Mahatma Education Society's
Pillai College of Arts, Commerce & Science
(Autonomous)
Affiliated to University of Mumbai
NAAC Accredited 'A' grade (3 cycles)
Best College Award by University of Mumbai
ISO 9001:2015 Certified

MAHATMA EDUCATION SOCIETY'S

PILLAI COLLEGE OF ARTS, COMMERCE & SCIENCE

(Autonomous)

NEW PANVEL

PROJECT REPORT ON

"Heart Disease Prediction"

IN PARTIAL FULFILLMENT OF

MASTERS OF DATA ANALYTICS

SEMESTER 1– 2023-24

PROJECT GUIDE

Prof. Shivapradeep Muthupandi

SUBMITTED BY: Rohan Uttam Khapane

ROLL NO: 3147

Introduction :

Cardiovascular diseases, including heart disease, represent a significant global health challenge, accounting for a substantial portion of morbidity and mortality worldwide. Understanding the factors that contribute to heart disease and identifying individuals at risk is crucial for effective prevention and intervention strategies. In this report, we delve into a comprehensive health and lifestyle dataset, with a particular focus on heart disease as the target variable.

The dataset at hand provides a wealth of information about individuals' general health, lifestyle choices, dietary habits, and various health-related factors. By analyzing this dataset, we aim to gain insights into the relationships between different variables and the prevalence of heart disease among the studied population. Our primary objective is to uncover patterns, risk factors, and potential predictors associated with heart disease, providing valuable information that can aid in both individual health management and public health policy development.

Dataset link:

https://www.kaggle.com/datasets/alphiree/cardiovascular-diseases-risk-prediction-dataset?datasetId=3475382&sortBy=voteCount

Using the above dataset will be performing all the Life Cycle phases of Data Science , i.e , Data Understanding, Data Preparation ,Data Visualization(EDA), Data Modelling, Model Evaluation.

Business Understanding and Model Deployment are also the phases , but as this in not an industry oriented project ,hence it is not included above.

The Below table specifies the name of the columns, their data types , the feature is categorical or numerical and their description

Column Name	Class of Column	Categorical /Numerical Values	Description
General_Health	character	Categorical	General health of a person categories as poor, fair, good and very good
Checkup	character	Categorical	About how long has it been since you last visited a doctor for a routine checkup
Exercise	character	Categorical	During the past month,other than your regular job,person participate in any physical activities or exercises such as running
Heart_Disease	character	Categorical	Person that reported having coronary heart disease
Skin_Cancer	character	Categorical	Whether the Person has a skin cancer or not

Other_Cancer	character	Categorical	Whether the person has any other type of cancer or not
Depression	character	Categorical	Whether the person has depression or not
Diabetes	character	Categorical	Whether the person has Diabetes or not
Arthritis	character	Categorical	Whether the person has Arthritis or not
Sex	character	Categorical	Gender of Person
Age_Category	character	Categorical	Age of Person
Height_.cm.	numeric	Numerical	Height of Person in cm
Weight_.kg.	numeric	Numerical	Weight of Person in kg
BMI	numeric	Numerical	Body Mass Index of Person
Smoking_History	character	Categorical	Whether the person is smoker of not
Alcohol_Consumption	numeric	Numerical	Alcohol Consumption of person in a week
Fruit_Consumption	numeric	Numerical	Number of Fruit Consumed by person in week
Green_Vegetables_Consumption	numeric	Numerical	Green vegetable consumption of a person
FriedPotato_Consumptioin	numeric	Numerical	Fried Potato consumption of a person

Loading the required libraries

```
> library(ggplot2)
> library(dplyr)
> library(caret)
> library(caTools)
```

Importing the data using read.csv()

```
> #---Import data-----------------------------------------------------------------------------------------------------------
> getwd()
[1] "D:/PG_SEM_1/Research Methodology/Research_paper"
> # setwd('D:/PG_SEM_1/Foundation of Data Science/PROJECT')
> cvd<-read.csv('D:\\PG_SEM_1\\Foundation of Data Science\\PROJECT\\CVD_cleaned.csv',header=TRUE,stringsAsFactors=FALSE)
```

#Data Understanding

head(cvd) give the first 5 row in dataset

```
> head(cvd)#show first 5 rows of the dataset
  General_Health                Checkup Exercise Heart_Disease Skin_Cancer Other_Cancer Depression Diabetes Arthritis    Sex Age_Category
1           Poor Within the past 2 years       No            No          No           No         No       Yes Female        70-74
2      Very Good    Within the past year       No           Yes          No           No         No       Yes  No Female        70-74
3      Very Good    Within the past year      Yes            No          No           No         No       Yes  No Female        60-64
4           Poor    Within the past year      Yes           Yes          No           No         No       Yes  No   Male        75-79
5           Good    Within the past year       No            No          No           No         No        No  No   Male          80+
6           Good    Within the past year       No            No          No           No        Yes        No Yes   Male        60-64
  Height_.cm. Weight_.kg.   BMI Smoking_History Alcohol_Consumption Fruit_Consumption Green_Vegetables_Consumption FriedPotato_Consumption
1         150       32.66 14.54             Yes                   0               30                           16                       12
2         165       77.11 28.29              No                   0               30                            0                        4
3         163       88.45 33.47              No                   4               12                            3                       16
4         180       93.44 28.73              No                   0               30                           30                        8
5         191       88.45 24.37             Yes                   0                8                            4                        0
6         183      154.22 46.11              No                   0               12                           12                       12
```

tail(cvd) give the last 5 rows in dataset

```
> tail(cvd)#show last 5 rows of the dataset
       General_Health                Checkup Exercise Heart_Disease Skin_Cancer Other_Cancer Depression
308849           Good Within the past 5 years      Yes            No          No           No         No
308850      Very Good    Within the past year      Yes            No          No           No         No
308851           Fair Within the past 5 years      Yes            No          No           No         No
308852      Very Good     5 or more years ago      Yes            No          No           No        Yes
308853      Very Good    Within the past year      Yes            No          No           No         No
308854      Excellent    Within the past year      Yes            No          No           No         No
                                  Diabetes Arthritis    Sex Age_Category Height_.cm. Weight_.kg.   BMI Smoking_History
308849                                  No        No   Male        55-59         168       58.97 20.98              No
308850                                  No        No   Male        25-29         168       81.65 29.05              No
308851                                 Yes        No   Male        65-69         180       69.85 21.48              No
308852 Yes, but female told only during pregnancy        No Female        30-34         157       61.23 24.69             Yes
308853                                  No        No   Male        65-69         183       79.38 23.73              No
308854                                  No        No Female        45-49         160       81.19 31.71              No
       Alcohol_Consumption Fruit_Consumption Green_Vegetables_Consumption FriedPotato_Consumption
308849                   0               16                           12                       0
308850                   4               30                            8                       0
308851                   8               15                           60                       4
308852                   4               40                            8                       4
308853                   3               30                           12                       0
308854                   1                5                           12                       1
```

name(cvd) give the name of column present in dataset

```
> names(cvd) #displays the names of all the column present in the dataset
 [1] "General_Health"              "Checkup"                     "Exercise"                    "Heart_Disease"
 [5] "Skin_Cancer"                 "Other_Cancer"                "Depression"                  "Diabetes"
 [9] "Arthritis"                   "Sex"                         "Age_Category"                "Height_.cm."
[13] "Weight_.kg."                 "BMI"                         "Smoking_History"             "Alcohol_Consumption"
[17] "Fruit_Consumption"           "Green_Vegetables_Consumption" "FriedPotato_Consumption"
```

dim(cvd)gives the now of row and no of column in dataset

```
> dim(cvd)#show the dimensions(no. of rows and no. of columns) of the dataset
[1] 308854     19
```

sapply(cvd,class) show the class of all the column present in dataset

```
> sapply(cvd,class) #show the class of all column of data set in a simple format
      General_Health             Checkup            Exercise       Heart_Disease
         "character"           "character"         "character"         "character"
        Skin_Cancer        Other_Cancer          Depression            Diabetes
         "character"           "character"         "character"         "character"
          Arthritis                 Sex        Age_Category         Height_.cm.
         "character"           "character"         "character"           "numeric"
        Weight_.kg.                 BMI     Smoking_History Alcohol_Consumption
          "numeric"             "numeric"         "character"           "numeric"
  Fruit_Consumption Green_Vegetables_Consumption FriedPotato_Consumption
          "numeric"             "numeric"           "numeric"
```

summary(cvd) show the summary of all the columns of the dataset

```
> summary(cvd)#show the summary of all columns of dataset
 General_Health       Checkup            Exercise         Heart_Disease      Skin_Cancer        Other_Cancer        Depression
 Length:308854      Length:308854      Length:308854      Length:308854      Length:308854      Length:308854      Length:308854
 Class :character   Class :character   Class :character   Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character


   Diabetes          Arthritis             Sex            Age_Category       Height_.cm.     Weight_.kg.          BMI
 Length:308854      Length:308854      Length:308854      Length:308854      Min.   : 91.0   Min.   : 24.95   Min.   :12.02
 Class :character   Class :character   Class :character   Class :character   1st Qu.:163.0   1st Qu.: 68.04   1st Qu.:24.21
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :170.0   Median : 81.65   Median :27.44
                                                                             Mean   :170.6   Mean   : 83.59   Mean   :28.63
                                                                             3rd Qu.:178.0   3rd Qu.: 95.25   3rd Qu.:31.85
                                                                             Max.   :241.0   Max.   :293.02   Max.   :99.33
 Smoking_History    Alcohol_Consumption Fruit_Consumption Green_Vegetables_Consumption FriedPotato_Consumption
 Length:308854      Min.   : 0.000      Min.   :  0.00    Min.   :  0.00               Min.   :  0.000
 Class :character   1st Qu.: 0.000      1st Qu.: 12.00    1st Qu.:  4.00               1st Qu.:  2.000
 Mode  :character   Median : 1.000      Median : 30.00    Median : 12.00               Median :  4.000
                    Mean   : 5.096      Mean   : 29.84    Mean   : 15.11               Mean   :  6.297
                    3rd Qu.: 6.000      3rd Qu.: 30.00    3rd Qu.: 20.00               3rd Qu.:  8.000
                    Max.   :30.000      Max.   :120.00    Max.   :128.00               Max.   :128.000
```

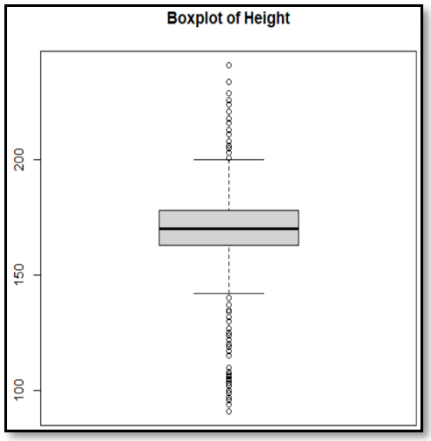
#Preparation of Data

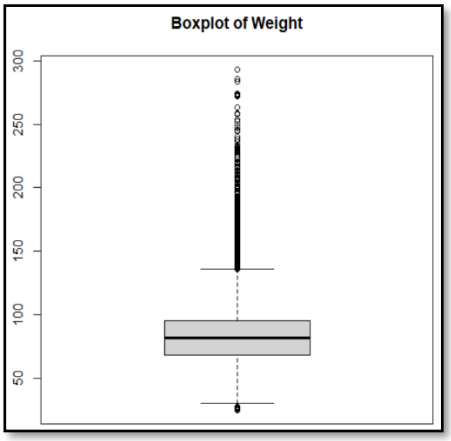
Check the Null values in each columns

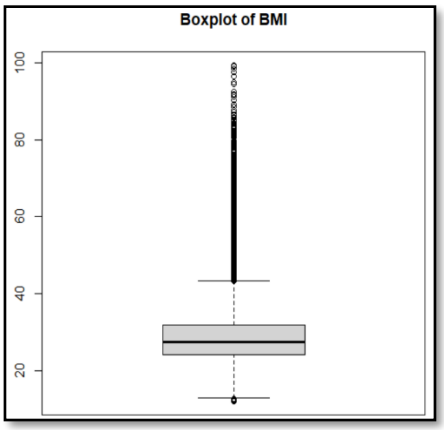
```
> #----Check NA in dataset----------------------------------------------------------------------------------------------------
> colSums(is.na(cvd)) #give all the null values present in the column seperatly
             General_Health                    Checkup                   Exercise                Heart_Disease
                          0                          0                          0                          0
               Skin_Cancer                Other_Cancer                 Depression                   Diabetes
                          0                          0                          0                          0
                 Arthritis                        Sex                Age_Category                Height_.cm.
                          0                          0                          0                          0
               Weight_.kg.                        BMI             Smoking_History         Alcohol_Consumption
                          0                          0                          0                          0
         Fruit_Consumption Green_Vegetables_Consumption      FriedPotato_Consumption
                          0                          0                          0
```

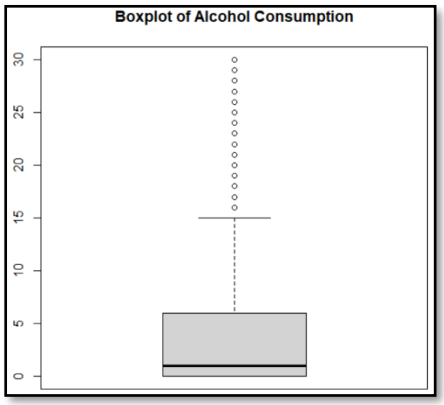
Boxplot of Numeric variables

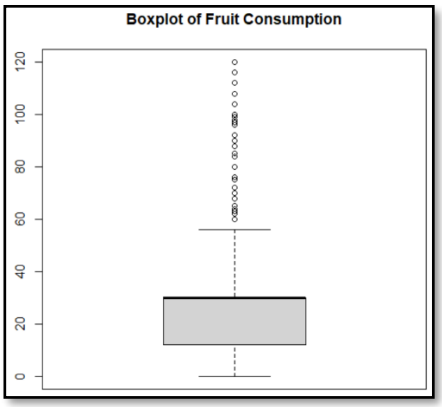
```
#---Boxplot-----------------------------------------------------------------------------------------------------------
boxplot(cvd$Height_.cm.,horizontal = FALSE,main="Boxplot of Height")
boxplot(cvd$Weight_.kg.,horizontal=FALSE,main="Boxplot of Weight")
boxplot(cvd$BMI,horizontal = FALSE,main="Boxplot of BMI")
boxplot(cvd$Alcohol_Consumption,horizontal = FALSE,main="Boxplot of Alcohol Consumption")
boxplot(cvd$Fruit_Consumption,horizontal=FALSE,main="Boxplot of Fruit Consumption")
boxplot(cvd$Green_Vegetables_Consumption,horizontal = FALSE,main="Boxplot of Green Vegetables Consumption")
boxplot(cvd$FriedPotato_Consumption,horizontal=FALSE,main="Boxplot of Fried Potato Consumption")
```

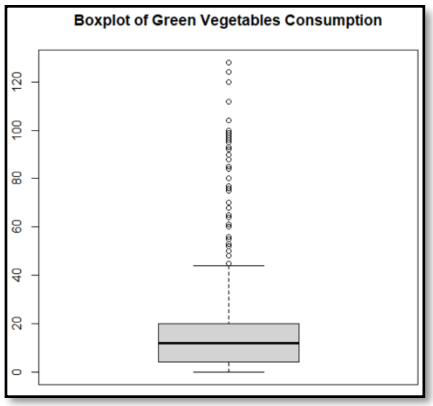






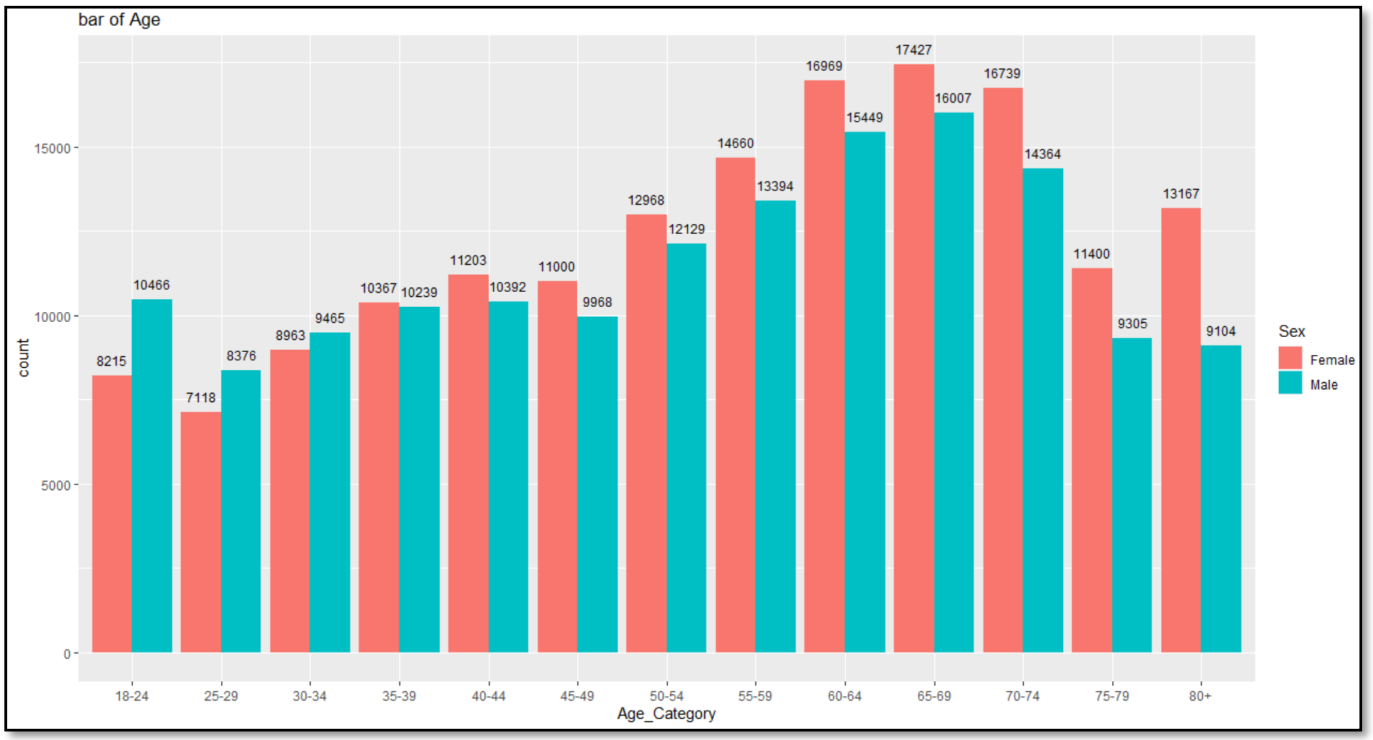




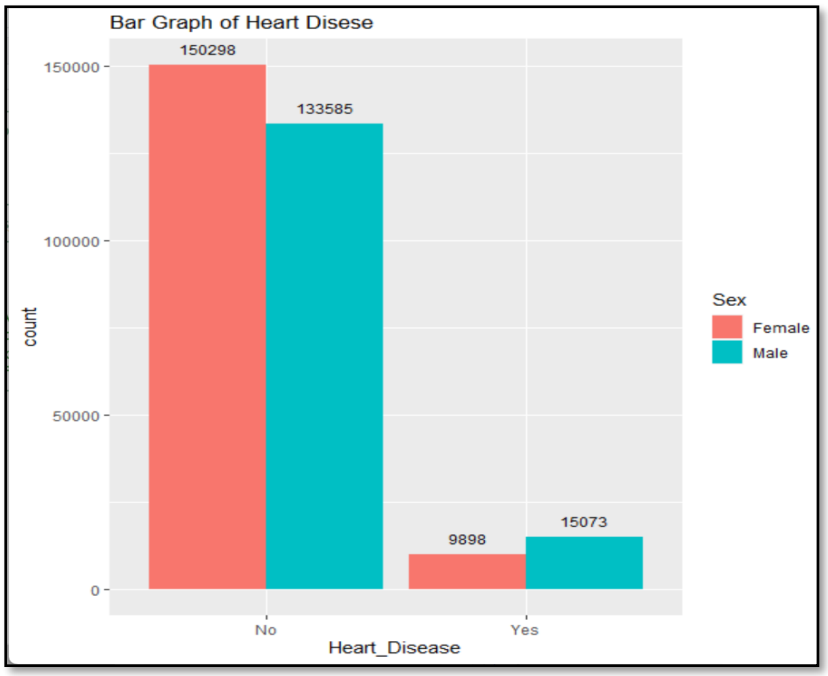


#Data Visualization

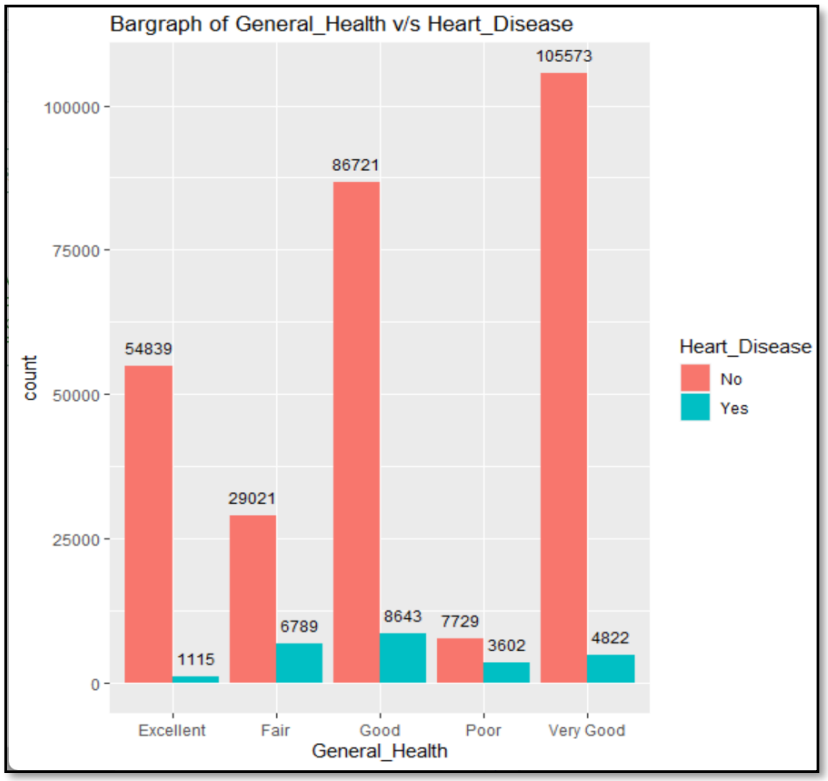
```
ggplot(cvd,aes(x=Age_Category,fill=Sex))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('bar of Age')
```



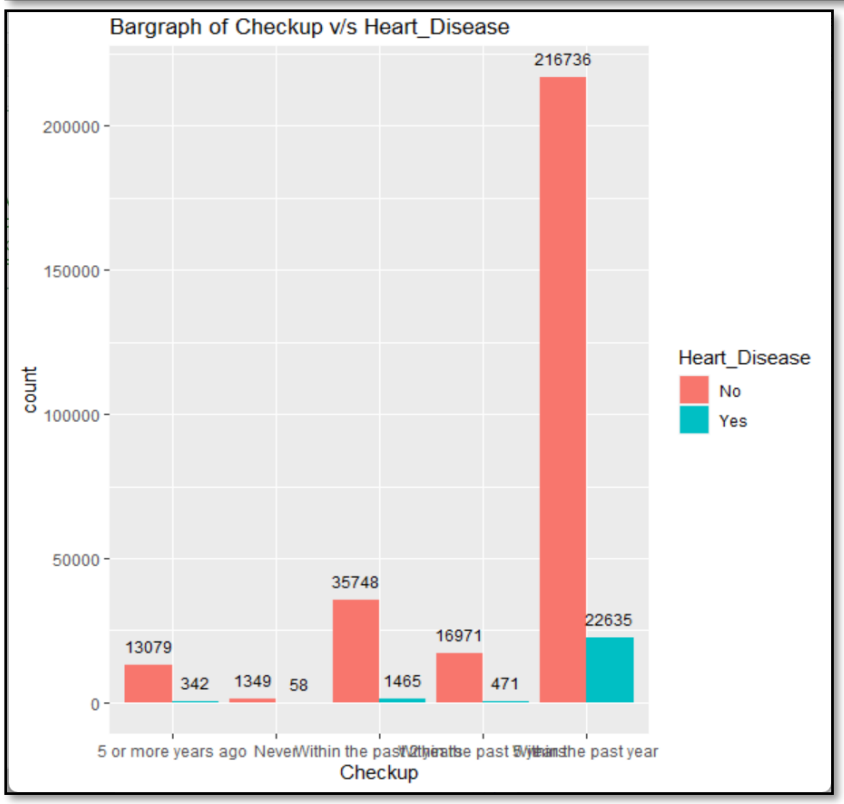
```
ggplot(cvd,aes(x=Heart_Disease,fill=Sex))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1,)+
  ggtitle('Bar Graph of Heart Disese')
```



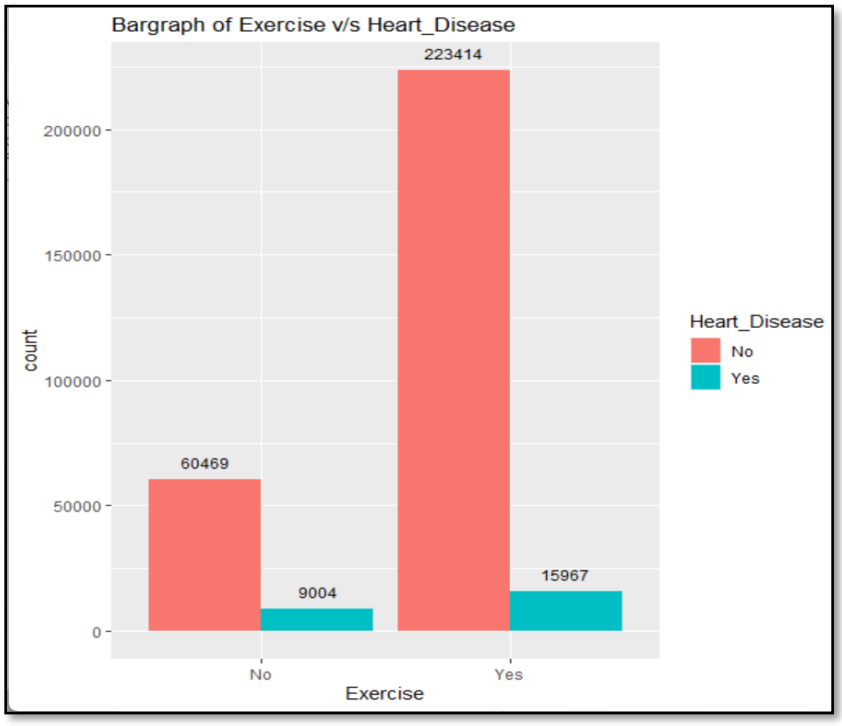

```
ggplot(cvd,aes(x=General_Health,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of General_Health v/s Heart_Disease')
```



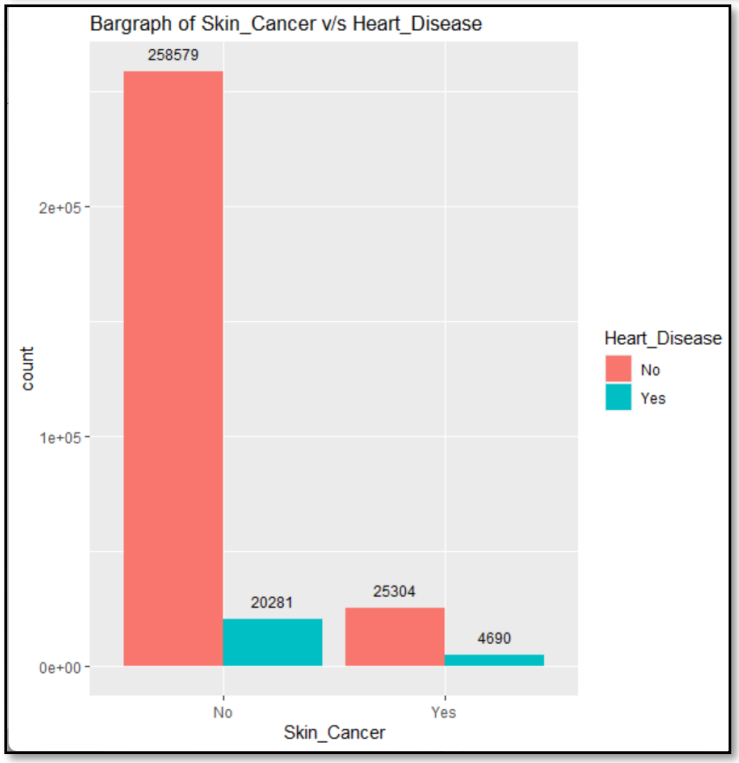
```
ggplot(cvd,aes(x=Checkup,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of Checkup v/s Heart_Disease')
```



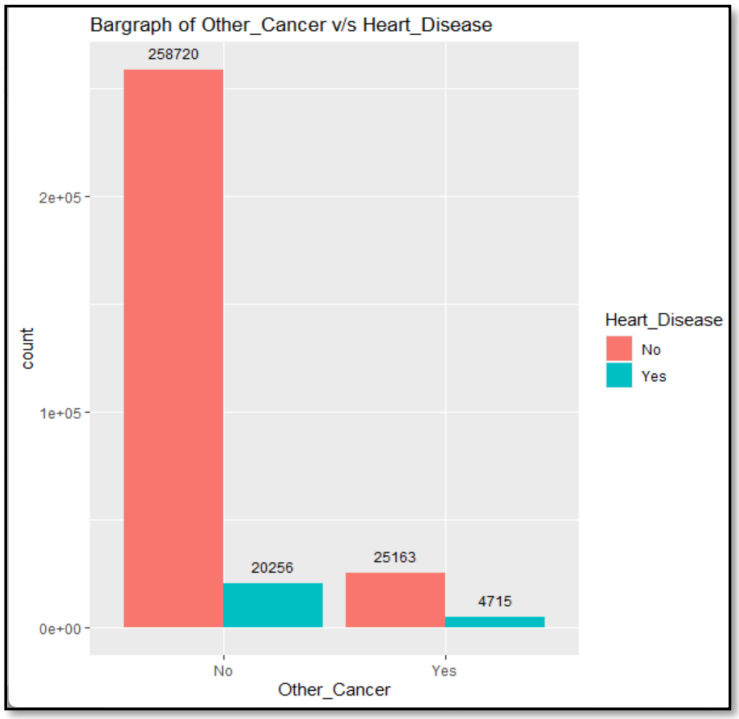

```
ggplot(cvd,aes(x=Exercise,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of Exercise v/s Heart_Disease')
```



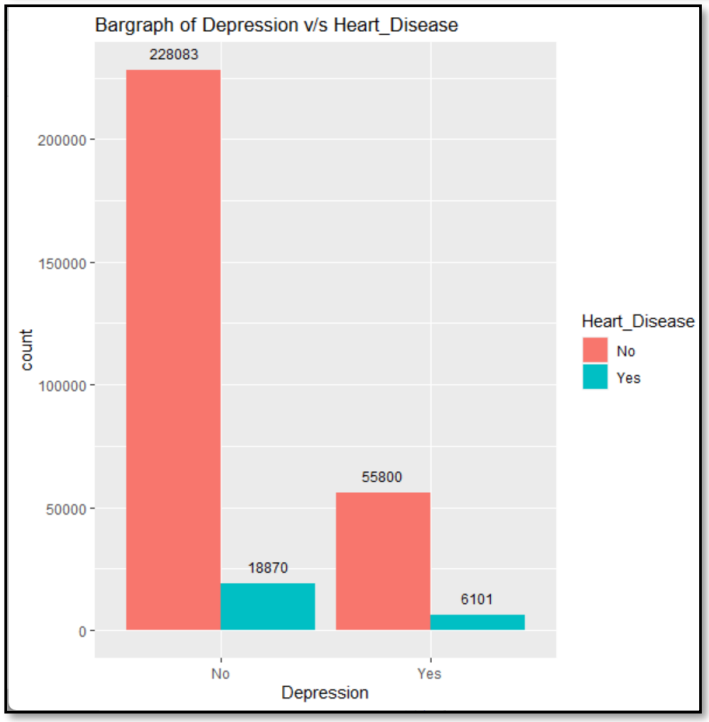
```
ggplot(cvd,aes(x=Skin_Cancer,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of Skin_Cancer v/s Heart_Disease')
```



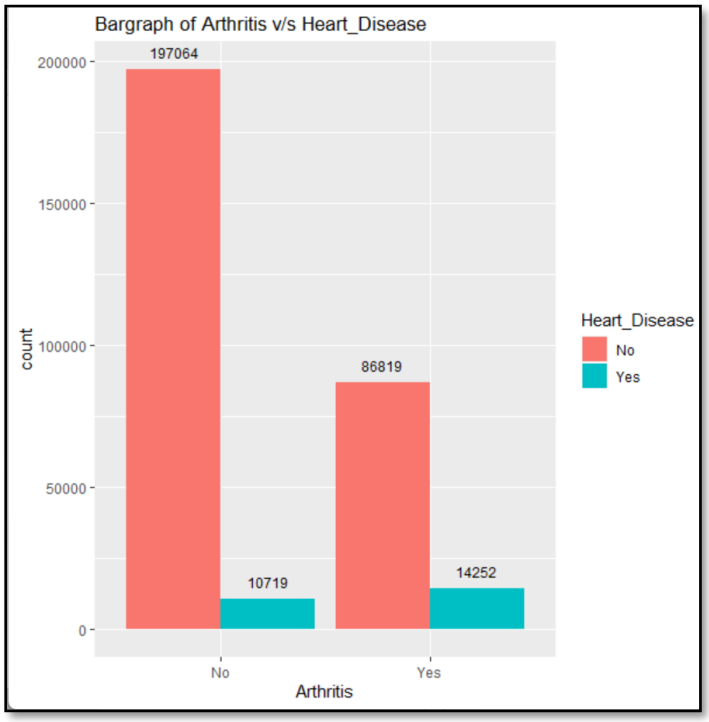

```
ggplot(cvd,aes(x=Other_Cancer,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of Other_Cancer v/s Heart_Disease')
```



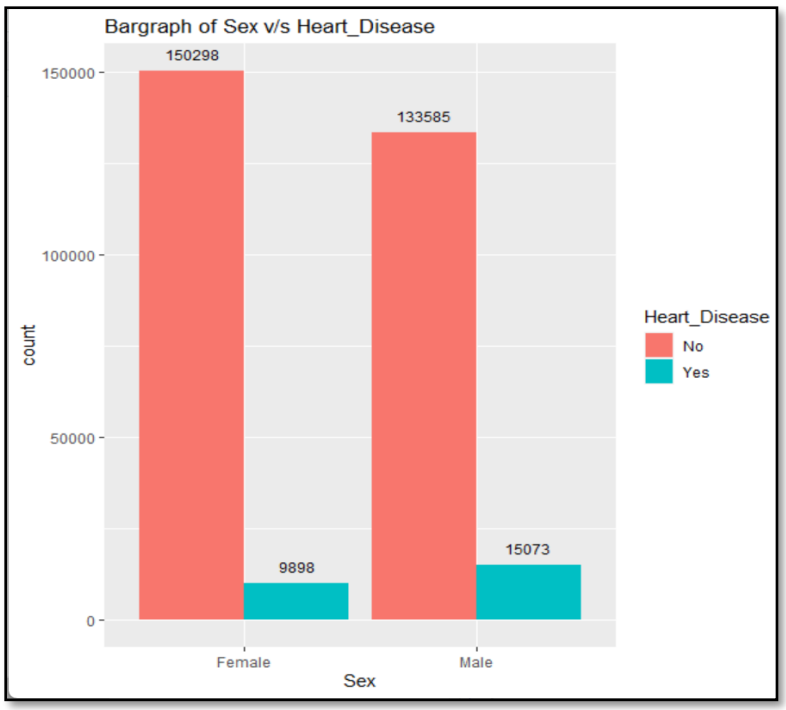
```
ggplot(cvd,aes(x=Depression,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of Depression v/s Heart_Disease')
```



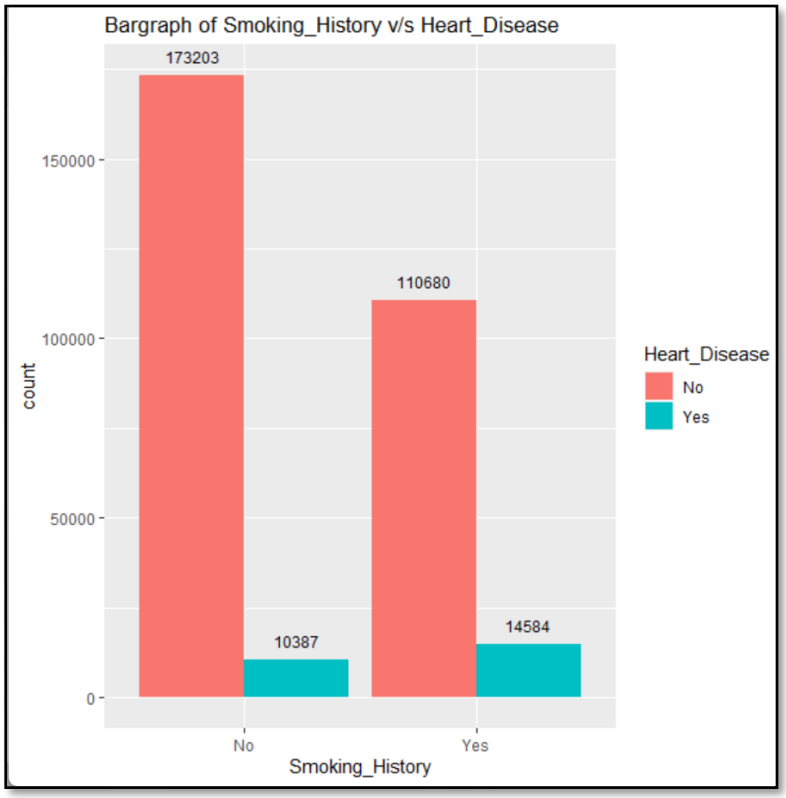

```
ggplot(cvd,aes(x=Arthritis,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of Arthritis v/s Heart_Disease')
```



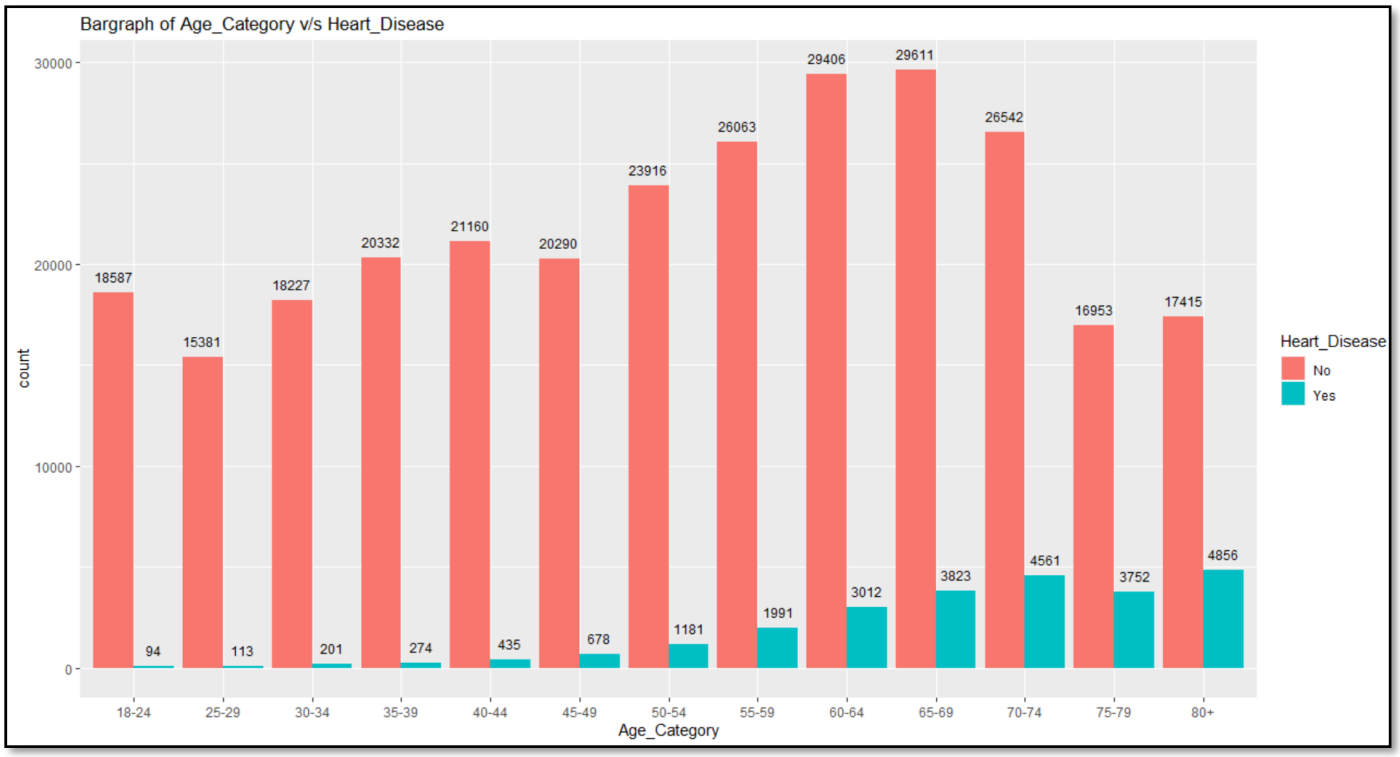
```
ggplot(cvd,aes(x=Sex,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of Sex v/s Heart_Disease')
```




```
ggplot(cvd,aes(x=Smoking_History,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of Smoking_History v/s Heart_Disease')
```



```
ggplot(cvd,aes(x=Age_Category,fill=Heart_Disease))+
  geom_bar(position='dodge')+
  geom_text(stat='count',aes(label=after_stat(count)),position=position_dodge(0.90),size=3,vjust=-1)+
  ggtitle('Bargraph of Age_Category v/s Heart_Disease')
```



#Model Building

The target variable is Heart_Disease .

> #Model

> a_cvd<-cvd #copied tha dataset into a_cvd

converted the Heart_Diseasee (target)variable in to 0 and 1

> a_cvd['Heart_Disease']<-as.numeric(as.factor(a_cvd$Heart_Disease))-1

> set.seed(123) #for generating same set of data in train & test data

> split<-sample.split(a_cvd$Heart_Disease,SplitRatio = 0.80) # spliting the data in 80:20 ratio in train and test data

> table(split)

split

FALSE TRUE

61771 247083

> train_d<-a_cvd[split,] #making train data for modeling containing 80% data

> test_d<-a_cvd[!split,] #making test data for model building containing 20% data

> ##Logistic Regression Model As the target variable is categorical (Yes or NO) format

>#glm(target_variable~independent_variables,data=datasetname,family=binomial(link="logit"))as target variable is Yes or No

> lrm_model<-glm(Heart_Disease~.,data=train_d,family = binomial(link = "logit"))

```
> #---Model building------------------------------------------------------------------------------------------------------------------------------------
> #Model
> a_cvd<-cvd #copied tha dataset into a_cvd
> dim(a_cvd)
[1] 308854     19
> # converted the Heart_Diseasee (target)variable in to 0 and 1
> a_cvd['Heart_Disease']<-as.numeric(as.factor(a_cvd$Heart_Disease))-1
> ## for logistic regression the target varible needs to in 0 or 1 format
> #table(a_cvd['Heart_Disease'])
> set.seed(123) #for generating same set of data in train & test data
> split<-sample.split(a_cvd$Heart_Disease,SplitRatio = 0.80) # spliting the data in 80:20 ratio in train and test data
> #why a_cvd$Heart_Disease ,as hear_disease is the target variable
> table(split)
split
 FALSE   TRUE
 61771 247083
> train_d<-a_cvd[split,] #making train data for modeling containing 80% data
> test_d<-a_cvd[!split,] #making test data for model building containing 20% data
> #dim(train1_d)
> #dim(test1_d)
> ##Logistic Regression Model As the target variable is categorical (Yes or NO) format
> #glm(target_variable~independent_variables,data=datasetname,family=binomial(link="logit"))as target variable is Yes or No
> lrm_model<-glm(Heart_Disease~.,data=train_d,family = binomial(link = "logit"))
```

> summary(lrm_model) #gives the summary of the model

```
Call:
glm(formula = Heart_Disease ~ ., family = binomial(link = "logit"),
    data = train_d)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8629  -0.4124  -0.2365  -0.1229   3.6468

Coefficients:
                                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                                   -6.319e+00  5.499e-01 -11.492  < 2e-16 ***
General_HealthFair                             1.722e+00  3.990e-02  43.160  < 2e-16 ***
General_HealthGood                             1.076e+00  3.781e-02  28.453  < 2e-16 ***
General_HealthPoor                             2.264e+00  4.485e-02  50.488  < 2e-16 ***
General_HealthVery Good                        5.325e-01  3.867e-02  13.769  < 2e-16 ***
CheckupNever                                   3.610e-01  1.683e-01   2.145 0.031939 *
CheckupWithin the past 2 years                 2.407e-01  7.127e-02   3.377 0.000733 ***
CheckupWithin the past 5 years                 1.442e-01  8.386e-02   1.719 0.085526 .
CheckupWithin the past year                    5.312e-01  6.500e-02   8.172 3.03e-16 ***
ExerciseYes                                   -1.865e-02  1.837e-02  -1.015 0.310030
Skin_CancerYes                                 1.400e-01  2.203e-02   6.354 2.09e-10 ***
Other_CancerYes                                4.155e-02  2.178e-02   1.908 0.056367 .
DepressionYes                                  2.395e-01  2.024e-02  11.834  < 2e-16 ***
DiabetesNo, pre-diabetes or borderline diabetes 1.374e-01  4.620e-02   2.974 0.002938 **
DiabetesYes                                    5.413e-01  1.894e-02  28.583  < 2e-16 ***
DiabetesYes, but female told only during pregnancy 1.528e-01  1.206e-01   1.267 0.205320
ArthritisYes                                   2.606e-01  1.716e-02  15.192  < 2e-16 ***
SexMale                                        8.312e-01  2.346e-02  35.439  < 2e-16 ***
Age_Category25-29                              3.296e-01  1.539e-01   2.141 0.032262 *
Age_Category30-34                              5.575e-01  1.408e-01   3.960 7.49e-05 ***
Age_Category35-39                              6.722e-01  1.346e-01   4.995 5.89e-07 ***
Age_Category40-44                              1.051e+00  1.273e-01   8.256  < 2e-16 ***
Age_Category45-49                              1.425e+00  1.233e-01  11.559  < 2e-16 ***
Age_Category50-54                              1.733e+00  1.201e-01  14.426  < 2e-16 ***
Age_Category55-59                              2.088e+00  1.183e-01  17.650  < 2e-16 ***
Age_Category60-64                              2.337e+00  1.174e-01  19.904  < 2e-16 ***
Age_Category65-69                              2.566e+00  1.171e-01  21.910  < 2e-16 ***
Age_Category70-74                              2.817e+00  1.171e-01  24.056  < 2e-16 ***
Age_Category75-79                              3.063e+00  1.177e-01  26.030  < 2e-16 ***
Age_Category80+                                3.340e+00  1.175e-01  28.434  < 2e-16 ***
Height_.cm.                                   -4.895e-03  3.149e-03  -1.555 0.120050
Weight_.kg.                                   -1.464e-04  2.853e-03  -0.051 0.959081
BMI                                            2.397e-03  8.232e-03   0.291 0.770921
Smoking_HistoryYes                             4.027e-01  1.657e-02  24.309  < 2e-16 ***
Alcohol_Consumption                           -9.745e-03  1.022e-03  -9.532  < 2e-16 ***
Fruit_Consumption                              2.008e-05  3.475e-04   0.058 0.953920
Green_Vegetables_Consumption                   7.571e-04  5.921e-04   1.279 0.201024
FriedPotato_Consumption                       -8.199e-04  9.760e-04  -0.840 0.400886
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 138783  on 247082  degrees of freedom
Residual deviance: 109421  on 247045  degrees of freedom
AIC: 109497

Number of Fisher Scoring iterations: 7
```

```
> #The model is trained using Training data
> #Procedure for testing Model
> ##IN testingdata all the independent variables required to predict the model are been taken from the testing dataset
> ##except target varible as needs to be predicted by the model
> testingdata<-test_d[,!names(test_d) %in% "Heart_Disease"]
```

```
> dim(testingdata)
[1] 61771    18
> pred<-predict(lrm_model,newdata =testingdata ,type = "response")#predicting the output of the testingdata using trained lrm_model
> #pred
> predicted_labels <- ifelse(pred >= 0.5, 1, 0)#taking a threshold of 0.5 if predicted value>= 0.5 asign value as 1 or else asing value as 0
> table(predicted_labels)
predicted_labels
    0     1
61137   634
```

#MODEL EVALUATION

```
> #---Model Evaluation--------------------------------------------
> #Model Evaluation
> accuracy <- sum(predicted_labels == test1_d) / length(test1_d)
> #Print accuracy
> cat("Accuracy:", accuracy, "\n")
Accuracy: 0.9192825
```

Accuracy of Model is 0.91928

```
> #confusion Matrix
> pred1 <- as.factor(predicted_labels)
> test1_d1 <- as.factor(test1_d)
> conf_matrix <- confusionMatrix(pred1,test1_d1)
> cat("Confusion Matrix:\n")
Confusion Matrix:
> print(conf_matrix)
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 56464  4673
         1   313   321

               Accuracy : 0.9193
                 95% CI : (0.9171, 0.9214)
    No Information Rate : 0.9192
    P-Value [Acc > NIR] : 0.4567

                  Kappa : 0.0976

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.99449
            Specificity : 0.06428
         Pos Pred Value : 0.92357
         Neg Pred Value : 0.50631
             Prevalence : 0.91915
         Detection Rate : 0.91409
   Detection Prevalence : 0.98974
      Balanced Accuracy : 0.52938

       'Positive' Class : 0
```

Confusion matrix

		Predicted Values	
		0	1
Actual values	0	56464	313
	1	4673	321

Logistic regression graph

```
> #Visualization of model
> plot(roc_obj,main="Roc Curve")
```