Project Review-2

Project Title	:	Data In Action - Telematics Data
Project Guide	:	Dr.Mamatha H.R
Project Team	:	PES1201801087 Anjana V Murthy PES1201800368 G Sree Pranavi PES1201800797 Harshitha Batta

Expected Outcome at the end of the project/Contest

The prime objective of the project is to ensure the user's safety and provide emergency assistance while trying to avoid varied types of road accidents based on the research and insights drawn from the provided fleet data.

We aim to detect unexpected situations and alert the driver before hand using the alarm details with respect to location, time and other determining factors.

As a part of future work, we would like to implement real time diagnostics on the vehicles to calculate traffic intensity, time, etc using live telematics data.

Considering the significance of each of the alert types, the following can be expected from the model.

- PCW which is activated when the car senses a pedestrian, can be used to find out the location and time of the highest density of pedestrians in order to take some safety considerations.

- The driver is alerted with HMW if the threshold distance between the vehicle and its preceding vehicle is crossed.

- FCW detects imminent collisions and alerts upto 2.7 secs before collision by calculating time to collision.

If the above 2 alert types occur frequently for a particular vehicle, it can be identified as rash driving. In case the alerts are observed repetitively at a particular location as well, we can consider that location endure dense traffic.

- If HB occurs recursively at a particular location irrespective of time and previous alarm, that location can be assumed to be causing some physical geographical troubles.

- If a HB follows FCW/ HMW or occurs irrespective of any of the alarms recursively for a particular vehicle (that is no emergency), it can be assumed to be rash driving. If stoppage follows HB, the intensity can be identified to be quite high as the speed drops suddenly.

- The fuel consumption increases and eventually the mileage decreases in case of harsh braking. So, studying the fuel data can be helpful in determining such disturbances and rash driving and can be used in further studies of fuel economy.

As a part of future work, we would like to implement real time diagnostics on the vehicles to calculate traffic intensity, time, etc using live telematics data.

Work Plan For the coming weeks

- 90% Milestone:

  Final Prediction Model.
  After analysing all the trends related to the alert types a prediction model shall be developed to match the expected outcomes. A spatial database is also planned to be implemented as a part of the model once all the constraints are satisfied.

- Final submission:

  Final prediction model.

Progress as on Date

- <u>Data Cleaning</u> :

  Cleaning and analysis of the dataset has been carried out to make the data efficient by replacing the null values with useful data.

1. In the alarmType sheet, we noticed that the last few rows had the latitude and longitude interchanged, we verified it and swapped the values.

2. In the FuelInfo sheet, some of the null values were replaced with computed results from other columns and few rows were removed.

- Post research, we learnt that each degree of latitude is approximately 69 miles (111 kilometers) apart and a degree of longitude is widest at the equator with a distance of 69.172 miles (111.321 kilometers).

  So, now if we consider two locations(latitude and longitude) taken from the dataset, from the sheet ‘alarmType’.

latitude	longitude
10.442667961120605	76.2600317138672
10.442667961120605	76.2600317138672

  So, for the above two pairs of latitude and longitude, if we calculate the distance between them, it would be less than a kilometer. So, we rounded the latitudes and longitudes to two digits after the decimal point, so that we could analyse the data of a particular area more accurately.

- Statistical conclusions :

After examining the datasets, the inferences drawn are (the latitudes and longitudes are rounded off to 0 decimals at this step for overall minimal validation, it has been performed with more precision later),

1. Each alarm type is compared with location and AM/PM:

   - ❑ PCW is maximum at 8 lat and 77 long and higher in AM.
   - ❑ FCW is maximum at 8 lat and 77 long and higher in AM.
   - ❑ HMW is maximum at 8 lat and 76 long and higher in PM.
   - ❑ HB is maximum at 8 lat and 77 long and higher in PM.
   - ❑ STOPPAGE is maximum at 8 lat and 76 long and higher in AM.

2. At each hour of the day, the frequency of each alarm type is compared irrespective of location. In general,

- Frequency of PCW is more in broad daylight (mid day).This may be because of high traffic situation in the morning compared to night time.

- In case of FCW frequency it is observed that it is alerted the maximum in the morning hours from dawn to noon.

- The HMW frequency doesn’t show much variation, it is high for most of the night along with the day hours similar to FCW when compared to the other alert types.

- It is also observed that frequency of HB and stoppage is not significantly affected by these factors.

3. At each hour of the day, the frequency of each alarm type is compared respective of location. In general,

- HMW is highest at 8 lat and 76 long around 1AM and 2AM.
- FCW is recorded the most between 8 lat and 9 lat 77 long around 10 and 11AM.
- PCW is max at 10 AM around 8 lat and 77 long.

This is just a minimal analysis to get an overall understanding. The model will be developed to obtain real time trends based on the dataset and predict the condition before hand.

- <u>Code samples for statistical analysis</u> :

The following are some snippets of the code used for obtaining the statistical conclusions explained in the previous slides.

Alarm type vs rounded off location coordinates

```
print("\n \n FCW \n")
df1 = pd.DataFrame(df,columns=['latitude','Longitude'])
df1=df.loc[df.alarmType=="FCW"]
df2 = pd.DataFrame(df1,columns=['latitude','longitude'])
x=df2['latitude_rounded_off']= np.floor(df2['latitude'])
y=df2['longitude_rounded_off']= np.floor(df2['longitude'])
print('Lats rounded off',set(x))
print('Longs rounded off',set(y))
f = df2.pivot_table(index=['latitude_rounded_off','longitude_rounded_off'],aggfunc='size')
print("\n rounded off locs vs fcw \n",f)
r = df1.pivot_table(index=['AM/PM'],aggfunc='size')
print("\n am/pm vs FCW",r)

print("-----------------------------------")
```

Alarm types wrt time and rounded off location coordinates

```python
print("\n \n Alarmtype vs locs vs time with locprecision = full \n")
hr=df['hour_rounded_off']= np.floor(df['recorded_at_hour'])
c = df.pivot_table(index=['latitude','longitude','hour_rounded_off','AM/PM','alarmType'],aggfunc='size')
pd.set_option("display.max_rows",df.shape[0] + 1)
print("\n",c)

print("-----------------------------------")

print("\n \n Alarmtype vs locs vs time with hrprecision = full \n")
df['latitude_rounded_off']= round(df['latitude'],3)
df['longitude_rounded_off']= round(df['longitude'],3)
c = df.pivot_table(index=['latitude_rounded_off','longitude_rounded_off','recorded_at_hour','AM/PM','alarmType
pd.set_option("display.max_rows",df.shape[0] + 1)
print("\n",c)
```

Alarm types wrt speed and vehicle id

```python
print("\n \n Alarmtype vs speed \n")
c = df.pivot_table(index=['alarmType','speed'],aggfunc='size')
pd.set_option("display.max_rows",df.shape[0] + 1)
print("\n",c)

print("-----------------------------------")

print("\n \n HB vs vehicle id \n")
df1 = pd.DataFrame(df,columns=['deviceId'])
df1=df.loc[df.alarmType=="HB"]
df2 = pd.DataFrame(df1,columns=['deviceId'])
pd.set_option("display.max_rows",df.shape[0] + 1)
p = df2.pivot_table(index=['deviceId'],aggfunc='size')
print(p)

print("-----------------------------------")
```

- Visualizations :

1. FREQUENCY OF ALARM TYPE WRT VEHICLE ID

   - ❑ We plotted bar graphs, for the vehicles with respect to the number of times each alarm type is activated.
   - ❑ There are 5 graphs, each pertaining to an alarm type.

2. FREQUENCY OF ALARM TYPE WRT TIME(AM/PM)

   - ❑ Based on the number of times the alarm type was activated, we plotted graphs with respect to time i.e AM/PM. To find out whether the density of traffic of pedestrians and cars, is more at daylight or at night.
   - ❑ There are 5 graphs, each pertaining to an alarm type.

VISUALISATION
FOR
FREQUENCY OF ALARM TYPE W.R.T VEHICLE ID

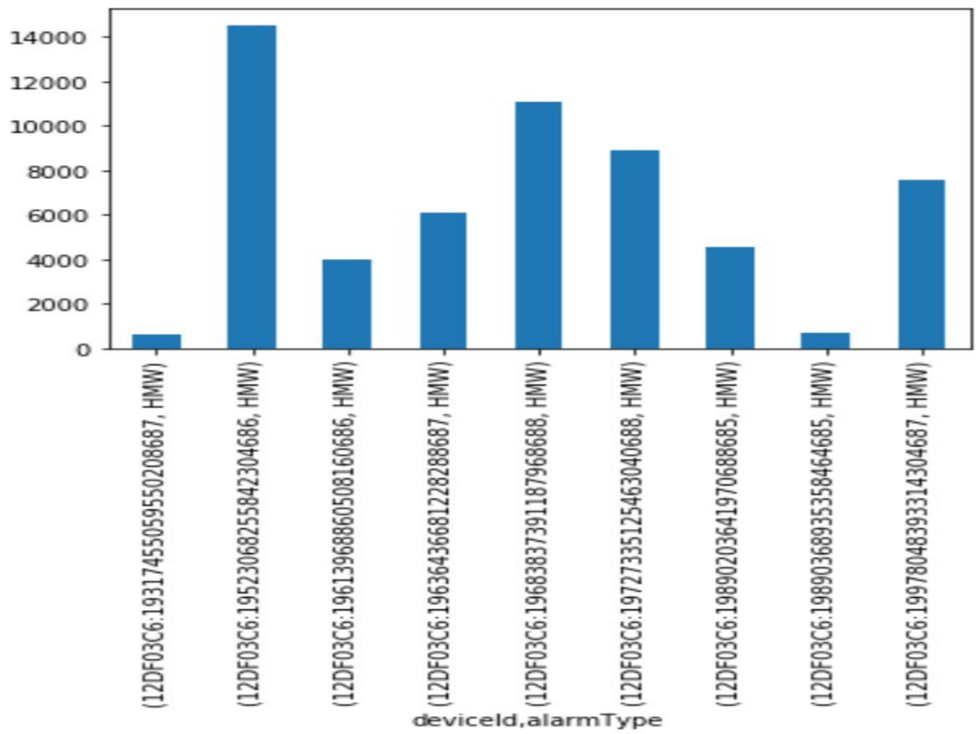
- Insights from the graphs:

  Let us assume,

  12DF03C6:19317455059550208687 as vehicle 1
  12DF03C6:19523068255842304686 as vehicle 2
  12DF03C6:19613968860508160686 as vehicle 3
  12DF03C6:19636436681228288687 as vehicle 4
  12DF03C6:19683837391187968688 as vehicle 5
  12DF03C6:19727335125463040688 as vehicle 6
  12DF03C6:19890203641970688685 as vehicle 7
  12DF03C6:19890368935358464685 as vehicle 8
  12DF03C6:19978048393314304687 as vehicle 9

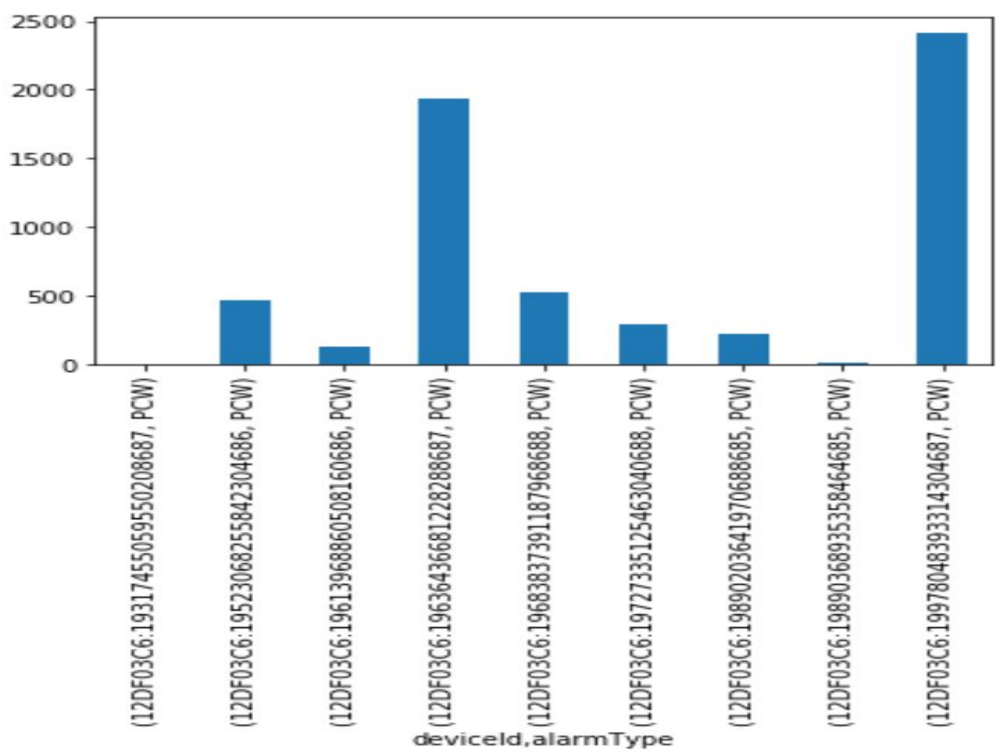
  Based on the graphs illustrated below,
  We can conclude that the frequency of alarm types is the highest for vehicles 2,4,5 indicating that the driver has to take more precautions and safety considerations while driving.Therefore a warning can be given to drivers of these vehicles.

HMW

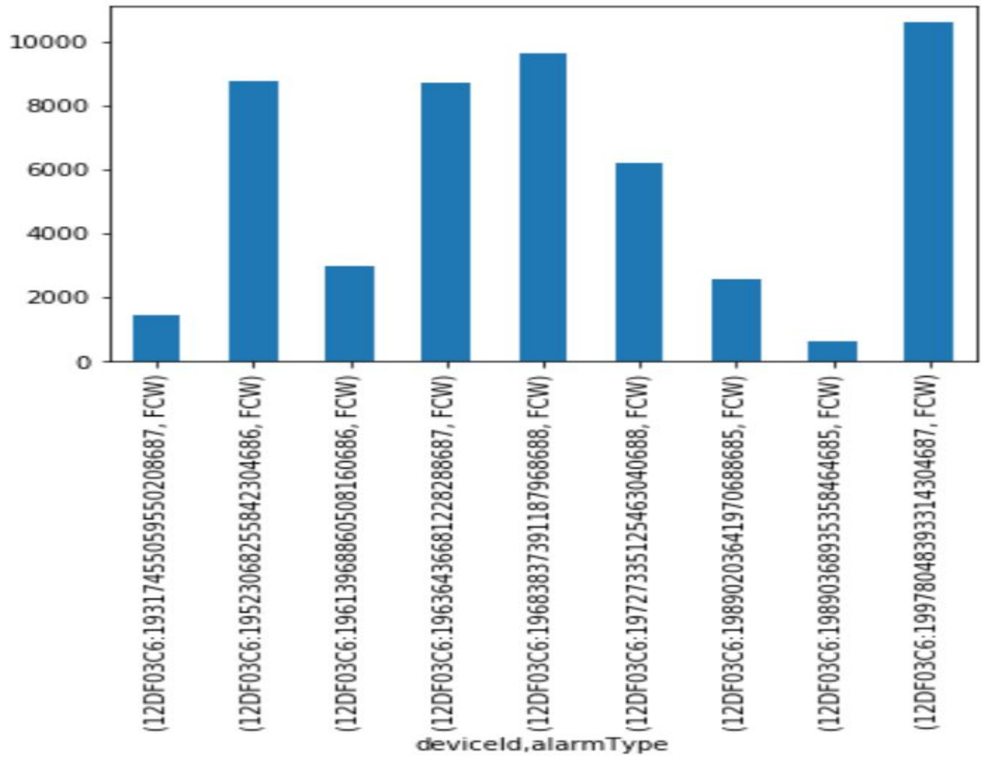


14000
12000
10000
8000
6000
4000
2000
0
(12DF03C6:1931745505955020868 7, HMW)
(12DF03C6:1952306825584230468 6, HMW)
(12DF03C6:1961396886050816068 6, HMW)
(12DF03C6:1963643668122828868 7, HMW)
(12DF03C6:1968383739118796868 8, HMW)
(12DF03C6:1972733512546304068 8, HMW)
(12DF03C6:1989020364197068868 5, HMW)
(12DF03C6:1989036893535846468 5, HMW)
(12DF03C6:1997804839331430468 7, HMW)
deviceId,alarmType

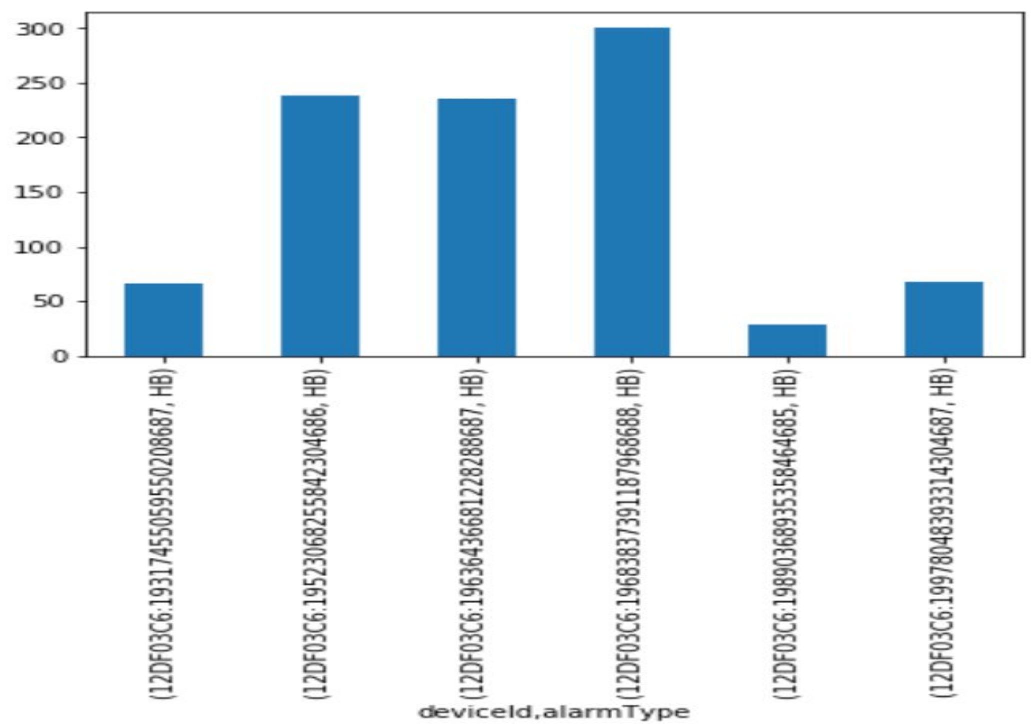
PCW


2500
2000
1500
1000
500
0
(12DF03C6:19317455059550208687, PCW)
(12DF03C6:19523068255842304686, PCW)
(12DF03C6:19613968860508160686, PCW)
(12DF03C6:19636436681228288687, PCW)
(12DF03C6:19683837391187968688, PCW)
(12DF03C6:19727335125463040688, PCW)
(12DF03C6:19890203641970688685, PCW)
(12DF03C6:19890368935358464685, PCW)
(12DF03C6:19978048393314304687, PCW)
deviceId,alarmType

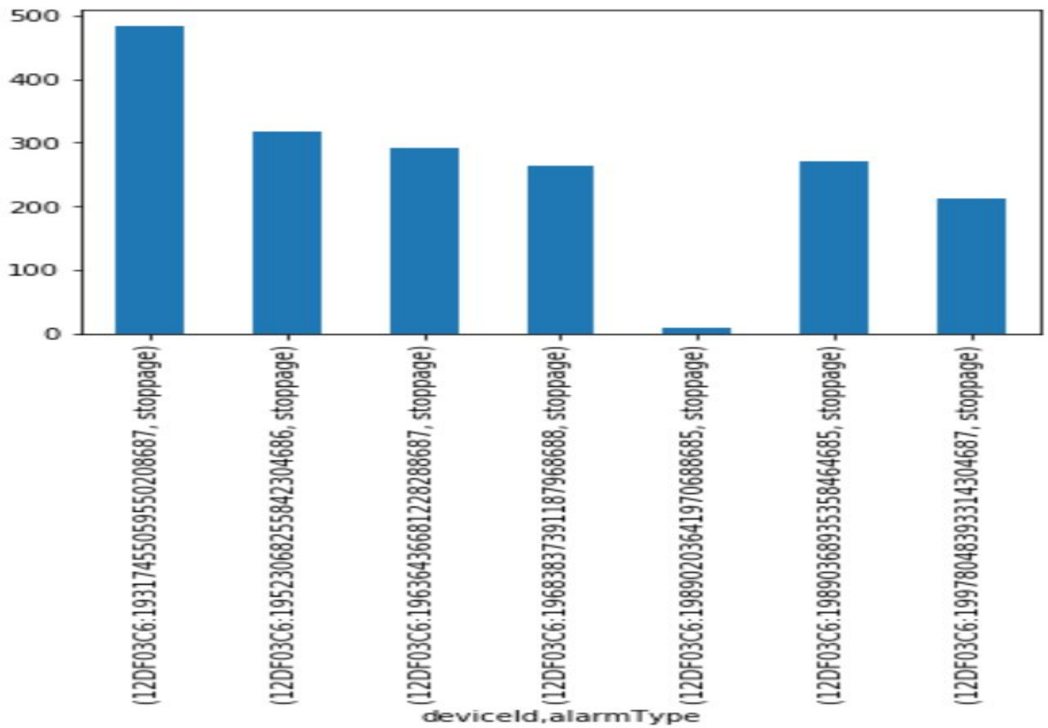
FCW


10000
8000
6000
4000
2000
0
(12DF03C6:19317455059550208687, FCW)
(12DF03C6:19523068255842304686, FCW)
(12DF03C6:19613968860508160686, FCW)
(12DF03C6:19636436681228288687, FCW)
(12DF03C6:19683837391187968688, FCW)
(12DF03C6:19727335125463040688, FCW)
(12DF03C6:19890203641970688685, FCW)
(12DF03C6:19890368935358464685, FCW)
(12DF03C6:19978048393314304687, FCW)
deviceId,alarmType

HB


300
250
200
150
100
50
0
(12DF03C6:1931745505955020208687, HB)
(12DF03C6:1952306825558423004686, HB)
(12DF03C6:1963643666812282888687, HB)
(12DF03C6:1968383739118796868688, HB)
(12DF03C6:1989036893535846464685, HB)
(12DF03C6:1997804839331430046887, HB)
deviceId,alarmType

STOPPAGE



500
400
300
200
100
0
(12DF03C6:1931745505955020868 7, stoppage)
(12DF03C6:1952306825558423046 86, stoppage)
(12DF03C6:1963643668122828868 7, stoppage)
(12DF03C6:1968383739118796868 8, stoppage)
(12DF03C6:1989020364197068868 5, stoppage)
(12DF03C6:1989036893535846468 5, stoppage)
(12DF03C6:1997804839331430468 7, stoppage)
deviceId,alarmType

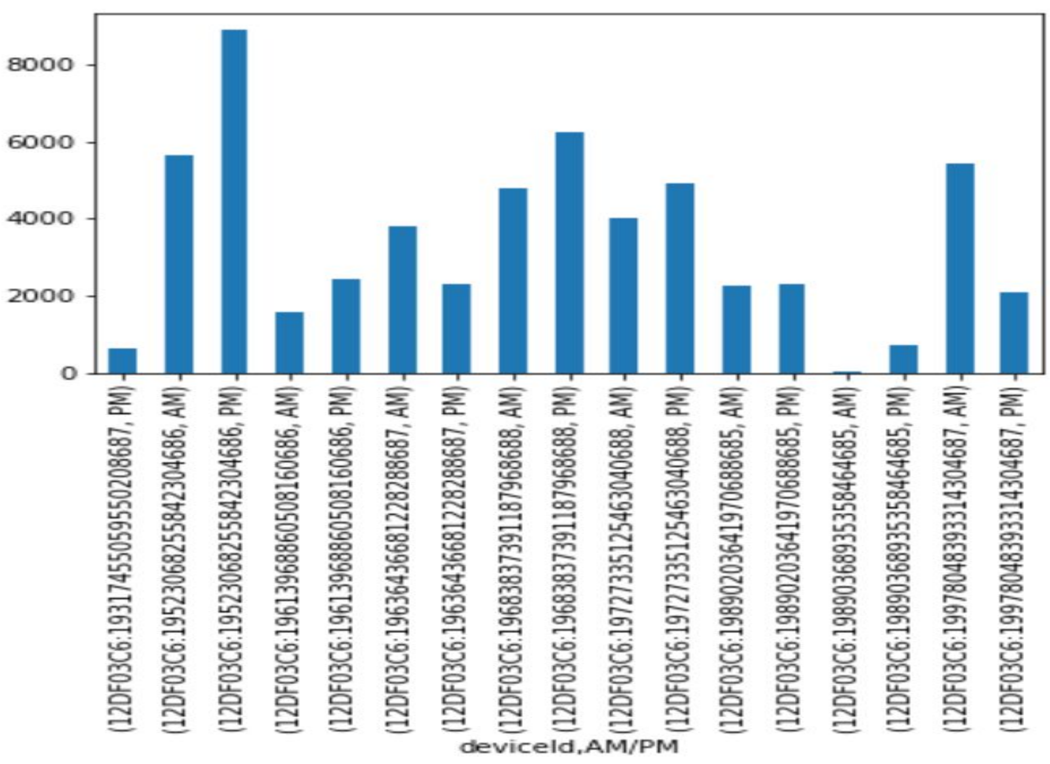
VISUALISATION
FOR
FREQUENCY OF ALARM TIME W.R.T TIME(AM/PM)

- Insights from the graphs :

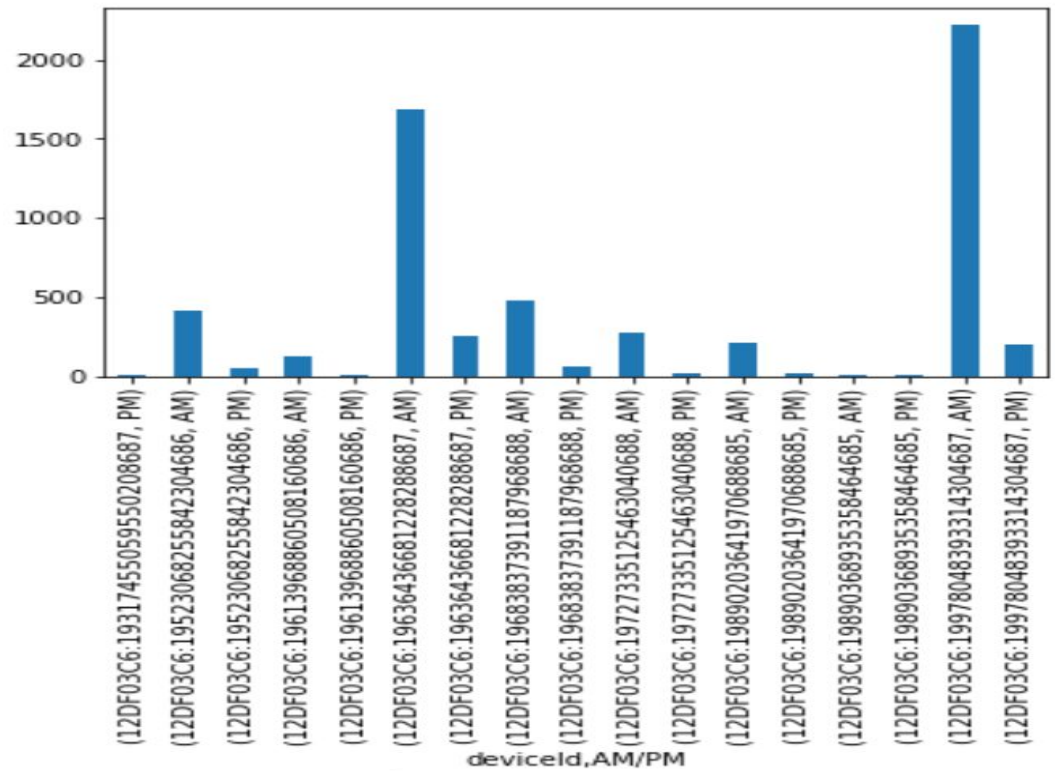
From the following visualizations some insights regarding each of the alarm types have been drawn.

- HB and Stoppage are not significantly related to the shift of the day.
- HMW does not vary much, is activated during most of time, quite frequently.
- FCW is observed to be high during the daylight, till the noon.
- PCW is recorded the most in the morning but is not continuously active.

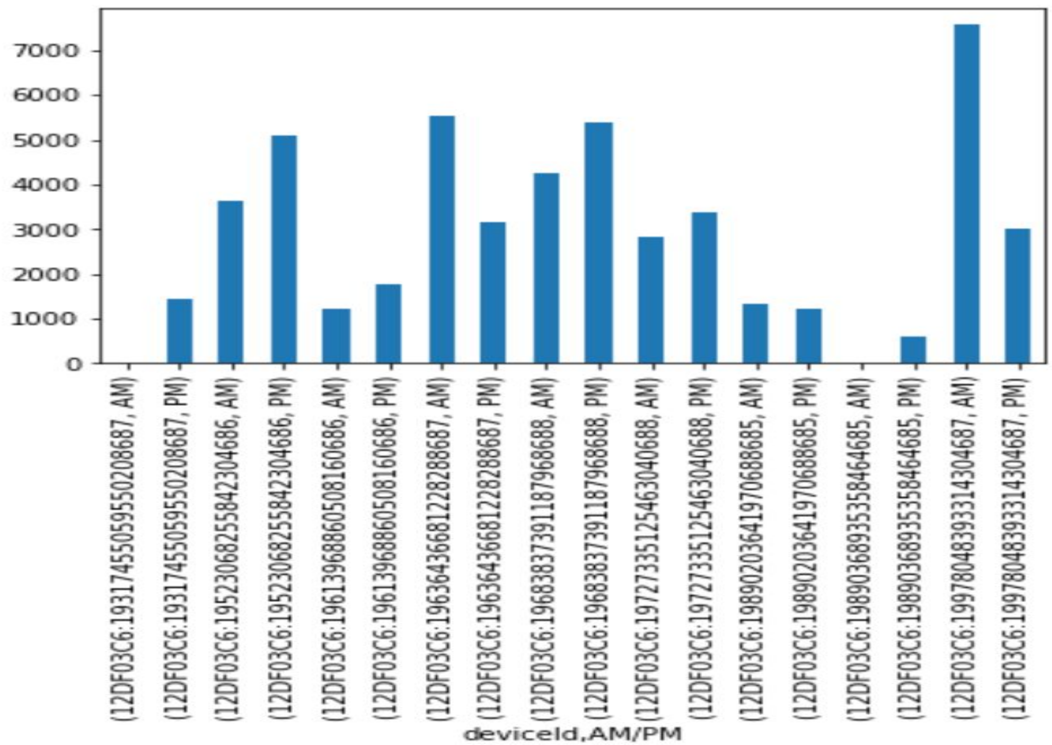
HMW


8000
6000
4000
2000
0
(12DF03C6:19317455059550208687, PM)
(12DF03C6:195230682555842304686, AM)
(12DF03C6:195230682555842304686, PM)
(12DF03C6:19613968860508160686, AM)
(12DF03C6:19613968860508160686, PM)
(12DF03C6:19636436681228288687, AM)
(12DF03C6:19636436681228288687, PM)
(12DF03C6:19683837391187968688, AM)
(12DF03C6:19683837391187968688, PM)
(12DF03C6:19727335125463040688, AM)
(12DF03C6:19727335125463040688, PM)
(12DF03C6:19890203641970688685, AM)
(12DF03C6:19890203641970688685, PM)
(12DF03C6:19890368935358464685, AM)
(12DF03C6:19890368935358464685, PM)
(12DF03C6:19978048393314304687, AM)
(12DF03C6:19978048393314304687, PM)
deviceId,AM/PM

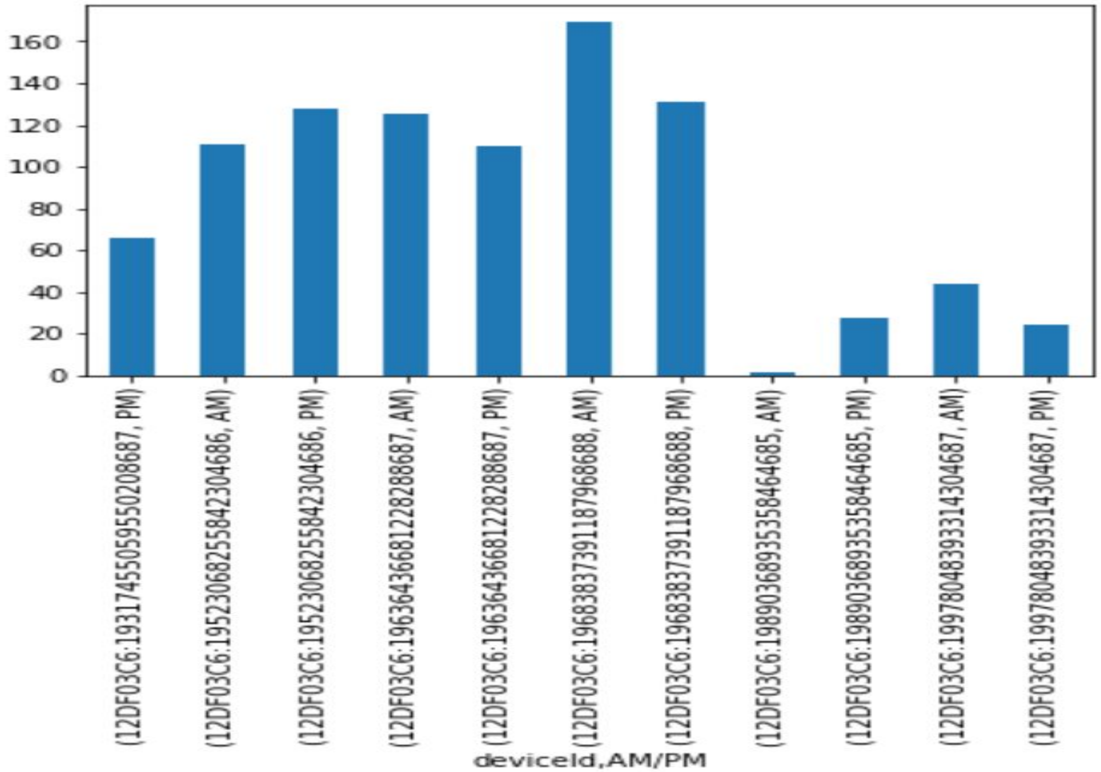
PCW


2000
1500
1000
500
0
(12DF03C6:19317455059550208687, PM)
(12DF03C6:19523068255842304686, AM)
(12DF03C6:19523068255842304686, PM)
(12DF03C6:19613968860508160686, AM)
(12DF03C6:19613968860508160686, PM)
(12DF03C6:19636436681228288687, AM)
(12DF03C6:19636436681228288687, PM)
(12DF03C6:19683837391187968688, AM)
(12DF03C6:19683837391187968688, PM)
(12DF03C6:19727335125463040688, AM)
(12DF03C6:19727335125463040688, PM)
(12DF03C6:19890203641970688685, AM)
(12DF03C6:19890203641970688685, PM)
(12DF03C6:19890368935358464685, AM)
(12DF03C6:19890368935358464685, PM)
(12DF03C6:19978048393314304687, AM)
(12DF03C6:19978048393314304687, PM)
deviceId,AM/PM

FCW

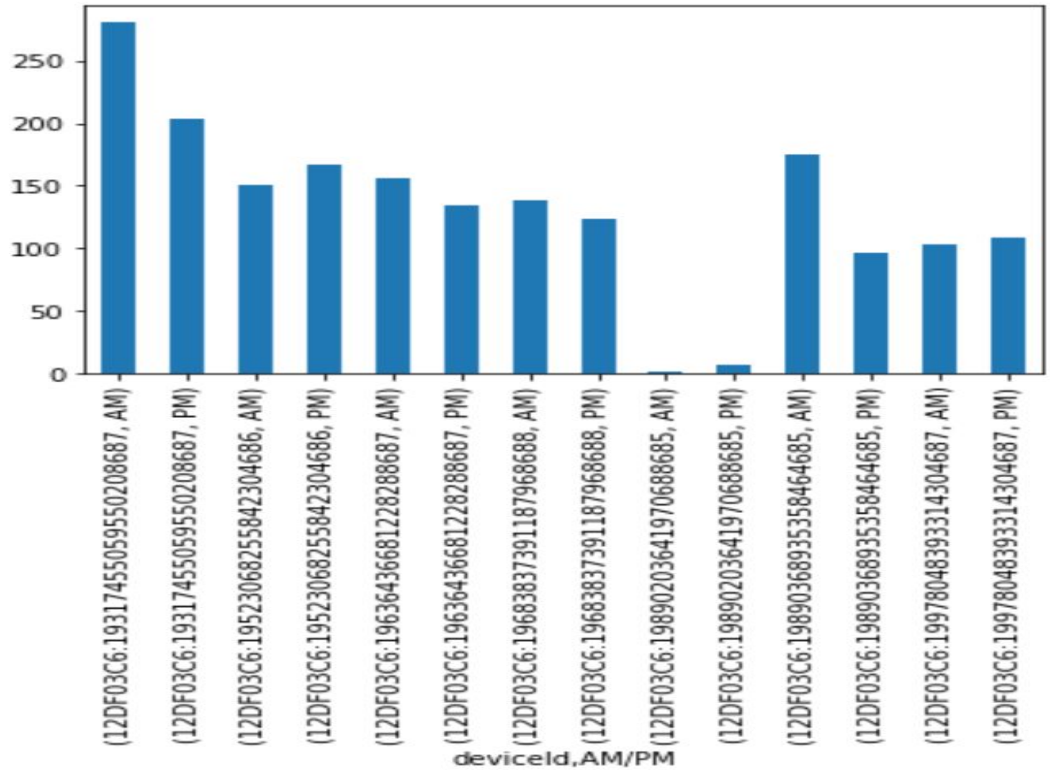

7000
6000
5000
4000
3000
2000
1000
0
(12DF03C6:19317455059550208687, AM)
(12DF03C6:19317455059550208687, PM)
(12DF03C6:19523068255842304686, AM)
(12DF03C6:19523068255842304686, PM)
(12DF03C6:19613968860508160686, AM)
(12DF03C6:19613968860508160686, PM)
(12DF03C6:19636436681228288687, AM)
(12DF03C6:19636436681228288687, PM)
(12DF03C6:19683837391187968688, AM)
(12DF03C6:19683837391187968688, PM)
(12DF03C6:19727335125463040688, AM)
(12DF03C6:19727335125463040688, PM)
(12DF03C6:19890203641970688685, AM)
(12DF03C6:19890203641970688685, PM)
(12DF03C6:19890368935358464685, AM)
(12DF03C6:19890368935358464685, PM)
(12DF03C6:19978048393314304687, AM)
(12DF03C6:19978048393314304687, PM)
deviceId,AM/PM

HB



160
140
120
100
80
60
40
20
0
(12DF03C6:19317455059550208687, PM)
(12DF03C6:19523068255842304686, AM)
(12DF03C6:19523068255842304686, PM)
(12DF03C6:19636436681228288687, AM)
(12DF03C6:19636436681228288687, PM)
(12DF03C6:19683837391187968688, AM)
(12DF03C6:19683837391187968688, PM)
(12DF03C6:19890368935358464685, AM)
(12DF03C6:19890368935358464685, PM)
(12DF03C6:19978048393314304687, AM)
(12DF03C6:19978048393314304687, PM)
deviceId,AM/PM

<u>STOPPAGE</u>



250
200
150
100
50
0
(12DF03C6:193174550595502086887, AM)
(12DF03C6:193174550595502086887, PM)
(12DF03C6:195230682558423046886, AM)
(12DF03C6:195230682558423046886, PM)
(12DF03C6:196364366812282886887, AM)
(12DF03C6:196364366812282886887, PM)
(12DF03C6:196838373911879686888, AM)
(12DF03C6:196838373911879686888, PM)
(12DF03C6:198902036419706886885, AM)
(12DF03C6:198902036419706886885, PM)
(12DF03C6:198903689353584646885, AM)
(12DF03C6:198903689353584646885, PM)
(12DF03C6:199780483933143046887, AM)
(12DF03C6:199780483933143046887, PM)
deviceId,AM/PM

- Code samples for graphical visualization :

The following are some snippets of the code used for obtaining the conclusions from the graphs explained in the previous slides.

```
fcw_time = pd.pivot_table(df3,index=['deviceId','AM/PM'],aggfunc='size')
fcw_time
fcw_time.plot(kind='bar')
```

```
hmw_device = pd.pivot_table(df4,index=['deviceId','alarmType'],aggfunc='size')
hmw_device
```

```
deviceId                             alarmType
12DF03C6:19317455059550208687        HMW            640
12DF03C6:19523068255842304686        HMW          14535
12DF03C6:19613968860508160686        HMW           3974
12DF03C6:19636436681228288687        HMW           6103
12DF03C6:19683837391187968688        HMW          11072
12DF03C6:19727335125463040688        HMW           8900
12DF03C6:19890203641970688685        HMW           4569
12DF03C6:19890368935358464685        HMW            728
12DF03C6:19978048393314304687        HMW           7543
dtype: int64
```

```
hmw_device.plot(kind='bar');
```

Prediction Model

Our model is going to ensure the users' safety by helping them avoid road accidents, with the data we have obtained.
We aim to detect the various situations and alert the driver before hand using the alarm details with respect to location, time and other determining factors and avoid unwanted circumstances.

After good literature survey, already detailed in the previous milestone, it is observed that this kind of work has not been done earlier with telematics data where insights regarding the traffic are drawn. General conclusions are related to the driver's behaviour. So now, this novel model focuses on the external factors.

The model firstly determines the trends of the occurences of the different alarm types with respect to date, time and location coordinates. If a user has to travel from location A to C via B, the model predicts the external conditions like heavy pedestrian traffic, traffic jams, irregular physical geography, etc at location B when the driver is in A based on the computed trends and maximum - minimum frequencies of alarms from the data. This helps the driver stay alerted in time and avoid unwanted chaos. The model will also compute the trends with respect to the vehicles to differentiate between rash driving and probable traffic troubles.

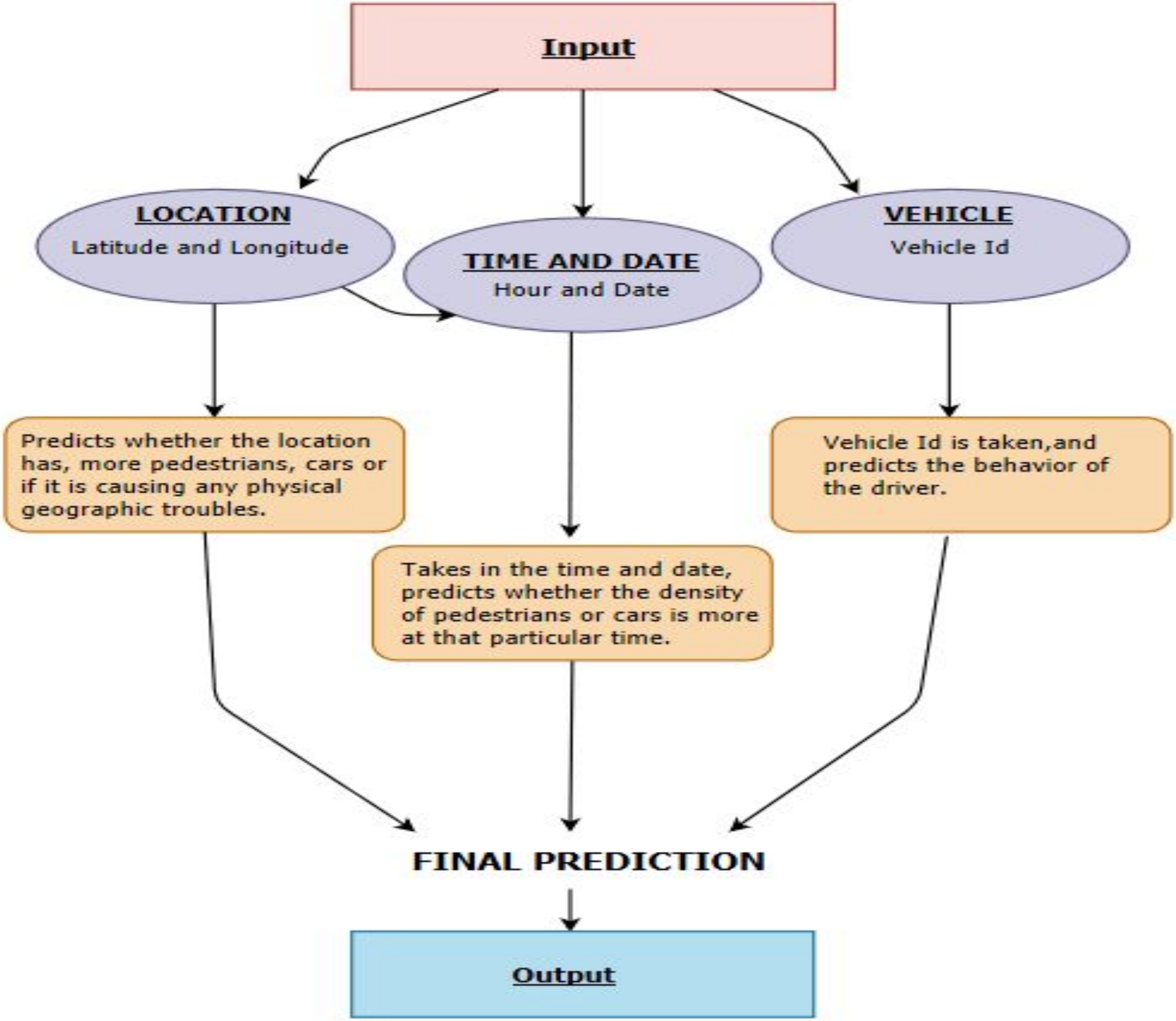
Prototype of the model (Initial)

- The model is trained with the preprocessed data set (80%) and gives the final prediction as the output. The training process involves analysis of trends of the alert type with respect to location and time. The prototype is discussed in detail with a block diagram in the following slide. A snippet of sample code is also added.

- Once the model is trained, the testing is done using 20% data with location coordinates and time as inputs.

- The initial plan shall accept location and time as inputs to predict the disturbances at the location.

- The next level development in the project would take inputs from GPS to predict the required outputs with an interface of a geographical map.


Input
LOCATION
Latitude and Longitude
TIME AND DATE
Hour and Date
VEHICLE
Vehicle Id
Predicts whether the location has, more pedestrians, cars or if it is causing any physical geographic troubles.
Takes in the time and date, predicts whether the density of pedestrians or cars is more at that particular time.
Vehicle Id is taken,and predicts the behavior of the driver.
FINAL PREDICTION
Output

- Sample code for the model:

  We started coding for the final prediction model, a snippet of the initial code is in the following slide.

  Once the location is taken as the input, it will check whether there is a high density of population, cars ahead or if there are geographical troubles in the surrounding area.

  The following snippet of code is just for the minimal idea, it will be improvised eventually in the course of time with detailed updates and improvements to fit in with the model's requirements.

```python
p=list()
c=list()
w=list()
ab=list()

for i in range(len(df)):
    p.append(round(df.loc[i,"latitude_rounded_off"],2))
    c.append(round(df.loc[i,"longitude_rounded_off"],2))
    w.append(df.loc[i,"alarmType"])

ab=list(zip(p,c))

info=dict()

for i in range(len(ab)):
    info[ab[i]]=w[i]

print("Enter the Latitude")
lat=float(input())
print("Enter the Longitude")
long=float(input())

lat1=[round(lat,2)]
long1=[round(long,2)]
blah=list(zip(lat1,long1))

for key in info.keys():
    if([key]==blah):
        if(info[key]=='FCW'):
            print("Imminent collisions in this Area! Watch Out")
        if(info[key]=='PCW'):
            print("Watch Out! Pedestrians are Ahead")
        if(info[key]=="HMW"):
            print("A lot of traffic here Buddy!")
        if(info[key]=="HB"):
            print("Geographic Troubles in this area.")
```

Any obstacles/challenges and any Assistance Required

Assistance in understanding all the attributes of the data set,

1. What is recorded_at_value in alert data?

Key Deliverables for the next milestone

- For the next milestone we aim to be ready with the final prediction model based on the prototype and to put our idea into effect.

- The model will be able to predict the troubles at a particular location in time and suggest the driver for an improved travel experience.

Summary

- A novel idea, unique from most of the other works based on fleet data (as per the literature survey) is developed for user safety post analysis of the data.
- The purpose of the model is avoid road mishaps by alerting the driver with the probable circumstances in time. Useful insights based on the trends of the alert types wrt the deciding factors - time, date, location coordinates will be drawn from time to time to keep the driver alerted about the forthcoming location.
- By the next milestone, we aim to be ready with the model to put our novel idea for driver safety into working.