Accident and Insurance Project

Panchatcharam M

Associate Professor

Department of Mathematics and Statistics,
IIT Tirupati

Vehicle Database

- Have a look at the VehiclDataSet.zip file
- This data contains information about the policy and accident damages
- It has the following fields

  - Gender
  - Age
  - Driving_License
  - Region_Code
  - Previously_Insured
  - Vehicle_Age
  - Vehicle_Damage
  - Annual_Premium
  - Policy_Sales_Channel
  - Vintage

- Find records without driving license
- Find all cars with less than 1 year age
- Find all cars with damages and more than 1 year age
- Find the cars with damages and premiums more than 40k
- Find the records with age below 25 and has 2 year olde damaged car with premium above 25k

- Create 5 problem statements to display more information about the analysis of this statement

- Assume that you are employed as an insurance manager, what would be your strategy, to increase the premium for increasing the policy, for example, will you increase the policy for age above 25 or above 40 etc
- Do a correlation study of each variable

- Assume that you are employed as a data analyst to analyse this data.
  - Read the content column
  - Identify each keywords in the content column and create emotions column and provide the emotions provided by each passenger
  - List the most good emotions flights and least good emotions reviews
  - Find the details where discussed about seat discomfort, flight attender discomfort, luggage issues, entertainment issues etc.

Vehicle Database

- Have a look at the Accident.zip file
- This data contains information about various accident details
- It has the many fields, a few are following

  - Accident Severity
  - Number of Vehicles
  - Day of Week
  - Number of Casualities
  - etc

- Find all accidents where more than 5 vehicles involved
- Find the most number of accidents day of the week
- Find the least number of accidents day of the week
- Find the largest collision with more causalities
- Find the total number of accidents involved in junctions
- Check traffic light condition

- Provide a categorical data analysis
- Provide a pairplot using kde
- Find the correlation between each variable and create a heatmap
- Create the boxplot for each column
- Create the swarmplot for each column
- Calculate outliers and then provide simple statistics

- Assume that you are employed as a policy maker for the road construction and signal construction
  - Provide a detailed insight about the study
  - Provide a better policy decision so that government can provide better roads
- Provide geographical information, you can use geopandas to plot the geographically the number of accidents per district or state etc.
- Analyse the CausaltiesBig.csv file and VehiclesBig.csv file and provide your analysis.