**Predicting the “Driving Style”**Problem Description  
  
Road safety rules and regulations are designed to prevent the citizens from fatal incidents. Although policies are in place, we observe negligent behaviour of the drivers which lead to serious injuries or death crashes. It is of utmost interest of the authorities to understand and analyse human behaviour to take necessary corrective and preventive actions.   
  
The stakeholders are the citizens, road transport authorities, Insurers and Researchers/Data service providers. In order to design a driving assistance system there is a need to get an understanding of the data on the driving patterns and broadly distinguish bad drivers from good ones. This in turn will benefit Insurers in analysing underwriting risks, prevent frauds and designing No-claim-discount systems (NCD systems), etc. Additionally, the concerned authorities will need insights to design benchmarks for qualifications and driver licensing regulations, etc.   
   
About Data:   
  
Every single vehicle is observed at various time stamps, to record the details of trips made, traffic conditions, vehicle details like length, weight, no of axles of the vehicle, road conditions, lanes switched, weather conditions etc. along with the driving styles are recorded.   
   
Objective:   
**You are expected to create an analytical and modelling framework to predict the driving style of each id categorizing into 1,2,3  and also obtain the actionable top 10 data insights (patterns) for “1” class using the tree based algorithms.  
“1” indicates : “Aggressive”,   
“2” indicates : ”Normal” and   
“3” indicates : “Vague”**

**Attribute details**

V1 Date & time of data

V2 Length of vehicle in cm

V3 Lane of the road of the vehicle(1 &2)

V4 Speed of the vehicle (kph)

V5 weight of vehicle in kg

V6 Number of axles

V7 ID of the preceding vehicle

V8 Speed of the preceding vehicle (kph)

V9 Weight of the preceding vehicle (kg)

V10 Length of preceding vehicle (cm)

V11 Time gap with the preceeding vehicle in seconds

V12 Weather details-Air temperature

V13 Precipitation-(Clear,rain, snow)

V14 Precipitation intensity- none, low, moderate, high

V15 Relative humidity-

V16 Wind direction-0 to 360)

V17 Wind speed in m/s

V18 Condition of the road wrt weather

**Visualization Tips**

Important Note: No points for "effort" of putting a plot which is illegible and meaningless. DO NOT throw plots only for completing visualization exercise. This will work against you. Every plot you present should be easy to read and should convey a message. If the message is not evident, write that down clearly under the plot. It is okay to explore the data with several plots but present only the plots which are telling a story.

**General Notes:**

1 "Clearly understand difference between following plots

1. Bar plot,

2. Box plot,

3. Histogram,

4. Scatter plot

5. Line plot

Decide which one you need before you draw. Most of the technical visualization can be covered by just these types. If you need anything beyond these, think again do you really need it or you are over complicating it?"

2 Readable legends, axis labels, plot titles, tick sizes, are essential of any plot. If your audience cannot read it, then the plot serves no purpose at all.

3 Default plotting commands in Python and R generate barebone plot with poor labels and font size. You almost always would have to write additional lines of codes to make the plot readable with figure and axis options. Python users consider using matplotlib along with seaborn. Some functionalities are stronger in seaborn library while some are not as good

4 Do not blame the tool for giving a bad plot. Tool will give exactly the plot which you ask for and in shape, size and color you ask. You just need to know how to ask!

5 Do not try to present too much information or too many variables in single plot. It is rarely useful for any purpose. 3-D plots for technical visualization are very poor choice as they are very hard to read.

6 Chose the color (and need for it) wisely. Several powerful plots can be presented just by using black and white scale. There is no additional points for making your chart colorful. If you have colors, it better carries a meaning without making it more confusing

7 "Some common mistakes seen in INSOFE presentations (other than poor font sizes and axis labels):

1. Box plot for describing frequency or count (Hint: Box plots do not give count. They only give spread)

2. Boxplot of different columns/variables with different scale on same axis. (Hint: If you cannot read range of any one box in your plot, then you are probably doing something wrong)

3. Not thinking through if the actual number or percentage of total would be better choice of bar chart. (Hint: There is no right answer but think about the story you want to tell. Think which representation would exactly tell the story you want to tell)

4. Overuse of colors when instead a simple B&W scale would have done the job.