**Heart Disease Patients**

**Team:** CS\_13

Team member 1: مريم ابراهيم رشدي ابراهيم

ID: 20191700623

Team member 2: مريم محمود عبد القادر محمد سويلم

ID: 20191700635

Team member 3: عمرو عزت حنفي محمد

ID: 20191700428

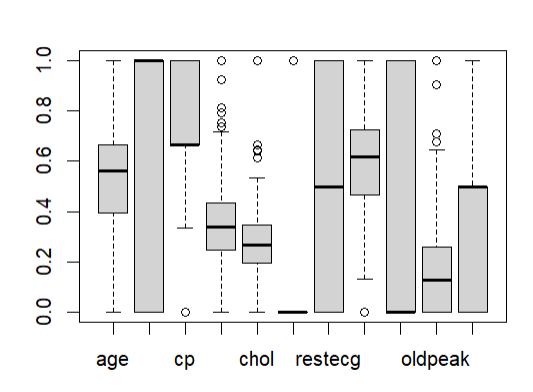
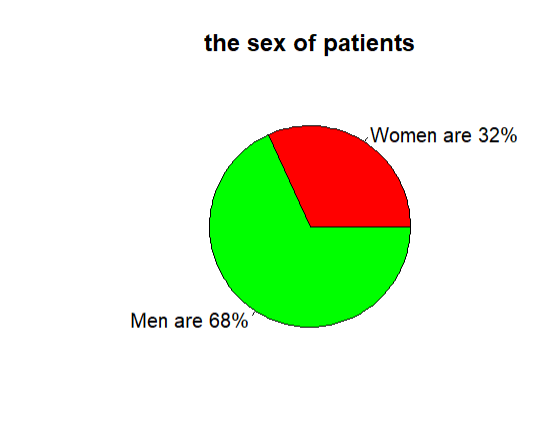
Team member 4: محمد اسامة عبدالراضي عفيفي

ID: 20191700501

**Preprocessing:**

* Fill missing values in data with median.
* Drop (ID) column as it’s not useful.
* Applied Feature scaling on data, to make it range between (0-1).
* Used feature correlation visualization to make a good decision in feature selection.

**Visualization:**

* Boxplot for data: 
* Pie Char

**Models:**

1. K-means
2. DBscan

**Evaluation Methods:**

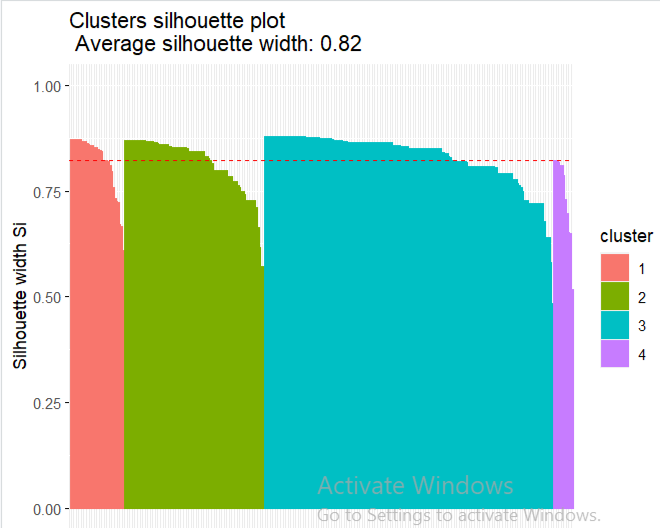
-The Silhouette score was used as a metric to evaluate the quality of clustering results and determine the optimal number of clusters. It measures how well each data point fits into its assigned cluster compared to other clusters. A higher Silhouette score indicates that the data points are well-clustered, while a lower score suggests overlapping or poorly separated clusters.

-Elbow Method was used to determine the number of centroids(k) to use in a k-means clustering algorithm.  In this method to determine the k-value we continuously iterate for k=1 to k=10 (Here n is the hyperparameter that we choose as per our requirement). For every value of k, we calculate the within-cluster sum of squares (WCSS) value.

**Trials:**

Average silhouette score in **K-means** for different features combination:

* (Thalach-slope-CP): 0.48 with 5 centers
* (age-Old peak- CP): 0.33
* (age-Old peak- CP- slope):0.4
* (thalach-slope): 0.65 with 4 clusters.
* (thalach-slope):0.65 with 3 clusters.
* (fbs-trestbps-chol-restecg-sex):0.74 with 7 clusters
* (fbs-trestbps-chol-restecg):0.75 with 4 clusters
* (trestbps-fbs-recetecg): 0.81 with 4 clusters
* (trestbps-fbs -sex):0.82 with 5 clusters

****

Average silhouette score in **DBscan** for different features combination:

* (age-Old peak- CP): 0.33, eps:0.2, with 5 clusters
* (trestbps-fbs-recetecg): 0.78

eps:0.17, with 5 clusters

* (trestbps-fbs-recetecg-sex): 0.79,

eps:0.27, with 5 clusters.

* (trestbps-fbs -sex):0.82 , eps:0.25, 5 clusters

A picture containing text, screenshot, line, diagram

Description automatically generated

A picture containing text, screenshot, diagram, plot

Description automatically generated