

Predicting a waiter's tip amount in Dollars based on the customers total bill, gender(sex), time of the day, if the customer is a smoker or not and size of the group.

QIYU HUANG

Dec 21st 2020

## **Github**

Code and data supporting this analysis is available at: <https://github.com/cybernaily/Final.git>

## **Keywords**

Linear Regression, Robust Algorithms, Data Analysis, Descriptive Analytics, p-value, t-statistic, R2 value, significance level, supervised learning.

## **Abstract.**

An Analysis of Waiters tip amounts given the customers total bill, gender(sex), time of the day, smoker and size of the group. This study seeks to analyse the factors that influence the tipping behavior of customers: The likelihood of a customer giving a tip and the size of the tip. 244 eating events were recorded in a restaurant used for this analysis. The data was reported in a collection of case studies for business statistics. The report indicates that the average tipping rate is 7.048% of the total bill. The overall analysis demonstrates that customers tipping decisions and tip sizes are functions of their social interest. It is safe to conclude that customers consider tipping more of a social norm rather than self-interested rational behavior.

## **Specifying the Research Question**

Is it possible to predict a waiters amount of tip in dollars based on the customers total bill, gender(sex), time of the day, smoker and size of the group

## **Introduction**

Tipping is the norm in restaurants all over the world. According to Lynn et al. (1993) The amount of tip varies depending on the nature of the customer and quality of service received. According to Azar (2007), the tip amount ranges from 15% to 20% of the total bill for excellent service. The aim of this project is to use machine learning, collaborative data filtering and Supervised learning to solve tabular data problems. We will analyse a dataset with 244 records with one waiters tips over a few months working in a restaurant.

Studying customers tipping behavior can be helpful in measuring customer satisfaction (Rathore, 2015) and in turn restaurant managers can leverage such information to improve the quality of their services. Based on previous studies, tip size is reportedly increased by friendly service, good suggestions, excellent food, prompt delivery of the main course and check, a self-introduction by the waiter, and receiving separate

checks(“Introduction: The promise of collaborative public service delivery,” 2019). The tip is decreased by waiting a long time for a beverage and being seated in a bad location.

Servers also show that there are no differences in tipping behavior. However, they have expressed that the effort required for some social groups in order to receive the same amount of tip is higher(Toporek, 2015). If this is the case then they would be more motivated to engage in cost-based statistical discrimination against groups that require more effort to serve.

The independent variable is the amount of tip in Dollars. Dependent variables are:

The customers total bill. Customers gender(sex). Time of day. Is the customer a smoker or not? The size of the group

## Methodology

**Thorough Data Cleaning** In this step, we clean the dataset by checking for missing values. None was found. we also check for duplicates. Only one record was found but we shall use it since it is very likely to have duplicate records based on the variables in the data. We then ensure that each column has the correct datatype in preparation for further analysis and modelling.

**Univariate analysis** Here we check the distribution of each and every column in a bid to understand how each varies. We also analyse the measures of central tendency for numerical columns and vlaue counts for categorical columns.

**Bivariate Analysis** Analysing the relationships between variables in pairs. Some of the pairs include: Total\_bill vs the tip amount, Gender distribution, gender vs smoker

**Regression Modelling:** modelling the data using a multiple Linear Regression model.

**Conclusion** Documenting the findings, reports and references. Was the analysis successful? Was the data sufficient? Was the statistical approach appropriate? How relevant is the analysis and who are the beneficiaries?

**Data Source and relevance** Data: <https://www.kaggle.com/jsphyg/tipping>

One data set will be used to investigate the factors that influence how customers tip in restaurants. The dataset is from kaggle which is a credible data source for data science projects. The data has 244 rows and 7 columns which is enough to train and test the results of our model. One waiter recorded information about each tip he received over a period of a few months working in one restaurant.

Can you predict the tip amount?

## Model

A Multiple Linear Regression was used for this analysis since we have more than one predictor variable. The coefficients then indicate which variables affect the dependent variable the most. Postive values depict a positive correlation while negative coefficients indicate low or no correlation. The coefficients are then used to predict the response variable. The general equation for a multiple linear regression is

$$y = a + b_1x_1 + b_2x_2 + \dots b_nx_n.$$

y is the response variable.

a, b<sub>1</sub>, b<sub>2</sub>...b<sub>n</sub> are the coefficients.

x<sub>1</sub>, x<sub>2</sub>, ...x<sub>n</sub> are the predictor variables.

## Data Cleaning

```
library("readr")
df <- read_csv("tips.csv")
head(df)
```

### Loading the dataset

```
##   total_bill  tip    sex smoker day   time size
## 1    16.99  1.01 Female   No  Sun Dinner    2
## 2    10.34  1.66   Male   No  Sun Dinner    3
## 3    21.01  3.50   Male   No  Sun Dinner    3
## 4    23.68  3.31   Male   No  Sun Dinner    2
## 5    24.59  3.61 Female   No  Sun Dinner    4
## 6    25.29  4.71   Male   No  Sun Dinner    4
```

```
tips_df <- data.frame(df)
head(tips_df)
```

### Previewing the top of the dataset

```
##   total_bill  tip    sex smoker day   time size
## 1    16.99  1.01 Female   No  Sun Dinner    2
## 2    10.34  1.66   Male   No  Sun Dinner    3
## 3    21.01  3.50   Male   No  Sun Dinner    3
## 4    23.68  3.31   Male   No  Sun Dinner    2
## 5    24.59  3.61 Female   No  Sun Dinner    4
## 6    25.29  4.71   Male   No  Sun Dinner    4
```

```
summary(tips_df)
```

### Previewing the summary of the dataset

```
##   total_bill      tip      sex      smoker
## Min.   : 3.07  Min.   : 1.000  Length:244  Length:244
## 1st Qu.:13.35  1st Qu.: 2.000  Class :character  Class :character
## Median :17.80  Median : 2.900  Mode  :character  Mode  :character
## Mean   :19.79  Mean   : 2.998
## 3rd Qu.:24.13  3rd Qu.: 3.562
## Max.   :50.81  Max.   :10.000
##   day      time      size
## Length:244  Length:244  Min.   :1.00
## Class :character  Class :character  1st Qu.:2.00
## Mode  :character  Mode  :character  Median :2.00
##                                     Mean  :2.57
##                                     3rd Qu.:3.00
##                                     Max.   :6.00
```

### Properties of the dataset

```
dim(tips_df)
```

### Dimensions

```
## [1] 244 7
```

```
#The dataframe has 244 row entries and 7 columns
```

```
colnames(tips_df)
```

### Column Names

```
## [1] "total_bill" "tip" "sex" "smoker" "day"  
## [6] "time" "size"
```

```
#The seven column names are:
```

```
sapply(tips_df, class)
```

### Column data types

```
## total_bill tip sex smoker day time  
## "numeric" "numeric" "character" "character" "character" "character"  
## size  
## "integer"
```

## Data Cleaning

### Missing values

```
#Checking the sum of missing values per column
```

```
colSums(is.na(tips_df))
```

```
## total_bill tip sex smoker day time size  
## 0 0 0 0 0 0 0
```

```
#there are no missing values in the data
```

### Duplicates

```
duplicated_rows <- tips_df[duplicated(tips_df),]  
duplicated_rows
```

```
## total_bill tip sex smoker day time size  
## 203 13 2 Female Yes Thur Lunch 2
```

```
#there is one duplicate entry in the data
```

```
#we shall retain the duplicate record because based on the coulmm values it is possible to have similar
```

## Checking the appropriate datatypes for each column

```
sapply(tips_df, class)

## total_bill      tip      sex      smoker      day      time
## "numeric"    "numeric" "character" "character" "character" "character"
##      size
## "integer"
```

## Univariate analysis

```
#install.packages("pacman")
figure_nums <- captioner::captioner(prefix = "Fig")
```

```
#Checking customer spending in the restaurant (Bill)
```

```
# mean
mean(tips_df$total_bill)
```

### Total Bill

```
## [1] 19.78594
```

```
# median
median(tips_df$total_bill)
```

```
## [1] 17.795
```

```
# mode
x <- tips_df$total_bill
#sort(x)
names(table(x))[table(x)==max(table(x))]
```

```
## [1] "13.42"
```

```
#each of the values printed below appear thrice in the dataset
```

```
#distribution
hist(x, col=c("darkorange"))
```

Most of the customers spend between 10 and 25 dollars in the restaurant

The users spend an average 19.78594 dollars for their meals.

The modal amount spent on the site “13.42”

The median amount spent is 17.795.

The distribution above is right-skewed.

```
# mean
mean(tips_df$tip)
```

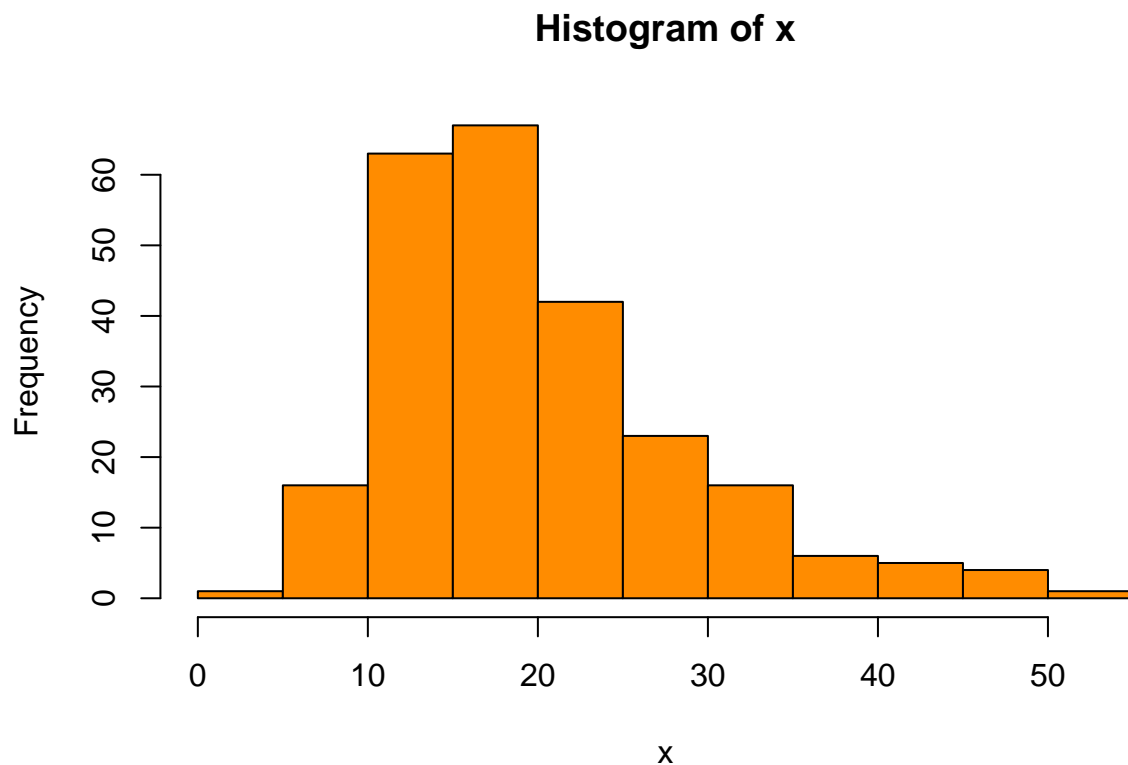


Figure 1: Plot1: Histogram of Total bill

#### Tip

```
## [1] 2.998279
```

```
# median
```

```
median(tips_df$tip)
```

```
## [1] 2.9
```

```
# mode
```

```
a <- tips_df$tip
```

```
#sort(x)
```

```
names(table(a))[table(a)==max(table(a))]
```

```
## [1] "2"
```

```
#distribution
```

```
hist(a, col=c("pink"))
```

The average amount of tips is 2.998279 dollars

The modal tip is 2 dollars

The median tip is 2.9 dollars.

The distribution above is right-skewed.

The highest frequency is 0-5 dollars.

The highest tip given is 7 dollars.

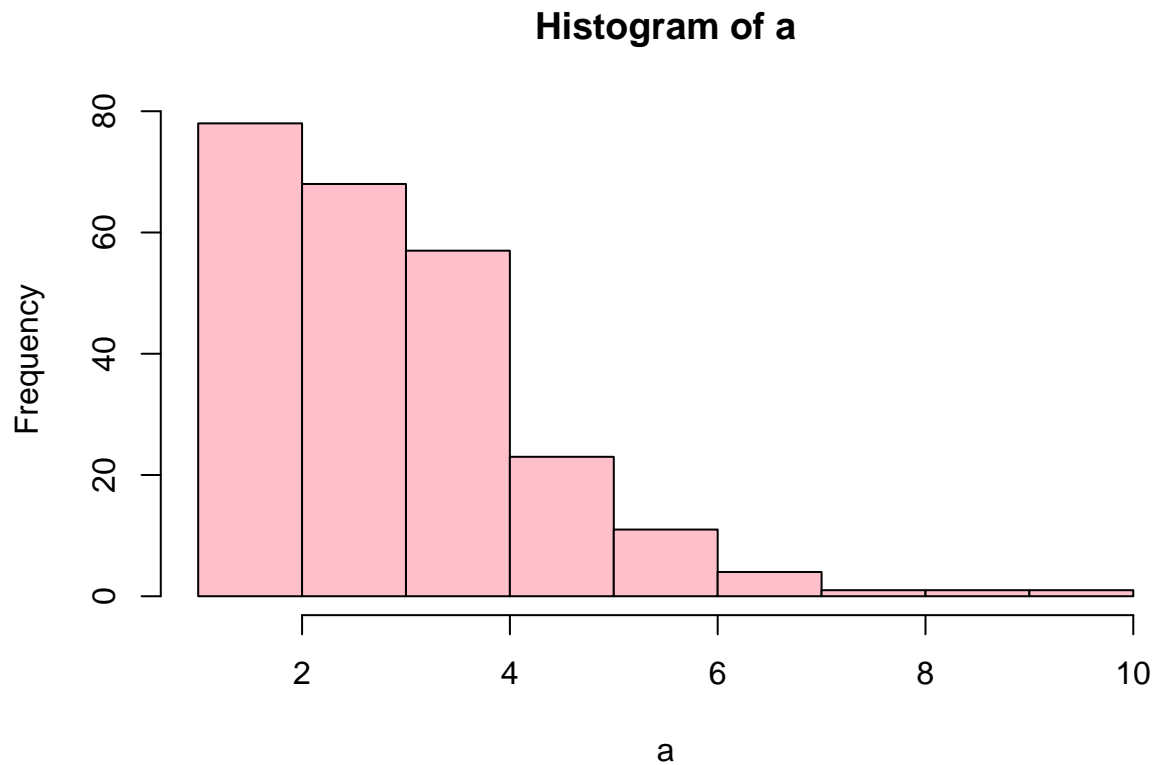


Figure 2: Plot2: Histogram of Tips

#### Sex

```
#gender of the user
# measures of central tendency
```

```
unique(factor(tips_df$sex))
```

#### Male

```
## [1] Female Male
```

```
## Levels: Female Male
```

```
gender_df <- table(tips_df$sex)
```

```
#distribution
```

```
barplot(gender_df, main="Gender Distribution",col=c("darkgreen"),xlab="Gender")
```

The gender distribution is not equal. The males are twice the females.

This also means that most people who eat in restaurants are male.

```
#This column indicates whether a customer is a smoker or not.
```

```
unique(factor(tips_df$smoker))
```

#### Smoker

```
## [1] No Yes
```

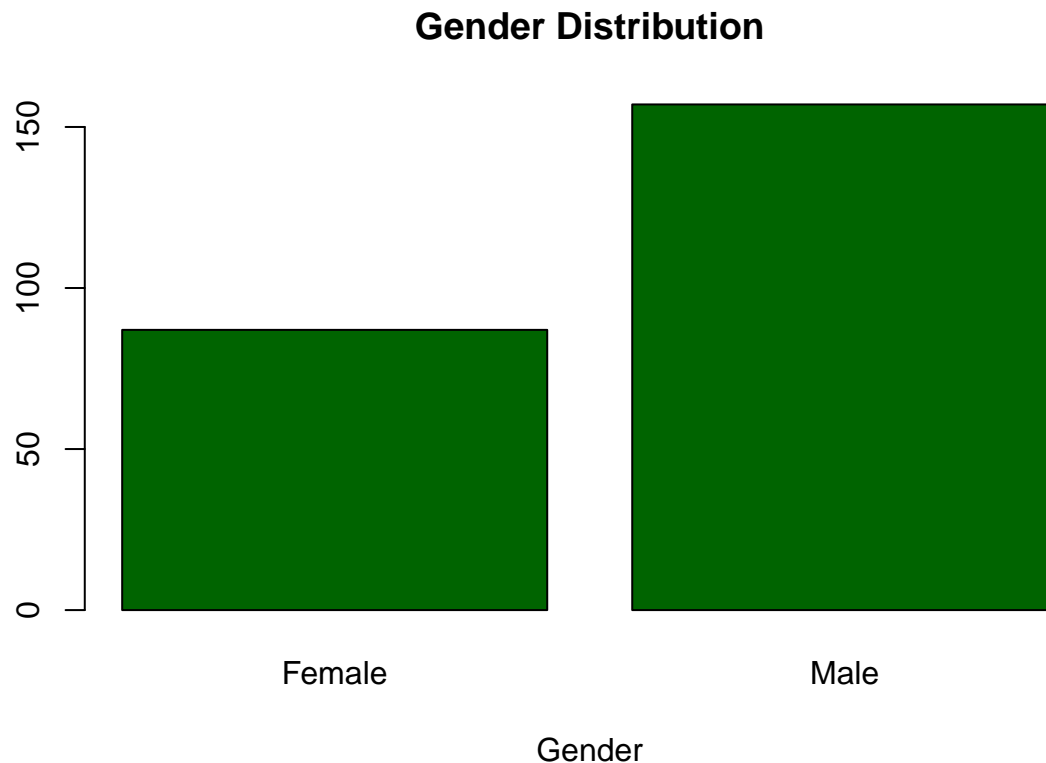


Figure 3: Plot3: Barplot of Gender

## Levels: No Yes

```
smoker_df <- table(tips_df$smoker)
#distribution
barplot(smoker_df, main="Smoker Distribution",col=c("brown"),xlab="Smoker")
```

Most of the customers who visit the restaurant are not smokers and the number of non-smokers are twice the number of smokers.

*#Day of the week*

```
unique(factor(tips_df$day))
```

Day

## [1] Sun Sat Thur Fri

## Levels: Fri Sat Sun Thur

```
day_df <- table(tips_df$day)
#distribution
barplot(day_df, main="Day of the Week",col=c("yellow"),xlab="Day")
```

The data was obtained from Thursday through to Sunday

Friday has the least number of customers.

Saturday has the most number of customers.

Saturday and Sunday are the busiest days.



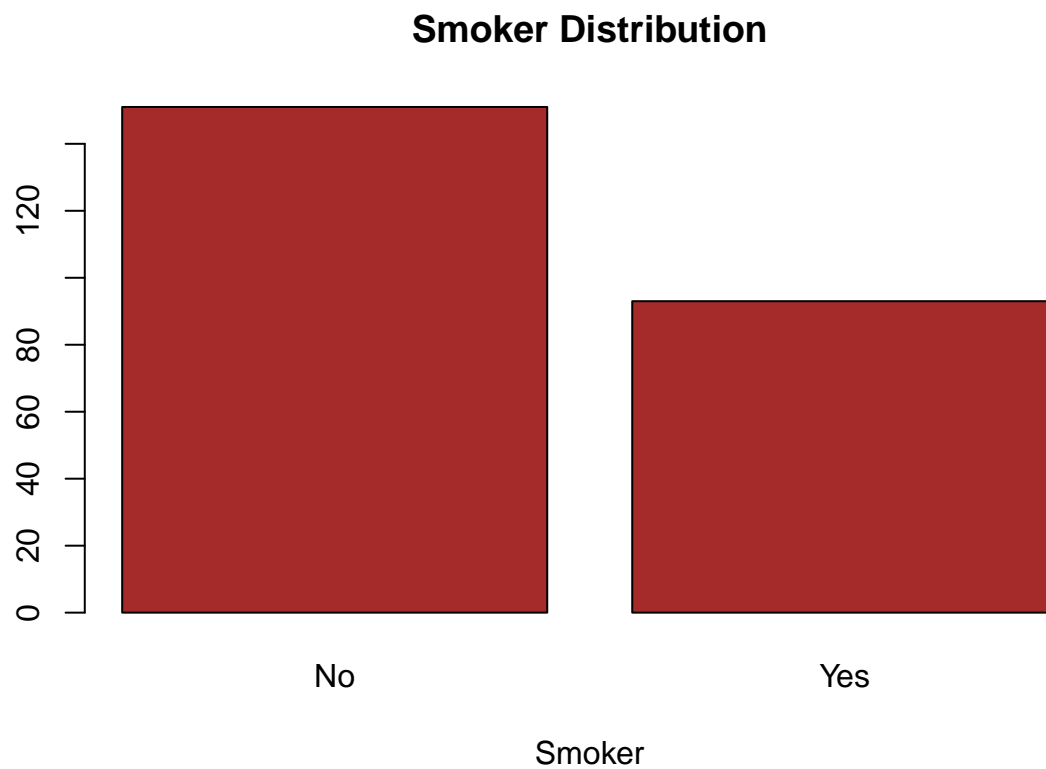


Figure 4: Plot4: Barplot of Smokers

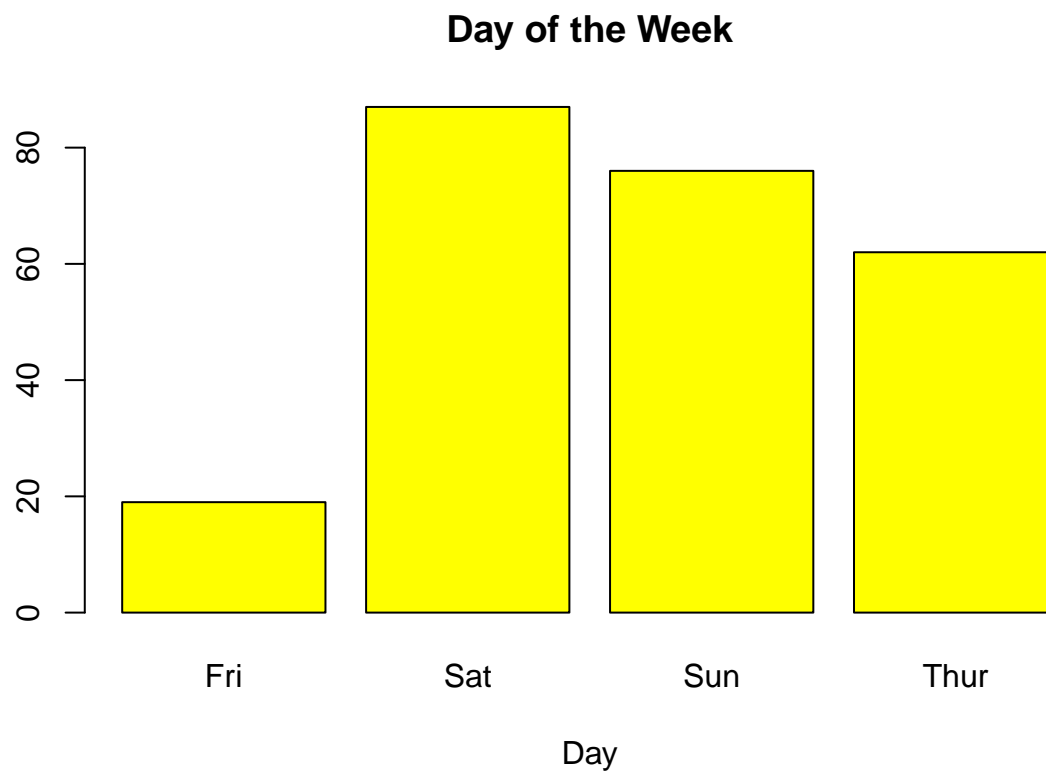


Figure 5: Plot5: Barplot of Days

```
#Time of Day
```

```
unique(factor(tips_df$time))
```

Time

```
## [1] Dinner Lunch
```

```
## Levels: Dinner Lunch
```

```
time_df <- table(tips_df$time)
```

```
#distribution
```

```
barplot(time_df, main="Time Distribution",col=c("darkblue"),xlab="Time")
```

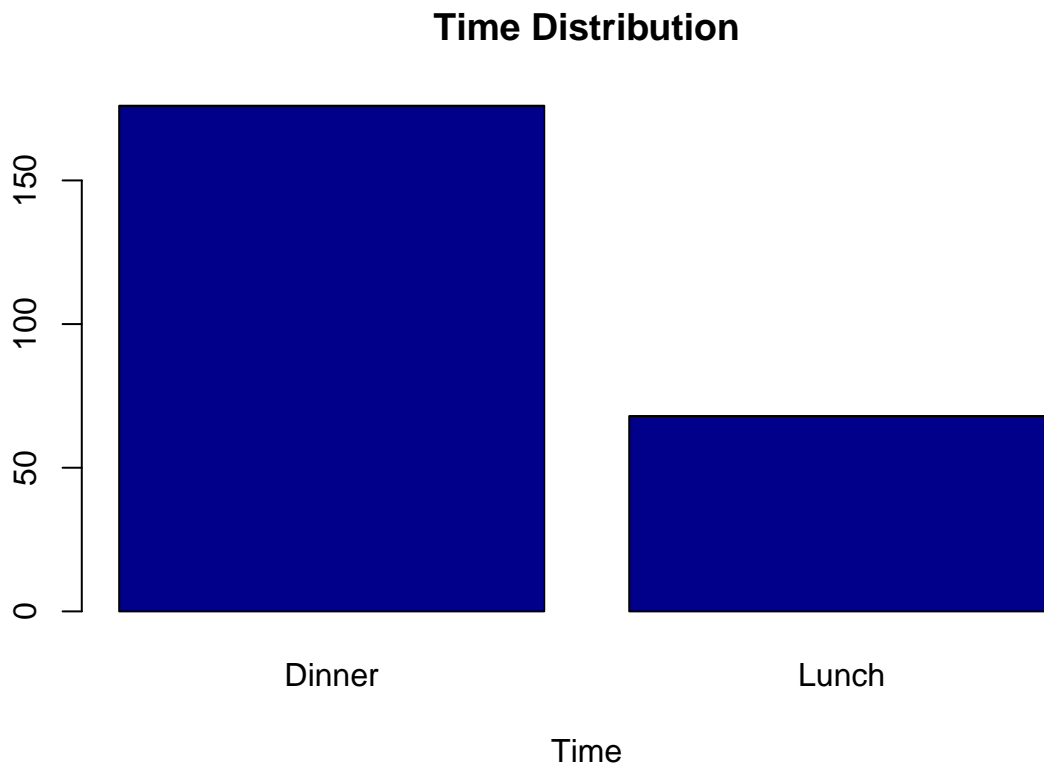


Figure 6: Plot6: Barplot of Time

Most of the clients have dinner at the restaurant as opposed to having lunch there.

```
unique(factor(tips_df$size))
```

Size

```
## [1] 2 3 4 1 6 5
```

```
## Levels: 1 2 3 4 5 6
```

```
size_df <- table(tips_df$size)
```

```
#distribution
```

```
barplot(size_df, main="Size Distribution",col=c("purple"),xlab="size")
```

The most popular number for the amount of diners is 2.

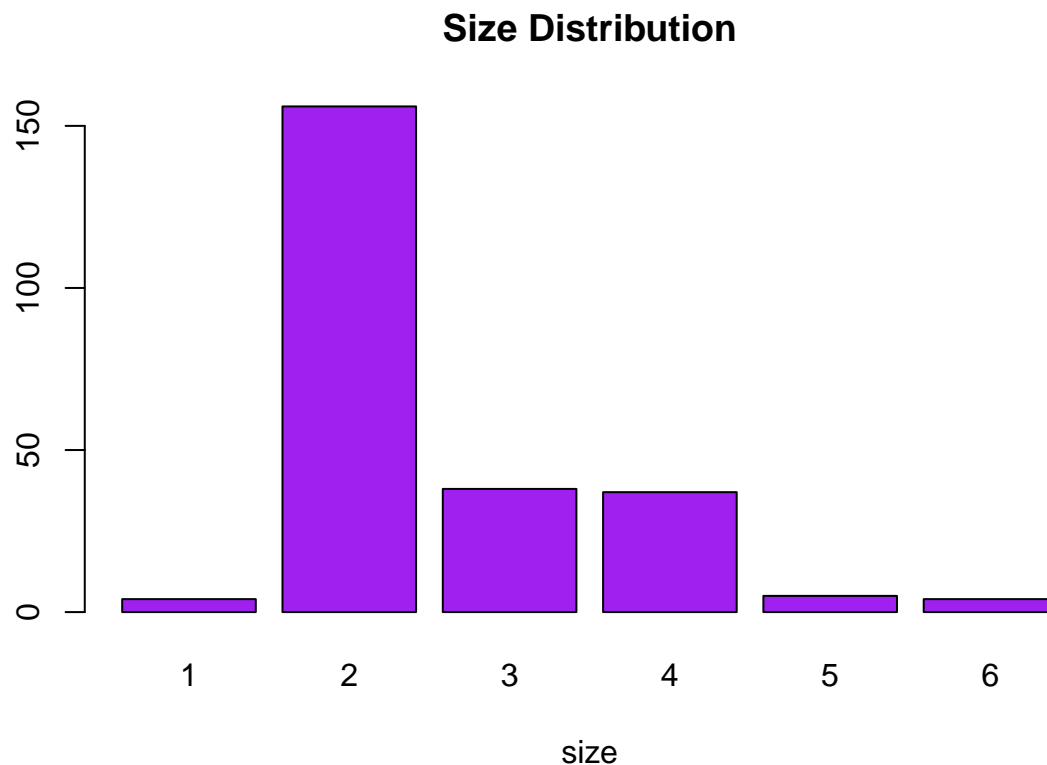


Figure 7: Plot7: Barplot of Size

very few customers visit the restaurant by themselves.

Groups of 3 and 4 are abit common.

there are no groups for more than six customers in one sitting.

## Bivariate Analysis and Multivariate Graphical Data Analysis

```
#substettnng the data for further numerical analysis
tips_df2 <- subset(tips_df, select = c(total_bill,tip,size ))
```

```
head(tips_df2)
```

```
##  total_bill  tip size
## 1    16.99 1.01   2
## 2    10.34 1.66   3
## 3    21.01 3.50   3
## 4    23.68 3.31   2
## 5    24.59 3.61   4
## 6    25.29 4.71   4
```

## Correlation

```
#The default method is Pearson, but we can also compute Spearman or Kendall coefficients.
mydata = cor(tips_df2, method = c("spearman"))
```

```
mydata1= cor(tips_df2, method = c("kendall"))
mydata2= cor(tips_df2, method = c("pearson"))
```

```
mydata #spearman
```

```
##           total_bill      tip      size
## total_bill  1.0000000 0.6789681 0.6047911
## tip         0.6789681 1.0000000 0.4682679
## size        0.6047911 0.4682679 1.0000000
```

```
mydata1 #kendall
```

```
##           total_bill      tip      size
## total_bill  1.0000000 0.5171810 0.4843421
## tip         0.5171810 1.0000000 0.3781847
## size        0.4843421 0.3781847 1.0000000
```

```
mydata2 #pearson
```

```
##           total_bill      tip      size
## total_bill  1.0000000 0.6757341 0.5983151
## tip         0.6757341 1.0000000 0.4892988
## size        0.5983151 0.4892988 1.0000000
```

Using the 3 correlation coefficients to get the correlation between the features, we can see that the correlation is average in most cases.

This means that most of the variables are somewhat dependent of each other

Significance levels (p-values) can also be generated using the rcorr function which is found in the Hmisc package.

```
#mydata.coeff = mydata.rcorr$r
#mydata.p = mydata.rcorr$P
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
corrplot(mydata)
```

A default correlation matrix plot (called a Correlogram) is generated. Positive correlations are displayed in a blue scale while negative correlations are displayed in a red scale

There is average positive correlation between the variables in the data.

## The Plots below are scatterplots of a few pairs of variables

```
# Libraries
library(ggplot2)

# create data
amount_spent <- tips_df$total_bill
Tip <- tips_df$tip
data <- data.frame(amount_spent,Tip)
```

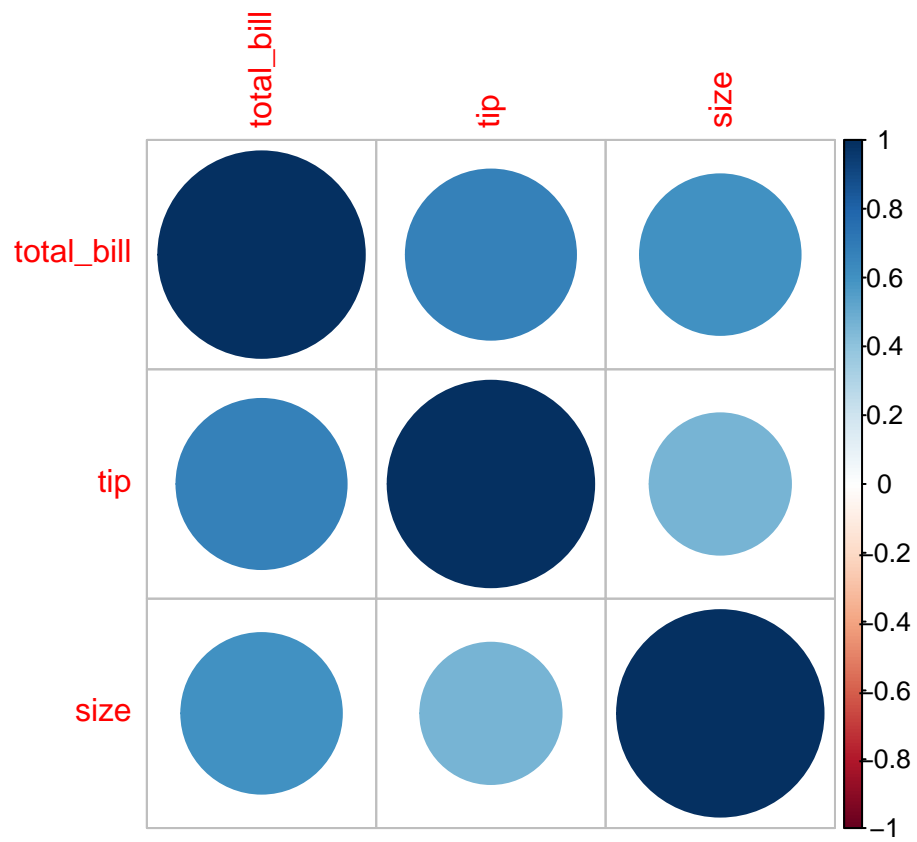


Figure 8: Plot8: Corrplot of Total\_bill, Tips and Size

```
# Plot
ggplot(data, aes(x=amount_spent, y=Tip)) + geom_point()
```

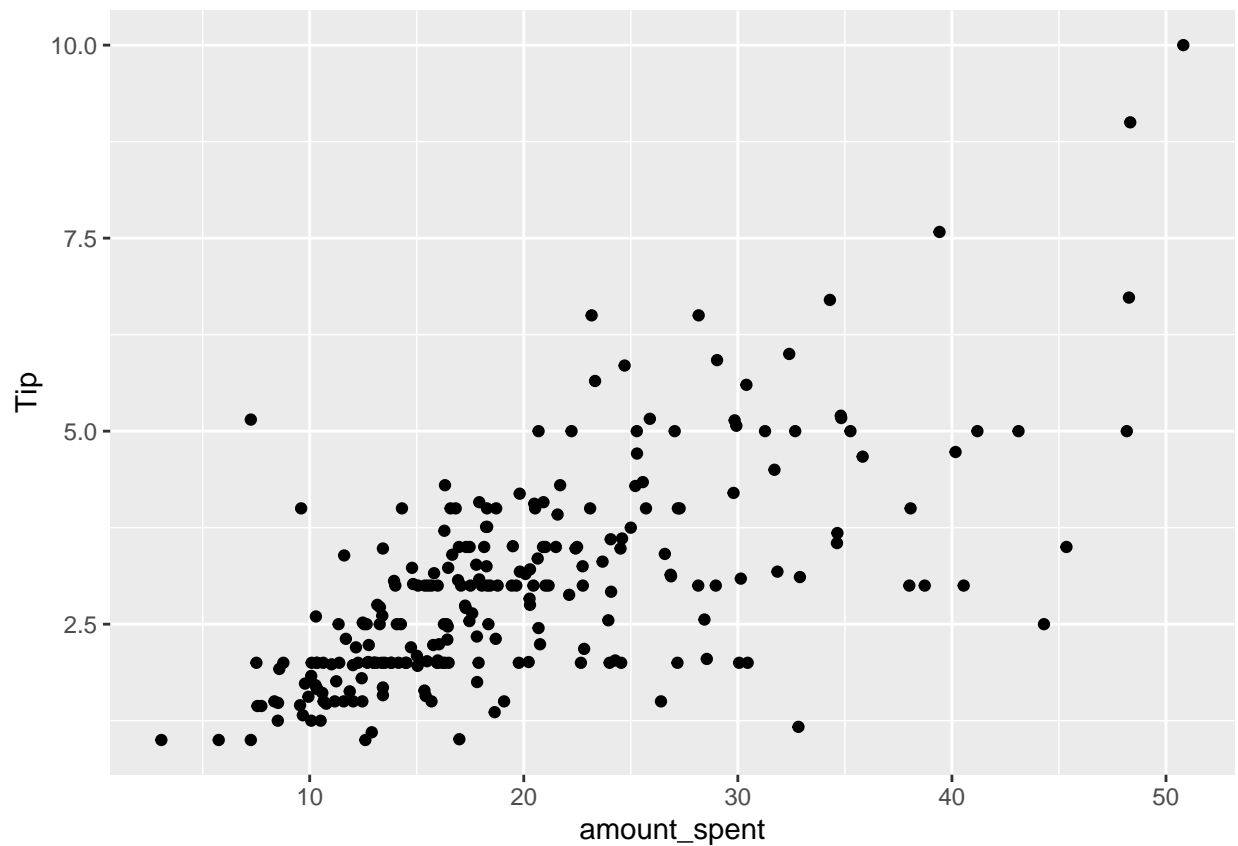


Figure 9: Plot9: Plot of Amount spent vs Tips

```
#positive non-linear correlation

avg_tip_percentage <- transform(tips_df, new = amount_spent / Tip)
mean(avg_tip_percentage[["new"]])
```

Amount spent in the restaurant vs amount of tip

```
## [1] 7.048932
```

```
#The average tipping rate is 7.048%
```

```
library(tidyverse)
```

sex VS Smoker

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v tibble  3.0.4    v dplyr   1.0.2
## v tidyr   1.1.2    v stringr 1.4.0
## v purrr   0.3.4    v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
#Male respondents who clicked on an add
dim(tips_df %>% filter(sex == 'Male', smoker == 'No'))

## [1] 97 7
#97

#Male respondents did not click on an add
dim(tips_df %>% filter(sex == 'Female', smoker == 'No'))

## [1] 54 7
#54

#Female respondents who clicked on an add
dim(tips_df %>% filter(sex == 'Male', smoker == 'Yes'))

## [1] 60 7
# 60

#Female respondents who clicked did not on an add
dim(tips_df %>% filter(sex == 'Female', smoker == 'Yes'))

## [1] 33 7
# 33

gender_vs_smoker <- c( 97 , 54 , 60 , 33 )

# barchart with added parameters
barplot(gender_vs_smoker, main = " gender_vs_smoker " , xlab = " Label " , ylab = " Count " ,
names.arg = c("Male&Non-smoker Female&Non-smoker Male&Smoker Female&Smoker"),
col = "darkred",
horiz = FALSE)
```

There are more male smokers than female smokers.

The number of male non-smokers is also high indicating a gender bias in the data.

The number of female non-smokers is higher than that of female smokers.

## Multivariate Analysis

```
# A glimpse of the data
library(dplyr)
glimpse(tips_df2)

## Rows: 244
## Columns: 3
## $ total_bill <dbl> 16.99, 10.34, 21.01, 23.68, 24.59, 25.29, 8.77, 26.88, 1...
## $ tip <dbl> 1.01, 1.66, 3.50, 3.31, 3.61, 4.71, 2.00, 3.12, 1.96, 3...
## $ size <int> 2, 3, 3, 2, 4, 4, 2, 4, 2, 2, 2, 4, 2, 4, 2, 2, 3, 3, 3,...
```

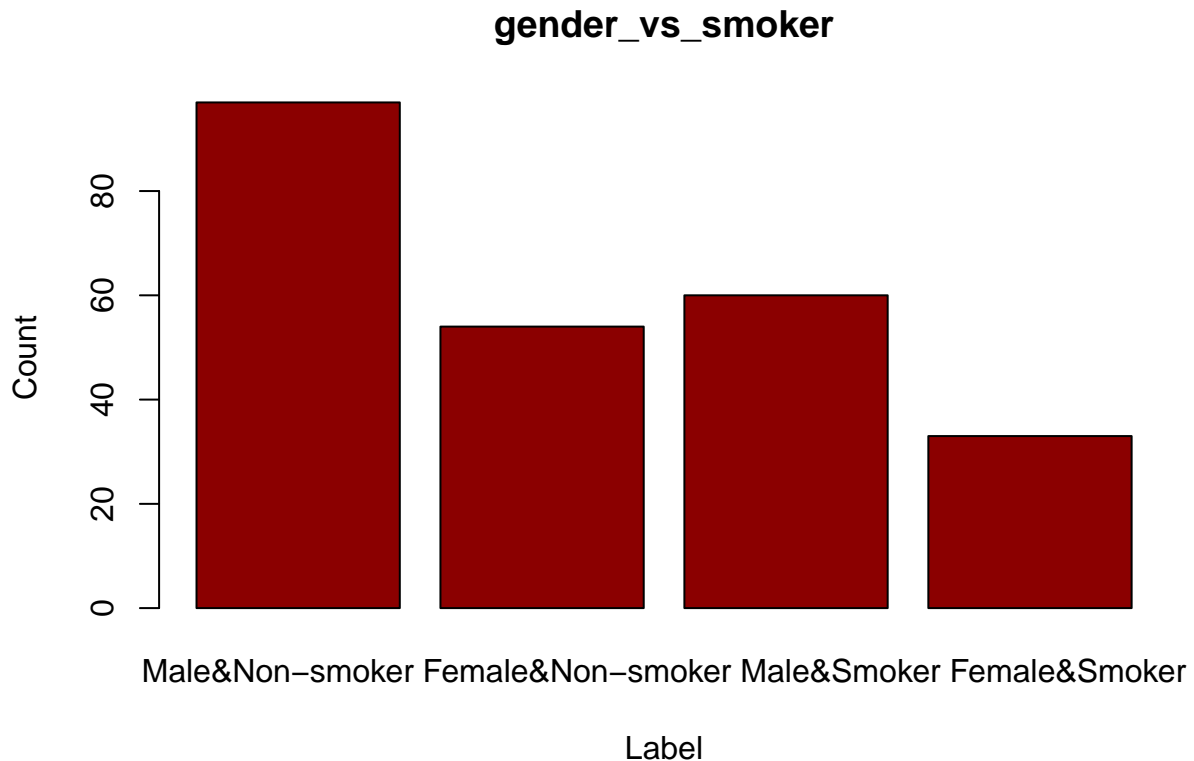


Figure 10: Plot10: Barplot of Gender vs Smoker

```
head(tips_df)
```

```
##   total_bill  tip    sex smoker day   time size
## 1    16.99  1.01 Female   No  Sun Dinner    2
## 2    10.34  1.66   Male   No  Sun Dinner    3
## 3    21.01  3.50   Male   No  Sun Dinner    3
## 4    23.68  3.31   Male   No  Sun Dinner    2
## 5    24.59  3.61 Female   No  Sun Dinner    4
## 6    25.29  4.71   Male   No  Sun Dinner    4
```

```
#subsetting the data
```

```
tips_df3 <- subset(tips_df, select = c(sex,smoker,day,time))
```

```
head(tips_df3)
```

```
##      sex smoker day   time
## 1 Female   No  Sun Dinner
## 2  Male   No  Sun Dinner
## 3  Male   No  Sun Dinner
## 4  Male   No  Sun Dinner
## 5 Female   No  Sun Dinner
## 6  Male   No  Sun Dinner
```

```
#converting the sex column to categorical variables.
```

```
#Code
```

```
tips_df4 <- as.data.frame(apply(tips_df3,2,function(x) {x<-as.numeric(factor(x,levels = unique(x)))}))
```



```
head(tips_df4)
```

```
##   sex smoker day time
## 1   1      1   1   1
## 2   2      1   1   1
## 3   2      1   1   1
## 4   2      1   1   1
## 5   1      1   1   1
## 6   2      1   1   1
```

*#confirming that we have the right number of unique values.*

*#Code*

*#how many unique items are in the sex column*

```
length(unique(unlist(tips_df4[c("sex")])))
```

```
## [1] 2
```

*#how many unique items are in the smoker column*

```
length(unique(unlist(tips_df4[c("smoker")])))
```

```
## [1] 2
```

*#hoe many unique items are in the day column*

```
length(unique(unlist(tips_df4[c("day")])))
```

```
## [1] 4
```

*#hoe many unique items are in the time column*

```
length(unique(unlist(tips_df4[c("time")])))
```

```
## [1] 2
```

*# horizontal merge*

```
d <- merge(tips_df2, tips_df4, all="true")
```

```
head(d)
```

```
##   total_bill  tip size sex smoker day time
## 1    16.99 1.01   2   1      1   1   1
## 2    10.34 1.66   3   1      1   1   1
## 3    21.01 3.50   3   1      1   1   1
## 4    23.68 3.31   2   1      1   1   1
## 5    24.59 3.61   4   1      1   1   1
## 6    25.29 4.71   4   1      1   1   1
```

## Modelling

```
head(d)
```

```
##   total_bill  tip size sex smoker day time
## 1    16.99 1.01   2   1      1   1   1
## 2    10.34 1.66   3   1      1   1   1
## 3    21.01 3.50   3   1      1   1   1
## 4    23.68 3.31   2   1      1   1   1
## 5    24.59 3.61   4   1      1   1   1
## 6    25.29 4.71   4   1      1   1   1
```

## Linear Regression

### Create Training and Test data

```
model <- lm(tip ~ total_bill+size+sex+smoker+day+time, data = d)
summary(model)
```

```
##
## Call:
## lm(formula = tip ~ total_bill + size + sex + smoker + day + time,
##     data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9279 -0.5547 -0.0852  0.5095  4.0425
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.689e-01  2.711e-02   24.67  <2e-16 ***
## total_bill   9.271e-02  5.799e-04  159.87  <2e-16 ***
## size        1.926e-01  5.428e-03   35.48  <2e-16 ***
## sex         -3.507e-14  8.860e-03    0.00      1
## smoker      -1.446e-14  8.871e-03    0.00      1
## day         -1.723e-14  6.261e-03    0.00      1
## time         8.902e-15  1.280e-02    0.00      1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.007 on 59529 degrees of freedom
## Multiple R-squared:  0.4679, Adjusted R-squared:  0.4678
## F-statistic: 8723 on 6 and 59529 DF, p-value: < 2.2e-16
```

The p-value is lower than the significance level which means that the model is statistically significant.

Next, we examine the coefficients table to obtain the estimate of regression beta coefficients and the associated t-statistic p-values

```
summary(model)$coefficient
```

```
##              Estimate Std. Error      t value      Pr(>|t|)
## (Intercept)  6.689447e-01  0.0271112043  2.467411e+01  9.569756e-134
## total_bill   9.271334e-02  0.0005799436  1.598661e+02  0.000000e+00
## size        1.925978e-01  0.0054283441  3.548003e+01  7.149936e-273
## sex         -3.506477e-14  0.0088596104 -3.957823e-12  1.000000e+00
## smoker      -1.446117e-14  0.0088708869 -1.630183e-12  1.000000e+00
## day         -1.722720e-14  0.0062613841 -2.751340e-12  1.000000e+00
## time         8.901614e-15  0.0127954921  6.956836e-13  1.000000e+00
```

The coefficients show us that total bill, time and size affect the size of the tip given while sex, smoker and day do not affect tip amounts

```
model2 <- lm(tip ~total_bill+size+time, data = d)
summary(model2)
```

```
##
## Call:
## lm(formula = tip ~ total_bill + size + time, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9279 -0.5547 -0.0852  0.5095  4.0425
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.689e-01  1.704e-02   39.26  <2e-16 ***
## total_bill   9.271e-02  5.799e-04   159.87  <2e-16 ***
## size         1.926e-01  5.428e-03    35.48  <2e-16 ***
## time        -6.531e-15  9.208e-03     0.00      1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.007 on 59532 degrees of freedom
## Multiple R-squared:  0.4679, Adjusted R-squared:  0.4678
## F-statistic: 1.745e+04 on 3 and 59532 DF,  p-value: < 2.2e-16

confint(model2)

##              2.5 %      97.5 %
## (Intercept)  0.63554615 0.70234333
## total_bill   0.09157667 0.09385000
## size         0.18195849 0.20323710
## time        -0.01804682 0.01804682

sigma(model2)/mean(d$tip)

## [1] 0.3359561
```

Our model has a 33% error rate

The value for R2 indicates the level of accuracy of the model. The R2 value is 46.79 for both models.

To increase the level of accuracy, we can train using more data, use a more robust algorithm such as gradient boost.

## Discussion

Our analysis indicates that it is possible to predict a waiters with 67% accuracy. The factors that influence the amount of tip the most are the total bill, size of the group and the time when the meal was taken. The average tipping rate is 7.048% of the total bill. Tips are an acceptable secondary source of income for servers thus one can calculate their earnings if the total bill is known. 67% is the confidence level and it is abit low. Therefore, some enhancements are necessary to increase the level of accuracy.

Conclusions:

The analysis was completed successfully. A multiple Linear regression was used to predict a waiters tip amount using other dependent variables. The coefficients of sex is -3.653905e-14 , that of being a smoker is -1.493559e-14 and for the day of week is -1.779964e-14. All the three coefficients are negative meaning that they do not affect the dependent variable(Amount of tip in dollars). The t-statistic for sex is -4.124228e-12, for that of smoker is -1.683664e-12 and day is -2.842765e-12 which further supports previous findings that the three variables have a negative relationship with the dependent variable.

**Weaknesses of the analysis.** The dataset had 244 records which means that the data may be insufficient for modelling.

Some of the variables such as sex, smoker and day do not influence tipping decisions.

**Future steps of the analysis.** using a larger dataset

Using variables that with a positive correlation towards the dependent variable.

## **References:**

Data: <https://www.kaggle.com/jsphyg/tipping>

Bryant, P. G. and Smith, M (1995) Practical Data Analysis: Case Studies in Business Statistics. Homewood, IL: Richard D. Irwin Publishing

Introduction: The promise of collaborative public service delivery. (2019). Collaboration in Public Service Delivery, 1-1. <https://doi.org/10.4337/9781788978583.00008>

Rathore, S. (2015). Capturing, analyzing, and managing word-of-Mouth in the digital marketplace. IGI Global.

Toporek, A. (2015). Be your customer's hero: Real-world tips and techniques for the service front lines. AMACOM.