



# BIG DATA PROJECT

## Truck Driver Risk Report

**Group 8 Members:** (Ranked by First Name)

- Fayad Hasan Khan
- Jielin Zhang
- Min Shi
- Rouzbeh Roshani
- Ruthwik Ranganath Bethapudi



# PROBLEM STATEMENT

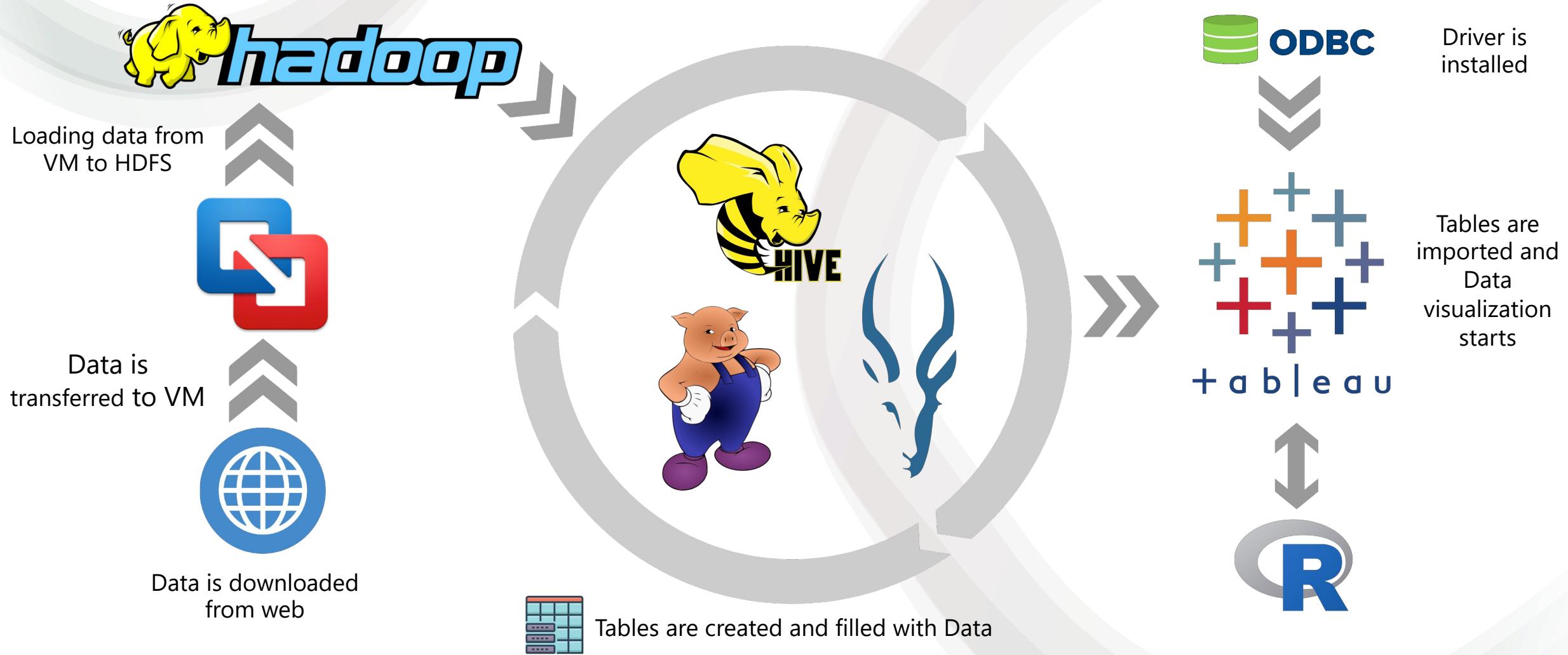
Accidents caused by large trucks remain one of the leading causes of injuries and deaths in the United States.

Hence, in order to capture the driving behavior of its truck drivers, ANT corporation has equipped its trucks with a device to log location and event data to increase safety.

# OBJECTIVES

- Identify dangerous commercial truck drivers
- Create an analytical dashboard for illustrations

# PROJECT FLOWCHART

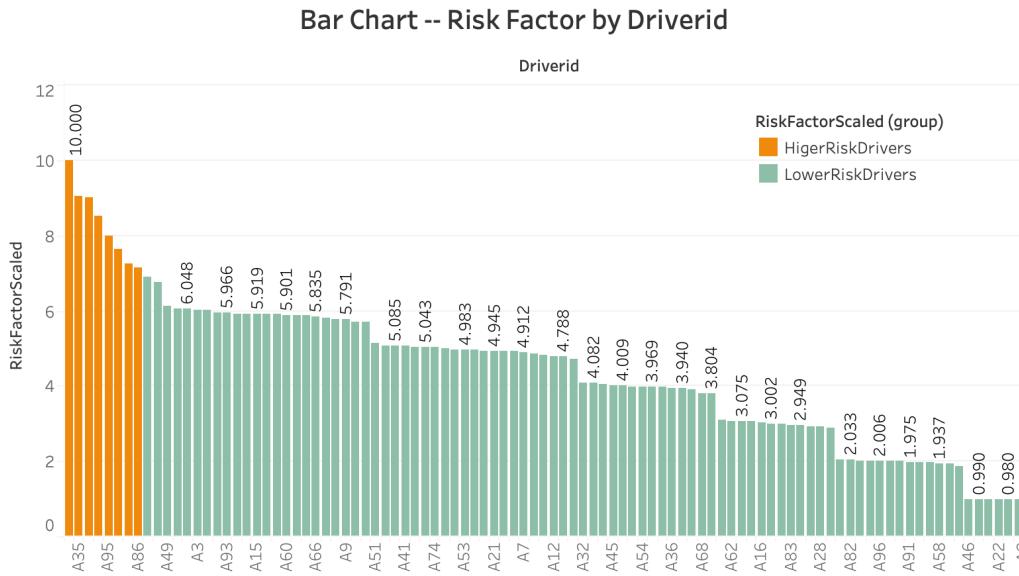


# TABLE RELATIONSHIP IN TABLEAU

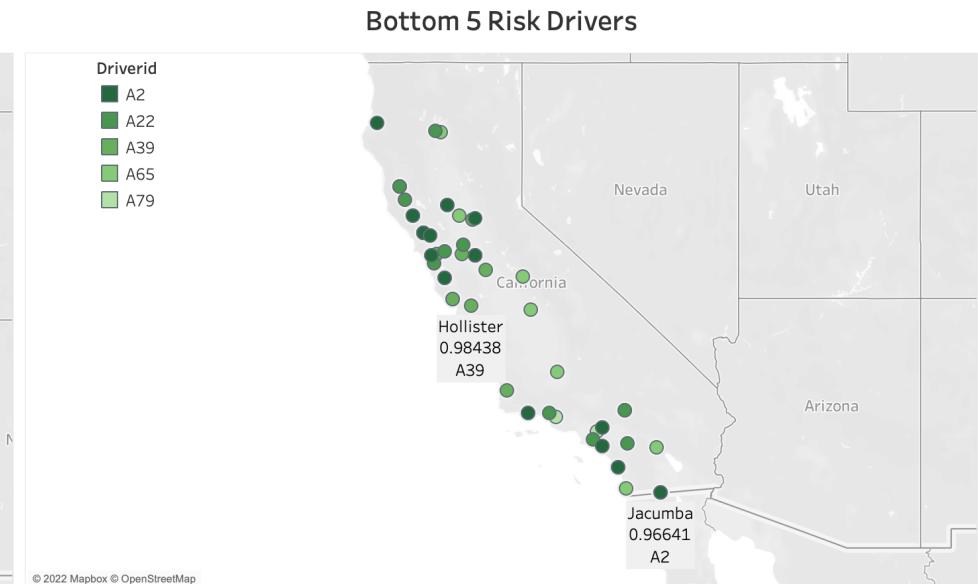
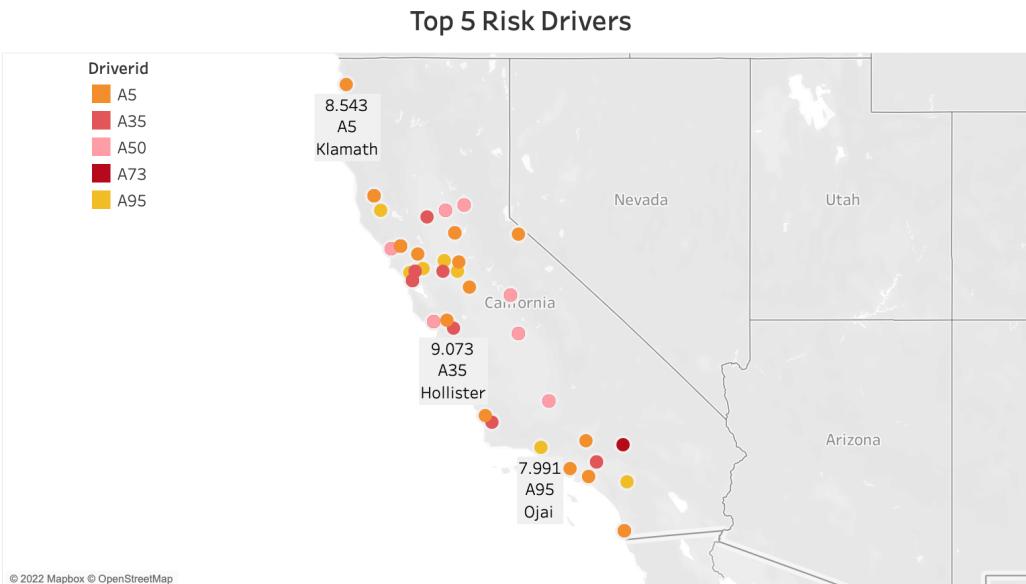
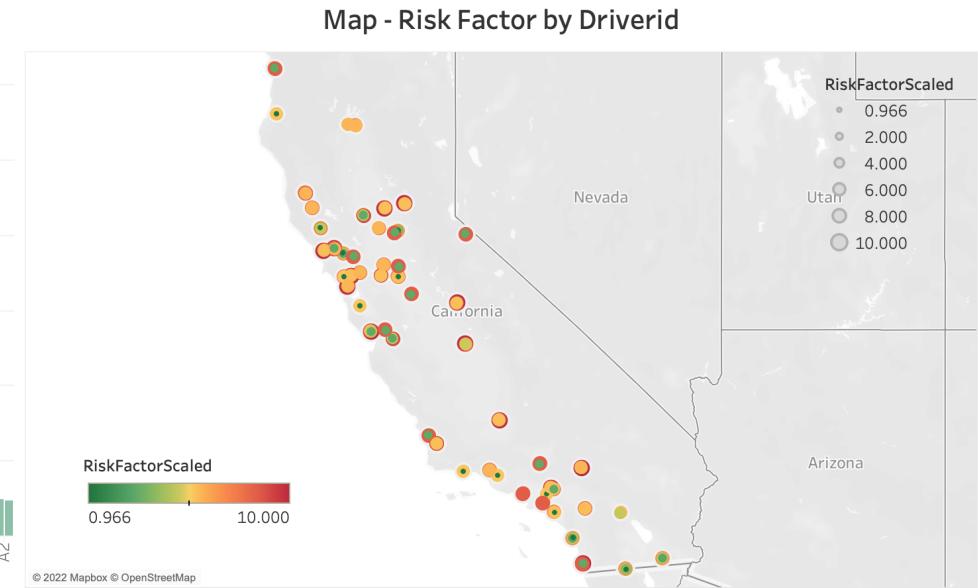


# ANALYSIS 1. Most and Least risky drivers

Most risky driver: A73  
(10 risk factor)

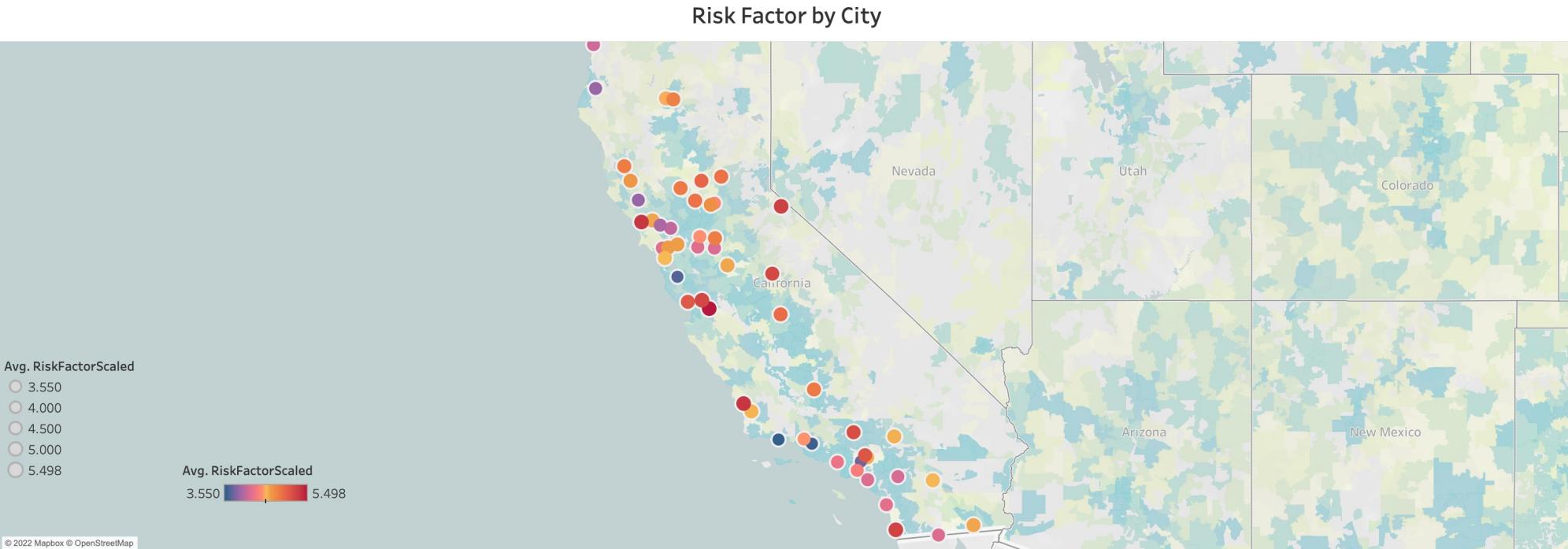
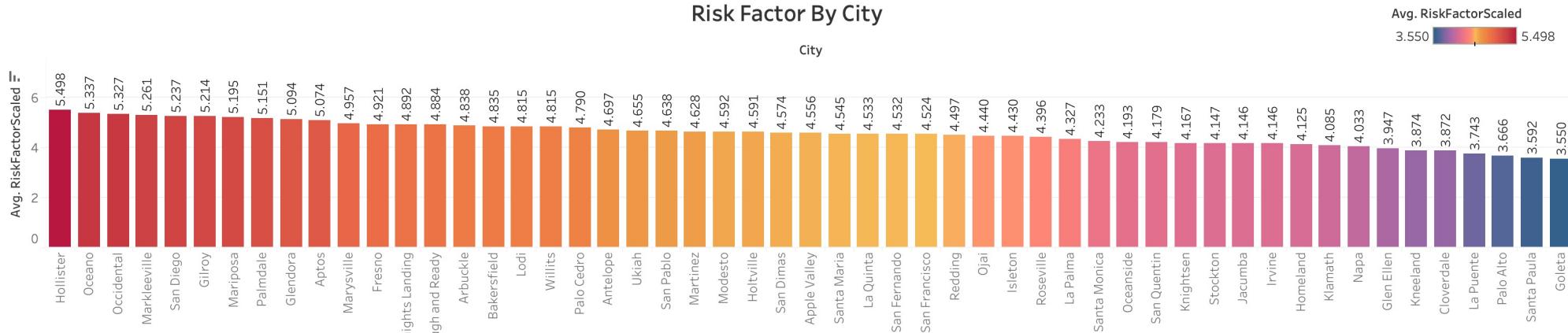


Least risky driver: A2  
(0.966 risk factor)



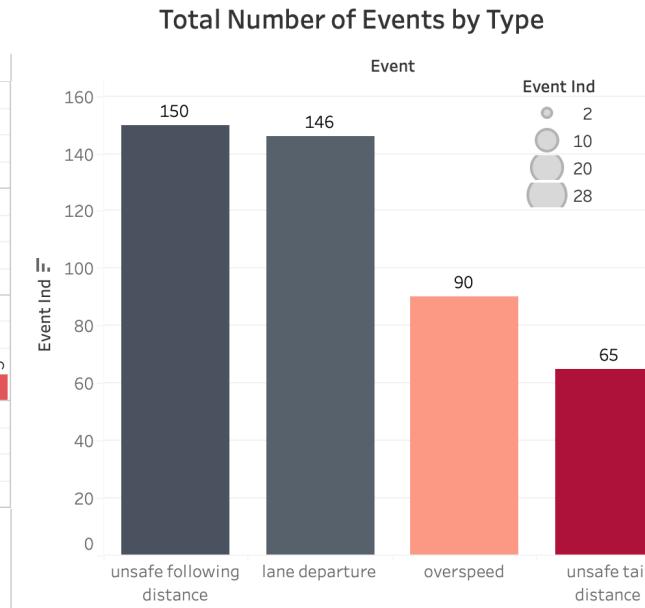
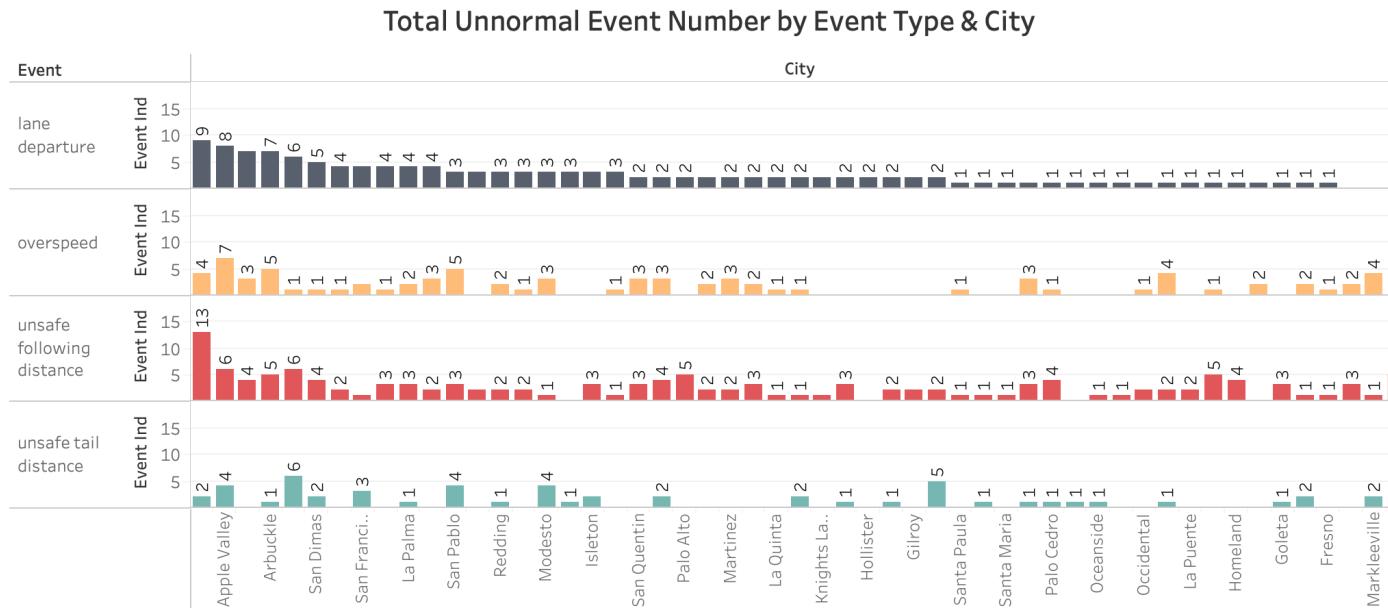
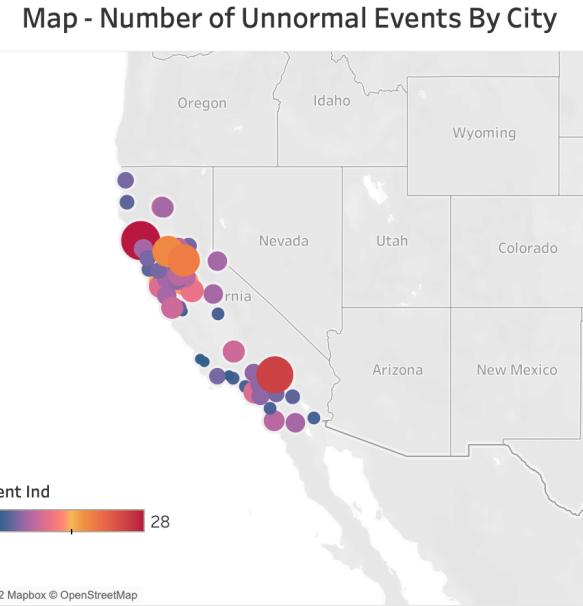
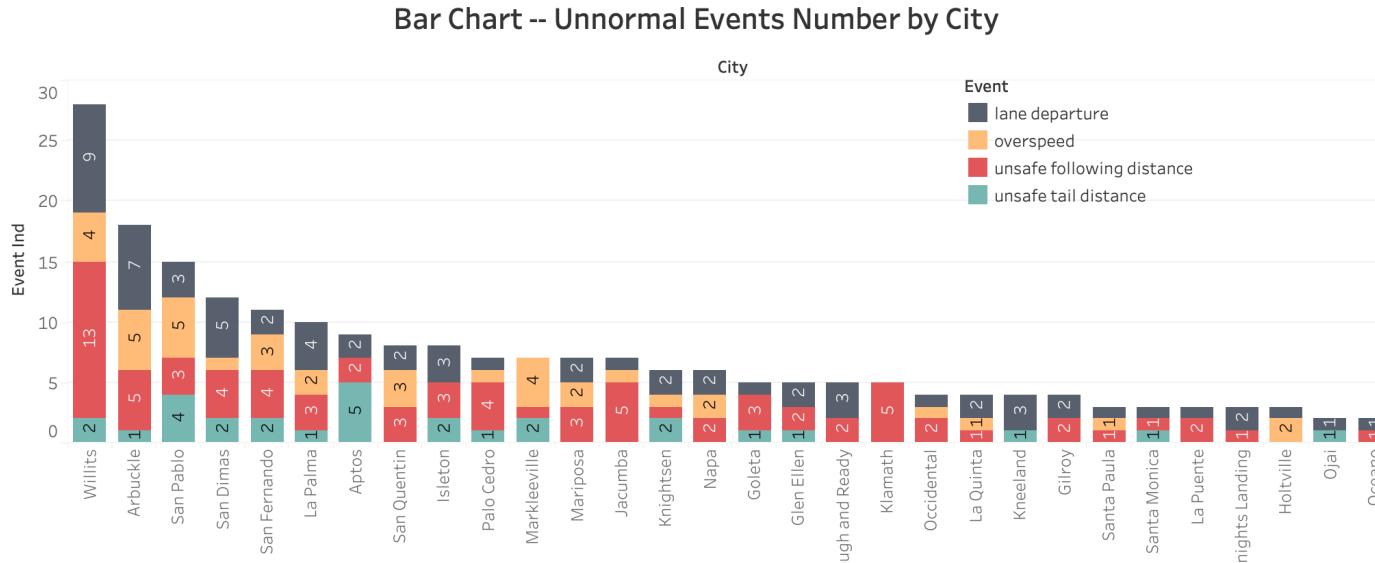
# ANALYSIS 2. Cities with the highest risk factor

City with highest average risk factor:  
Hollister  
(with risk factor 5.498)



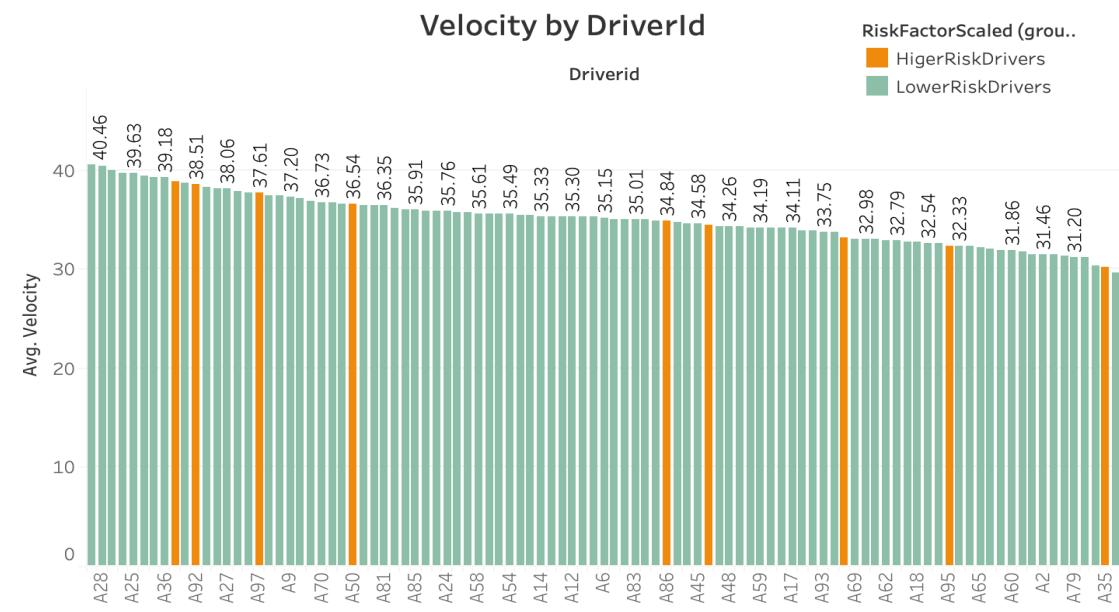
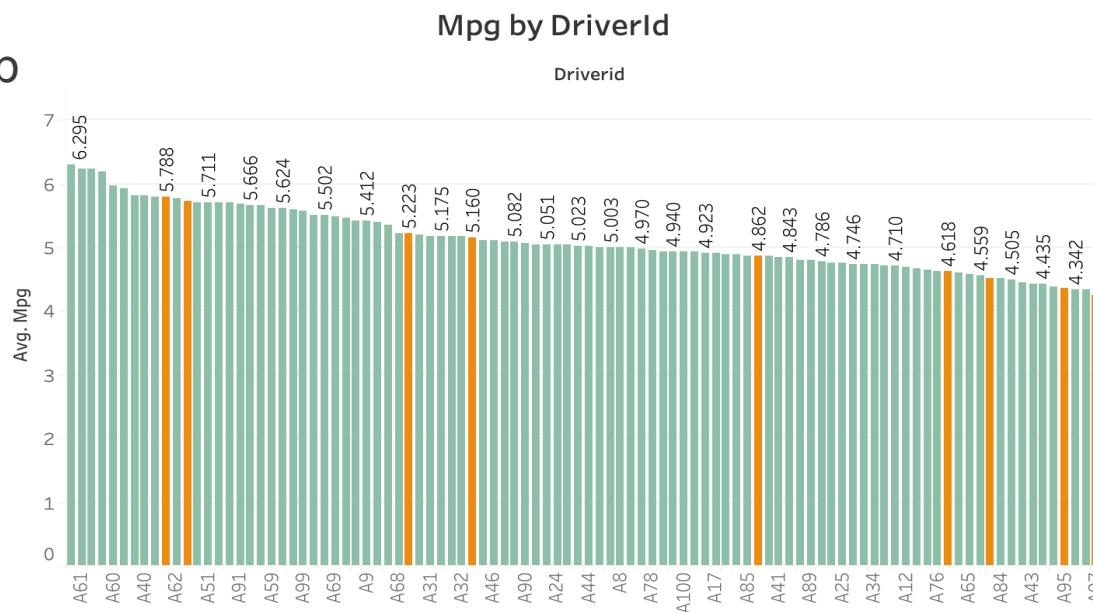
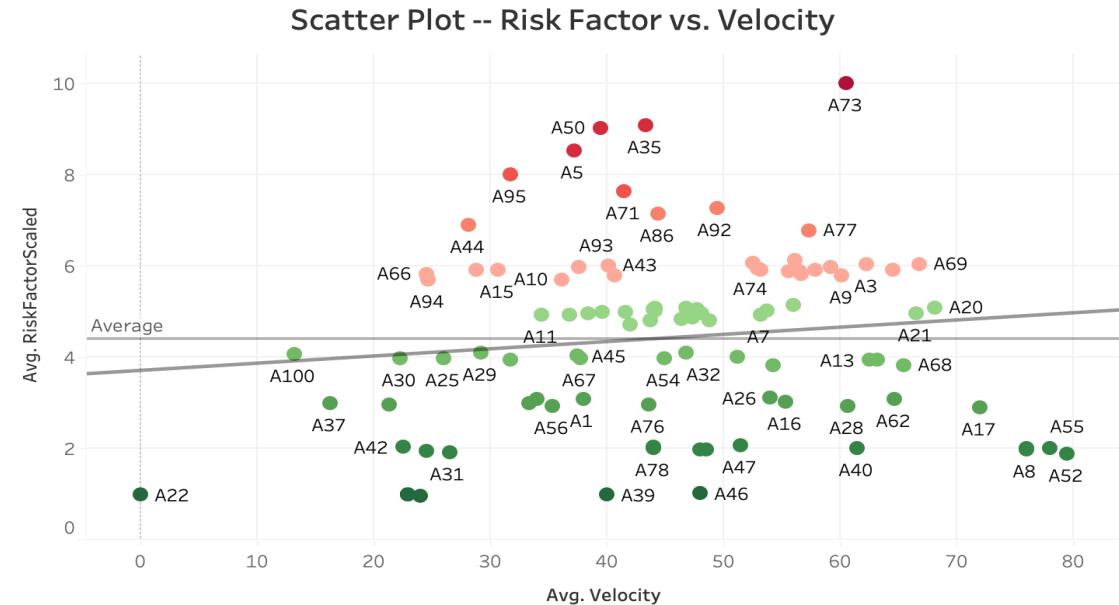
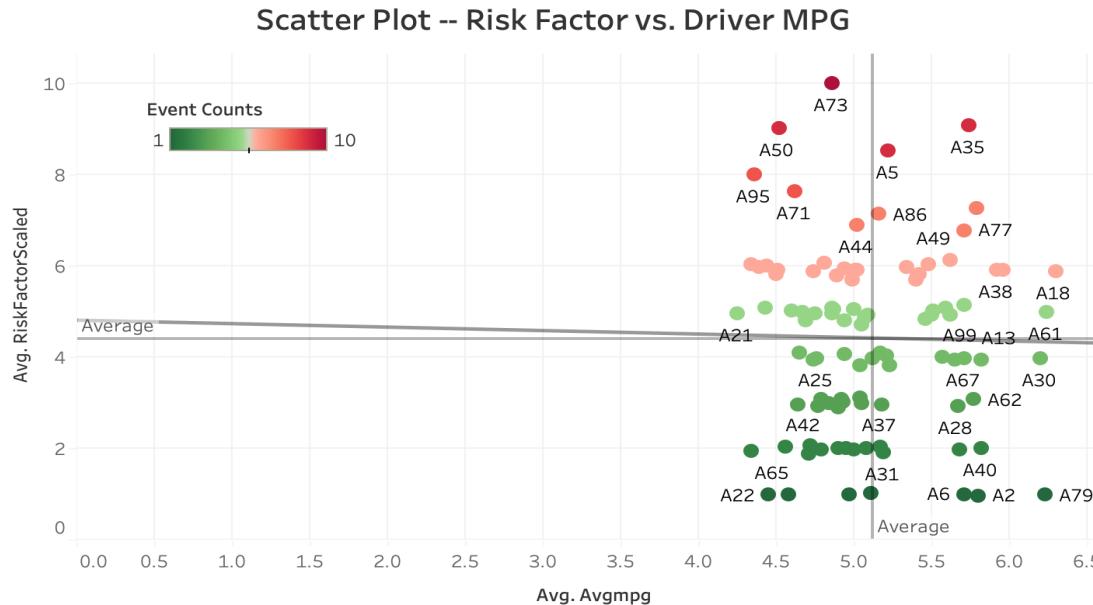
# ANALYSIS 3. Events and risk factor

Unsafe following distance causes highest number of abnormal event.



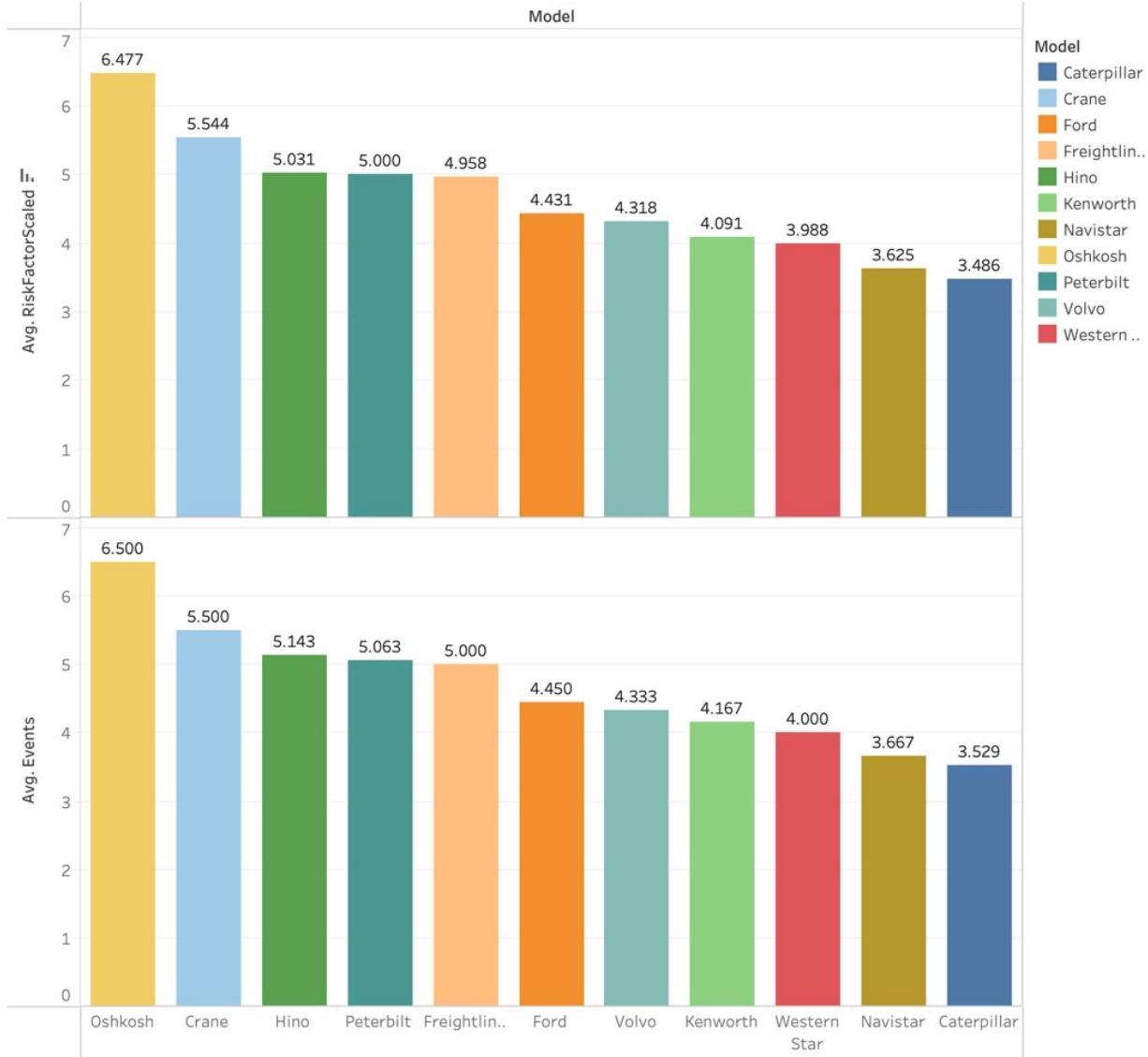
# ANALYSIS 4. MPG/Velocity and risk factor

MPG or Speed does not have obvious relationship with risk factor.

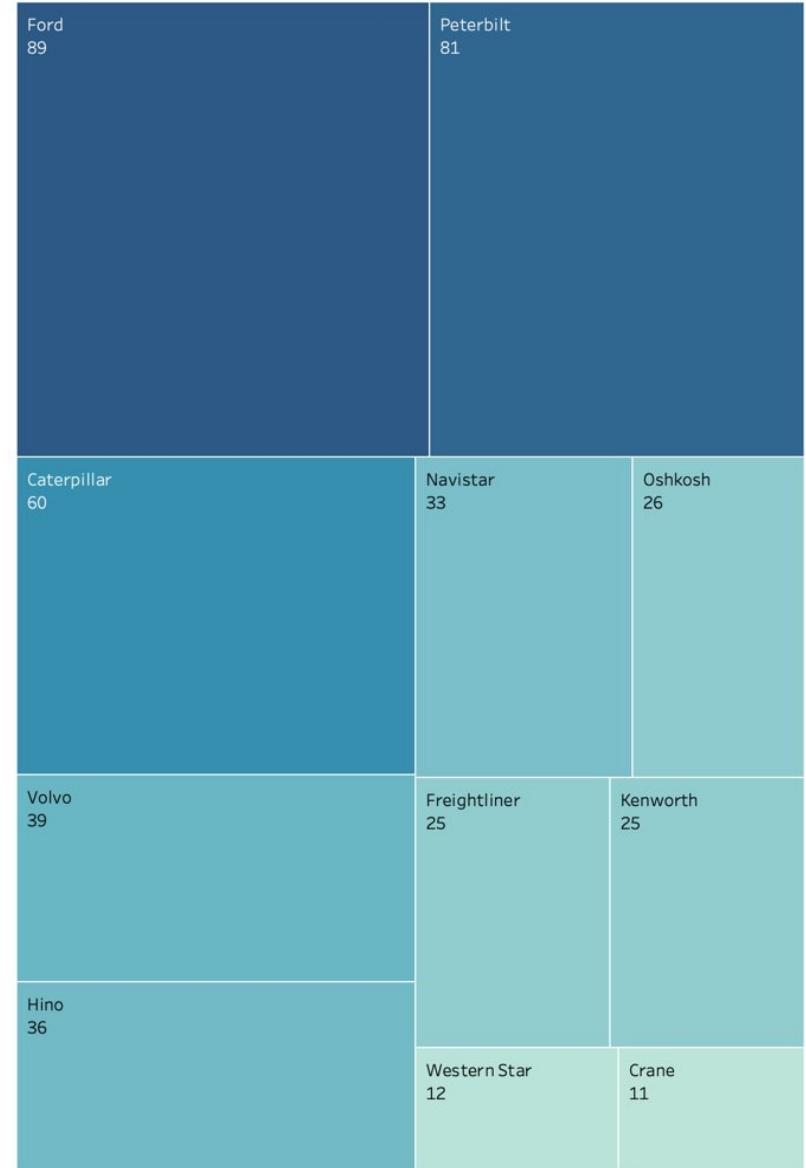


# ANALYSIS 5. Truck model and risk factor

Bar Chart -- Risk Factor & Number of Unnormal Events by Model



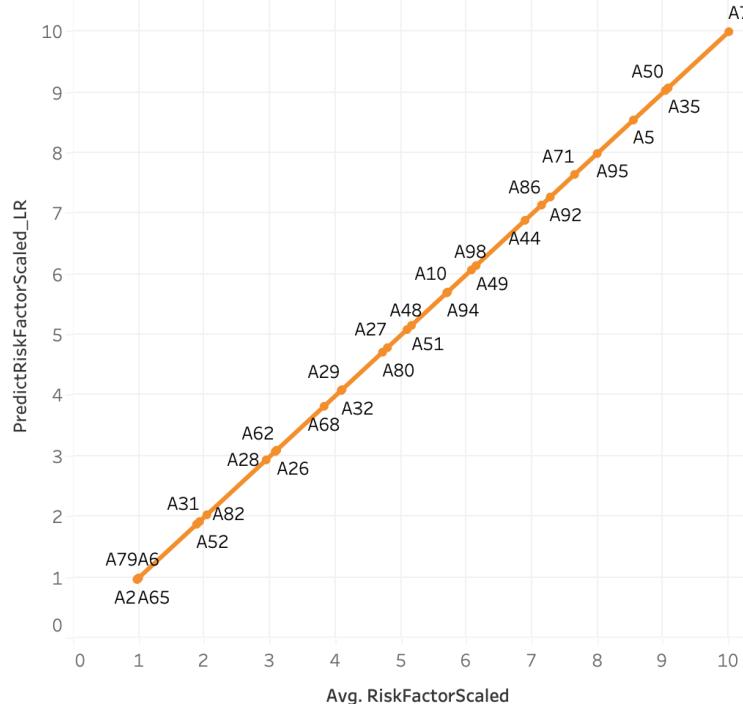
Number of Unnormal Events by Model



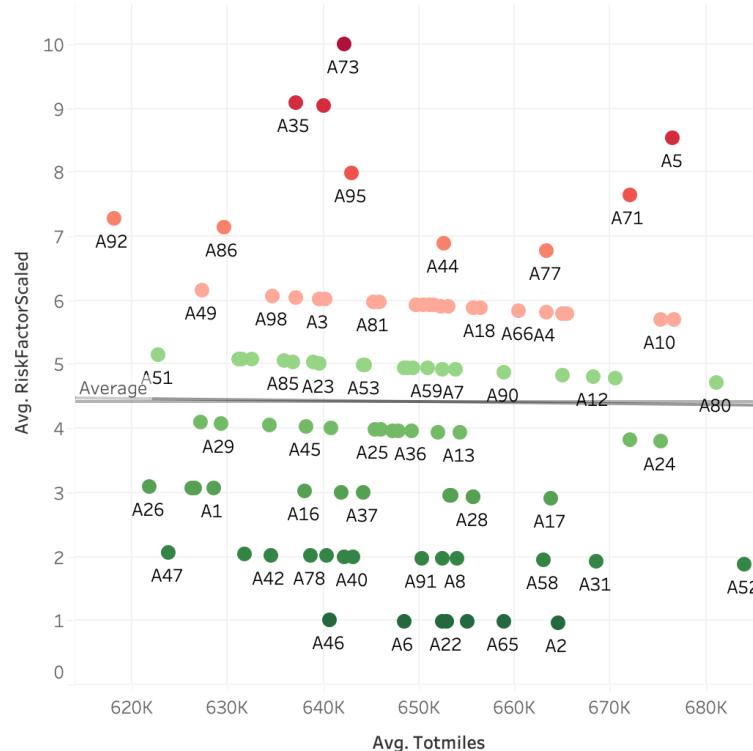
Different truck model has different frequency of unnormal events, thus different level of risk factor.

# ANALYSIS 6. Linear Regression and Correlation

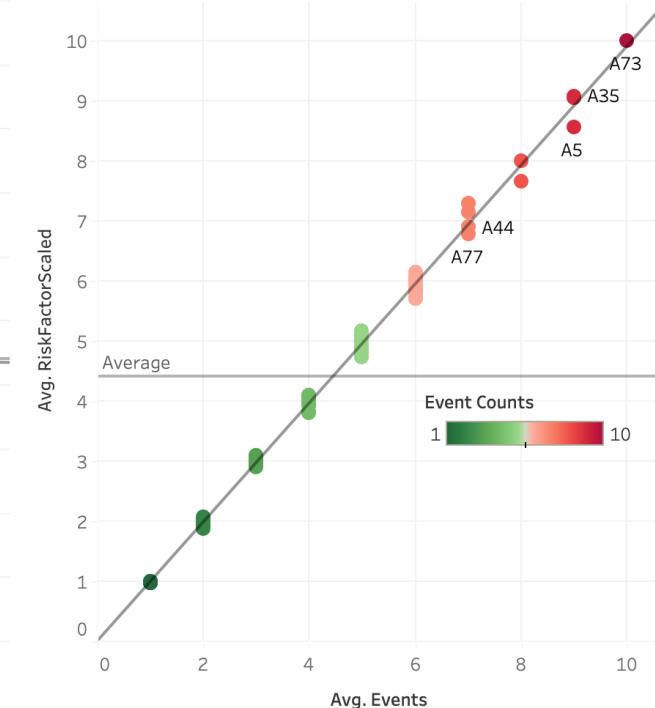
Predicted Risk Factor based on Linear Regression vs. Risk Factor Scaled



Scatter Plot -- Risk Factor vs. Total Miles



Scatter Plot -- Risk Factor vs. Number of Unnormal Events



```
SCRIPT_REAL("
y <- .arg1;
x1 <- .arg2;
x2 <- .arg3;
fit <- lm(y ~ x1 + x2, na.action=na.exclude);
fitted(fit);
",
SUM([RiskFactorScaled]), SUM(LN([Totmiles])), SUM([Events]))
```

Linear Regression Results in Stata

risk_stand~d	Coefficient	Std. err.	t	P> t
events	.9893698	.0022078	448.12	0.000
miles	-7.05e-06	3.08e-07	-22.85	0.000
_cons	4.575836	.1998879	22.89	0.000

## Inference

The predicted risk factor based on linear regression model ( $\text{risk factor} \sim \ln(\text{totmiles}) + \text{numberOfEvents}$ ) is highly consistent with the scaled risk factor.

There is no relationship between total miles and risk factor.

Stata results show that the coefficient of events is meaningful but the coefficient of total miles is nearly 0.

Overall, we could conclude that only the number of unnormal events affects risk factor.

# CHALLENGES

1. An outlier (A97) in riskfactor table skewed the results.

>> **Solution:** Scaled the riskfactor and exclude the outlier.

2. In the riskfactor table, we found the A19 driver's record was missing.

>> **Solution:** Checked A19 distinct events, there was only normal event.

```
SELECT DISTINCT event          Results | Chart  
FROM geolocation  
WHERE truckid = 'A19';           ↓  
                                event  
                                normal
```

3. After checking “geolocation” table, we found some abnormal values.

>> **Solution:** Need to talk to a SME for clarification.

truckid	driverid	event	latitude	longitude	city	state	velocity	event_ind	idling_ind
A40	A40	overspeed	37.957702	-121.29078	Stockton	California	77	1	0
A5	A5	unsafe following distance	41.526509	-124.038407	Klamath	California	33	1	0
A48	A48	overspeed	38.752124	-121.288006	Roseville	California	77	1	0
A94	A94	unsafe following distance	37.941325	-122.484704	San Quentin	California	0	1	1
A49	A49	lane departure	39.150171	-123.207783	Ukiah	California	41	1	0

## CONCLUSIONS and SUGGESTIONS

- Risk factors are not dependent on MPG/Velocity, it is perfectly correlated with count of unnormal events.
- Most unnormal event occurring city is Willits, it may be worth to analyze further on that, if local factors, such as traffic, road conditions, road maps are related to the risk factors.
- Top 5 Risky drivers can be counseled about unsafe driving and consequences and given more training if needed.
- Since Unsafe following distance is the most occurring event and it affects the risk factor most, an alert system can be installed on all the trucks which will notify the drivers if they have crossed the threshold of safe distance.