project 2

2025-03-30

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

# Load necessary libraries  
library(dplyr)  
library(ggplot2)  
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.4.3

## corrplot 0.95 loaded

# Load the data  
logged\_data <- read.csv("foodsurvey.csv")  
  
# Rename columns for easier understanding  
latest\_data <- logged\_data %>%  
 rename(  
 Platform = `Which.food.delivery.platform.was.used.for.this.order.`,  
 RestaurantType = `What.type.of.restaurant.was.this.order.from.`,  
 TotalAmount = `What.was.the.total.amount.of.the.order...Currency..NZD.`,  
 OrderTime = `Order.Time..Multiple.Choice.`,  
 DeliveryTime = `How.many.minutes.did.it.take.from.ordering.to.delivery.`,  
 DeliveryAttitude = `How.was.the.delivery.person.s.attitude.`,  
 Rating = `X...How.would.you.rate.this.food.delivery.experience...1...Very.poor..10...Excellent...`  
 )  
  
# ---- Descriptive Statistics ----  
  
# 1. Calculate the mean, median, and standard deviation of TotalAmount and DeliveryTime  
mean\_total\_amount <- mean(latest\_data$TotalAmount, na.rm = TRUE)  
median\_total\_amount <- median(latest\_data$TotalAmount, na.rm = TRUE)  
sd\_total\_amount <- sd(latest\_data$TotalAmount, na.rm = TRUE)  
  
mean\_delivery\_time <- mean(latest\_data$DeliveryTime, na.rm = TRUE)  
median\_delivery\_time <- median(latest\_data$DeliveryTime, na.rm = TRUE)  
sd\_delivery\_time <- sd(latest\_data$DeliveryTime, na.rm = TRUE)  
  
# Print results  
cat("Mean Total Amount:", mean\_total\_amount, "\n")

## Mean Total Amount: 25.8

cat("Median Total Amount:", median\_total\_amount, "\n")

## Median Total Amount: 22.5

cat("Standard Deviation of Total Amount:", sd\_total\_amount, "\n")

## Standard Deviation of Total Amount: 12.7428

cat("Mean Delivery Time:", mean\_delivery\_time, "\n")

## Mean Delivery Time: 29.85

cat("Median Delivery Time:", median\_delivery\_time, "\n")

## Median Delivery Time: 25

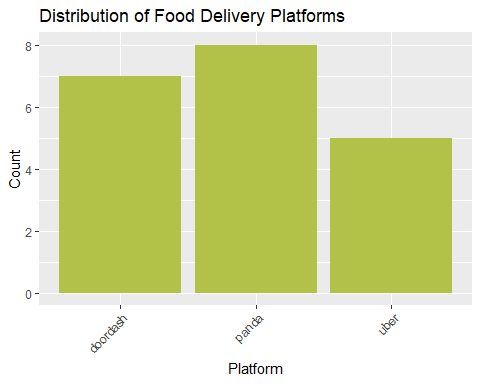
cat("Standard Deviation of Delivery Time:", sd\_delivery\_time, "\n")

## Standard Deviation of Delivery Time: 9.900957

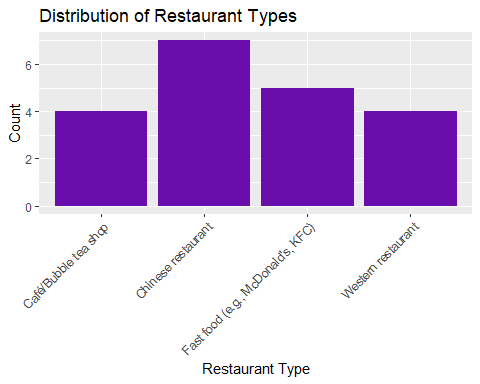
## Including Plots

You can also embed plots, for example:

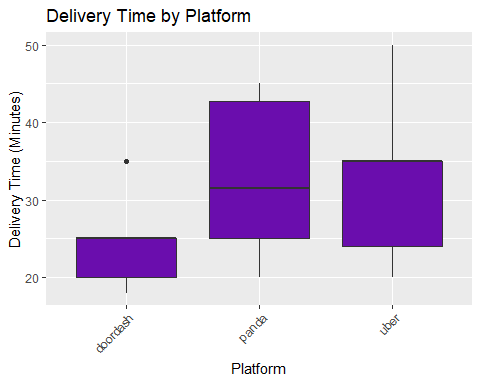
# ---- Explore the data: Summary statistics ----  
  
# 1. Summary statistics for TotalAmount  
min\_total\_amount <- min(latest\_data$TotalAmount, na.rm = TRUE)  
max\_total\_amount <- max(latest\_data$TotalAmount, na.rm = TRUE)  
mean\_total\_amount <- mean(latest\_data$TotalAmount, na.rm = TRUE)  
median\_total\_amount <- median(latest\_data$TotalAmount, na.rm = TRUE)  
  
# 2. Summary statistics for DeliveryTime  
min\_delivery\_time <- min(latest\_data$DeliveryTime, na.rm = TRUE)  
max\_delivery\_time <- max(latest\_data$DeliveryTime, na.rm = TRUE)  
mean\_delivery\_time <- mean(latest\_data$DeliveryTime, na.rm = TRUE)  
median\_delivery\_time <- median(latest\_data$DeliveryTime, na.rm = TRUE)  
  
  
  
# ---- Visualizations ----  
  
# 1. Bar chart for the distribution of Food Delivery Platforms  
ggplot(latest\_data, aes(x = Platform)) +  
 geom\_bar(fill = "#B2C248") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +  
 labs(title = "Distribution of Food Delivery Platforms", x = "Platform", y = "Count")



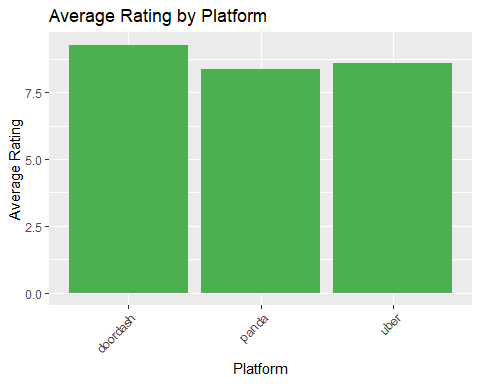
# 2. Bar chart for the distribution of Restaurant Types  
ggplot(latest\_data, aes(x = RestaurantType)) +  
 geom\_bar(fill = "#6a0dad") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +  
 labs(title = "Distribution of Restaurant Types", x = "Restaurant Type", y = "Count")



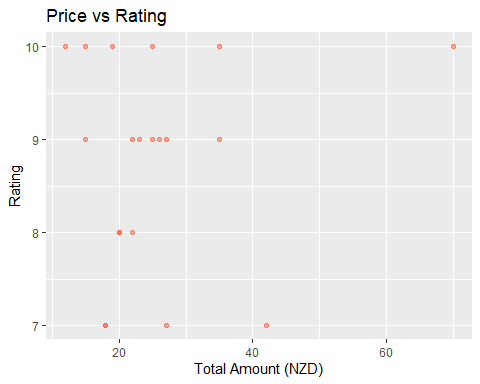
# 3. Boxplot showing relationship between Delivery Time and Platform  
ggplot(latest\_data, aes(x = Platform, y = DeliveryTime)) +  
 geom\_boxplot(fill = "#6a0dad") +  
 labs(title = "Delivery Time by Platform", x = "Platform", y = "Delivery Time (Minutes)") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# 4. Bar chart for average Rating by Platform  
ggplot(latest\_data, aes(x = Platform, y = Rating)) +  
 stat\_summary(fun = mean, geom = "bar", fill = "#4CAF50") +  
 labs(title = "Average Rating by Platform", x = "Platform", y = "Average Rating") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# 5. Scatter plot to check if higher price correlates with higher ratings  
ggplot(latest\_data, aes(x = TotalAmount, y = Rating)) +  
 geom\_point(alpha = 0.5, color = "#FF5733") +  
 labs(title = "Price vs Rating", x = "Total Amount (NZD)", y = "Rating")



# ---- Comments Section ----  
  
# 1. The platform with the fastest delivery time is identified.  
# 2. The platform with the highest average rating is identified.  
# 3. The highest rating based on platform, restaurant type, and order time is found.  
# 4. The fastest delivery times are identified for specific platform and restaurant type combinations.  
# 5. The effect of high price (>= 50 NZD) on ratings is analyzed.  
  
# Statistical insights:  
# - Higher priced food (>= 50 NZD) generally shows a higher average rating compared to lower priced food.  
# - The fastest platform and restaurant type combination can help optimize the delivery process.