ENEA Test Case

Data Science
23/11/2022

Georgia Tech

Client & Issue Raised

Georgia Tech

Client & Issue Raised

- Client: Australian Road Safety

- Issue Raised: How to reduce the number of accident while developing new infrastructures ?

  - Important question since the infrastructure development is really important in Australia.

  - Investment plan of AUD 100 billion *(Source: https://www.infrastructureaustralia.gov.au/australian-infrastructure-audit-2019-executive-summary)*

Georgia Tech

Solution Process

Solution Process (1/2)

- ACCIDENT Datasets provided freely by the Victorian Government (public)

- ACCIDENT Folder (crash from 2000 to 2020):
  - *ACCIDENT.csv -> basic accident details (time, severity, …)*
  - *PERSON.csv -> person based details (age, sex, …)*
  - *VEHICLE.csv -> vehicle based data (type, year of manufacturing, …)*
  - *ROAD_SURF_COND.csv -> (wet road, dry, icy, …)*
  - *ATMOSPHERIC_COND.csv -> (rain, winds, …)*
  - *NODE.csv*
  - …

- First inspection of these datasets:
  - Duplicates -> remove duplicated samples
  - Datatypes -> some conversion required (text format -> date format)
  - Missing Values -> drops and imputation (mean or median strat)

Georgia Tech

Solution Process (2/2)

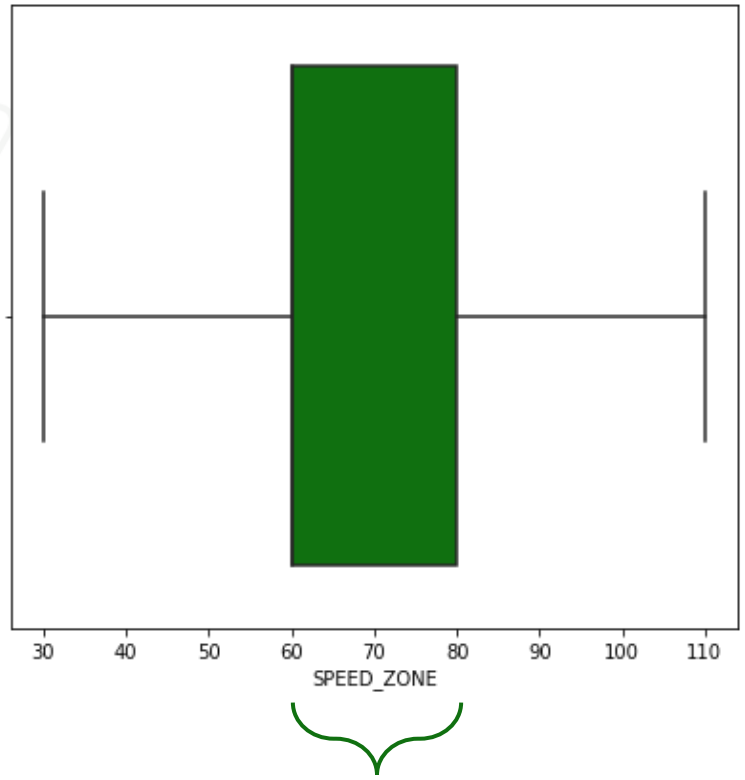
- Feature identification for Data Analysis.

- Creation of a new Dataset with new variables for Data Analysis

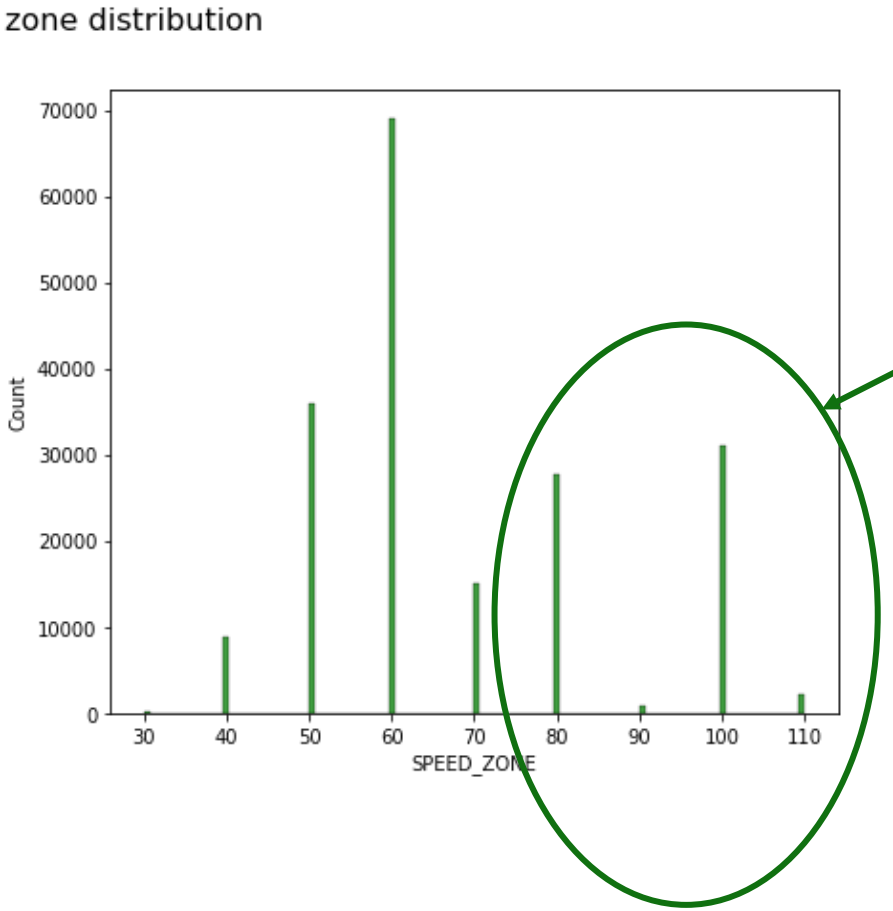
Data Analysis

Data Analysis (1/11)

- Speed Zone Distribution



50% of the crashes

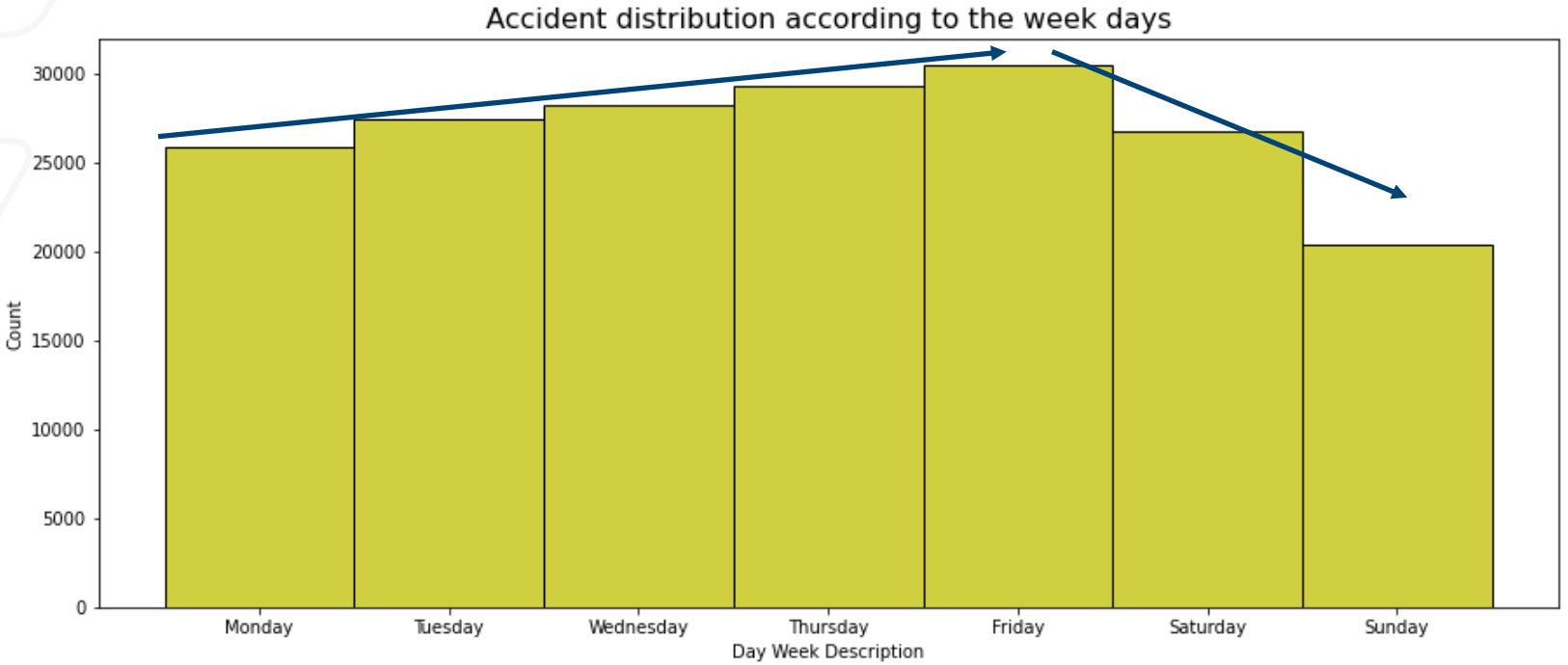


Important difference between 90 km/h and 100 km/h

Most of the major highways are limited to 80 km/h or 100 km/h

Data Analysis (2/11)

- Day Week Distribution



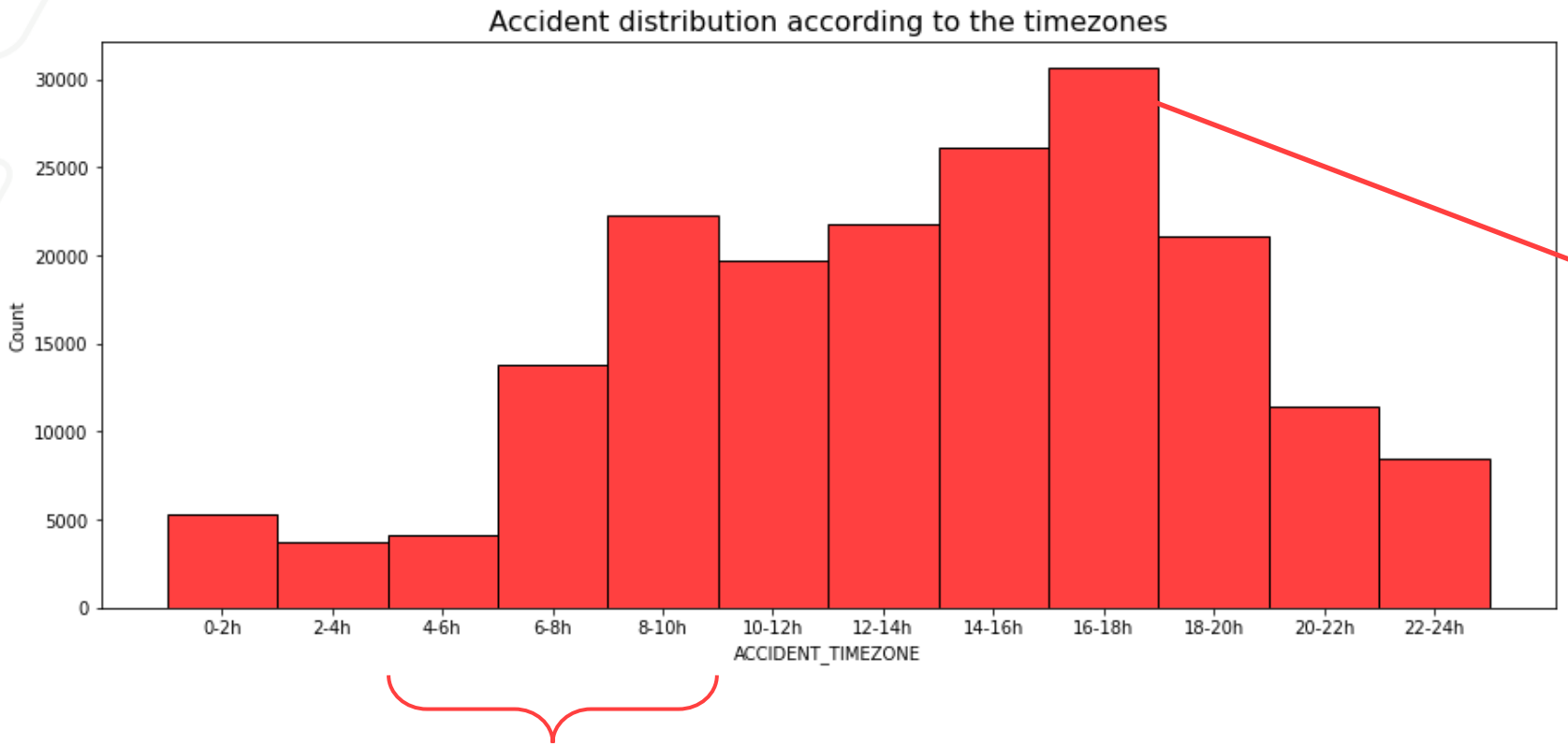
Relevant descrease on Saturdays and Sundays => people drive less during the weekend

Constant augmentation of crashes during working days (fatigue accumulation ?)

Georgia Tech

Data Analysis (3/11)

- Time Zone Distribution



Highest frequency between 16-18h

Rush hours corresponding to the end of the working day.

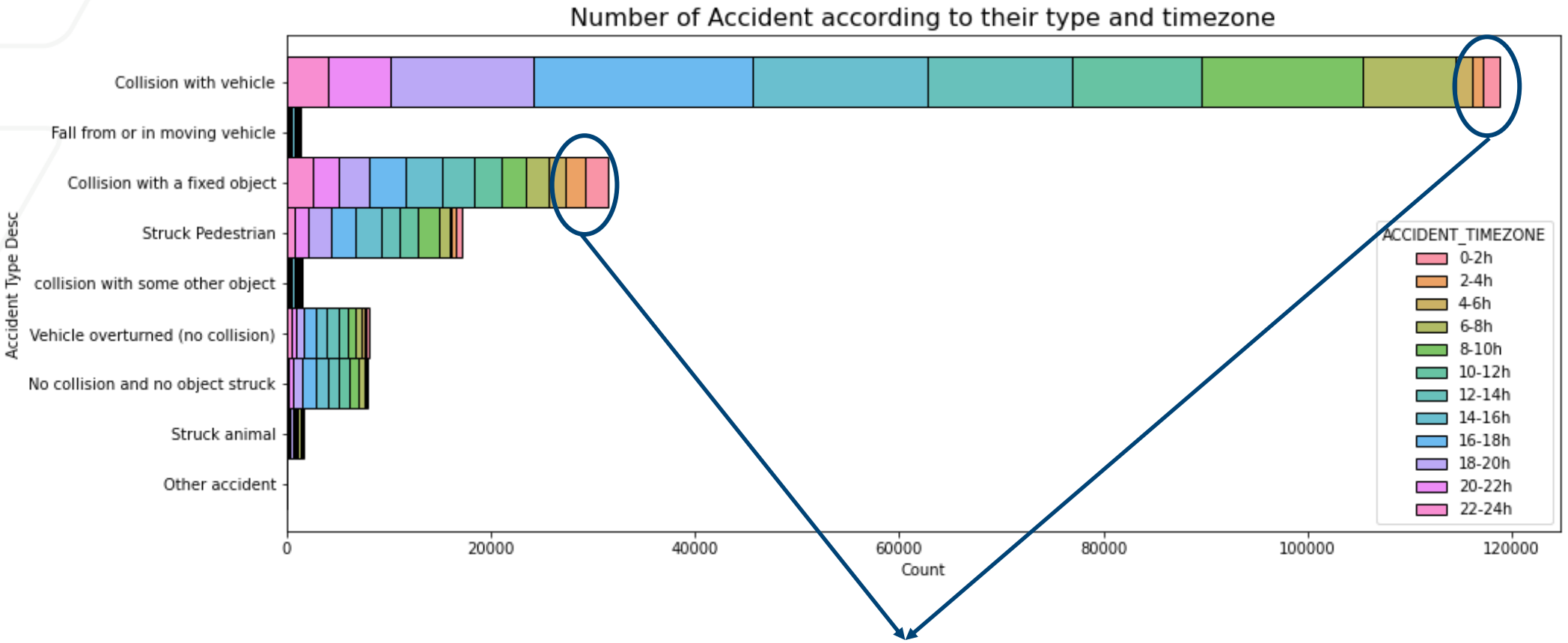
Significant rise => start of working hours

Georgia Tech

Data Analysis (4/11)

- Accident Type

Collision with vehicle main accident type

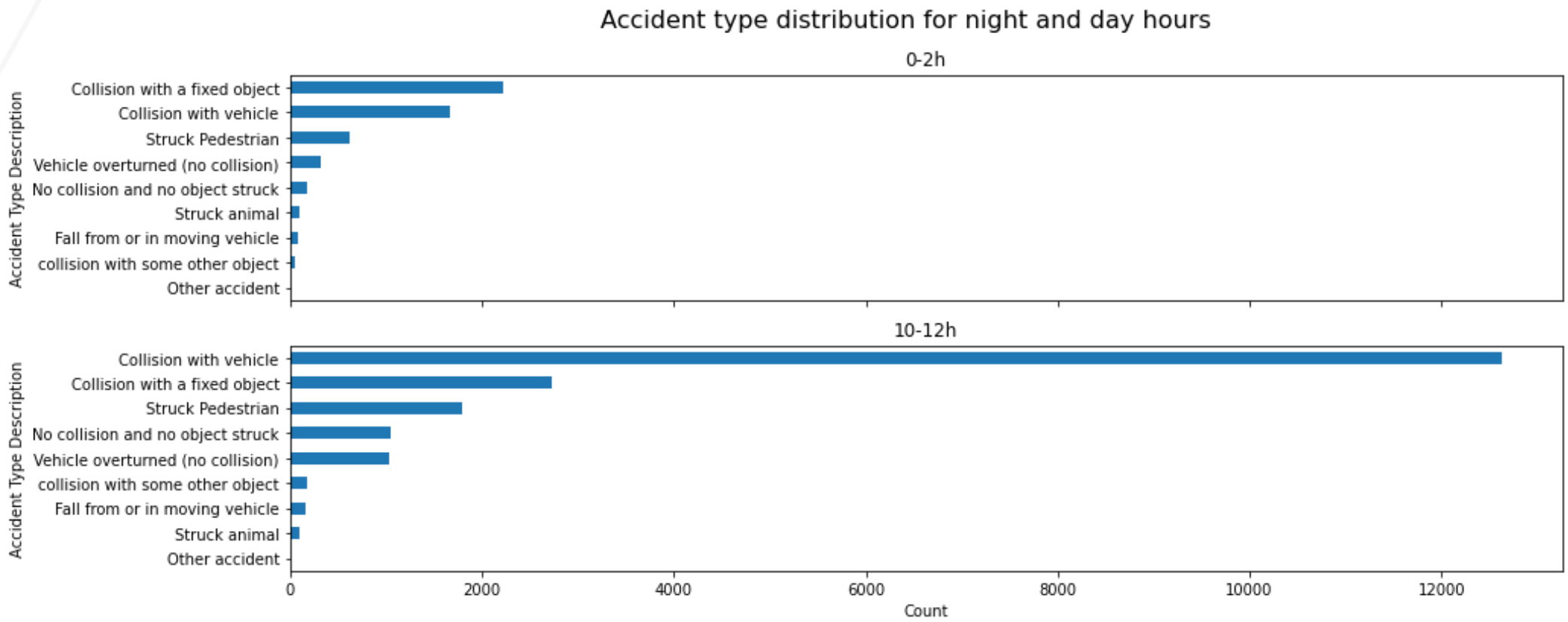


Collision with a fixed object for night hours

Georgia Tech

Data Analysis (5/11)

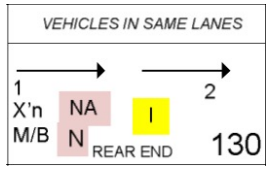
- Accident Type

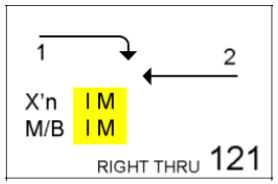


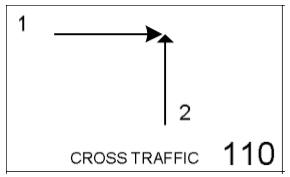
Georgia Tech

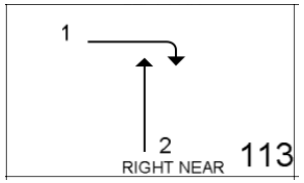
Data Analysis (6/11)

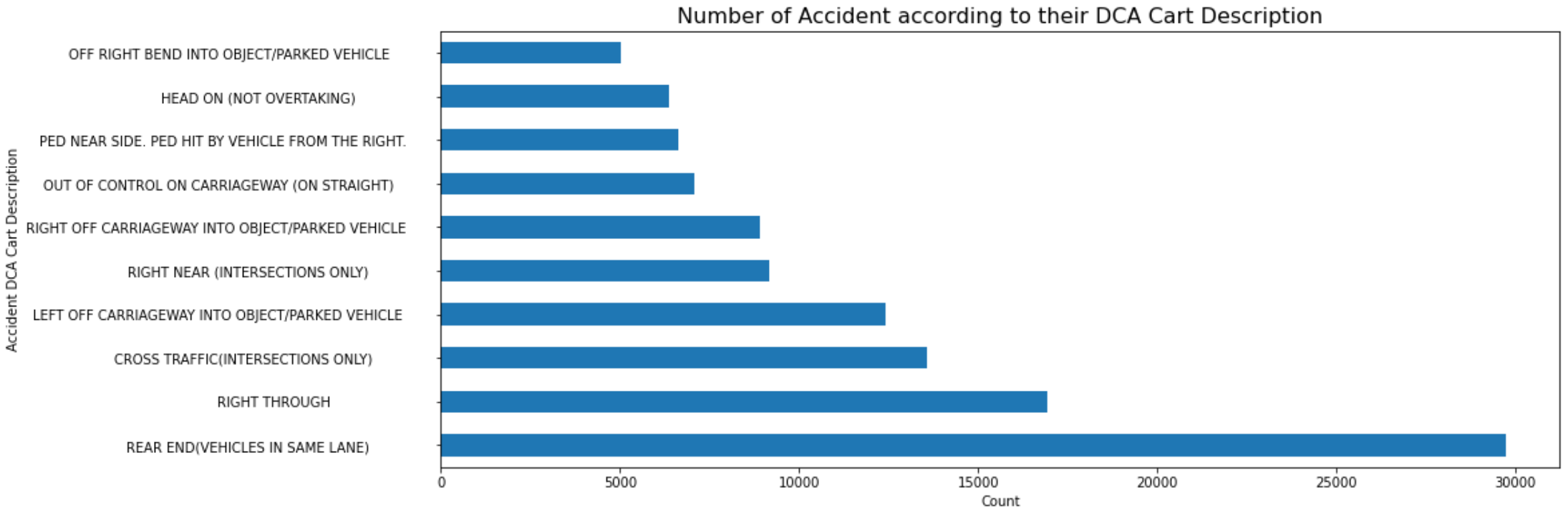
- DCA Cart





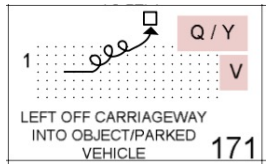


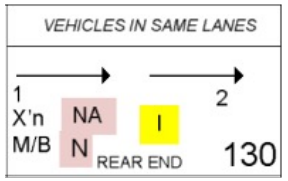


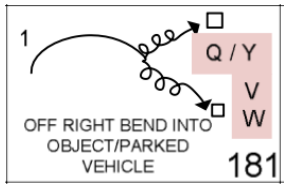


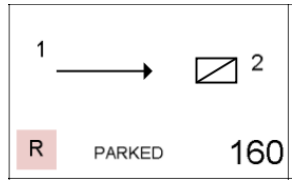
Data Analysis (7/11)

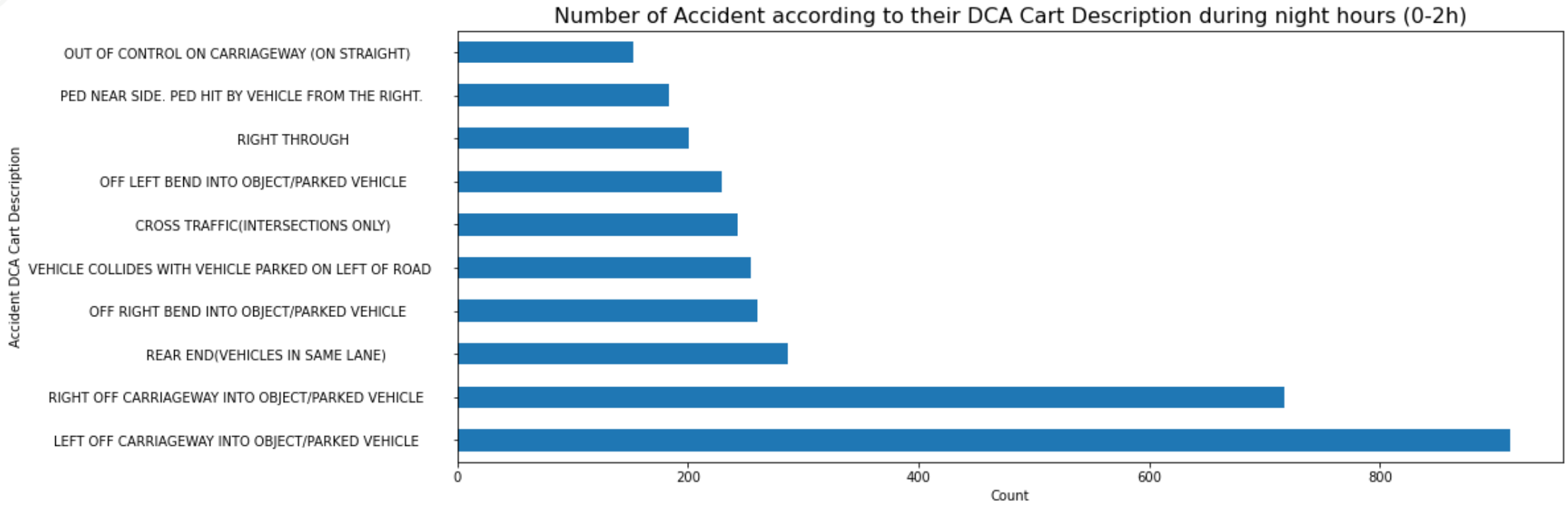
- DCA Cart





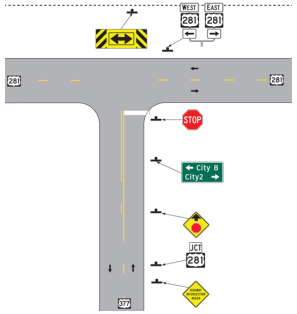


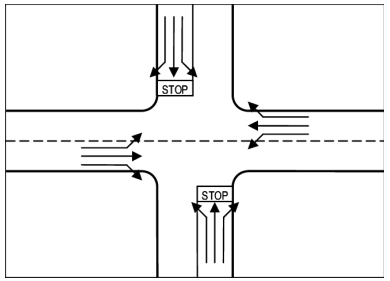


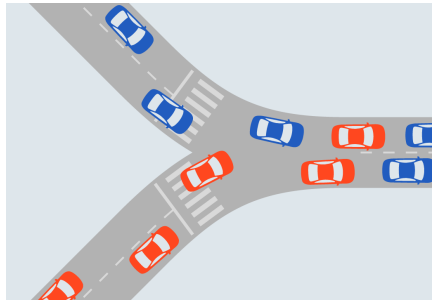


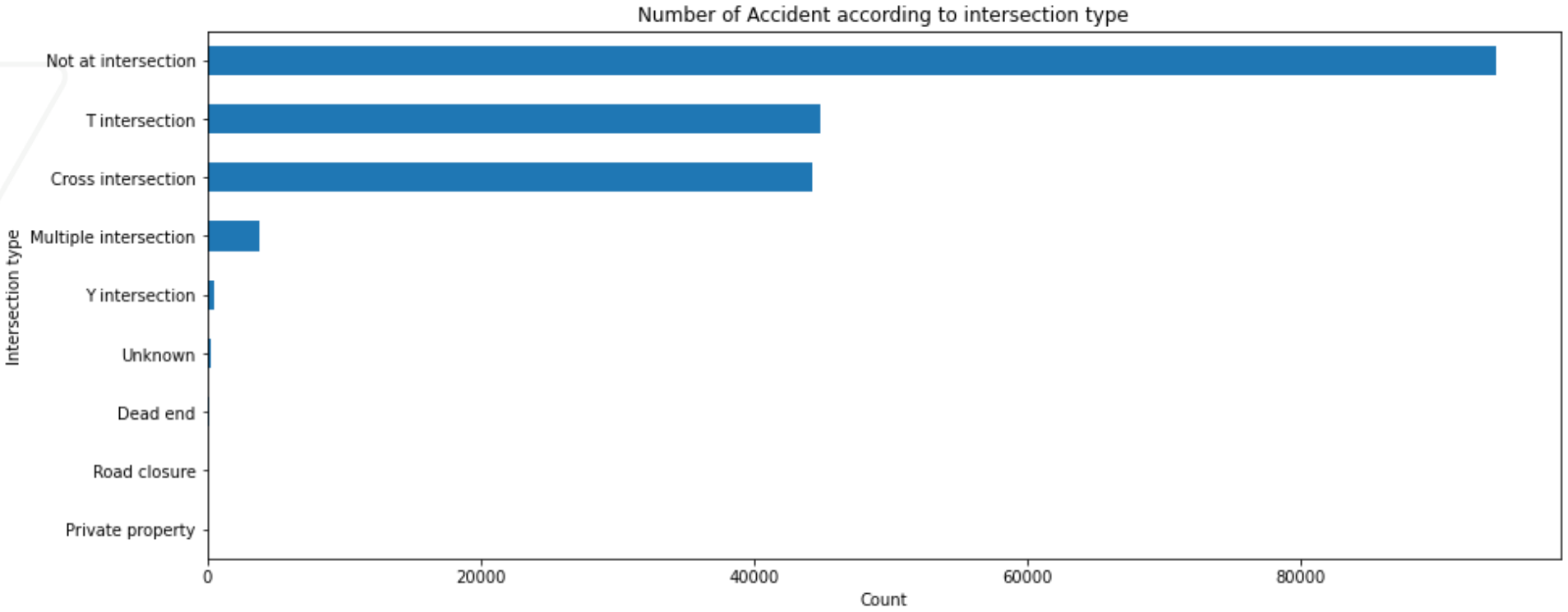
Data Analysis (8/11)

- Intersection type



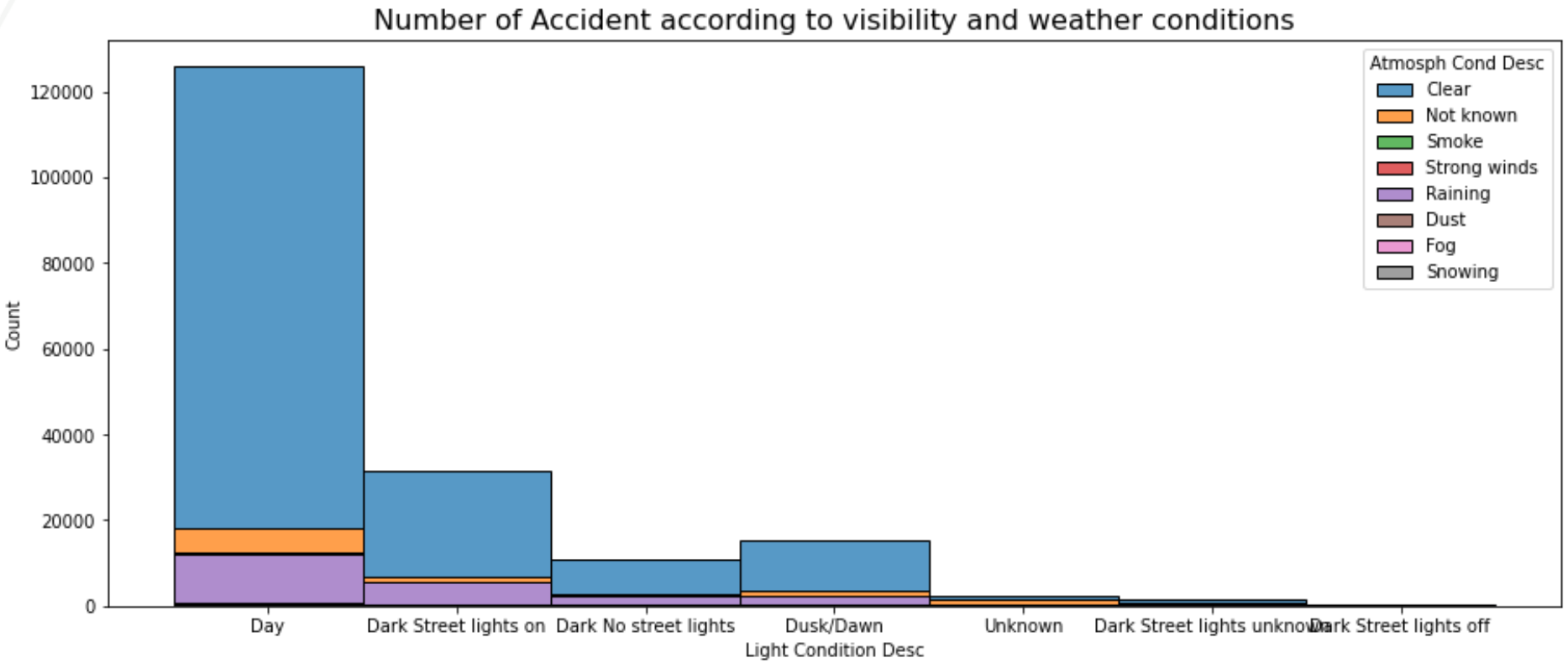






Data Analysis (9/11)

- Weather and Light conditions

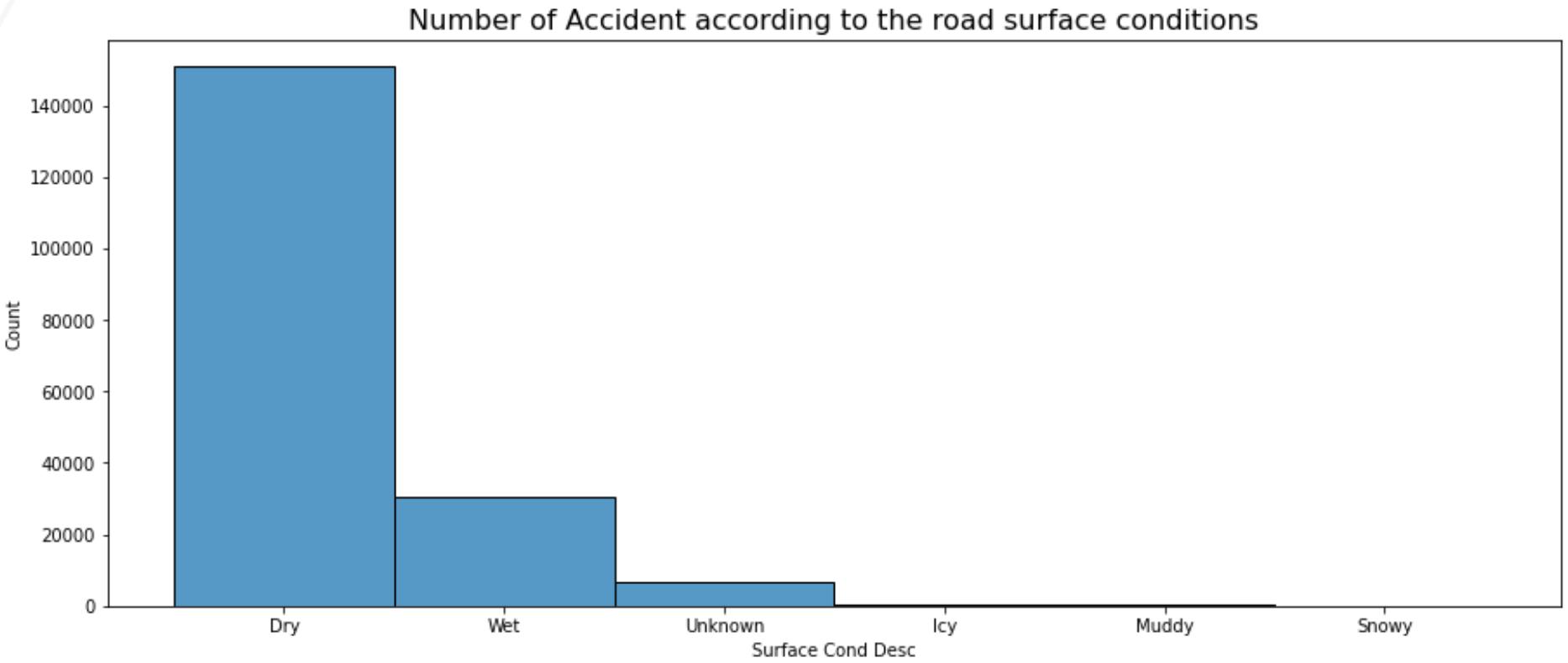


Most accidents occur during good weather and visibility conditions

Georgia Tech

Data Analysis (10/11)

- Road Surface conditions

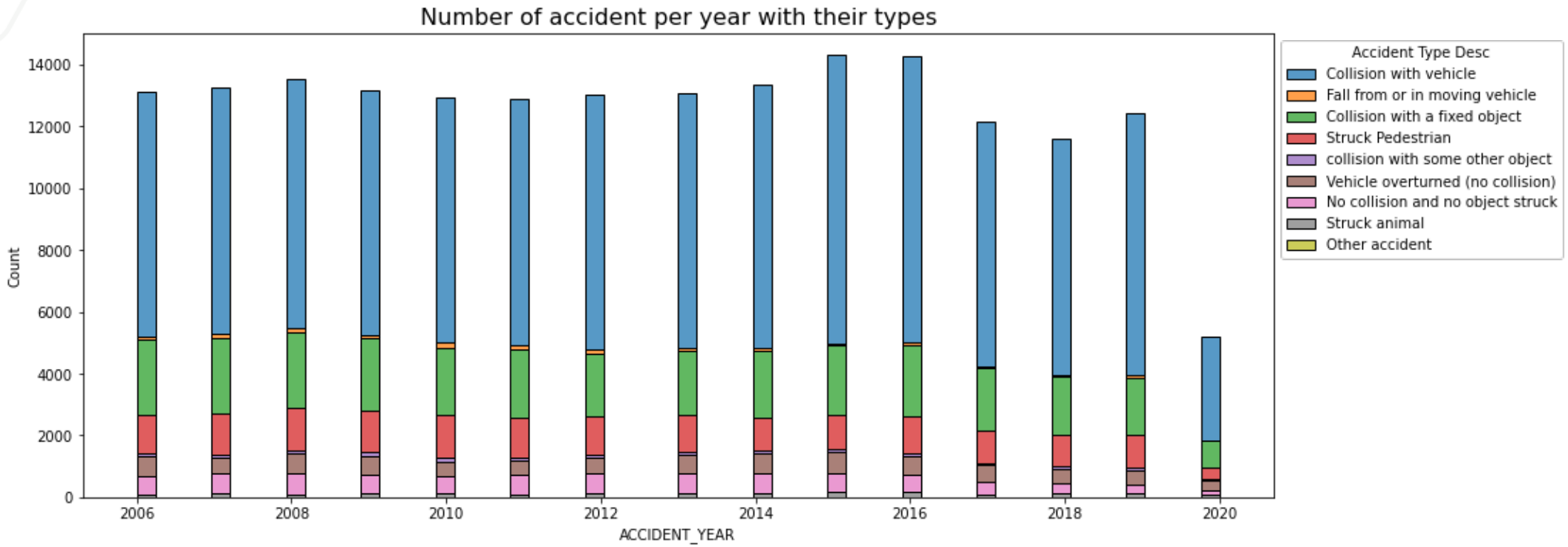


Most accidents occur during good road surface conditions

Georgia Tech

Data Analysis (11/11)

- Overall Trending



The number of accident is trending downwards (2020 huge drop because of COVID-19)

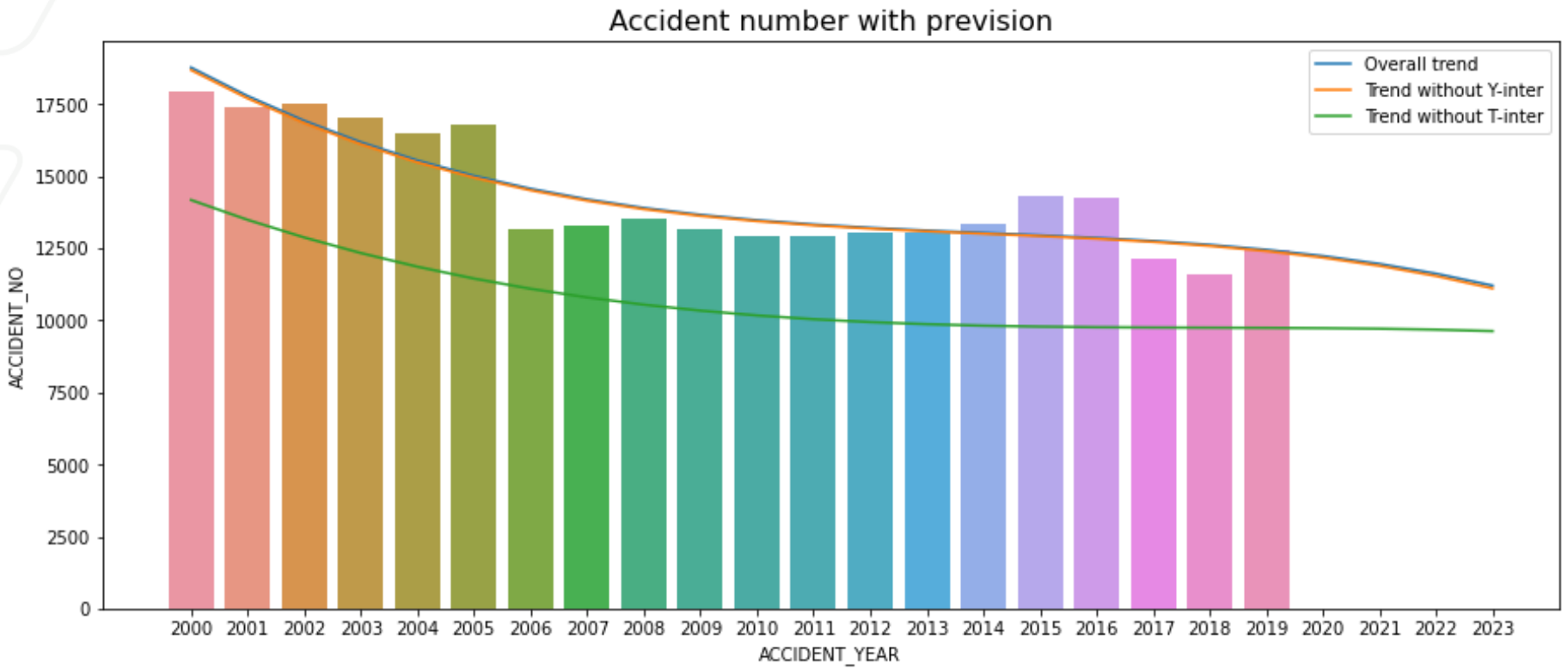
Georgia Tech

Insights using Machine Learning

Georgia Tech

Insights based on ML

- Polynomial Regression



The model:

- A 3-degree Polynomial Regression
- $R^2 = 0.80$
- RMSE = 866.43

The number of accidents tends to decrease for the following years

Georgia Tech

CONCLUSION

Georgia Tech

Conclusion

- Issue Raised: How to reduce the number of accident while developing new infrastructures ?
  - Important focus on road limited from 60 km/h to 80 km/h => Cities and some major highways
  - More vigilance during rush hours and Fridays (maybe no public works during those schedules)
  - More light panels for the night traffic
  - Develop as possible the number of Y-intersections => will have a huge impact for the following years compared to T and Cross intersections