BigData Term Project: Group 4

Risky Driver Analysis
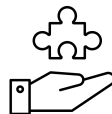
Satya Moturi Harish
Ravi Kumar Bommaji
Bhargava Madire Reddy
Sahana Mudduluru
Varun Bhavnani

AGENDA

Project Objectives

Process Flow

Data Analysis

Conclusion

Project Objective

Essential steps towards analyzing Risk factors

Project Perspective:

Processing the data to generate analytics reports and a visual presentation that enables:
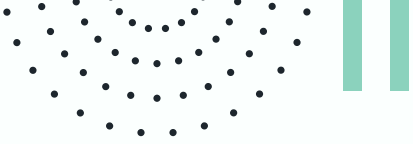
Decision-makers and drivers to reduce the risks.

Optimization of the business operations and improve profitability.

Improvement in driver safety, and increase efficiency.

Challenges

Business Challenges:

The most prevalent causes of injuries and fatalities in the US are due to the Large truck accidents.

Cost management and regulatory compliance.
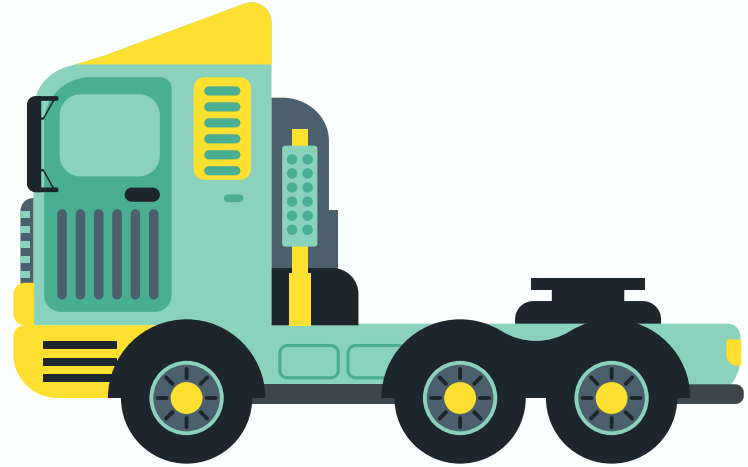
Technically Challenges:

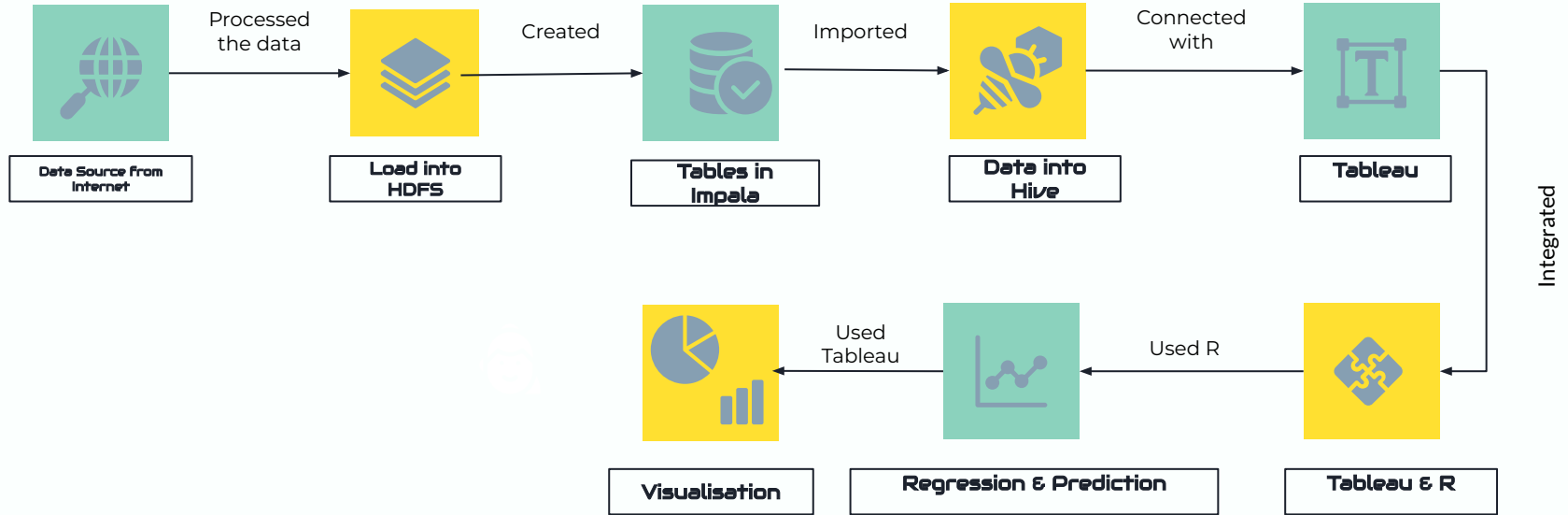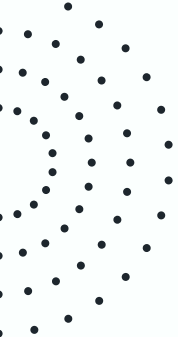Model development, validation and Interoperability.

Data collection and Integration, Data quality, and privacy concerns (Data Security)
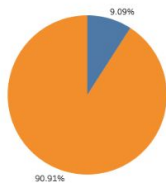
Process Flow

Process Flow

Box	Connector	Box
Data Source from Internet	Processed the data →	Load into HDFS
Load into HDFS	Created →	Tables in Impala
Tables in Impala	Imported →	Data into Hive
Data into Hive	Connected with →	Tableau
Tableau	Integrated ↓	Tableau & R
Tableau & R	Used R →	Regression & Prediction
Regression & Prediction	Used Tableau →	Visualisation

Data Analysis

Exploratory data analysis

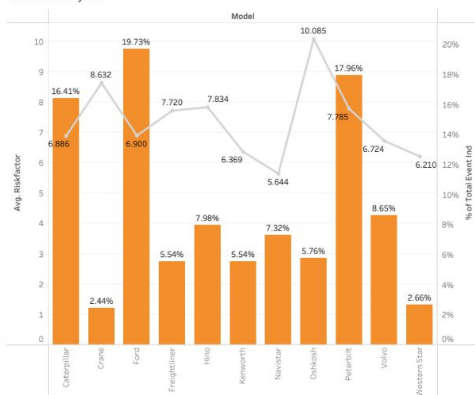Understanding the hindsight & insight from the data

High Level Analysis:

- We can notice the distribution of High Risky drivers and Low Risky drivers. There are 9% of the total drivers that have a tendency of rash driving.

- Lane departure and unsafe following are the two most prone events that leads to high risk.

- Hollister, Oceano & San Diego are the top cities that gives high average risk factors.

Exploratory data analysis

Understanding the hindsight & insight from the data

Risky Drivers Analysis:

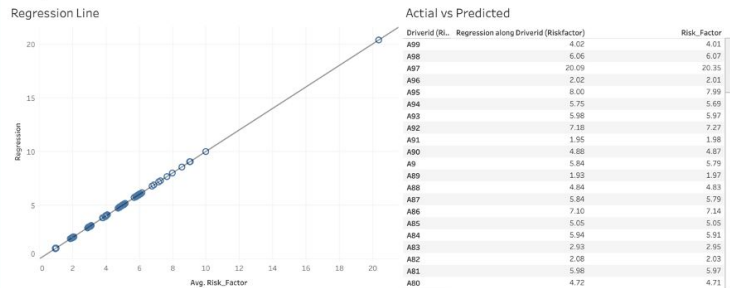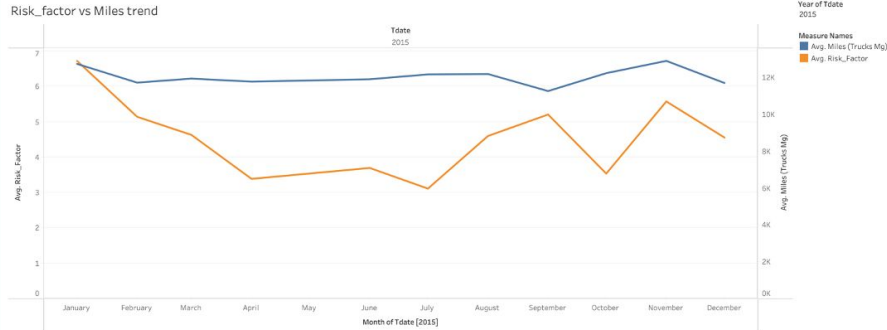- Ford and Peterbilt models are prone to accidents and needs the companies attention.

- A97 & A73 are the most unsafe drivers with 14 & 10 events respectively.

- We can get all the information for a particular driver with the overall table by using actions in tableau.

Risk Factor Prediction

Understanding the hindsight & insight from the data

Risk_factor vs Miles trend

Tdate
2015

Year of Tdate
2015

Measure Names
■ Avg. Miles (Trucks Mg)
■ Avg. Risk_Factor

Miles vs Mpg Trend

Regression Line

Actial vs Predicted

Driverid (Ri...	Regression along Driverid (Riskfactor)	Risk_Factor
A99	4.02	4.01
A98	6.06	6.07
A97	20.09	20.35
A96	2.02	2.01
A95	8.00	7.99
A94	5.75	5.69
A93	5.98	5.97
A92	7.18	7.27
A91	1.95	1.98
A90	4.88	4.87
A9	5.84	5.79
A89	1.93	1.97
A88	4.84	4.83
A87	5.84	5.79
A86	7.10	7.14
A85	5.05	5.05
A84	5.94	5.91
A83	2.93	2.95
A82	2.08	2.03
A81	5.98	5.97
A80	4.72	4.71

Trend Analysis:

- In Risk Factor and Miles, we can see that the trend of risk factor is not stationary. There are sudden spikes in the last 2 quarters of 2015.

- We can see the miles and mpg trend for the most risky driver: A97. This is an indication of rash driving.

- Regression Analysis predicted the Risk_factor based on events and total miles. Each had a significant effect on Risk Factors.

Regression   trucks (default.trucks)+ (default)
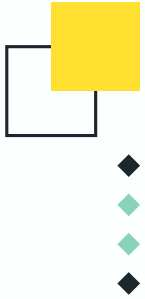
Results are computed along Table (across).

```
SCRIPT_REAL(
  "
  y <- .arg1;
  x1 <- .arg2;
  x2 <- .arg3;

  reg <- lm(y ~ x1 + x2)
  reg$fitted
  ", SUM([Risk_Factor]), SUM([Events]), SUM([Totmiles (Riskfactor)]))
```

Conclusion

- Identifying potential hazards that may lead to accidents is essential in any line of business.

- Through this risk analysis, we identified various factors such as driver fatigue, vehicle maintenance, and weather conditions that increase the risk of accidents.

- Data visualization (Using Tableau) is used to develop many models/visuals to compare the accidents by risky and non-risky drivers, accidents in major cities and companies, percentage of cause of accidents etc.

- This information can be used to develop strategies to increase company profits (cost-cutting), to mitigate potential risks, such as implementing regular vehicle maintenance schedules, providing driver training programs, and monitoring driver fatigue levels and hence can reduce the likelihood of accidents and ensure the safety of both the truck drivers and other road users.

Thank you!