health-care.R

mfgdiags1

2024-09-29

# Load necessary libraries  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
library(reshape2) # For melting the correlation matrix  
  
# Load the dataset  
data <- read.csv('healthcare\_dataset.csv')  
  
# 1. Summary Statistics for numerical columns  
summary\_stats <- summary(select(data, Age, Billing.Amount, Room.Number))  
print(summary\_stats)

## Age Billing.Amount Room.Number   
## Min. :13.00 Min. :-2008 Min. :101.0   
## 1st Qu.:35.00 1st Qu.:13241 1st Qu.:202.0   
## Median :52.00 Median :25538 Median :302.0   
## Mean :51.54 Mean :25539 Mean :301.1   
## 3rd Qu.:68.00 3rd Qu.:37821 3rd Qu.:401.0   
## Max. :89.00 Max. :52764 Max. :500.0

# 2. Frequency counts for categorical columns  
medical\_condition\_freq <- table(data$Medical.Condition)  
gender\_freq <- table(data$Gender)  
  
print("Medical Condition Frequency:")

## [1] "Medical Condition Frequency:"

print(medical\_condition\_freq)

##   
## Arthritis Asthma Cancer Diabetes Hypertension Obesity   
## 9308 9185 9227 9304 9245 9231

print("Gender Frequency:")

## [1] "Gender Frequency:"

print(gender\_freq)

##   
## Female Male   
## 27726 27774

# 3. Correlation matrix for numerical columns  
correlation\_matrix <- cor(select(data, Age, Billing.Amount, Room.Number), use = "complete.obs")  
print("Correlation Matrix:")

## [1] "Correlation Matrix:"

print(correlation\_matrix)

## Age Billing.Amount Room.Number  
## Age 1.0000000000 -0.003831942 -0.0007202144  
## Billing.Amount -0.0038319421 1.000000000 -0.0029427395  
## Room.Number -0.0007202144 -0.002942740 1.0000000000

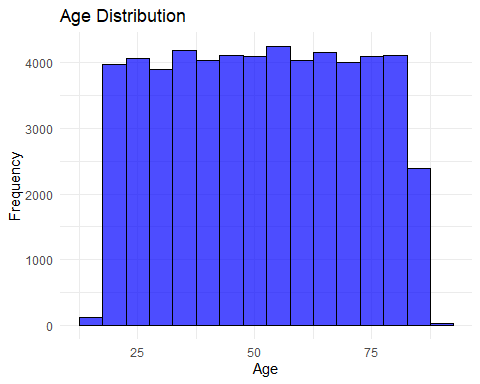
# 4. Checking for missing values  
missing\_values <- colSums(is.na(data))  
print("Missing Values:")

## [1] "Missing Values:"

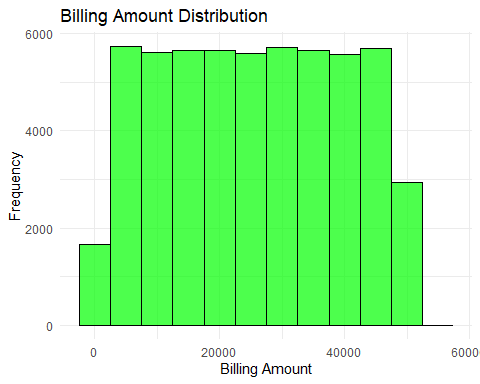
print(missing\_values)

## Name Age Gender Blood.Type   
## 0 0 0 0   
## Medical.Condition Date.of.Admission Doctor Hospital   
## 0 0 0 0   
## Insurance.Provider Billing.Amount Room.Number Admission.Type   
## 0 0 0 0   
## Discharge.Date Medication Test.Results   
## 0 0 0

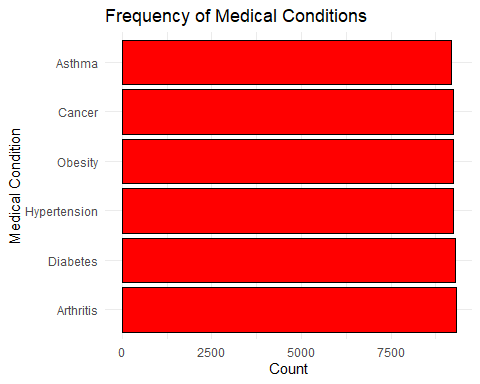
# 1. Age Distribution Histogram  
ggplot(data, aes(x = Age)) +  
 geom\_histogram(binwidth = 5, fill = "blue", color = "black", alpha = 0.7) +  
 labs(title = "Age Distribution", x = "Age", y = "Frequency") +  
 theme\_minimal()



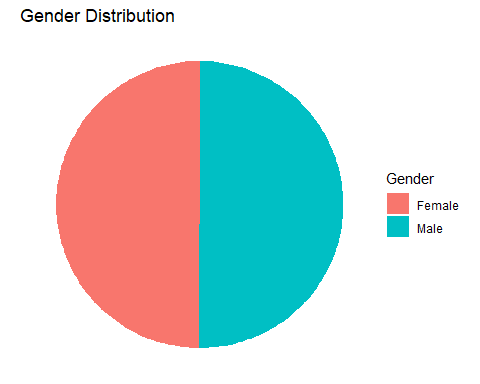
# 2. Billing Amount Distribution Histogram  
ggplot(data, aes(x = Billing.Amount)) +  
 geom\_histogram(binwidth = 5000, fill = "green", color = "black", alpha = 0.7) +  
 labs(title = "Billing Amount Distribution", x = "Billing Amount", y = "Frequency") +  
 theme\_minimal()



# 3. Medical Condition Frequency Bar Plot  
ggplot(data, aes(x = reorder(Medical.Condition, -table(Medical.Condition)[Medical.Condition]))) +  
 geom\_bar(fill = "red", color = "black") +  
 labs(title = "Frequency of Medical Conditions", x = "Medical Condition", y = "Count") +  
 theme\_minimal() +  
 coord\_flip()



# 4. Gender Distribution Pie Chart  
gender\_dist <- data %>%  
 group\_by(Gender) %>%  
 summarize(count = n()) %>%  
 mutate(percentage = count / sum(count) \* 100)  
  
ggplot(gender\_dist, aes(x = "", y = percentage, fill = Gender)) +  
 geom\_bar(width = 1, stat = "identity") +  
 coord\_polar("y") +  
 labs(title = "Gender Distribution") +  
 theme\_void() +  
 theme(legend.position = "right")



# 5. Correlation Heatmap  
# Melt the correlation matrix for ggplot  
melted\_corr <- melt(correlation\_matrix)  
  
ggplot(melted\_corr, aes(x = Var1, y = Var2, fill = value)) +  
 geom\_tile(color = "white") +  
 scale\_fill\_gradient2(low = "blue", high = "red", mid = "white",   
 midpoint = 0, limit = c(-1, 1), space = "Lab",   
 name = "Correlation") +  
 labs(title = "Correlation Matrix Heatmap") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, vjust = 1, size = 12, hjust = 1))

