Heart Disease Project

Panchatcharam M

Associate Professor

Department of Mathematics and Statistics,
IIT Tirupati

Heart Attack Database

- Have a look at the heart.csv file
- This data contains information about a few medical details
- It has the following fields
  - Age: Age of the patient
  - Sex: Sex of the patient
  - exang: exercise induced angina (1 = yes; 0 = no)
  - ca: number of major vessels (0-3)
  - cp: Chest Pain type chest pain type
    - Value 1: typical angina
    - Value 2: atypical angina
    - Value 3: non-anginal pain
    - Value 4: asymptomatic
  - trtbps: resting blood pressure (in mm Hg)
  - chol: cholestoral in mg/dl fetched via BMI sensor
  - fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
  - rest_ecg: resting electrocardiographic results
  - Value 0: normal
  - Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  - Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
  - thalach: maximum heart rate achieved
  - target: 0= less chance of heart attack 1= more chance of heart attack

- Fill the missing values by 0
- Find the number of unique records of each column
- 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

- Based on the correlation study and heatmap, what can you conclude?
  - Does any of the variable has a direct correlation?
  - List out the variables that has positive correlation and negative correlation
  - Predict the age for which the heart attack is common

- If you are assigned to provide an analysis of this dataset and provide recommendation to patient about chlostrol and blood sugar, what would be your recommendation?
- Create 5 problem statements and provide the answer for the same based on this dataset

Heart Disease Database

- Have a look at the HeartDisease.csv file
- This data contains heart disease information
- It has the following fields
  - Year
  - LocationAbbr
  - LocationDesc
  - GeographicLevel
  - DataSource Class
  - Topic
  - Data_Value
  - Data_Value_Unit
  - Data_Value_Type
  - Data_Value_Footnote_Symbol
  - Data_Value_Footnote
  - StratificationCategory1
  - Stratification1
  - StratificationCategory2
  - Stratification2
  - TopicID
  - LocationID
  - Location 1

- Remove all insufficient data details
- List out the Asian and Pacific Islander with Heart disease mortality above 400 per 1 Lakh population
- List out the white people with heart disease mortality in ascending order and alphabetical order of county
- Find male Hispanic, national level and state level data

- Assume that you are a medical advisor to the county or state or USA.
- Analyse the data and provide input about the county or location or state which has more heart disease and require further inspection report
- Create 10 problem statement based on this dataset and provide your solution.