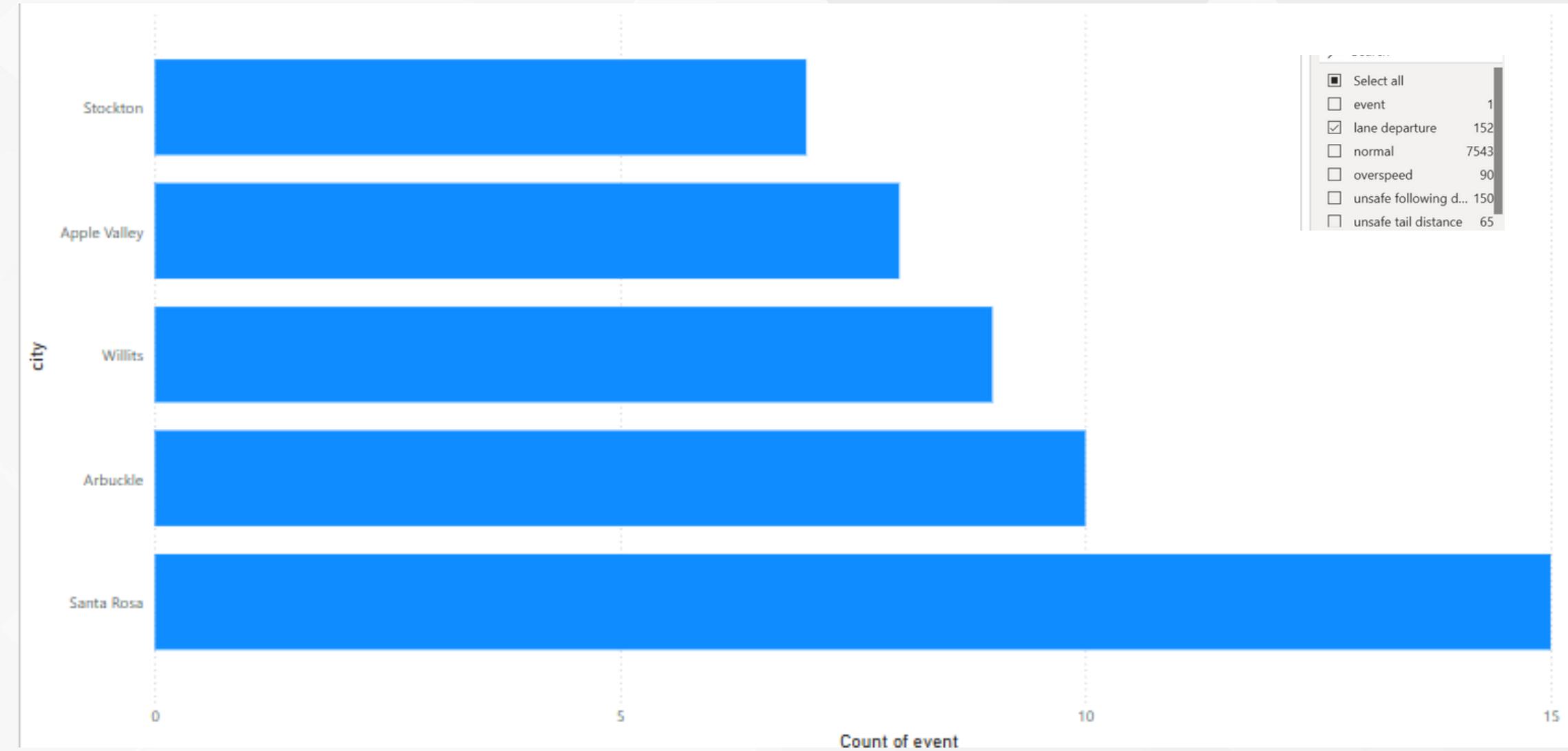
Top 10 States with the Most Speeding Tickets



TOP 5 CITIES BY OVERSPEED EVENT COUNT

Highlights the top 5 cities with the most recorded events (e.g., Santa Rosa, Arbuckle).

The chart updates per event type (e.g., Overspeed, Lane Departure) for focused insights.



Cities like Santa Rosa and Arbuckle show the highest counts, indicating priority zones for safety monitoring.

QUESTION - 4

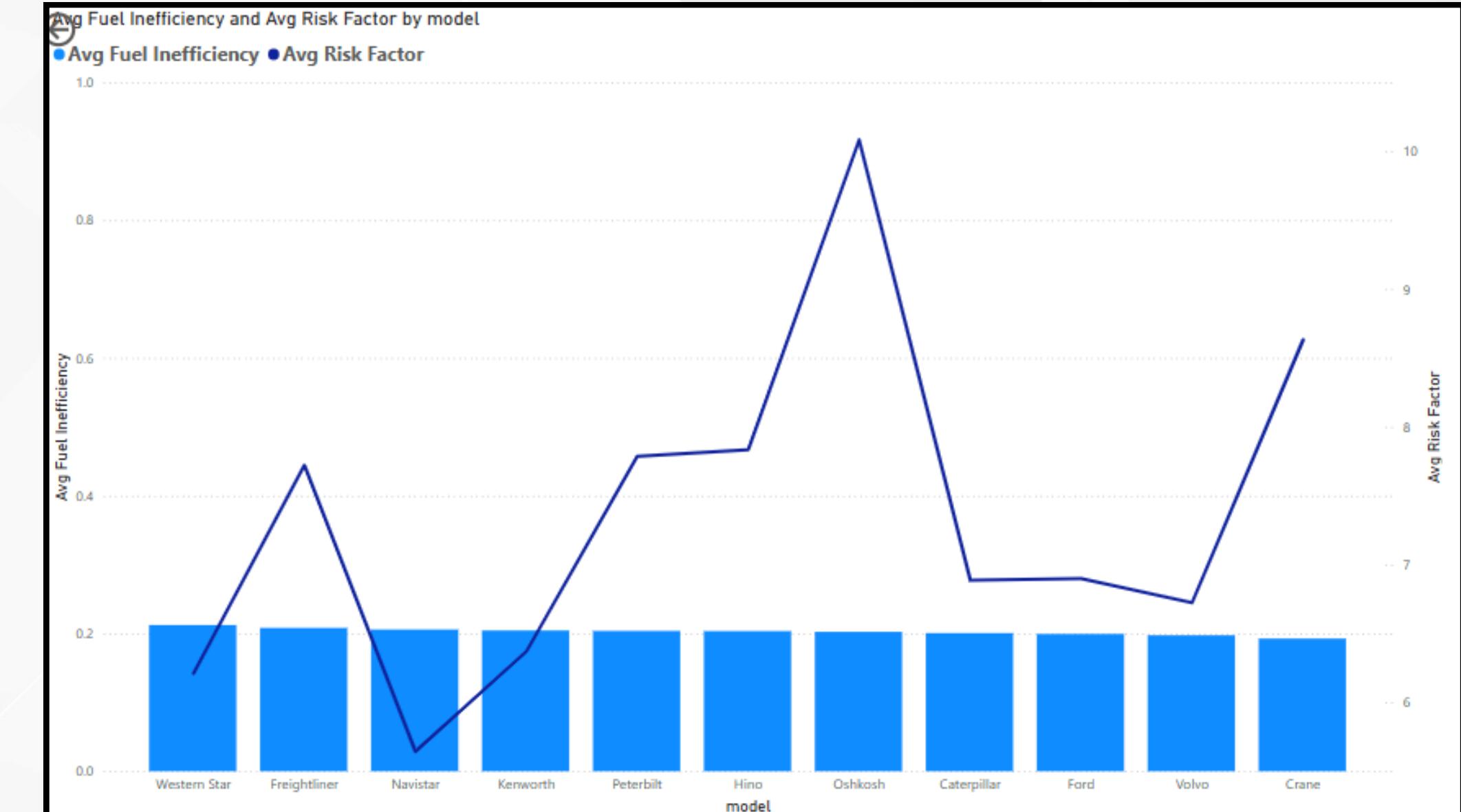
**How do truck models
compare in terms of fuel
inefficiency and average
driver risk factor?**



FUEL INEFFICIENCY & RISK FACTOR BY MODEL

Oshkosh shows the highest risk factor despite average fuel inefficiency.

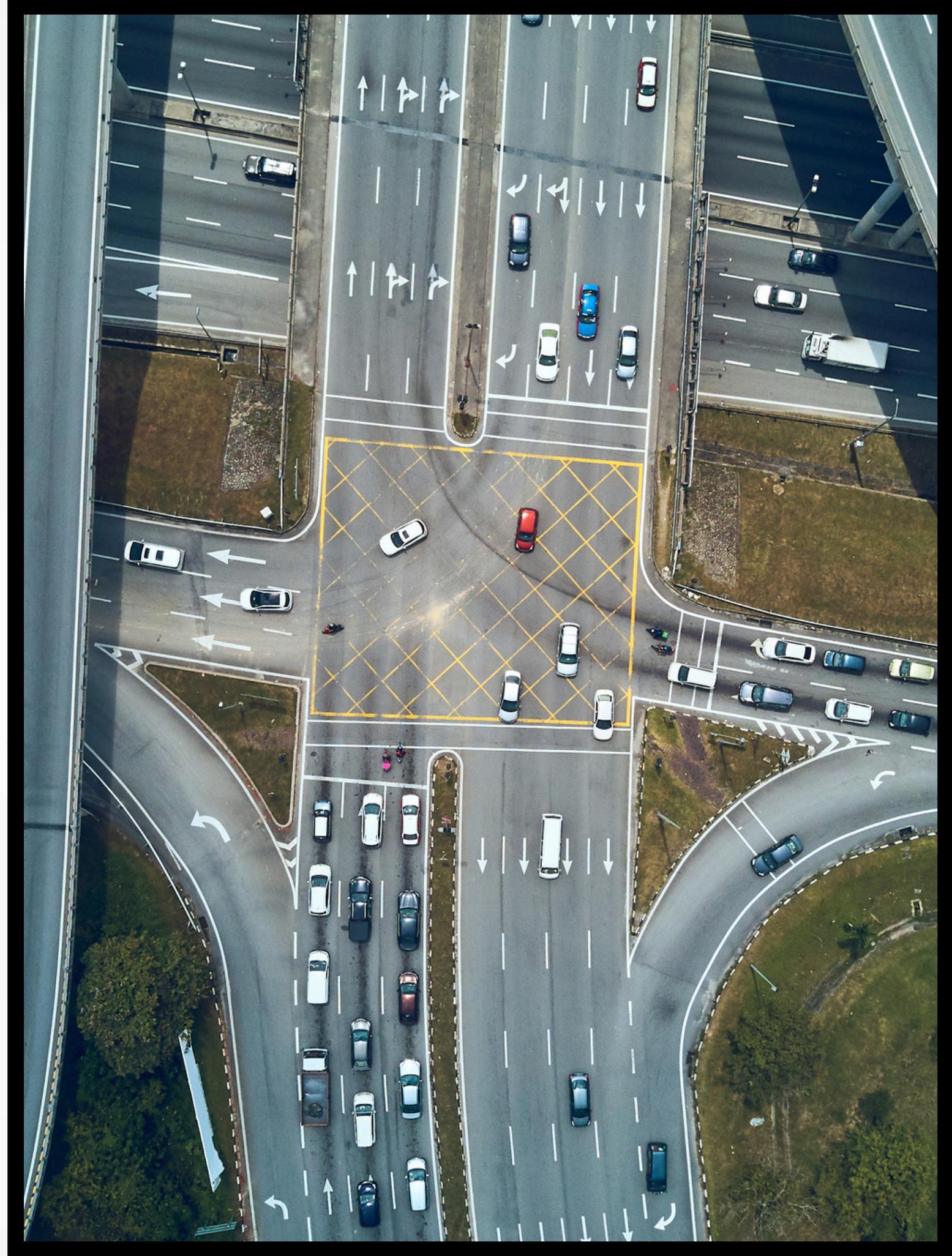
Freightliner and Crane have notable spikes in risk relative to peers.



Most other models maintain consistent risk and fuel patterns.

QUESTION - 5

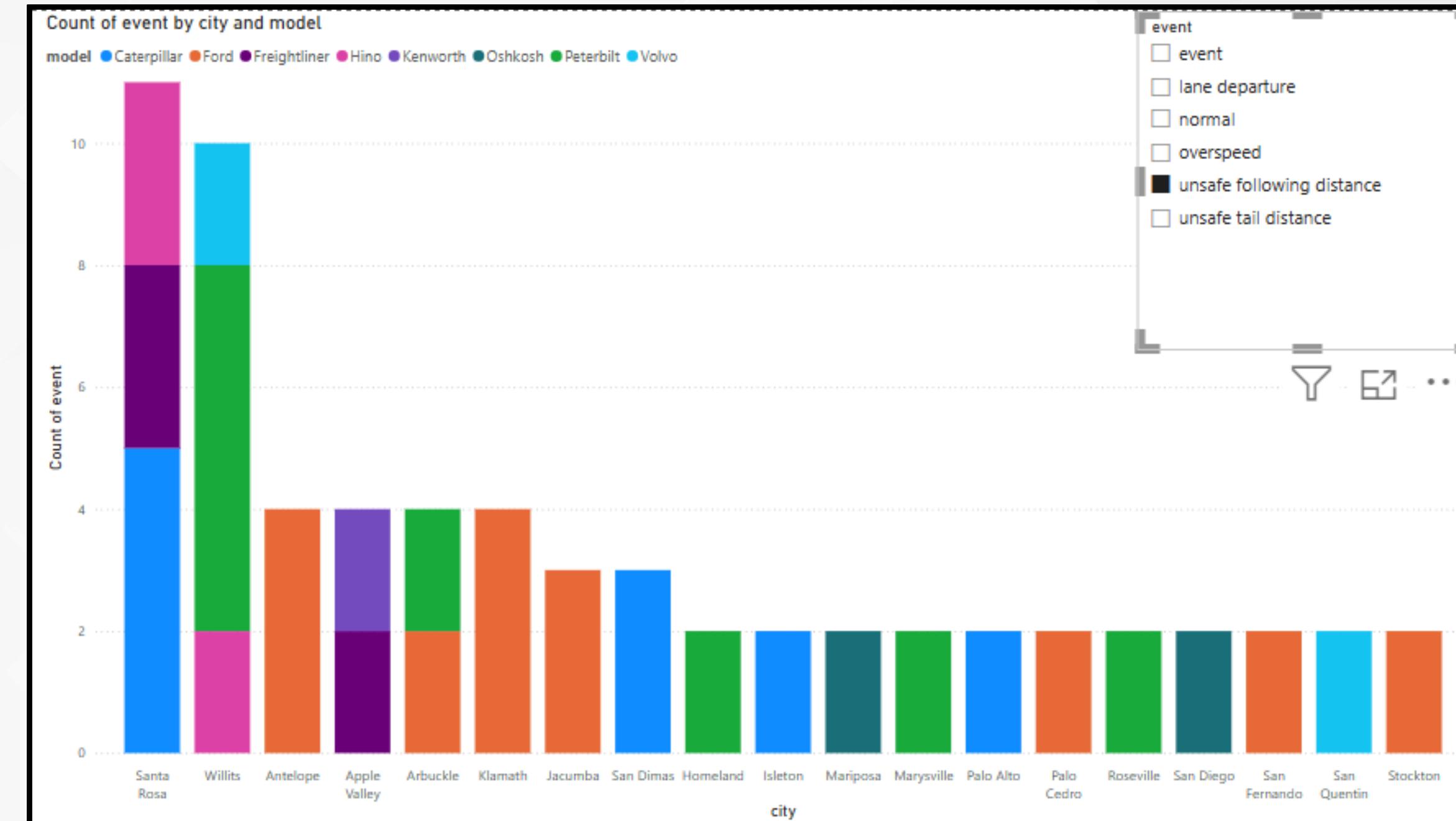
**Which cities and truck
models are most frequently
involved in safety-related
events?**



EVENT DISTRIBUTION: CITIES VS. TRUCK MODELS

Santa Rosa leads with the highest overspeed events, spanning multiple truck models.

Other cities have lower event frequency and fewer model types involved.



Cities like Willits and Arbuckle show moderate counts, dominated by specific models (e.g., Kenworth, Hino)

A photograph showing a massive traffic jam of trucks on a multi-lane highway. The trucks are lined up in both directions, stretching far into the distance. To the right of the highway, there is a railway track with several sets of parallel tracks. The surrounding area is a mix of green fields and some buildings in the far background. The sky is clear and blue.

THANK YOU!

QUESTIONS?