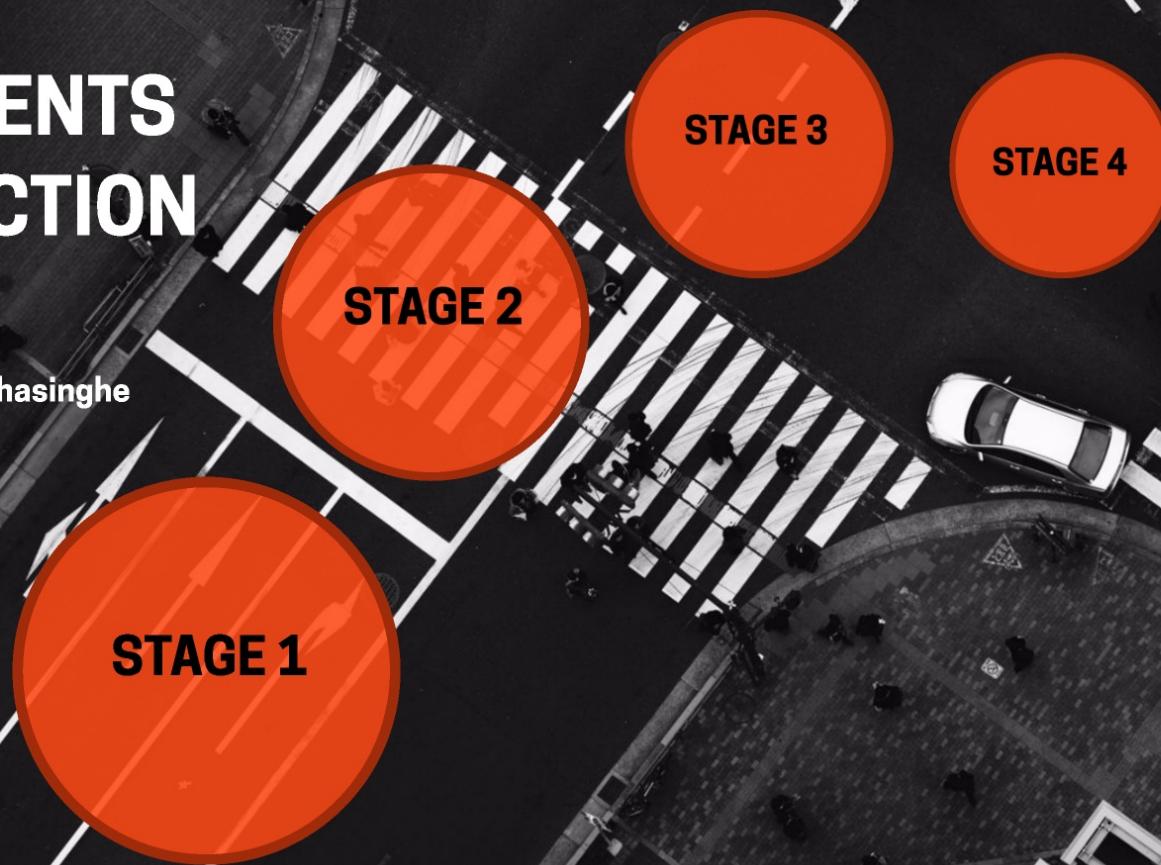


ROAD ACCIDENTS PREDICTION

By Winnie Wetthasinghe
04/06/2021





OVERVIEW

- Domain Context
Topic introduction
- Business Problem Definition
Business vs Data problem
- Data
- Values Delivery
EDA, Modelling and Results
- Conclusions & Future Recommendations
Takeaways & future tasks

Domain Context

Business Problem Definition

Stakeholders

Domain Context- Two Aspects



Leading causes of
deaths and injuries
worldwide

1

1.25 million
deaths per year

Public health problem
&
Socioeconomic
development issue

Do you know ?



2
Car insurance
companies
evaluate risk
factors when
estimating a quote
and making a claim

Get Your Instant Quote Now!

Full Name	Phone	Email
<input type="text" value="John Doe"/>	<input type="text" value="555-1234"/>	<input type="text" value="johndoe@example.com"/>
State		Date of Birth
<input type="text" value="California"/>		<input type="text" value="06/27/1995"/>
Gender	Tobacco Use	Health Class
<input checked="" type="radio"/> Female	<input checked="" type="radio"/> Yes	<input type="radio"/> No
Preferred plan		
Term Period	Coverage Amount	
<input type="text" value="20 year term level"/>	<input type="text" value="\$500,000"/>	<input type="text" value="\$0"/>
<input type="button" value="Display Quotes"/>		
<small>View Privacy Statement</small>		

Keep up with
changing
conditions?

Road Safety Strategies

Insurance risk factors

Business Problem

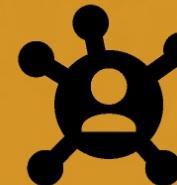
vs.

Data Problem

- How can we reduce road mortality and injury?
- What factors have most effect on mortality?

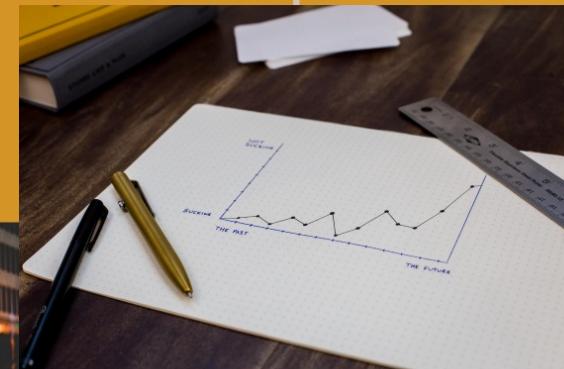


- What is the measure that can be used?
Severity of an accident
- Data Science project finds the most suitable ML model that can predict severe accidents.
- Apply reverse engineering to find most relevant factors that cause severe accidents.



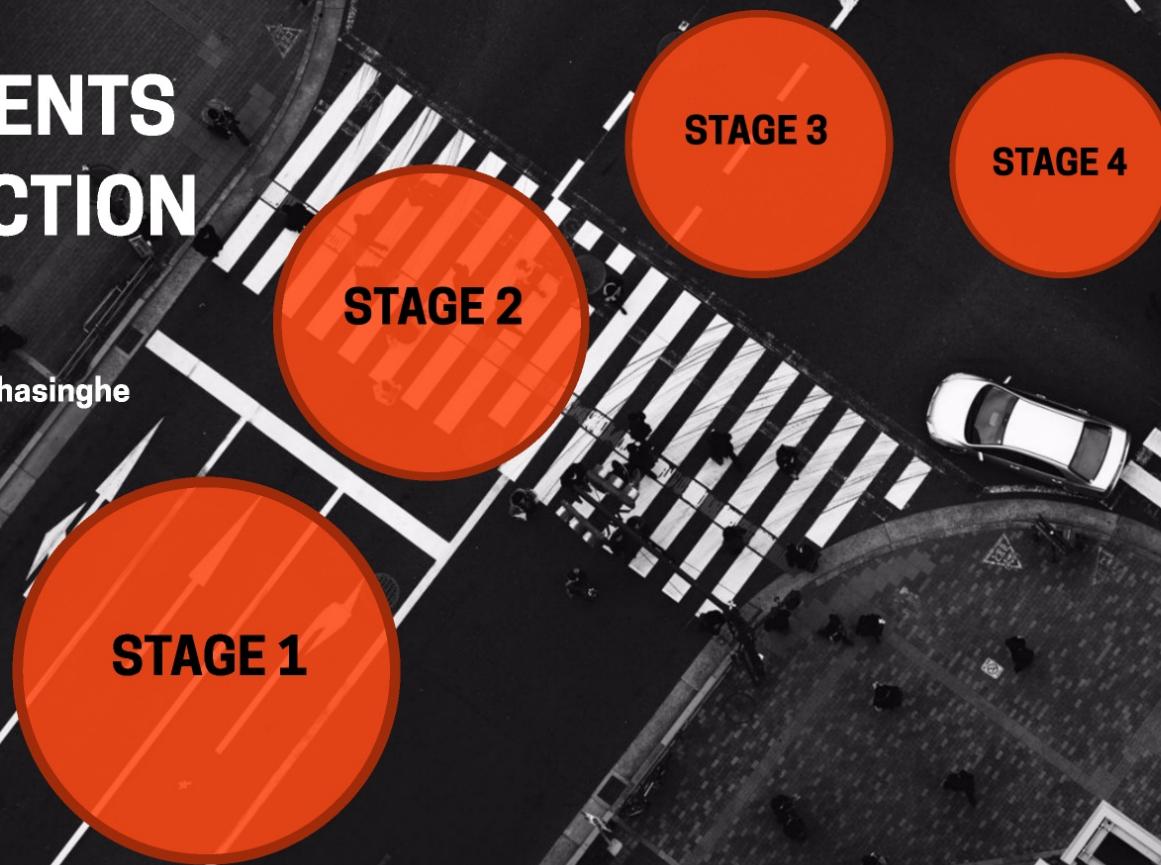
Stakeholders

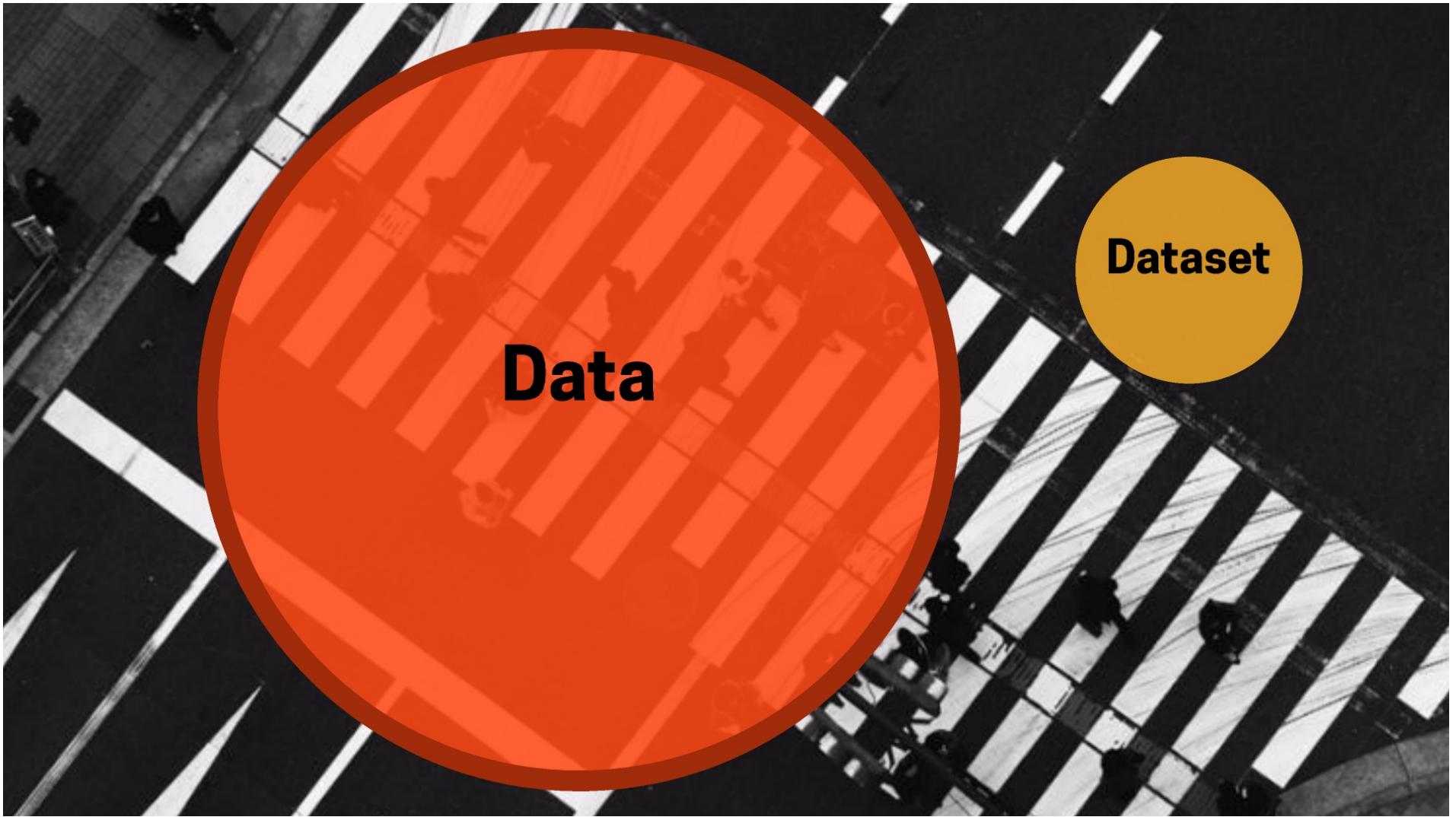
- State Police Departments.
- Departments of Infrastructure and Transport countrywide
- Car Insurance companies
- Research Companies
- Emergency hospital services (Ambulances)



ROAD ACCIDENTS PREDICTION

By Winnie Wetthasinghe
04/06/2021

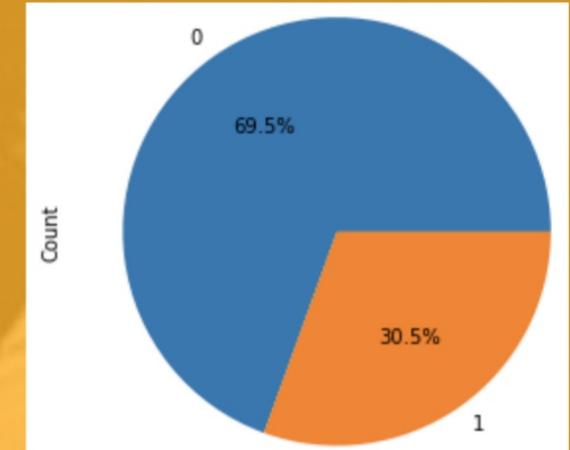
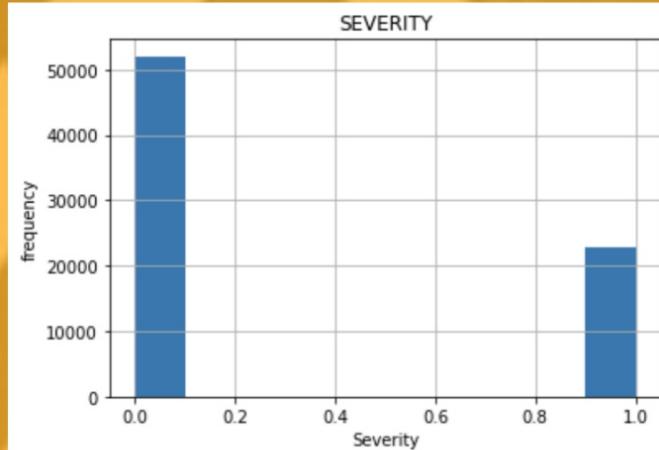




About dataset

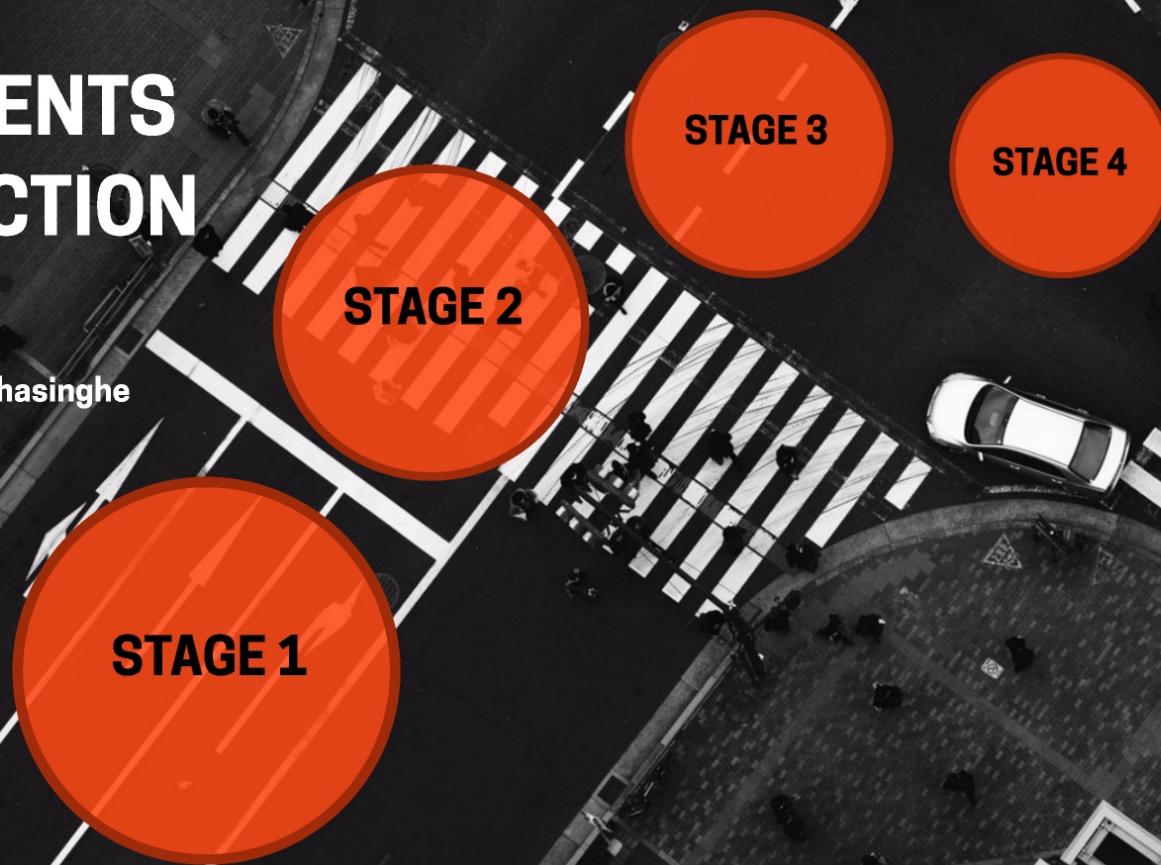
Target variable

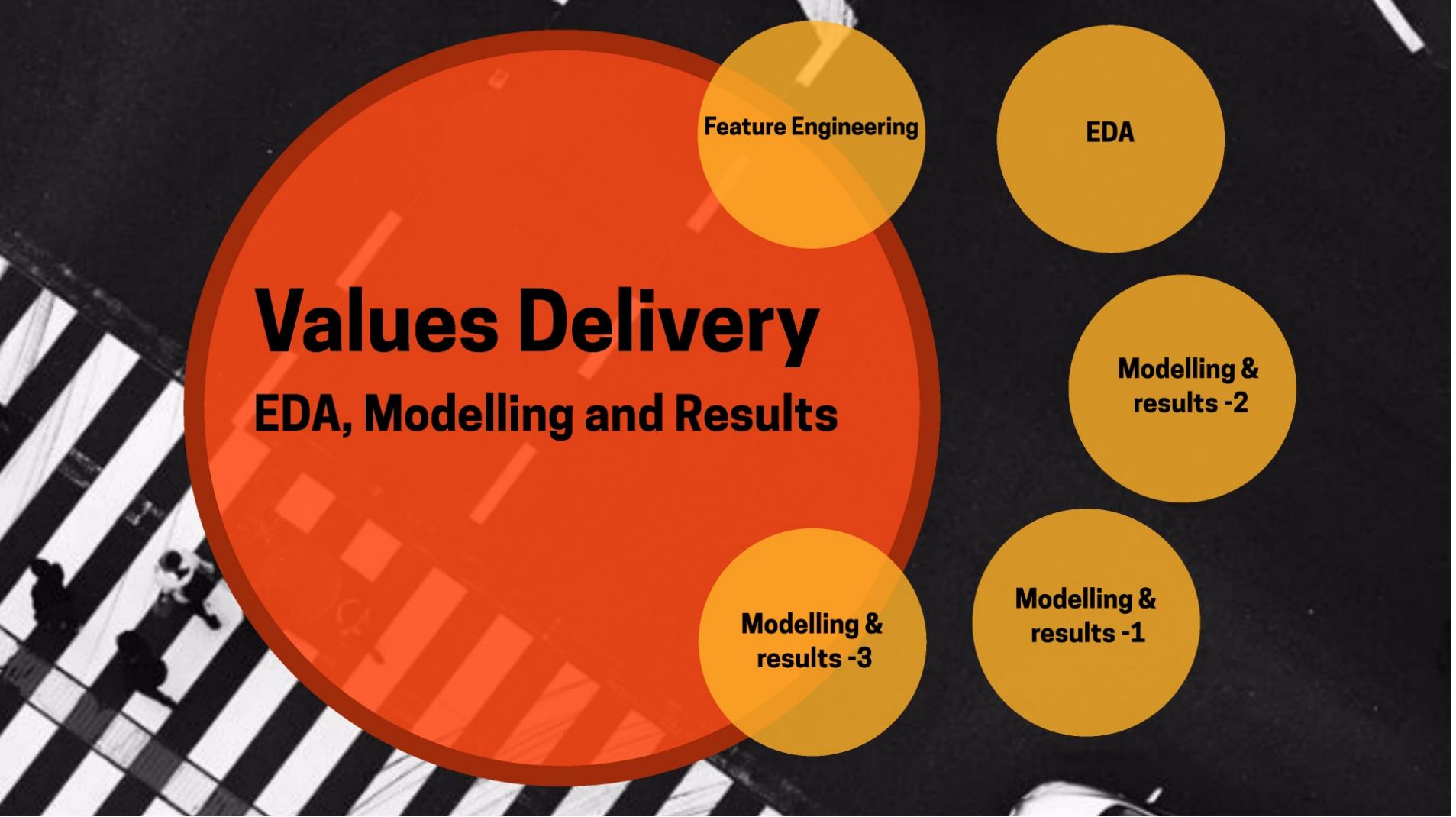
- Crash Statistics from 2013 - 2019 -Victoria State
- Original dataset from VicRoads.gov.au also available at Kaggle.com
- Features such as traffic composition, weather conditions, roads conditions, demography and drivers & passenger details etc.



ROAD ACCIDENTS PREDICTION

By Winnie Wetthasinghe
04/06/2021





Values Delivery

EDA, Modelling and Results

Feature Engineering

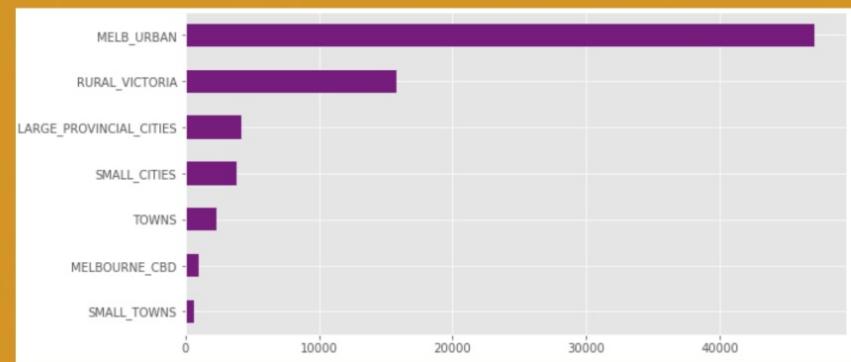
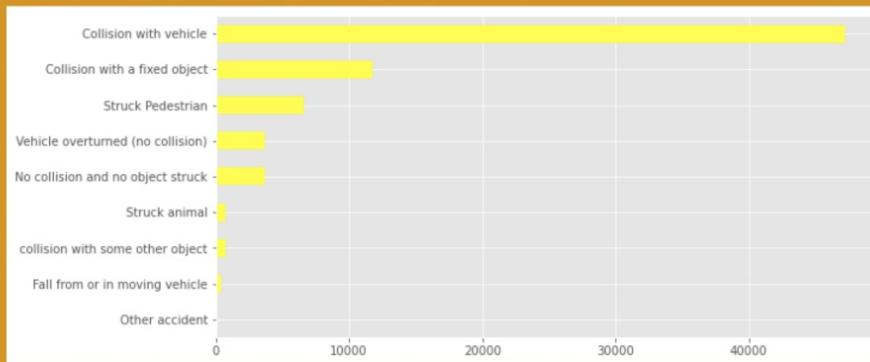
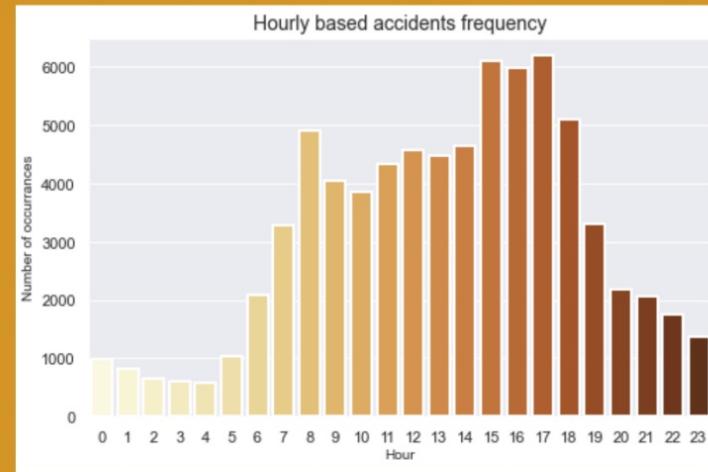
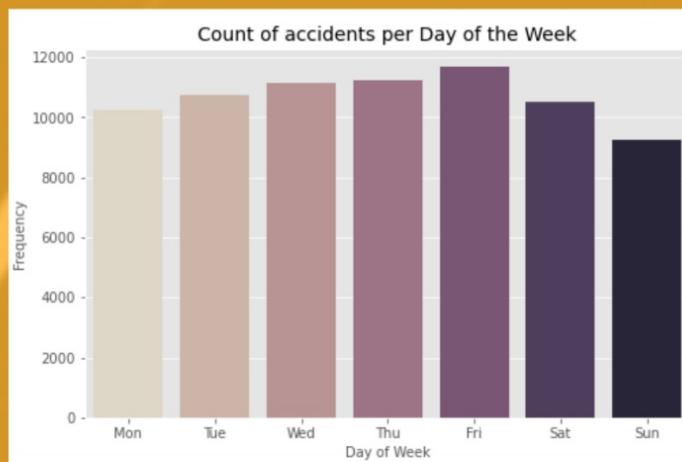
EDA

Modelling &
results -2

Modelling &
results -3

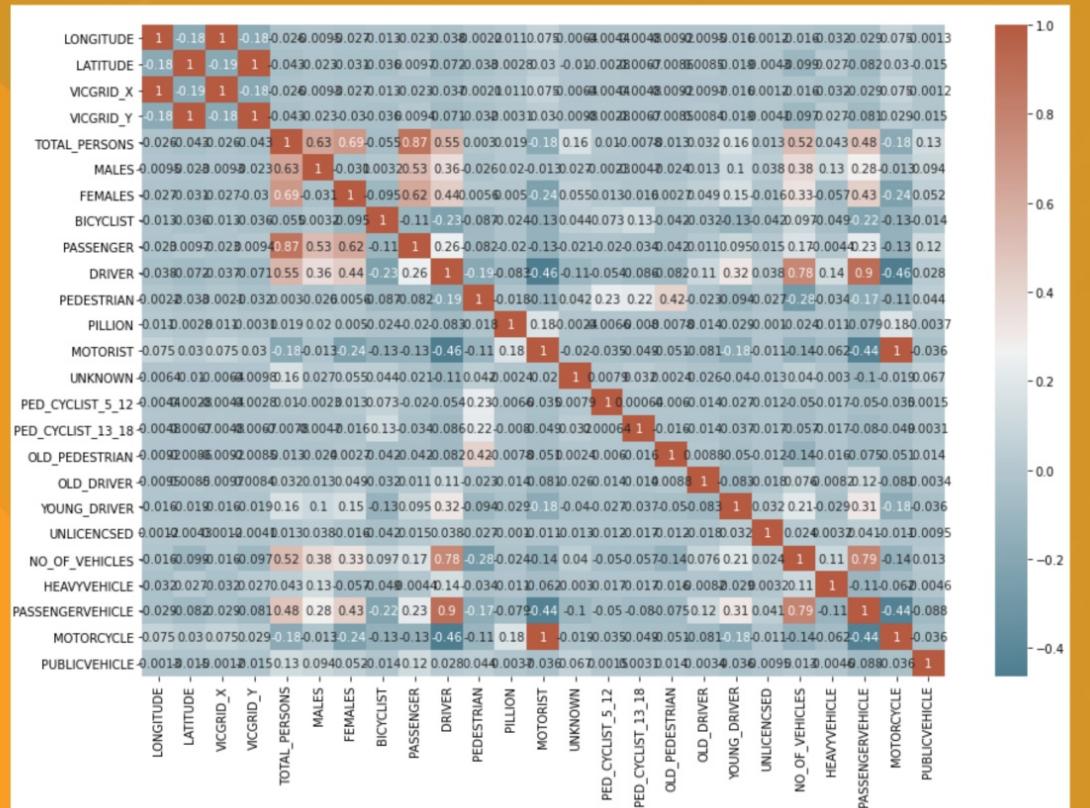
Modelling &
results -1

Some insights from EDA

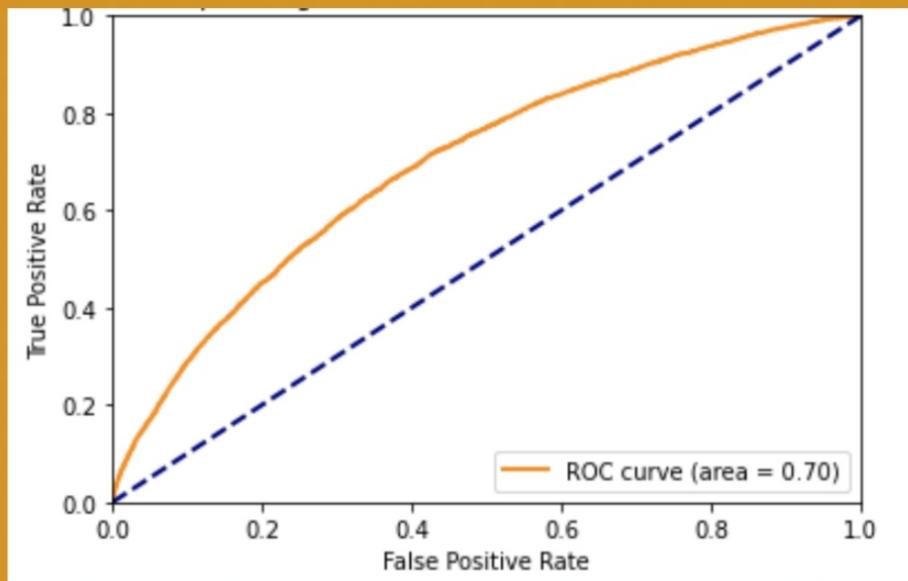


Feature Selection and Feature Engineering

- No interesting highly correlated features
- Most of the features are categorical in nature
- Introduced new features by producing dummy variables.

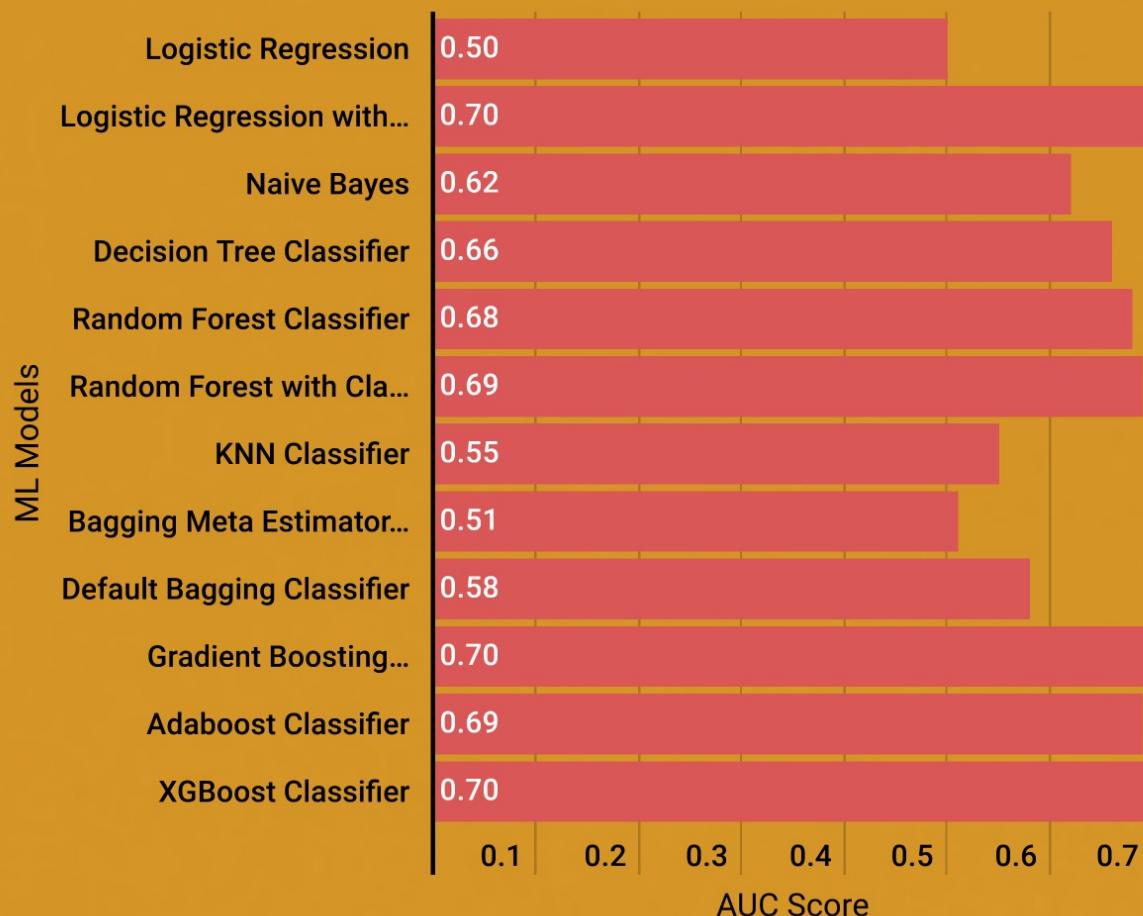


Model Evaluation with selected Performance Metric

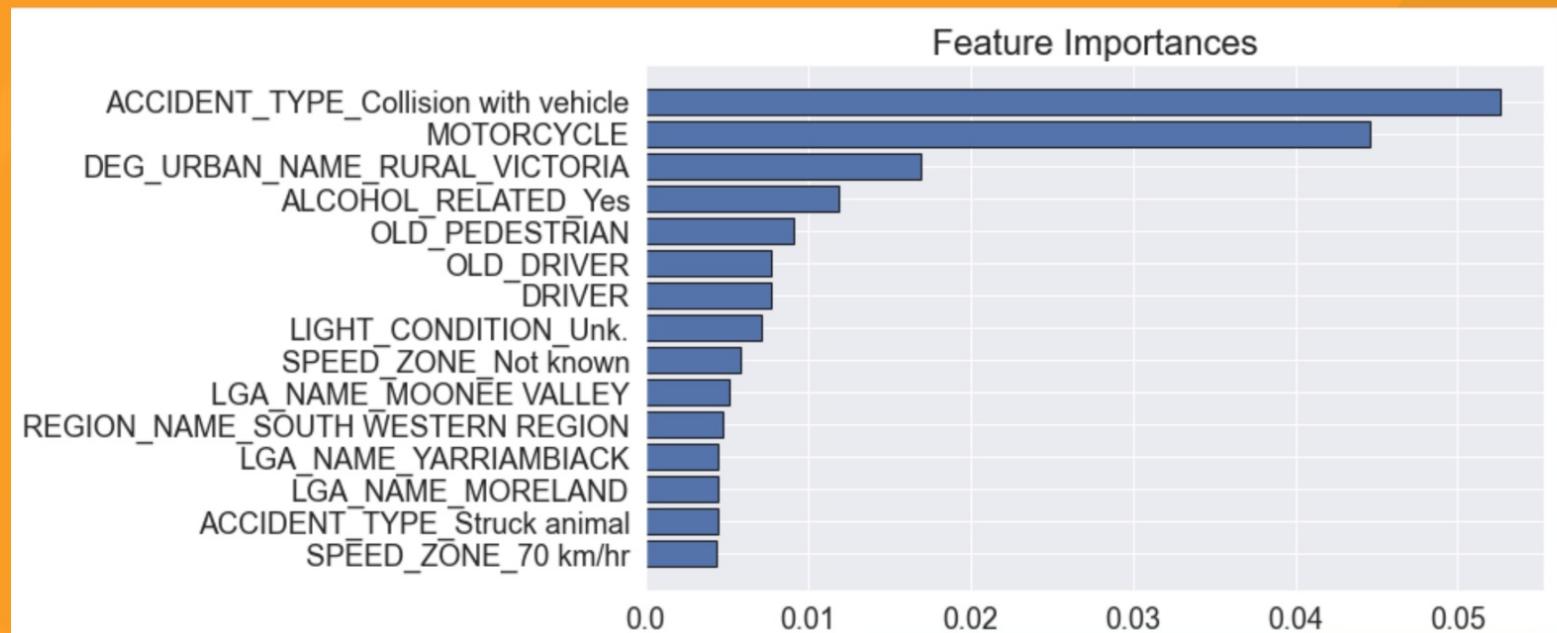


- ROC curve is the metric that assess the model ability to distinguish between binary (0 or 1) classes.
- Between 0.4 - 0.6 FPR and 0.6 - 0.8 TPR gives the best balance between TP and FP.
- TPR- Probability of detection.
- FPR-Probability of false alarm.

ML Model Performance

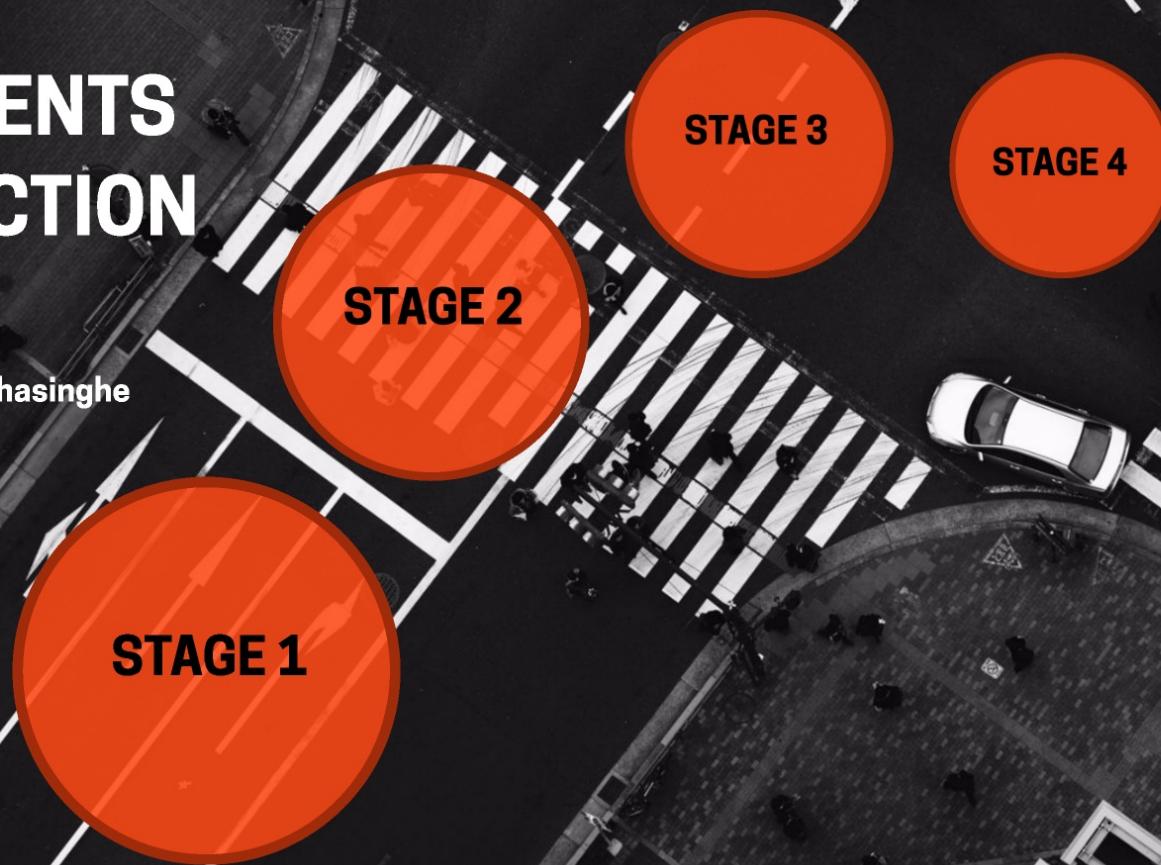


Selected Model Results



ROAD ACCIDENTS PREDICTION

By Winnie Wetthasinghe
04/06/2021



Conclusions & Future Recommendations

Takeaways & future tasks

Conclusions

Thank you!

Future Tasks

References

Conclusions

- Logistic Regression, Gradient Boost and XGBoost Classifiers perform well.
- Two most crucial contributing factors that are affecting the severity level of accidents are **collision with another vehicle** and accidents involved in **motorcycles**.



Future Tasks

- Perform more hyperparameter tuning for the selected models and feature engineering work to increase scores.
- Apply deep learning techniques to further implementations of models
- Incorporate NLP to detect recent traffic information such as social media related & open data sources to gather data.



References



- [https://reader.eelsevier.com/reader/sd/pii/S209575642030101X?token=3E750E36FF75A24665601D09A2597DC8572E83CDEC34DC69480A5D842B29DADE5E21437CB57FA76F0BDB39F57866F828&originRegion=us-east-1&originCreation=20210530224527](https://reader.elsevier.com/reader/sd/pii/S209575642030101X?token=3E750E36FF75A24665601D09A2597DC8572E83CDEC34DC69480A5D842B29DADE5E21437CB57FA76F0BDB39F57866F828&originRegion=us-east-1&originCreation=20210530224527)
- <https://github.com/NFaraji/An-analysis-of-Victoria-crash-data/blob/master/Insight-Crashes.ipynb>
- <https://www.kaggle.com/gaurav896/victoria-state-accident-dataset>



ROAD ACCIDENTS PREDICTION

By Winnie Wetthasinghe
04/06/2021

