

MORTALITY RISK FACTOR ANALYSIS

- Name:Gopalakrishnan Kumar IIT- B Alumnus
- Occupation: Consultant Data Scientist
- LinkedIn URL: <https://www.linkedin.com/in/gopalakrishnan-kumar-a73301110/>

Project Goals



Objectives

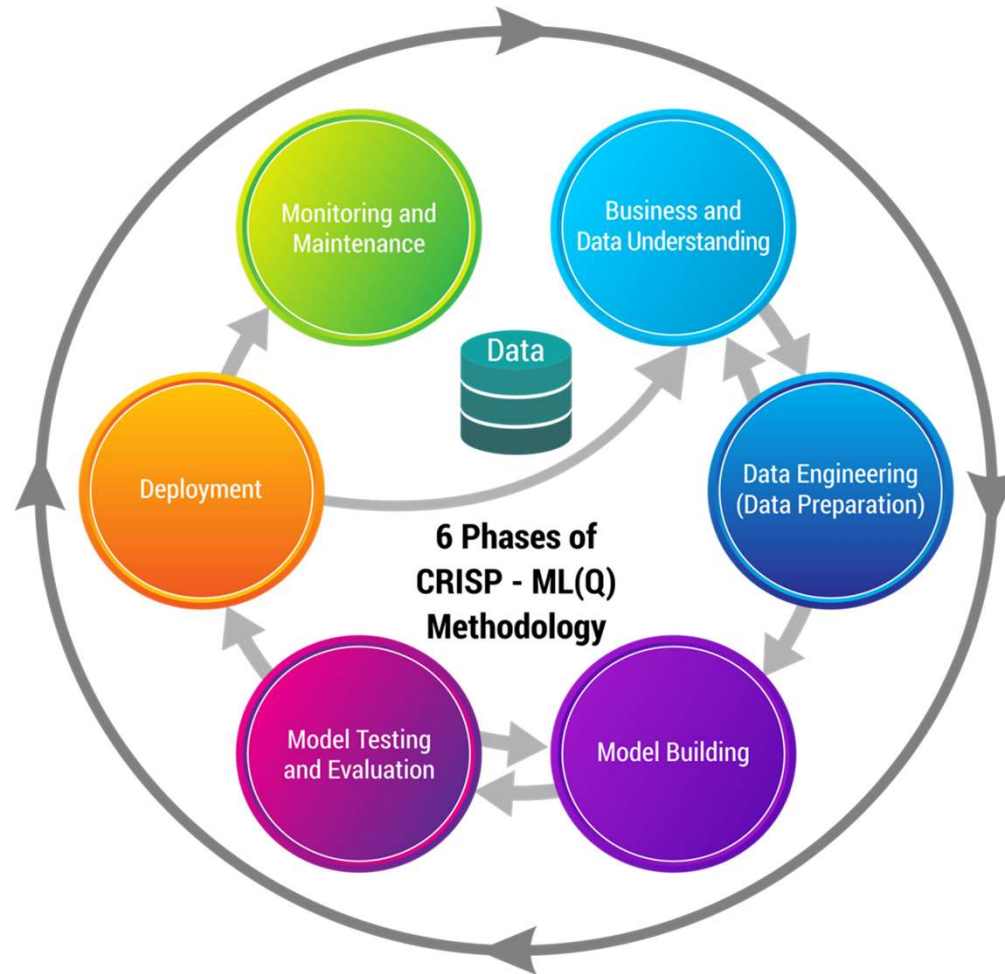
- Identify the factors which are influencing mortality risk



Constraints

- Limited dataset is not giving sufficient information
- Mortality Risk also depends on several factors such as intensity of disease, type of medication etc.

CRISP-ML(Q) Methodology



Technical Stacks



- **Microsoft Excel** - MS Excel is a commonly used Microsoft Office application. It is a spreadsheet program which is used to save and analyze numerical data. I extracted the data from the server and stored it in Excel for later use.



- **MySQL** - SQL (Structured Query Language) is a standardized programming language that's used to manage relational databases and perform various operations on the data. We used it to warehouse the data.



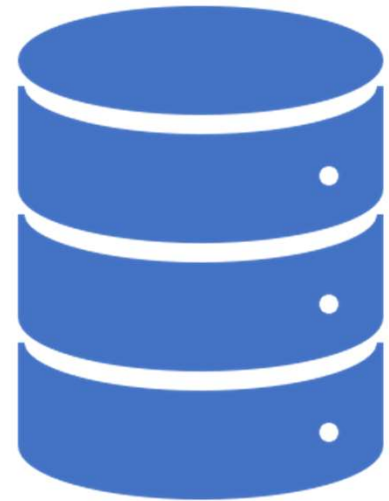
- **Spyder**- Spyder is an open-source cross-platform integrated development environment (IDE) for scientific programming in the Python language. Spyder integrates with a number of prominent packages in the scientific Python stack, including NumPy, SciPy, Matplotlib, pandas, IPython, SymPy and Cython, as well as other open-source software. I Used Spyder to develop python code



- **Tableau** - Tableau helps people see and understand data. Its visual analytics platform is transforming the way people use data to solve problems. I primarily used Tableau to make more enhanced visualizations on my data subsets.

Data Preparation

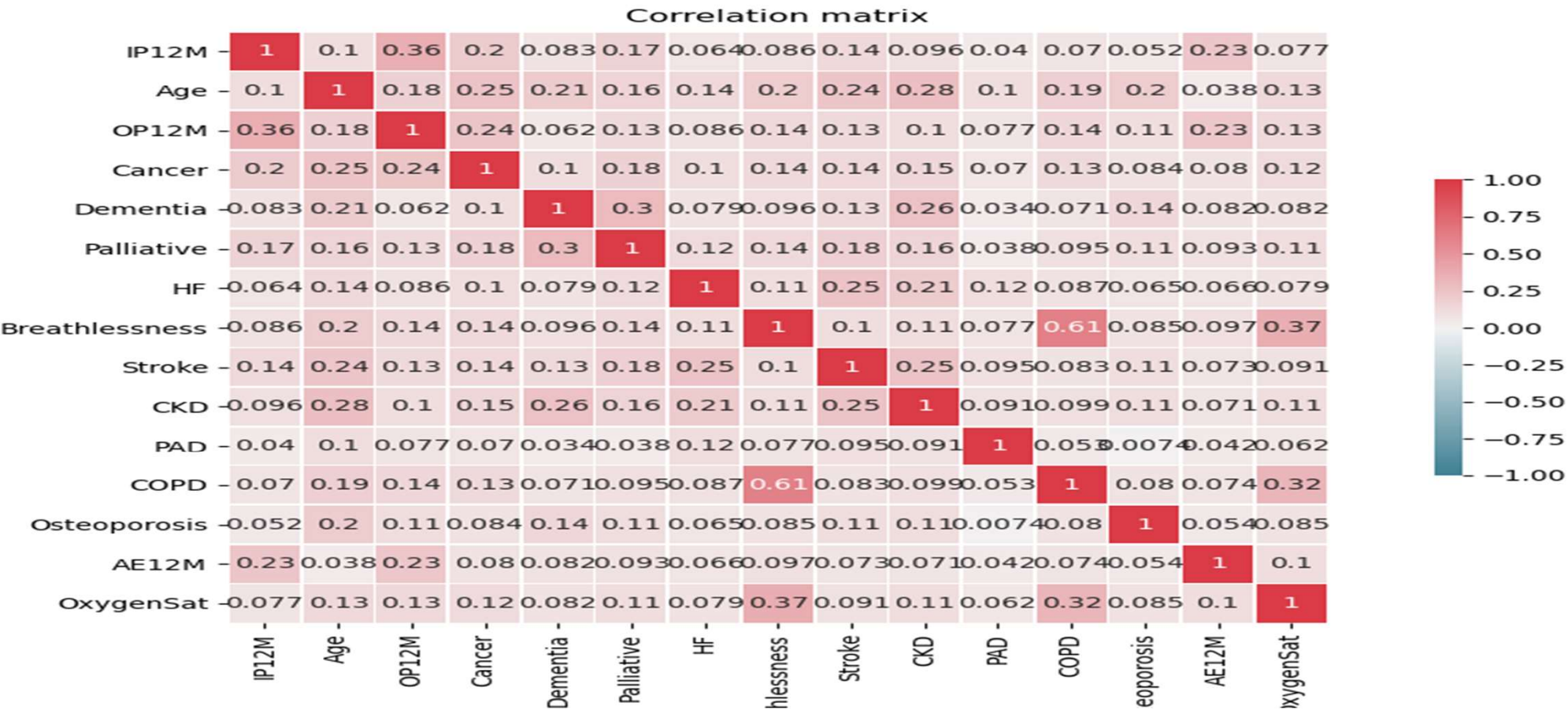
- Data used for training the model is genuine best known to our knowledge collected from clients for further analytics.
- Data cleansing
- And additional steps for data preparation
 - Dropped the irrelevant column.
 - Median imputation for missing data.
 - 2195 Duplicates detected.
 - Dropped a zero variance column.
 - We have an unnamed column.



Exploratory Data Analysis



Correlation Matrix of Top 15 Features:

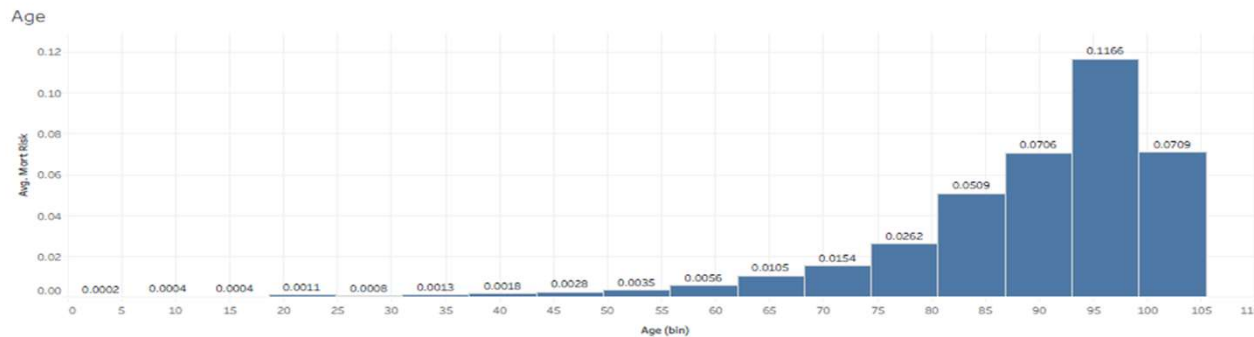


Bivariate Analysis (Factors Vs. Mortality Risk)

Bivariate analysis is analyzing two variables. categorical scatter plot is used to check for the correlation between two variables.

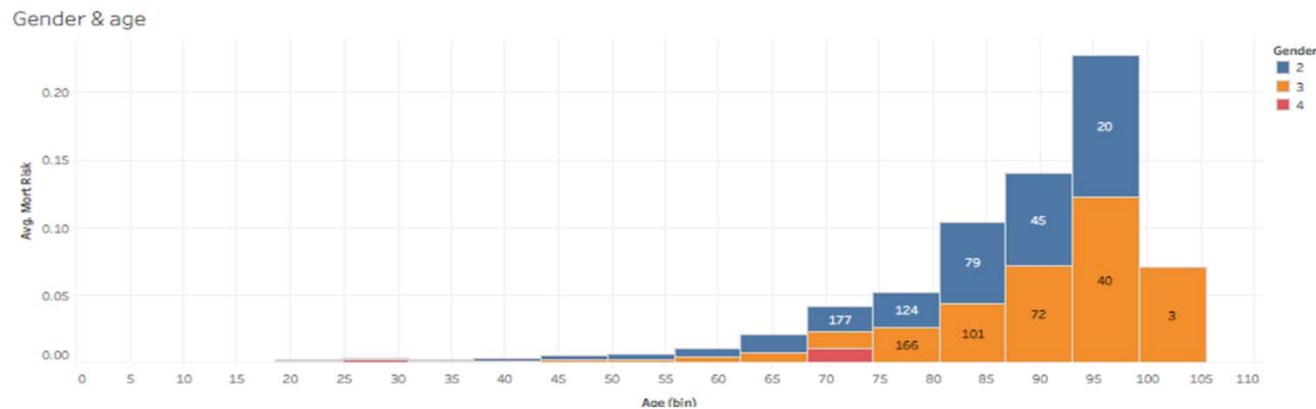
Age Vs. Mort_Risk:

People age has greater than 60 has high value of mortality risk as compared to person people age less than 60



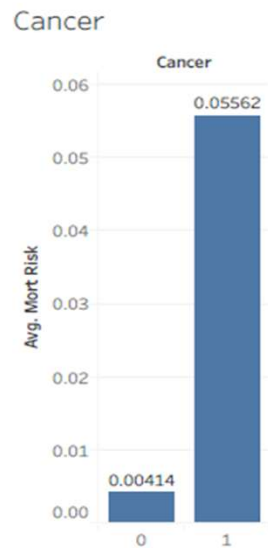
Age and Gender Vs. Mort_Risk:

Gender 3 people are effecting more than other genders with respect to age

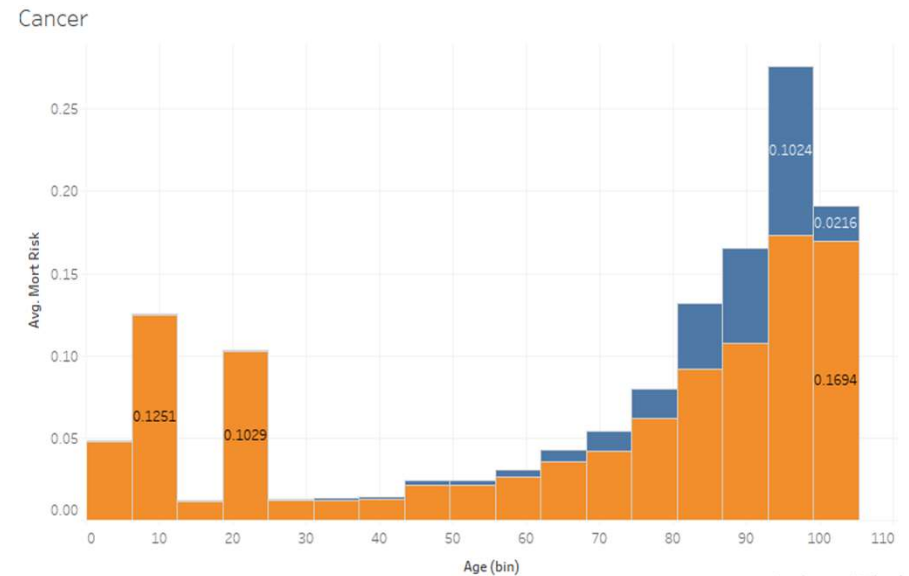


1 = present
0 = Not present

#Cancer VS Mortality Risk



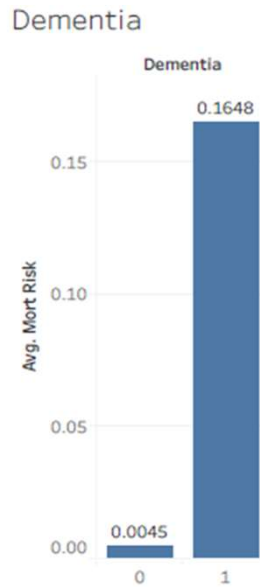
- Average Mortality Risk of people who is having cancer is 5.562%



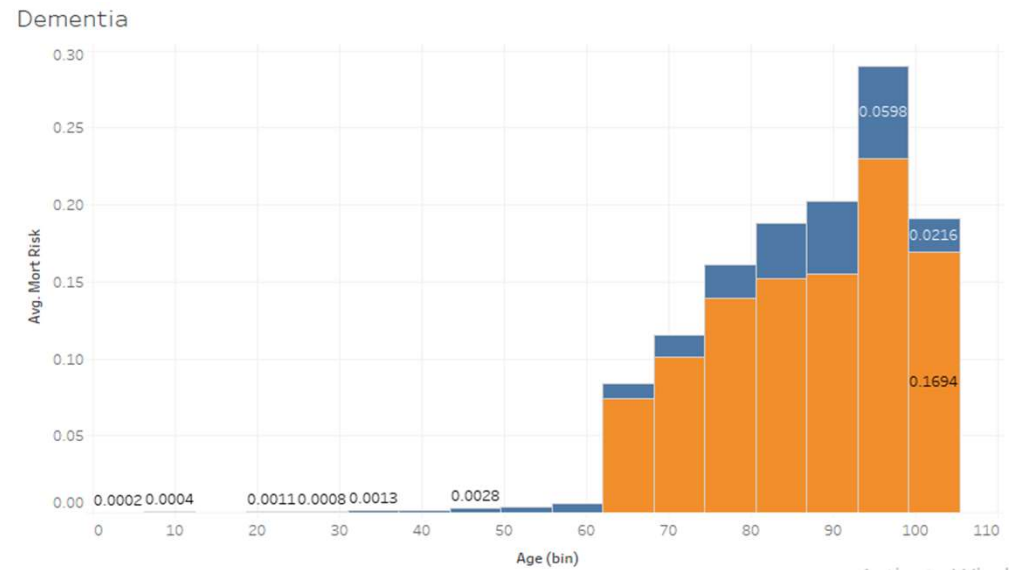
- Here age group less than 24yrs and greater than 74 yrs are having high mort_risk compared to the age group of 25 to 73.

1 = present
0 = Not present

#Dementia VS Mortality Risk



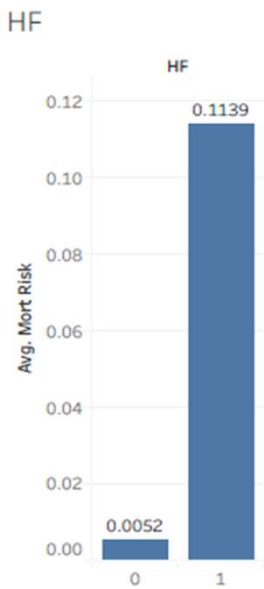
Average Mortality Risk of people who is suffering with dementia is 16.48%



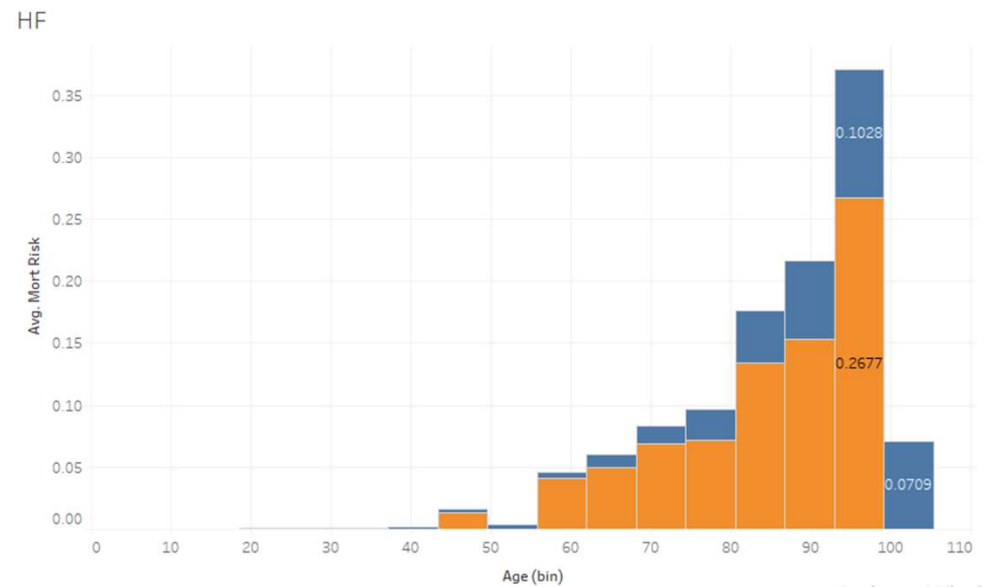
- Here age group greater than 63 yrs are having presence of disease and mort_risk and less than 63yrs people are having very less impact.

1 = present
0 = Not present

#HF VS Mortality Risk



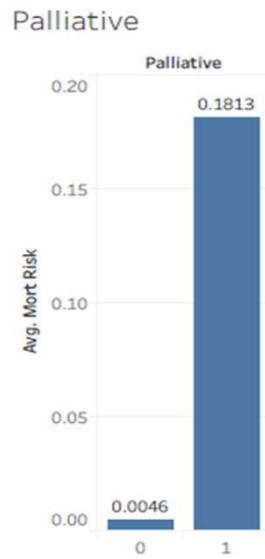
Average Mortality Risk of people who is having Heart Failure is 11.39%



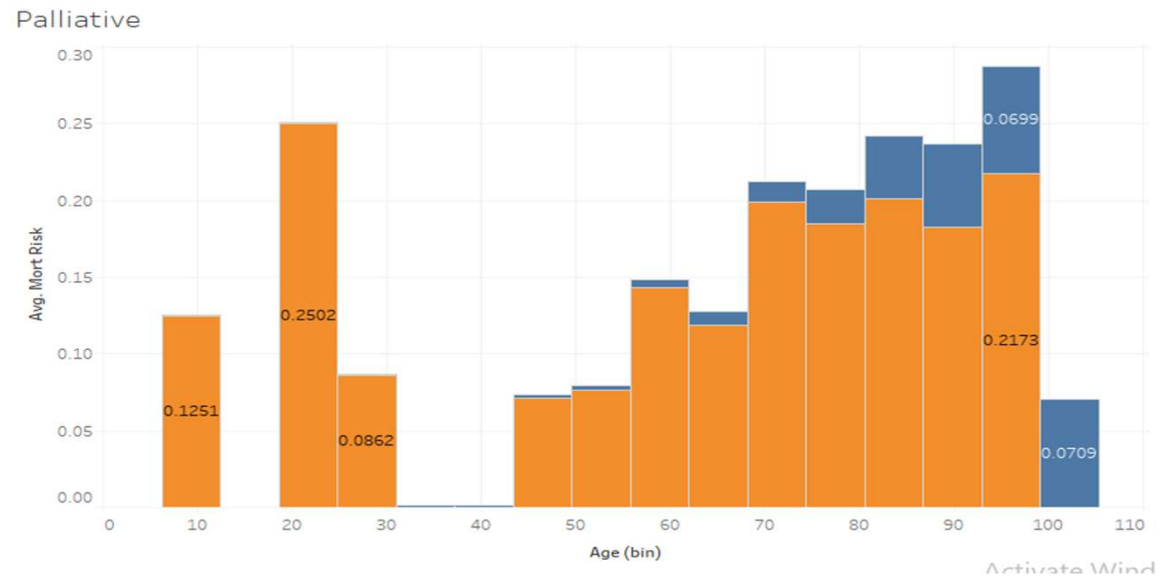
Here age group of more than 55yr having presence of disease & has high mortality risk.

1 = present
0 = Not present

#Palliative VS Mortality Risk

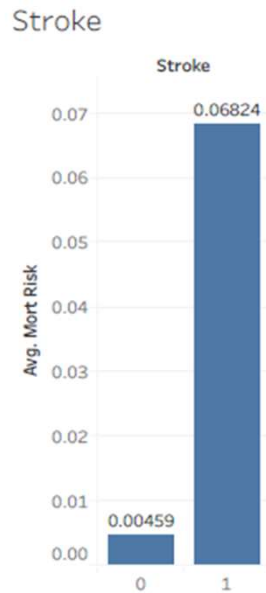


Average Mortality Risk for Palliative is 18.13%



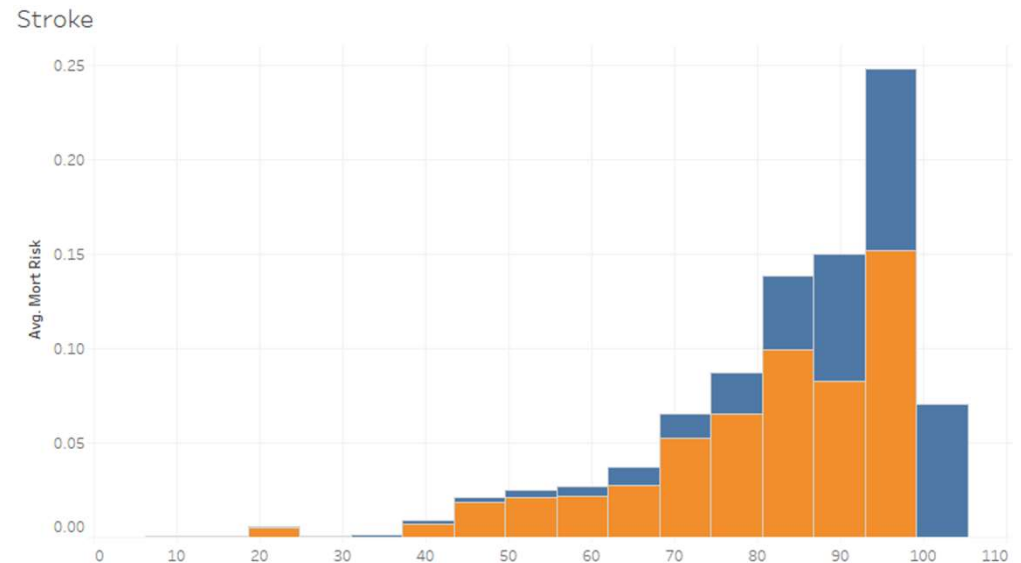
Palliative is having high impact irrespective of age group

#Stroke VS Mortality Risk



Average mortality risk for the people who is suffered from a stroke is 6.824%

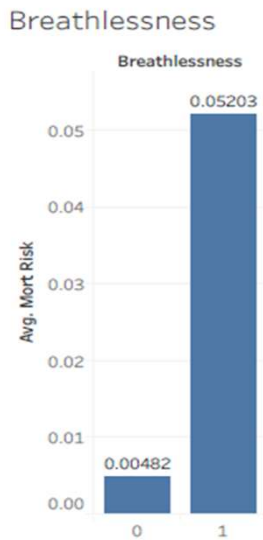
1 = present
0 = Not present



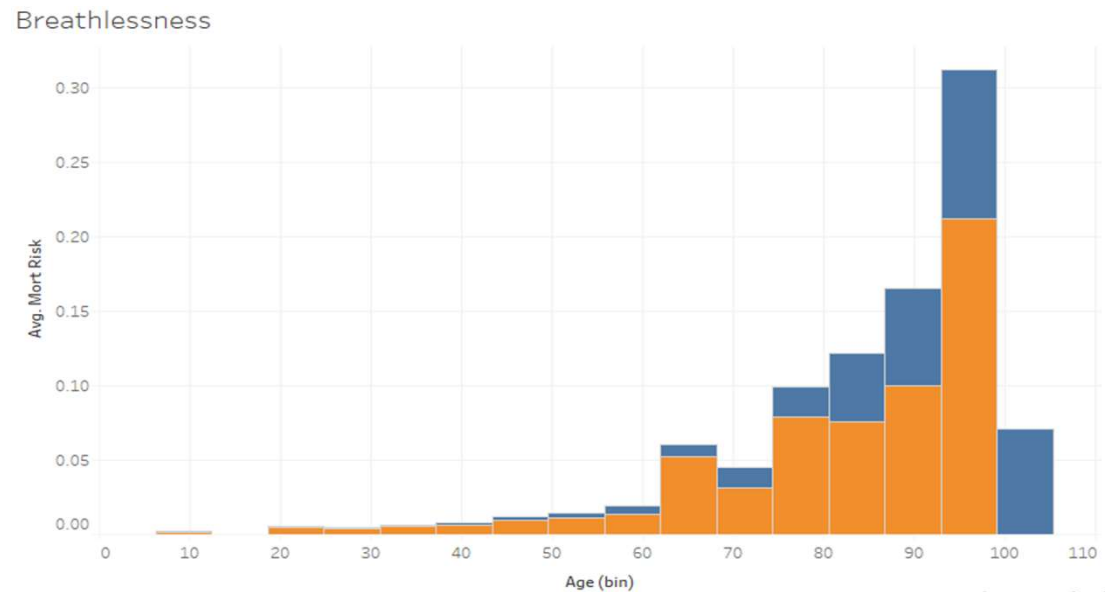
Stroke is present in the people majorly from the age of 38 and it is directly proportional to the age, when the age is increasing, mort risk wrt stroke is also increasing

1 = present
0 = Not present

#Breathlessness VS Mortality Risk



Average mortality risk for the people who is having breathlessness problem is 5.203%

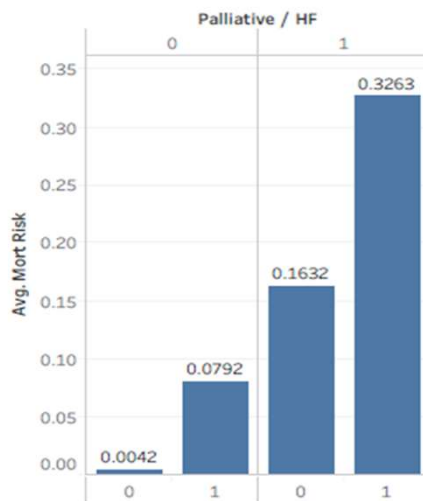


Mort risk for breathlessness is directly proportional to age

1 = present
0 = Not present

#Palliative & HF VS Mortality Risk

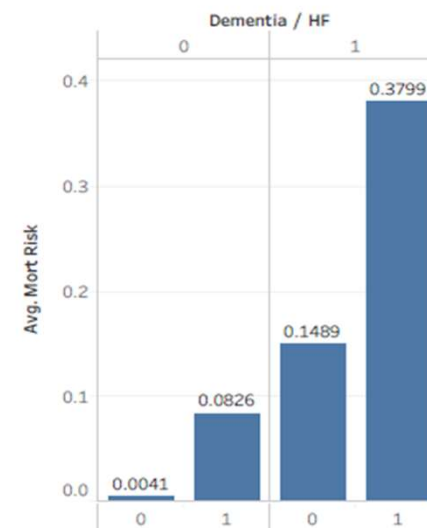
Palliative & HF



Average mort risk for the people who is having both palliative and HF is 32.63%

#Dementia & HF VS Mortality Risk

Dementia & HF

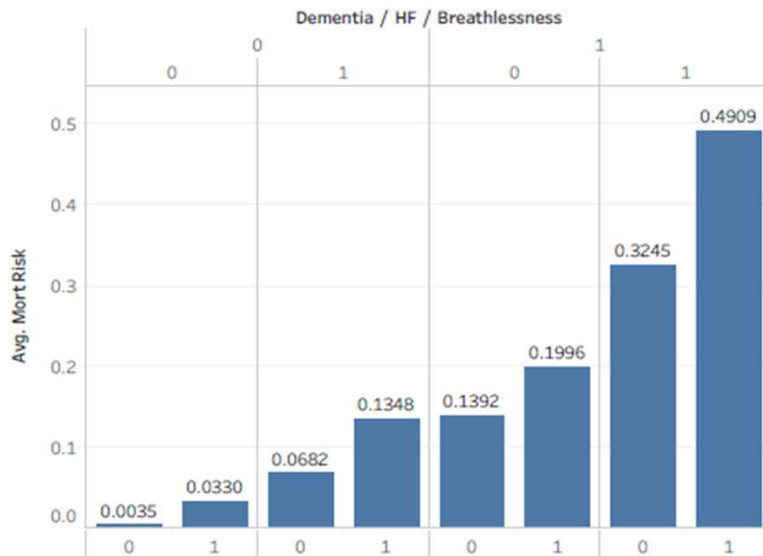


Average mort risk for the people who is having both Dementia and HF is 37.99%

1 = present
0 = Not present

#Dementia, HF & Breathlessness VS Mortality Risk

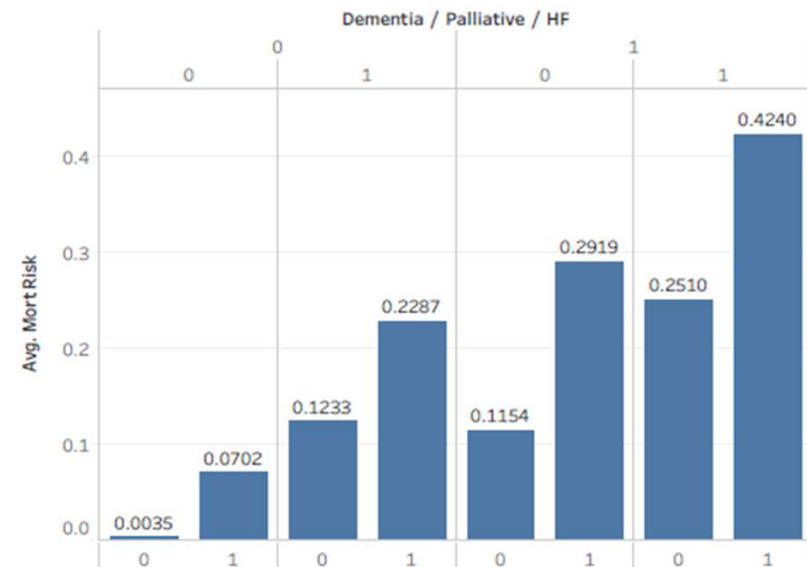
Dementia, HF & Breathlessness



Average mort risk for the people who is having Dementia, Breathlessness and HF is 49.09%

#Dementia, HF & Palliative VS Mortality Risk

Dementia, HF & Palliative

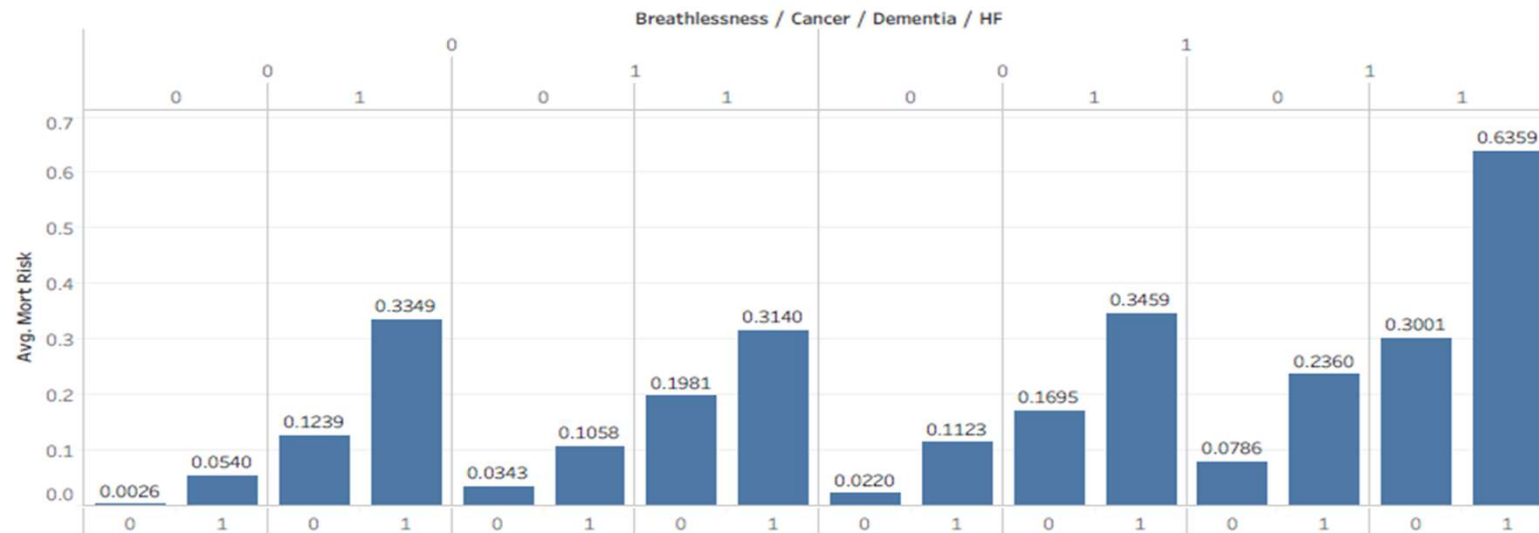


Average mort risk for the people who is having Dementia, Palliative and HF is 42.4%

1 = present
0 = Not present

#Dementia, HF, Breathlessness & Cancer VS Mortality Risk

Dementia, HF, Breathlessness & Cancer



Average mort risk for the people who is having Dementia, Breathlessness, HF & Cancer is 63.59%

Features Vs. Mortality Risk Tabular Values:

Single	Mort_Risk
Cancer	5.56%
Dementia	16.48%
Palliative	18.13%
HF	11.39%
Breathlessness	5.20%
Stroke	6.82%

combination of 2 vs mort_risk	Cancer	Dementia	Palliative	HF	Breathlessness	Stroke
Cancer		24.69%	21.68%	18.13%	11.35%	11.88%
Dementia	24.69%		27.99%	37.99%	23.85%	21.69%
Palliative	21.68%	27.99%		32.63%	24.88%	22.71%
HF	18.13%	37.99%	32.63%		18.96%	12.02%
Breathlessness	11.35%	23.85%	24.88%	18.96%		14.31%
Stroke	11.88%	21.69%	22.71%	12.02%	14.31%	
Average	17.55%	27.24%	25.98%	23.95%	18.67%	16.52%

combination of 3 vs mort_risk	Cancer	Dementia	Palliative	HF	Breathlessness	Stroke	Average
Dementia & Palliative	35.27%			42.40%	34.38%	32.72%	36.19%
Dementia & HF	42.13%		42.40%		49.09%	35.46%	42.27%
Palliative & HF	35.54%	42.40%			35.20%	29.40%	35.64%

combination of 4 vs mort_risk	Cancer	Dementia	Palliative	HF	Breathlessness	Stroke
Dementia & HF & Breathlessness	63.59%		49.09%			49.09%
Dementia & Palliative & HF	47.62%				49.09%	39.76%

Model Building



Model Description

Algorithms used:

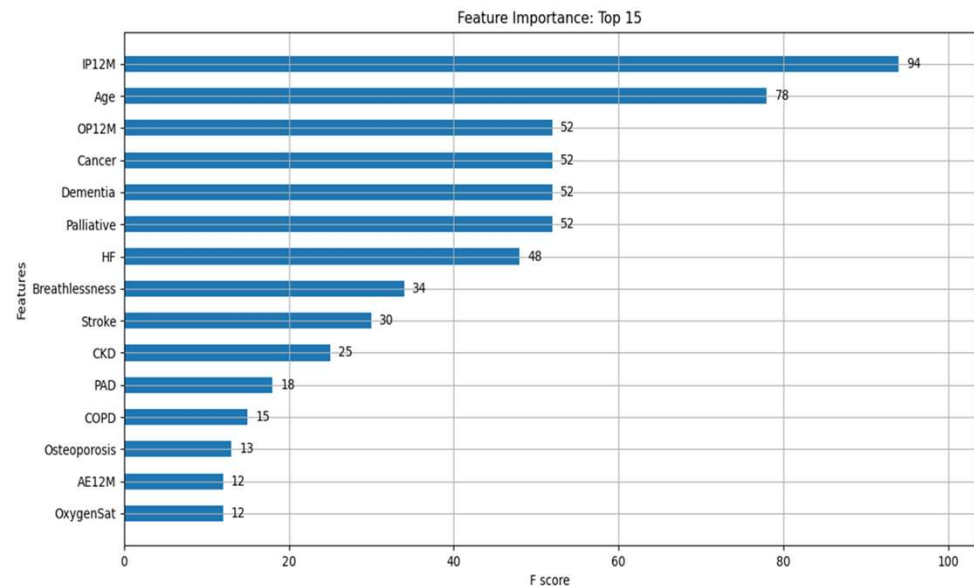
XGB Regression with max depth 3 to find feature importance.

Survival Analysis with in-patient hospital activity as the time component.

Descriptions:

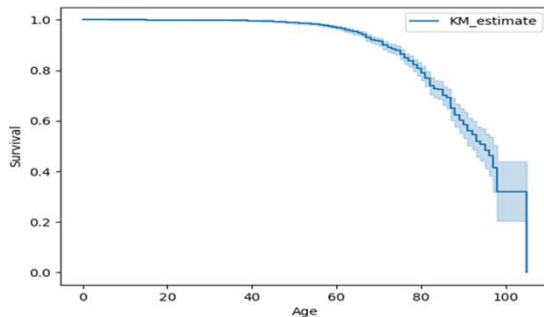
Split the data into training (80%) & testing (20%)

Feature Importance was obtained after using XGB Regression.

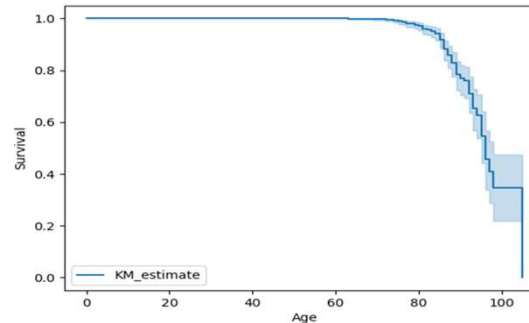


Survival Analysis

- Survival analysis is a branch of statistics for analyzing the expected duration of time until one event occurs, such as death in biological organisms and failure in mechanical systems.
- For this project, I have taken into account the Top 4 diseases from the feature importance plot as the combination of these is giving high mort risk:
 - Cancer
 - Dementia
 - Breathlessness
 - Heart Failure

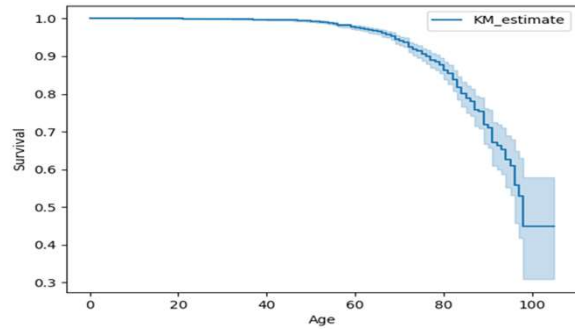


The plot alongside is the survival of cancer patients with respect to their age

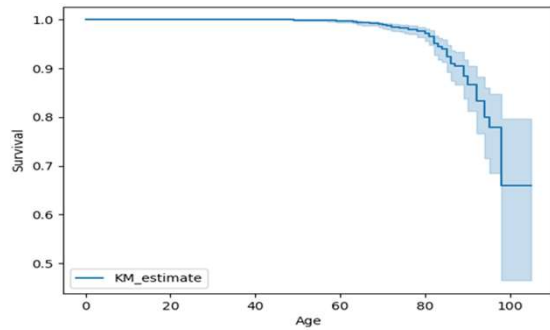


The plot alongside is the survival of Dementia patients with respect to their age

Survival Analysis



The plot alongside is the survival of Breathlessness patients with respect to their age



The plot alongside is the survival of HF patients with respect to their age

Deployment Strategy

- Web Based
- Mobile Based
- Cloud Based
- Enterprise Based





Thanks