DATA INCUBATOR

CAPSTONE PROJECT

On

PEDESTRIAN CRASH DATA ANALYSIS
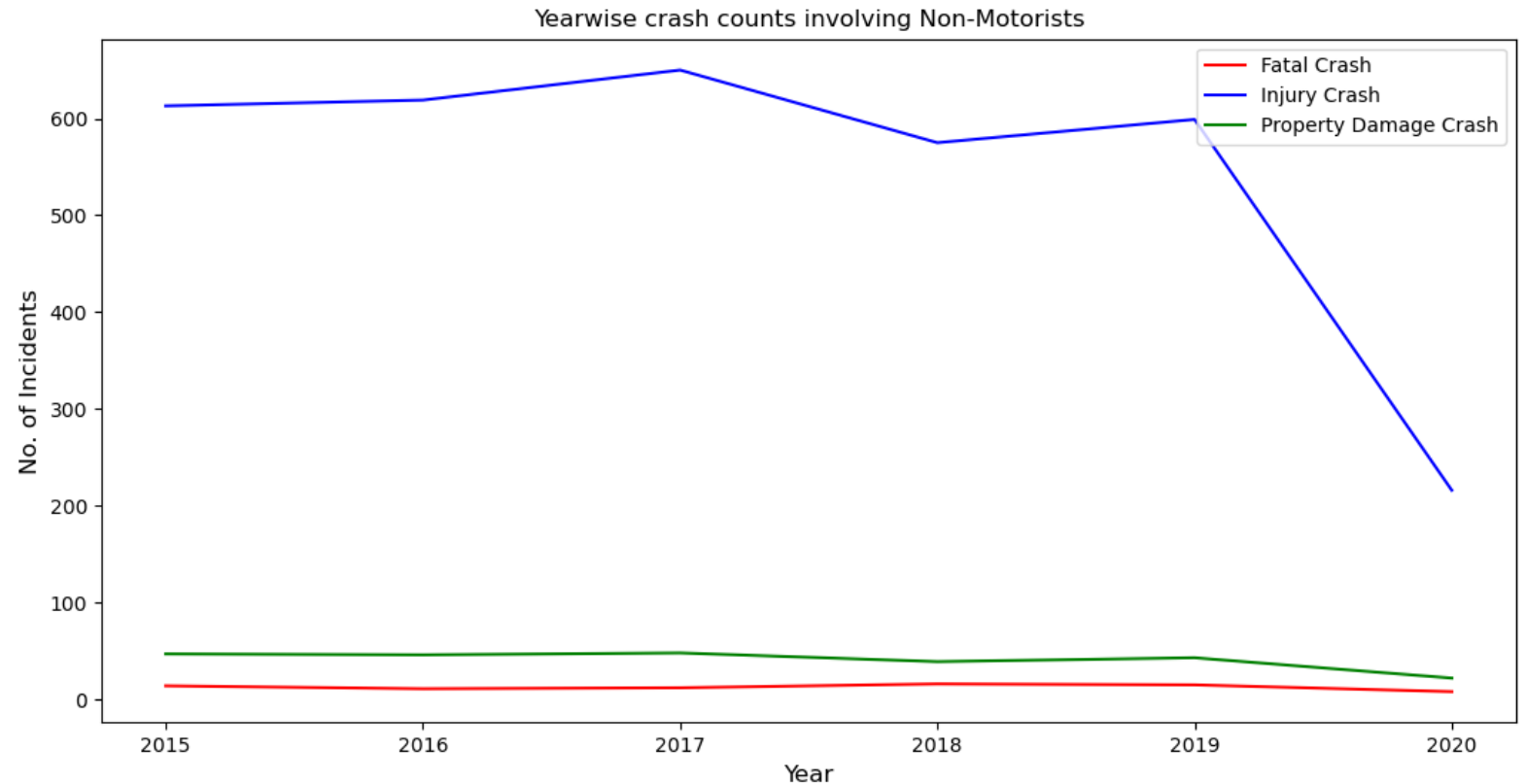
By

AZHAGAN (**AZY**) AVR

Motivation

- In the US non-motorist are given the right-of-way (ROW) at all instances, even if the motorist is in the ROW.

- It is an undeniable fact that the non-motorists are more susceptible to injuries than a motorist, hence the ROW rule.

- But with such a rule in place ignorance can also play a part while being a non-motorist and an undue advantage can be taken by a non-motorist.

- Hence, I am using a crash dataset to analyze the incidents involving motorists and non-motorists (Pedestrians and Bi-cyclists).
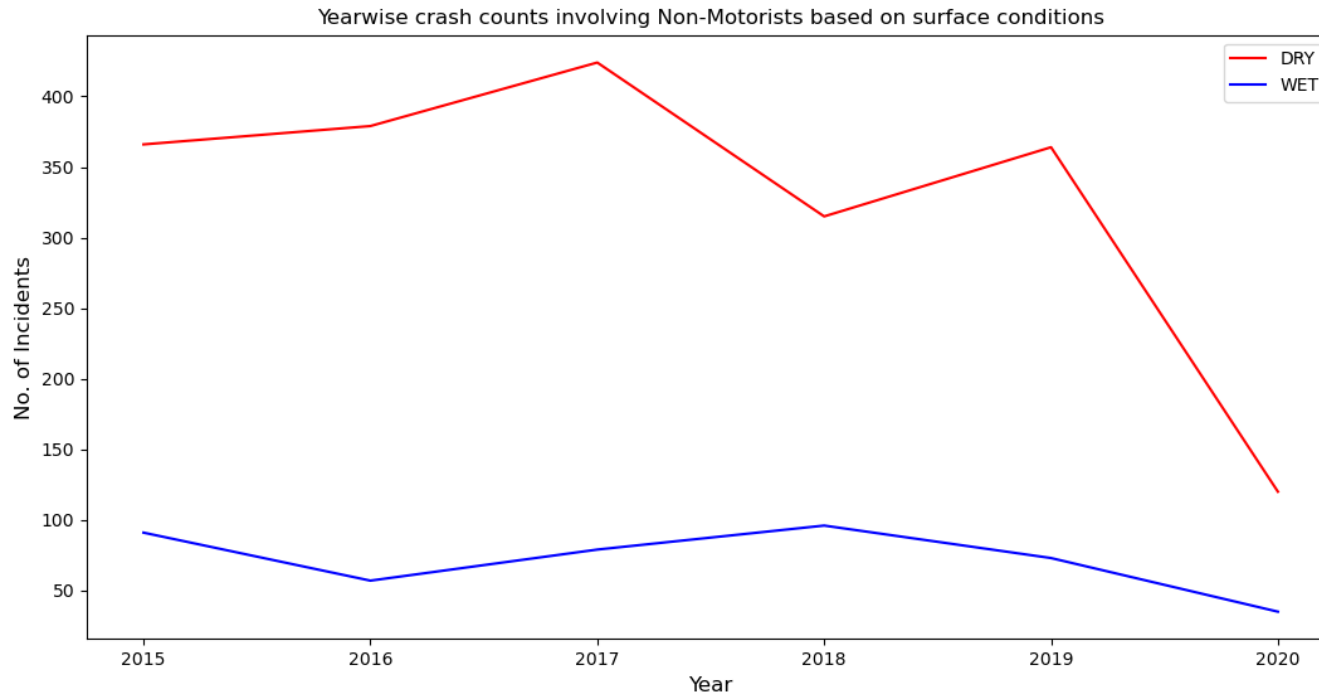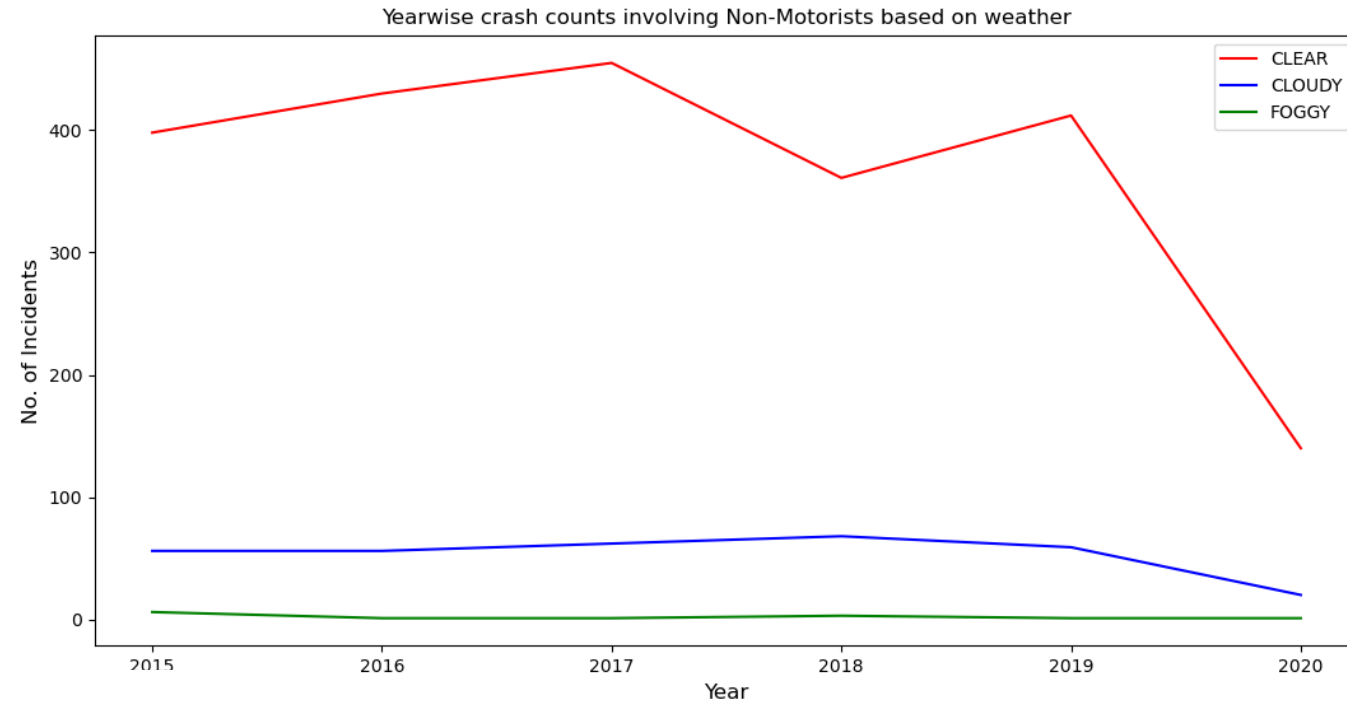
Crash Incidents

- The dataset consists of data from 2015 to 2020, for the city of Montgomery in Maryland.
- Crash types are classified as:
  - Fatal crash,
  - Property damage crash
  - Injury crash.
- All these crashes shown involve pedestrians.

Yearwise crash counts involving Non-Motorists

https://data.montgomerycountymd.gov/api/views/n7fk-dce5/rows.csv?accessType=DOWNLOAD

External Factors

WEATHER

- A common assumption is that foggy conditions, reduce visibility and crashes tend to increase.
- But in case of pedestrian related crashes, a greater number of crashes take place when the skies are clear.
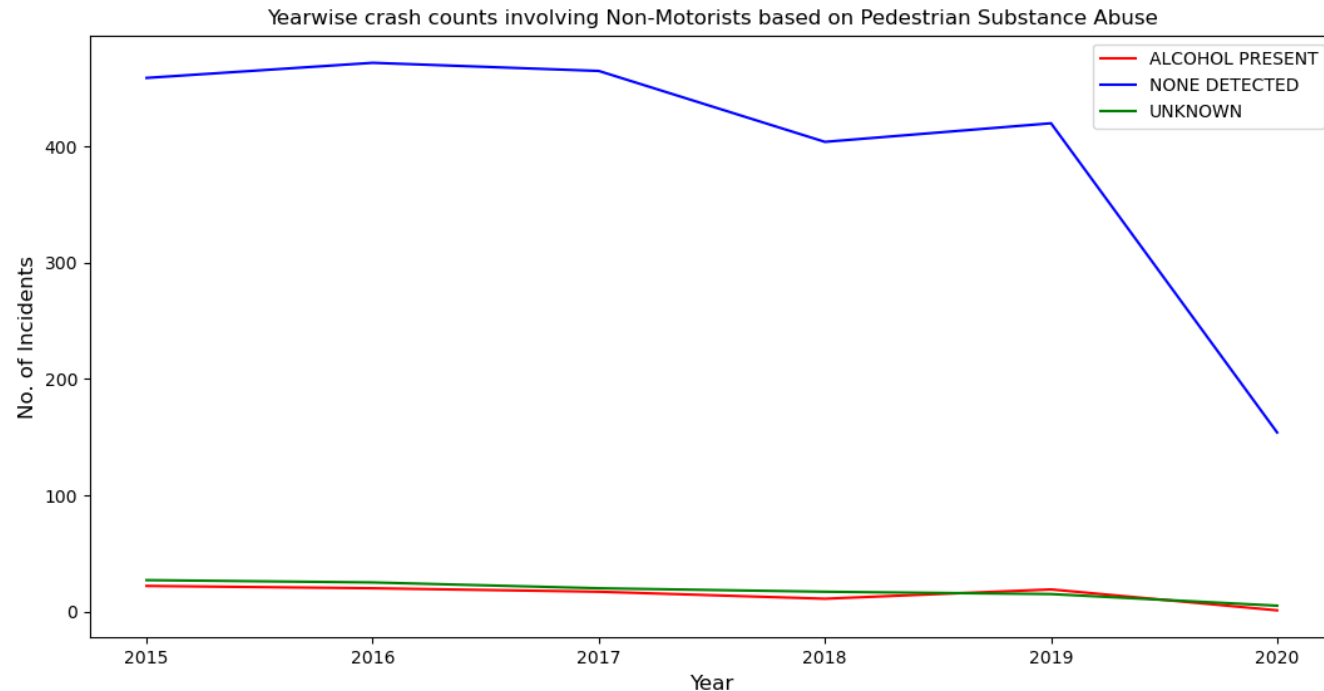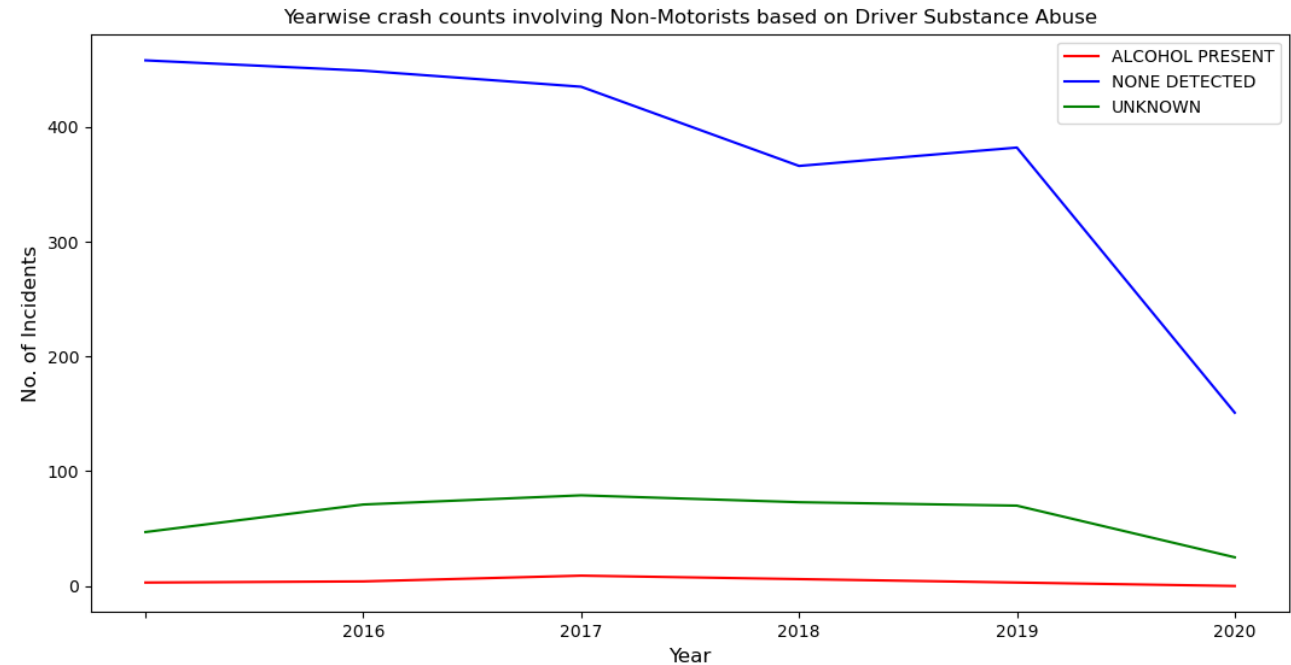

Yearwise crash counts involving Non-Motorists based on weather

Yearwise crash counts involving Non-Motorists based on surface conditions

SURFACE

- Likewise, more accidents occur on dry roads than wet roads.

Internal Factors

DRIVER UNDER INFLUENCE

- Also another common assumption is that, alcohol consumption might result in high number of crashes.

Yearwise crash counts involving Non-Motorists based on Driver Substance Abuse

Yearwise crash counts involving Non-Motorists based on Pedestrian Substance Abuse

PEDESTRIAN UNDER INFLUENCE

- Both pedestrian and driver substance abuse has very less impact, when compared to cases without substance abuse.

Conclusion

- Data is like a mirror that projects the reality.
- In this case the external factors like wet roads or foggy weather have least impact on pedestrian related crashes and internal factors like substance abuse have a similar effect.
- Hence it is a bit premature to say that this is the case for across the US.
- More in-depth research is needed to identify the causes of the pedestrian related crashes.

Future Scope

- Location based analysis like, pedestrian at signal or sidewalk or cross walk, etc.
- Geospatial information is also available in the dataset, which can be used to identify hotspots.
- Data.gov has close to 355 datasets related to pedestrian related crash data.
- Analyze data from different cities.
- Also would like to explore some predictive and regression techniques to analyze the data.

Tools Used

- Pandas
- Numpy
- Matplotlib

Thank You