- week3

  Detailed report on CAPSTONE:

A full report consisting of all of the following components (15 marks):

- Introduction where you discuss the business problem and who would be interested in this project.
- Data where you describe the data that will be used to solve the problem and the source of the data.
- Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
- Results section where you discuss the results.
- Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
- Conclusion section where you conclude the report.

2. A link to your Notebook on your Github repository pushed showing your code. (15 marks)

3. Your choice of a presentation or blogpost. (10 marks)

- **Background**
- Predict accident severity at each output area given datetime and driving conditions For each Police Force we build a prediction model with the following objective:

  Given date, time, weather, light and road conditions, predict accident severity within the operating geographic area of a police force and transportation department.

  Predicting theaccident severity at a specific location in the given dataset. Which helps the people travelling through the route by different vehicles, got awareness in that path. so that, we can reduce the accidents. Factors that influence the accident are as follows:
- Vehicles may be car, trucks, etc., .

- accidents may happen at different latitudes and longitudes.we have to find those based on city or province.
- Severity code and its description specifies the situation of the victim.
- victim vehicle and how many victims affected gives us how many are under that severity.
- Type of accident(hit pedestrician, hit cycle,etc.,).
- Junction type where most accidents were happened.
- Affect due to Weather conditions.
- Accident due to spped or not.
- Direction affected the collision.
- Which will be helpful to prevent accidents and reduce homeless people and also.
- Introduction | Business Undertanding
- In an effort to reduce the frequency of car collisions in a community, an algorithim must be developed to predict the severity of an accident given the current weather, road and visibility conditions. When conditions are bad, this model will alert drivers to remind them to be more careful.
- **Data Understanding**

  Data:
- Load file
- Display categorical features
- Inspect the features like Accident_Severity, Speed_Limit,etc.,
- Most accidents occur at low speed limits but they are more severe in proportion at higher speed limits (makes sense)
- Weather, road surface and light conditions
- Most accidents occur in normal weather, only Fog or High winds increase the proportion of severe accidents
- Darkness significantly increase the probability of severe accidents
- Bad Road surface conditions do not increase accident severity
- Junctions types
- The majority of accidents occur away from a junction
- Also the presence of a junction do not increase accident severity
- Display ordinal features
- Other Features(if any)

- Comparing weather, lightcond and roadcond and check whether they affect the accident.
- Geospatial visualisations
- Accident_Severity-GeoSpatial associations analysis
- As seen on the Geospatial visualisations, there is some association between localisation and accident severity.
- Specifically, given a certain accident_severity, determining latt/long is possible.
- The reverse is not true though: It seems difficult to predict accident severity given a certain location.

There are 194,674 perceptions and 38 factors in this informational index. Since we might want to distinguish the components that cause the mishap and the degree of seriousness, we will utilize SEVERITYCODE as our reliant variable Y, and attempt various blends of autonomous factors X to get the outcome. Since the perceptions are very enormous, we may need to sift through the missing esteem and erase the irrelevant sections first. At that point we can choose the factor which may have more effect on the mishaps, for example, address type, climate, street condition, and light condition.

- The objective Data to be anticipated under (SEVERITYCODE 1-prop harm 2-injury) mark.Severity codes are as follows:
- 0: Little to no Probability
- 1: very low probability- chance or property damage
  2: low probability- chance of injury

- **Methodology**
  Initially, I thought that Our data is now ready to be fed into machine learning models. We will use the following models:
- K-Nearest Neighbor (KNN)
  KNN will help us predict the severity code of an outcome by finding the most similar to data point within k distance.
- Decision Tree

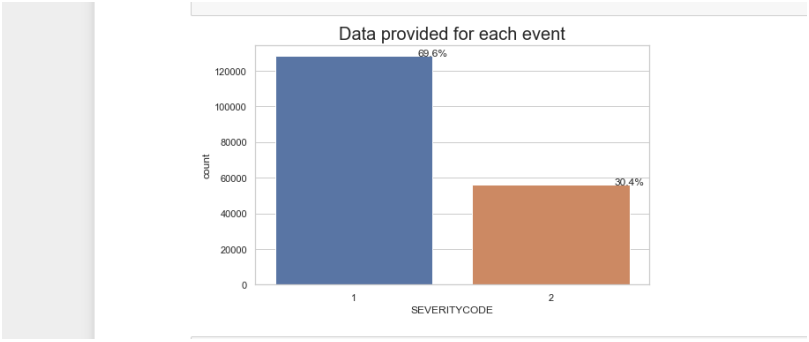
A decision tree model gives us a layout of all possible outcomes so we can fully analyze the concequences of a decision. It context, the decision tree observes all possible outcomes of different weather conditions.

- Logistic Regression

  Because our dataset only provides us with two severity code outcomes, our model will only predict one of those two classes. This makes our data binary, which is perfect to use with logistic regression.

  Note: But in reality, No algorithm is suitable to do predictions as described below. First we imported the data through read_csv(). I noticed that it had 194,674 rows and 38 columns. Therefore, we narrowed it down to 9 columns (‘Severity’, ‘X’, ‘Y’, ‘Location’, 'Vehcount’, ‘Weather’, ‘Roadcond’, ‘Lighdcond’ and ‘Hitparkedcar’) and delete the missing values, which made the final dataset with 184,167 observations and 9 variables.

- Since most of the variable were categorical, it was hard to make the regression model. So, in this study, we focused more on the graphical data and the value count for different categories. There were around 69.6% (2/3) level 1 accidents and 30.4% (1/3) level 2 accidents.



- **Results and Discussion**

  We generated the graphical information based on Seaborn library. The result showed that some locations did have more car accidents than the other places.

  After that, we checked about the weather, road, and light condition. We calculated the total number of car accidents under different situations. There was no significant evidence showed that they might be the reason for the accidents.

```
In [34]: df_weather_cond= df['WEATHER'].value_counts().to_frame()
         df_weather_cond.rename(columns={'WEATHER': 'value_counts'}, inplace=True)
         df_weather_cond.index.name = 'WEATHER'
         df_weather_cond
```

Out[34]:

WEATHER	value_counts
Clear	108833
Raining	31987
Overcast	27105
Unknown	13846
Snowing	888
Other	765
Fog/Smog/Smoke	553
Sleet/Hail/Freezing Rain	112
Blowing Sand/Dirt	49
Severe Crosswind	24
Partly Cloudy	5

```
In [35]: lightcond_counts = df['LIGHTCOND'].value_counts().to_frame()
         lightcond_counts.rename(columns={'LIGHTCOND': 'value_counts'}, inplace=True)
         lightcond_counts.index.name = 'LIGHTCOND'
         lightcond_counts.head(10)
```

Out[35]:

LIGHTCOND	value_counts
Daylight	113522
Dark - Street Lights On	47250
Unknown	12416
Dusk	5763
Dawn	2422
Dark - No Street Lights	1450
Dark - Street Lights Off	1145
Other	188
Dark - Unknown Lighting	11

```
In [36]: #roadcond
         roadcond_counts = df['ROADCOND'].value_counts().to_frame()
         roadcond_counts.rename(columns={'ROADCOND': 'value_counts'}, inplace=True)
         roadcond_counts.index.name = 'ROADCOND'
         roadcond_counts.head(10)
```

Out[36]:

ROADCOND	value_counts
Dry	121871
Wet	46009
Unknown	13795
Ice	1174
Snow/Slush	984
Other	116
Standing Water	102
Sand/Mud/Dirt	63
Oil	53

- **Conclusion**
- This venture and investigation are very useful for the Seattle transportation division and police office. Before I did the examination, I believed that possibly climate, street, and light condition may cause more mishaps, the outcomes demonstrated that it was not right. Nonetheless, we do make sense of that the mishaps are exceptionally identified with some particular areas.
- Besides, there are a few spots which has more mishaps during the dull time(more information might be given). For those spots, including lights may be a decent answer for decrease the crashes. Additionally, when more vehicles engaged with the

mishap, it appears to be that the degree of seriousness will increment. They may should be reacted quickly to spare life and wards.