D207 Code Utilized

library(dplyr)

library(dtplyr)

library(naniar)

library(tidyverse)

library(visdat)

library(stats)

med\_clean <- read\_csv(‘C://Users//lgben//OneDrive//Desktop//MSDA//D207 – Exploratory Data Analysis//med\_clean\_data.csv’)

*Exploring relationship between Complication Risk and Readmission*

comprisk\_readmis\_df <- med\_clean %>%

group\_by(Complication\_risk) %>%

summarize(ReAdmis)

comprisk\_readmis <- med\_clean %>%

group\_by(Complication\_risk) %>%

summarize(ReAdmis) %>%

table()

***Visualizing relationship between Complication Risk and Readmission***

ggplot(comprisk\_readmis\_df, aes(Complication\_risk, fill=ReAdmis)) +

geom\_bar()

-gives raw counts

ggplot(comprisk\_readmis\_df, aes(Complication\_risk, fill=ReAdmis)) +

geom\_bar(position=’fill’)

-gives proportions across all complication risks

***Performing Chi-Squared Test***

degreedom\_comprisk\_readmis <- (ncol(comprisk\_readmis) – 1) + (nrow(comprisk\_readmis) – 1)

comprisk\_readmis\_chisq <- chisq.test(comprisk\_readmis)

-output

Pearson’s Chi-squared test

data: comprisk\_readmis

X-squared = 0.15902, df=2, p-value = 0.9236

***Univariate EDA***

*Continuous Variables*

ggplot(med\_clean, aes(Initial\_days)) +

geom\_histogram(binwidth = 2)

ggplot(med\_clean, aes(VitD\_levels)) +

geom\_histogram(binwidth = 0.5)

*Categorical Variables*

ggplot(med\_clean, aes(Initial\_admin)) +

geom\_bar()

ggplot(med\_clean, aes(Marital)) +

geom\_bar()

***Bivariate EDA***

*Continuous Variables*

ggplot(med\_clean, aes(x=Income, y=Age)) +

geom\_point()

cor(x=med\_clean$Income, y=med\_clean$Income)

-output = 10.01222814

ggplot(med\_clean, aes(x=Initial\_days, y=TotalCharge)) +

geom\_point()

cor(x=med\_clean$Initial\_days, y=med\_clean$TotalCharge)

-output = 0.9876403

*Categorical Variables*

ggplot(med\_clean, aes(Complication\_risk, fill=HighBLood)) +

geom\_bar(position = ‘fill’)

ggplot(med\_clean, aes(Complication\_risk, fill=HighBlood)) +

geom\_bar()

chisq.test(x=med\_clean$Complication\_risk, y=med\_clean$HighBlood)

-output

Pearson’s Chi-squared test

data: med\_clean$Complication\_risk and med\_clean$HighBlood

X-squared = 7.8116, df = 2, p-value = 0.02013

ggplot(med\_clean, aes(BackPain, fill = Overweight)) +

geom\_bar()

ggplot(med\_clean, aes(BackPain, fill = Overweight)) +

geom\_bar(position = ‘fill’)

chisq.test(med\_clean$BackPain, y=med\_clean$Overweight)

-output

Pearson’s Chi-squared test

data: med\_clean$BackPain and med\_clean$Overweight

X-squared = 0.97206, df = 1, p-value = 0.3242

Sources Utilized:

Datacamp/R documentation for referencing chisq.test