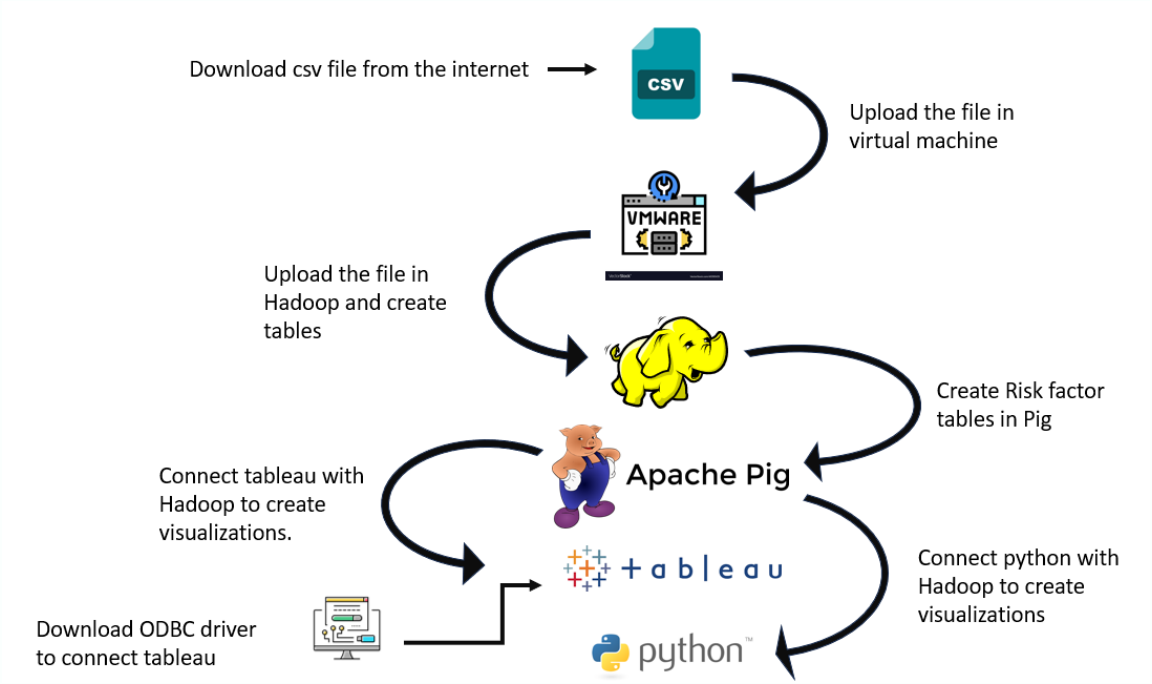
ANT Trucks Data Analysis

Problem Workflow

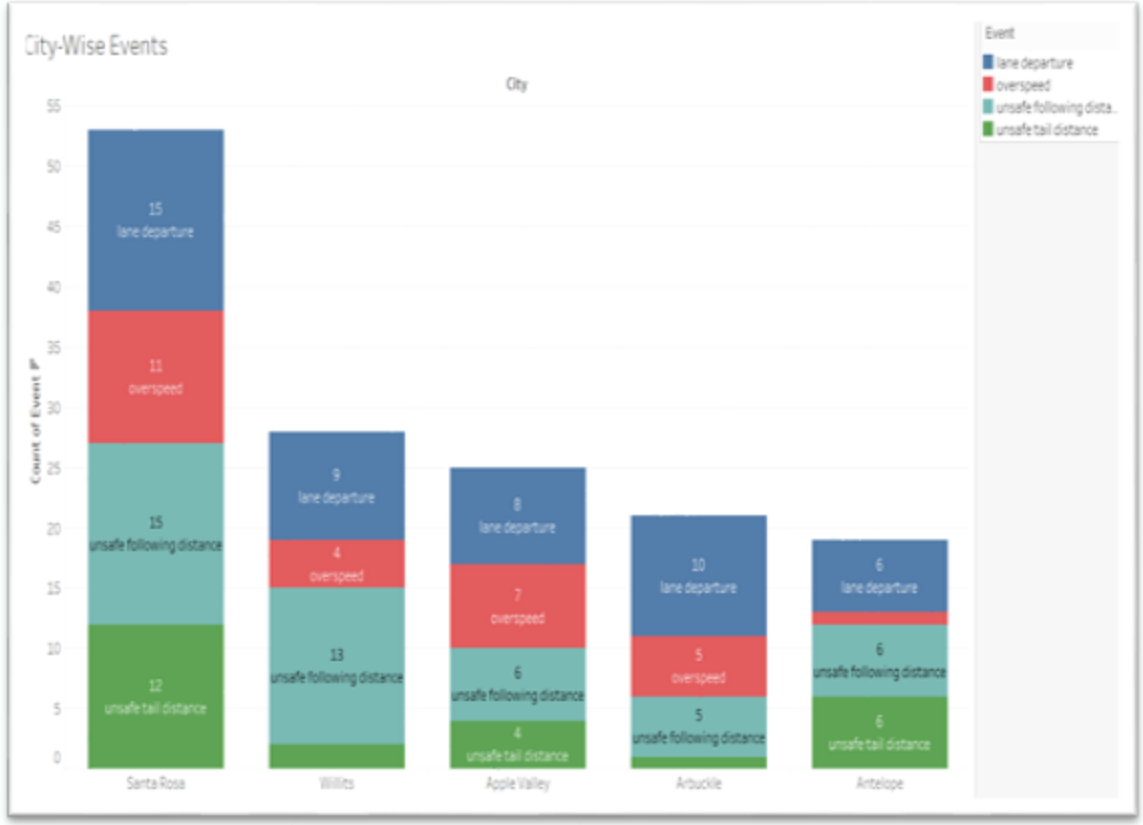
The objective of this presentation is to identify dangerous truck drivers nationwide and allow fleet manager to take necessary actions



City-wise events

Question:
Which cities see the most number of events and which events are more frequent than the others?

Answer:
- Santa Rosa, Willis, Apple Valley, Arbuckle and Antelope see most number of unsafe events.
- Lane departure seems to be the most frequently occurring in all the cities



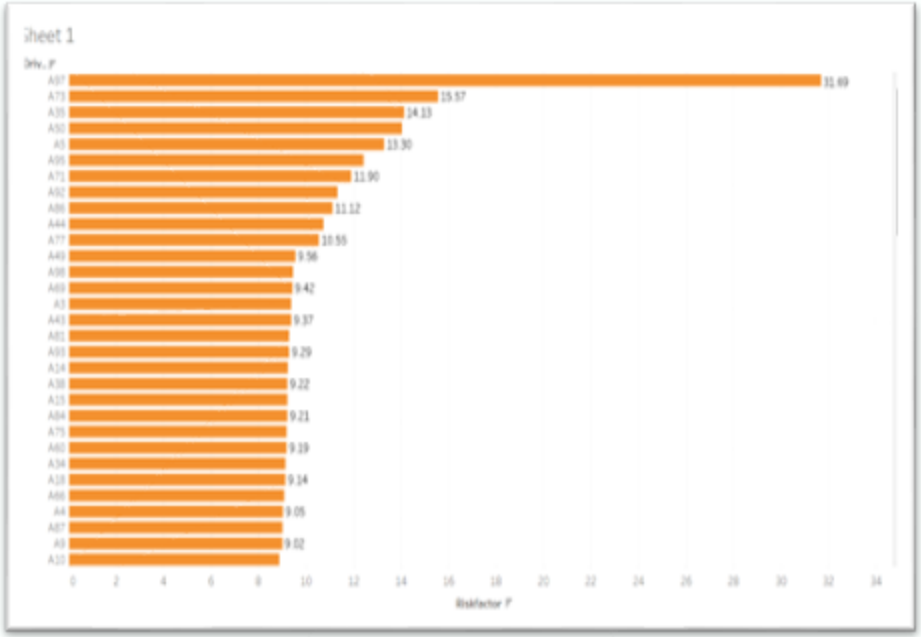
Identifying risky driving patterns based on event type and risk factor

Event type

Risky driver according to event type

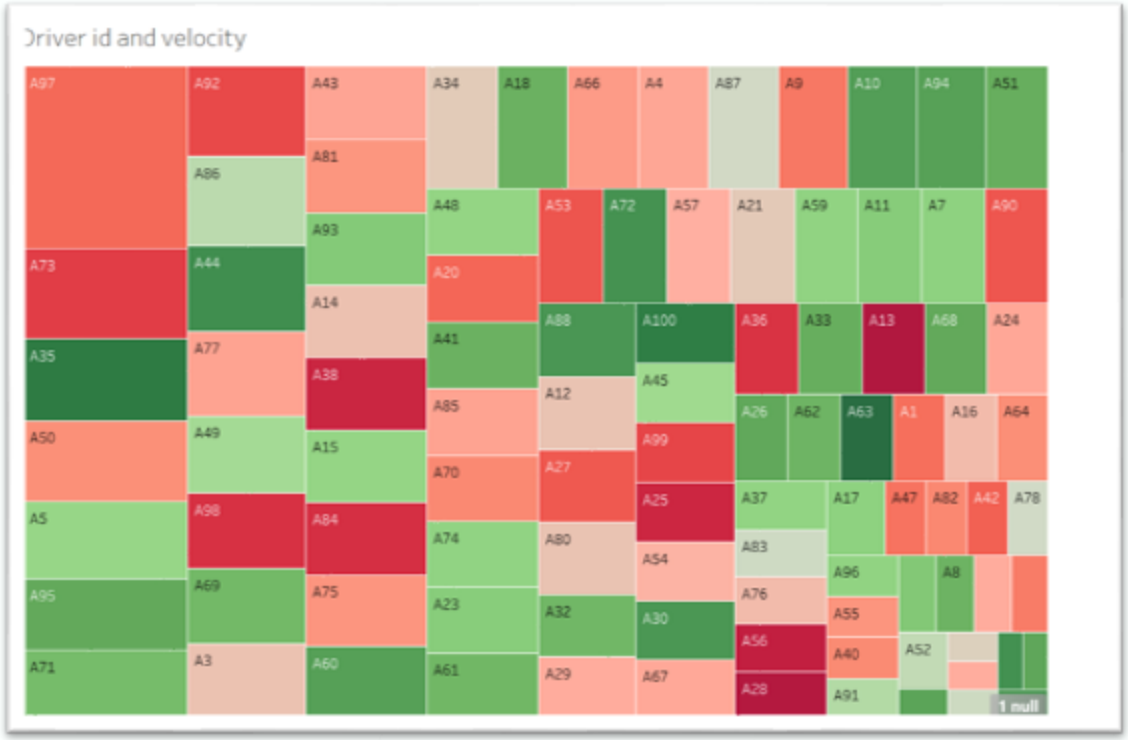
Driverid	lane departure	normal	overspeed	unsafe following di..	unsafe tail distance
A97	11	60	5	1	3
A95	5	72		1	2
A4	4	74	1	1	
A36	4	76			
A18	4	74	1		1
A94	3	74	1	2	
A92	3	73		4	
A87	3	74		2	1
A61	3	75		1	1
A54	3	76	1		
A50	3	71	2	3	1
A49	3	74	1	2	
A44	3	73		2	2

Risk factor

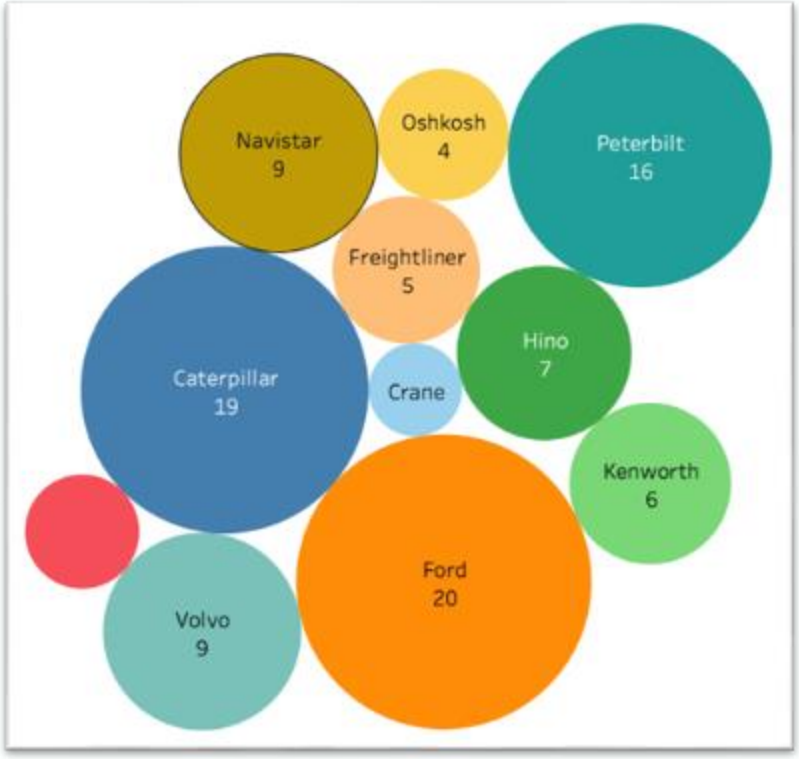


Drivers with high average velocities and risk factor

- The color represents average velocity for each drivers. Red representing higher velocity.
- Size represents the risk factor.
- Most of the drivers with high risk have high average velocities, however A35,A5,A95 do not have high average velocities even though they have high risk factor.



Popular Trucks & Risk Factor Score



Model	
Oshkosh	10.085
Crane	8.632
Hino	7.834
Peterbilt	7.785
Freightliner	7.720
Ford	6.900
Caterpillar	6.886
Volvo	6.724
Kenworth	6.369
Western Star	6.210
Navistar	5.644

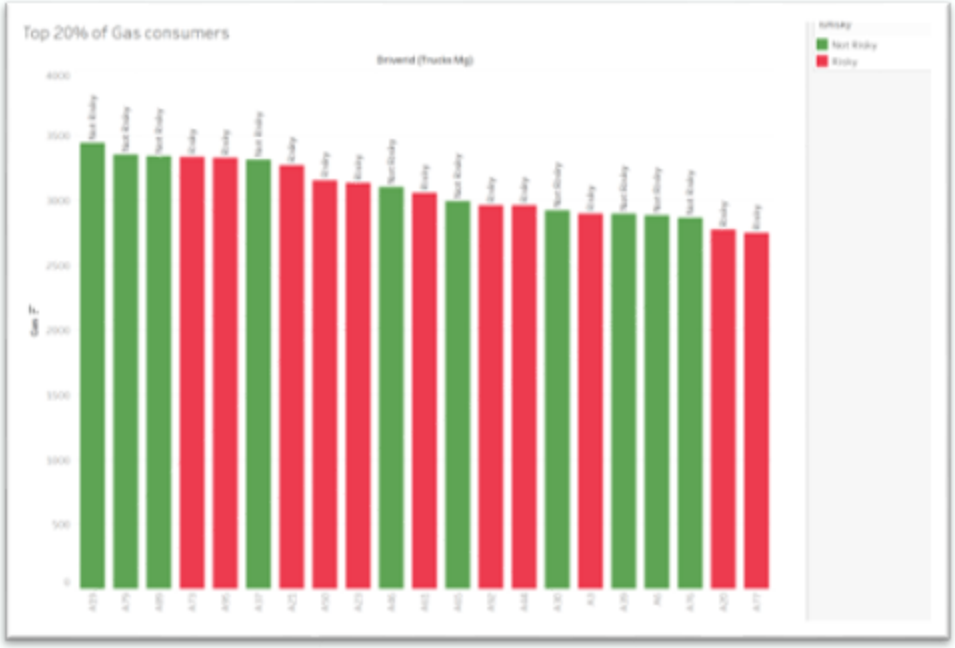
Trucks' total mileage & Risk Factor Score

Question:
Does truck mileage correlate to risky event?

Answer:
when we plot "total mileage" against "average riskfactor", no relationship was shown

Does gas consumption affect risk factor ?

- A correlation coefficient of approximately -0.02 suggests a very weak negative correlation between mileage and the risk factor in your data.
- The negative sign indicates a negative linear relationship, and the magnitude of the coefficient being close to zero implies a weak correlation.




```
# Correlation between the gas consumption and the dependent variable'riskfactor'
correlation = merged_data['Gas'].corr(merged_data['riskfactor'])

print(f'Correlation between Gas consumption and Risk Factor: {correlation}')
```

```
Correlation between Gas consumption and Risk Factor: -0.020264522260486865
```

Insights and recommendations

- From the mileage perspective, we don't see any correlation between mileage number and risk factor score.
- The current most used models seem to have low risk-factor score, which indicates a good sign from a safety perspective
- If budget permits, ANT can further increase the number of trucks used with lower risk factor score.
- For the trucks associated with high number of risky events such as overspeed, ANT can check if the engine of those models is allowed or forced to turn beyond its design limit.
- Truck drivers who tend to have a high-risk factor and have higher number of unsafe events could be given regular training and driving tests.