Level 3:

Task 1: Predictive Modeling

Build a regression model to predict the aggregate rating of a restaurant based on available features. Split the dataset into training and testing sets and evaluate the model's performance using appropriate metrics. Experiment with different algorithms (e.g., linear regression, decision trees, random forest) and compare their performance.

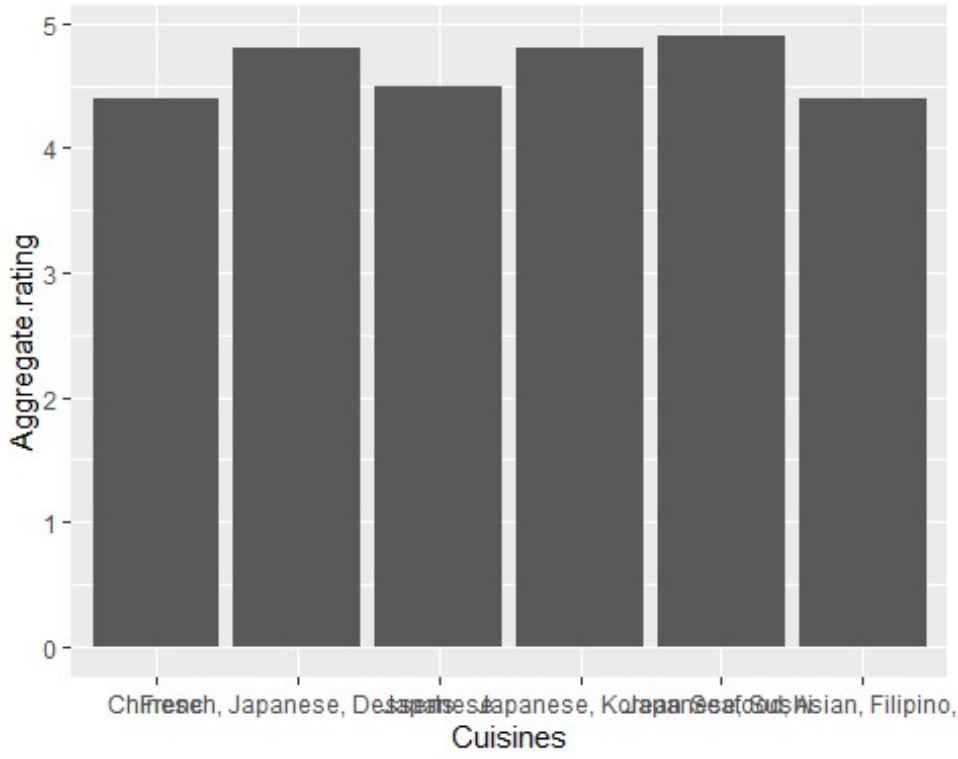
```
train_index<-sample(1:120,0.7*120)
x_train<-df$Encode_Has_Table_Booking[train_index]
y_train<-df$Aggregate.rating[train_index]
x_test<-df$Encode_Has_Table_Booking[-train_index]
y_test<-df$Aggregate.rating[-train_index]
df_train<-data.frame(x=x_train,y=y_train)
df_test<-data.frame(x=x_test,y=y_test)
lm_model<-function(df_train){
   beta1<-sum((df_train$x-mean(df_train$x))*(df_train$y-
mean(df_train$y)))/sum((df_train$x-mean(df_train$x))^2)
 beta0<-mean(df_train$y)-beta1*mean(df_train$x)
 return(c(x1=beta0,y1=beta1))
}
ans<-lm_model(df_train)
print("Slope and Intercept From Linear Regression Model")

## [1] "Slope and Intercept From Linear Regression Model"

ans

##        x1        y1
## 3.1110811 0.6794595

lr_predict<-function(ans,df_test)
{
 y_pred<-ans["x1"]+ans["y1"]*df_test$x
 return(data.frame(pred=y_pred))
}
ans1<-lr_predict(ans,df_test)

print("Prediction Of Aggregate Rating based On Features")

## [1] "Prediction Of Aggregate Rating based On Features"

glimpse(head(ans1))

## Rows: 6
## Columns: 1
## $ pred <dbl> 3.790541, 4.470000, 3.790541, 4.470000, 3.790541, 4.470000

#Performance Evaluation
mse<-mean((df_test$y-ans1$pred)^2)
```

```
mse<-sqrt(mse)

print("The Evaluated Performance Metrices based on Root Mean Square Error")

## [1] "The Evaluated Performance Metrices based on Root Mean Square Error"

mse

## [1] 1.923761
```

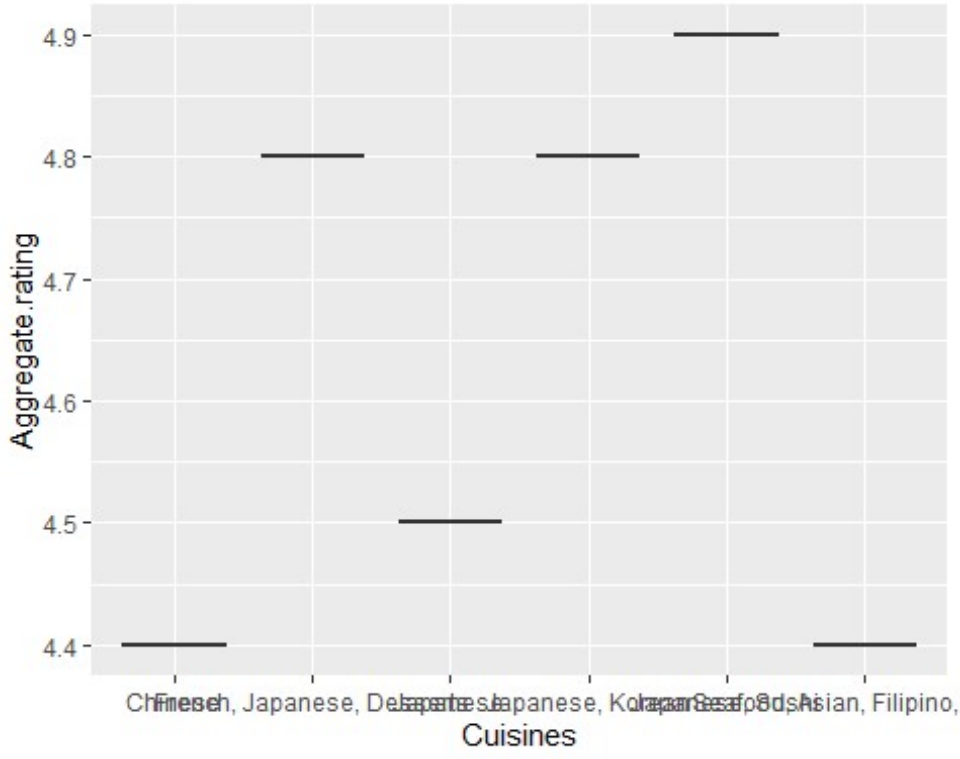
Task 2: Customer Preference

Analysis Analyze the relationship between the type of cuisine and the restaurant's rating. Identify the most popular cuisines among customers based on the number of votes. Determine if there are any specific cuisines that tend to receive higher ratings.

```
ans<-head(df)
ggplot(data =ans)+geom_bar(mapping = aes(x=Cuisines,y=Aggregate.rating),stat
= "identity")
```




```
ggplot(data =ans)+geom_boxplot(mapping = aes(x=Cuisines,y=Aggregate.rating))
```




```
print("Most Popular Cuisines")
```

```
## [1] "Most Popular Cuisines"
```

```
df %>% filter(Votes==max(Votes)) %>% summarise(Cuisines)
```

```
##                   Cuisines
## 1 Italian, American, Pizza
```

```
print("Specific Cuisines to receive high Ratings")
```

```
## [1] "Specific Cuisines to receive high Ratings"
```

```
df %>% filter(Aggregate.rating==max(Aggregate.rating)) %>% reframe(Cuisines)
```

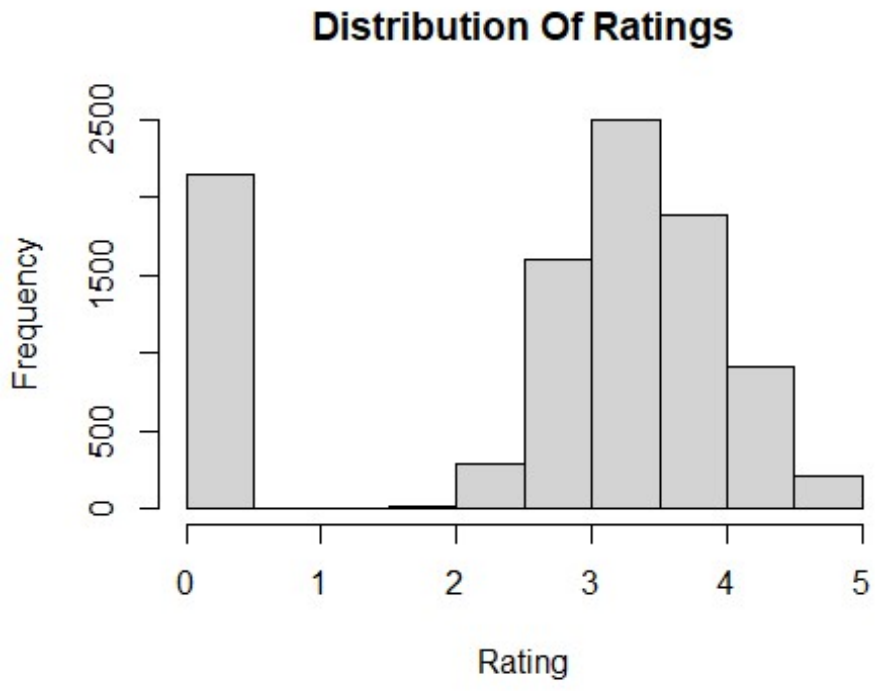
```
##                                 Cuisines
## 1                         Japanese, Sushi
## 2                 European, Asian, Indian
## 3                       Filipino, Mexican
## 4                           International
## 5                     Brazilian, Bar Food
## 6                     Brazilian, Bar Food
## 7           American, Caribbean, Seafood
## 8                                 Burger
## 9                BBQ, Breakfast, Southern
## 10                                 Asian
## 11               American, Coffee and Tea
## 12               Sandwich, Seafood, Cajun
```

```
## 13                       Pizza, Sandwich
## 14               American, Sandwich, Tea
## 15               American, BBQ, Sandwich
## 16               Burger, Bar Food, Steak
## 17                     Hawaiian, Seafood
## 18                              Japanese
## 19                         Italian, Deli
## 20                       European, German
## 21                    Indian, North Indian
## 22                     Continental, Indian
## 23                                  Indian
## 24                                  Indian
## 25            Cafe, North Indian, Chinese
## 26                               Fast Food
## 27 North Indian, European, Mediterranean
## 28                         Bakery, Desserts
## 29                             North Indian
## 30       Mexican, American, Healthy Food
## 31                             North Indian
## 32 European, Mediterranean, North Indian
## 33 European, Mediterranean, North Indian
## 34           Italian, Bakery, Continental
## 35                  North Indian, Chinese
## 36                  North Indian, Chinese
## 37                      Mughlai, Lucknowi
## 38   North Indian, South Indian, Mughlai
## 39 North Indian, European, Mediterranean
## 40                               Ice Cream
## 41                           Modern Indian
## 42                           Modern Indian
## 43  North Indian, Chinese, Mediterranean
## 44                       Sunda, Indonesian
## 45                         Sushi, Japanese
## 46                       Sunda, Indonesian
## 47                       Sunda, Indonesian
## 48                                Desserts
## 49                                Desserts
## 50                                   Steak
## 51                                 British
## 52                 Taiwanese, Street Food
## 53               American, Burger, Grill
## 54                                Chinese
## 55                European, Contemporary
## 56                                  Tapas
## 57                                 French
## 58                                Seafood
## 59                           World Cuisine
## 60                                    Cafe
## 61                               Bar Food
```

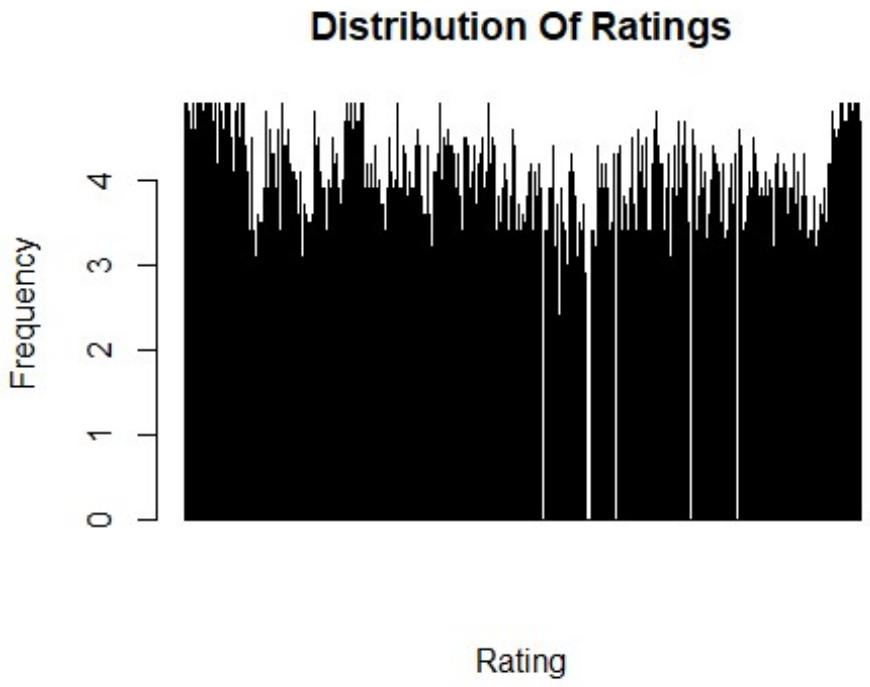
Task 3: Data Visualization

Create visualizations to represent the distribution of ratings using different charts (histogram, bar plot, etc.). Compare the average ratings of different cuisines or cities using appropriate visualizations. Visualize the relationship between various features and the target variable to gain insights.

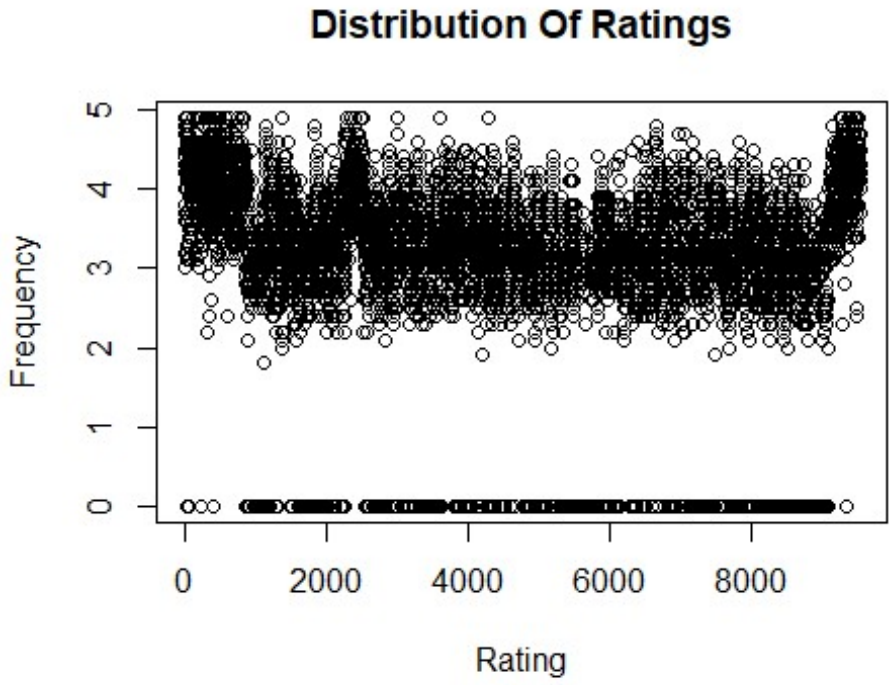
```
hist(df$Aggregate.rating,main = "Distribution Of Ratings",xlab =
"Rating",ylab = "Frequency")
```



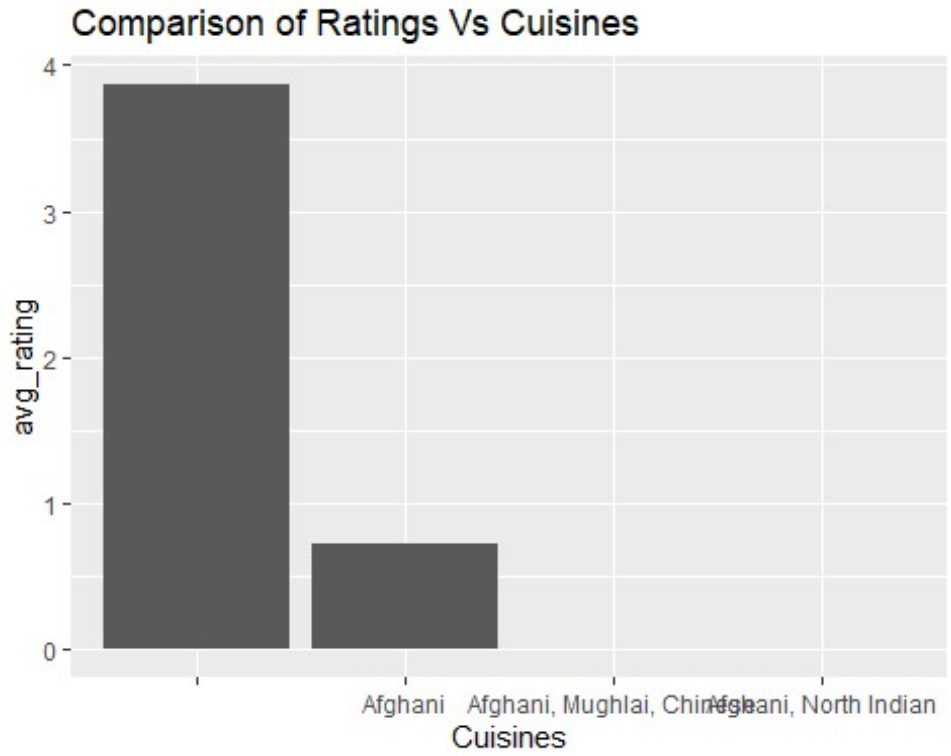

```
barplot(df$Aggregate.rating,main = "Distribution Of Ratings",xlab =
"Rating",ylab = "Frequency")
```



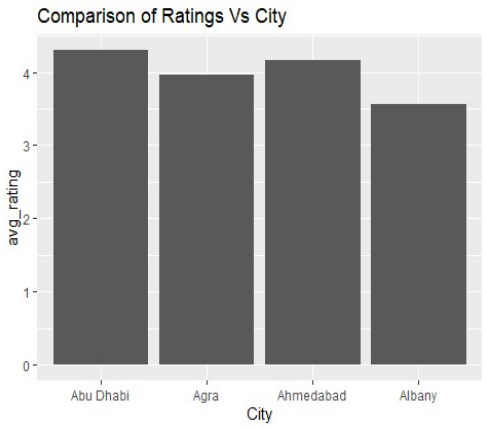

```
plot(df$Aggregate.rating,main = "Distribution Of Ratings",xlab =
"Rating",ylab = "Frequency")
```



```
avg_rate_diff_cuisine<-head(df %>% group_by(Cuisines) %>%
summarise(avg_rating=mean(Aggregate.rating)),4)

library(ggplot2)
ggplot(data=avg_rate_diff_cuisine,mapping=aes(x=Cuisines,y=avg_rating))+geom_
bar(stat="identity")+labs(title = "Comparison of Ratings Vs Cuisines")
```




```
avg_rate_diff_city<-head(df %>% group_by(City) %>%
summarise(avg_rating=mean(Aggregate.rating)),4)

ggplot(data=avg_rate_diff_city,mapping=aes(x=City,y=avg_rating))+geom_bar(sta
t="identity")+labs(title = "Comparison of Ratings Vs City")
```



```
ggplot(data=df)+geom_boxplot(mapping=aes(x=Cuisines,y=Aggregate.rating))+coor
d_cartesian(xlim = c(0,3))
```