WEBSITE AD-CLICK

Research Project by
ALI AHMAD

CONTENTS

Case Study

Objective

Univariate Analysis

Bivariate Analysis

Model Building and Outcome

Observations

Business Recommendations

CASE STUDY

This dataset is mainly about who is likely going to click on the advertisement.

It consists of 14 variables:-

VistID	Time_Spent	Age	Avg_Income
Internet_Usage	Ad_Topic	Country_Name	City_code
Male	Time_Period	Weekday	Month
	Year	Clicked	

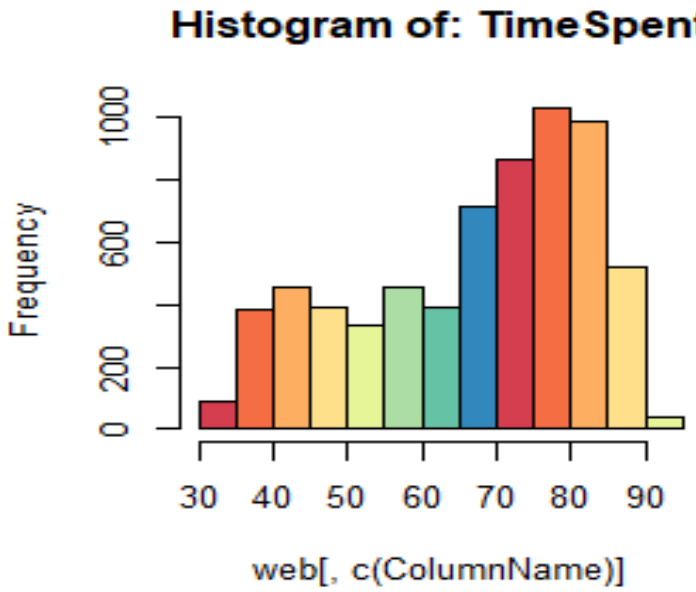
OBJECTIVE

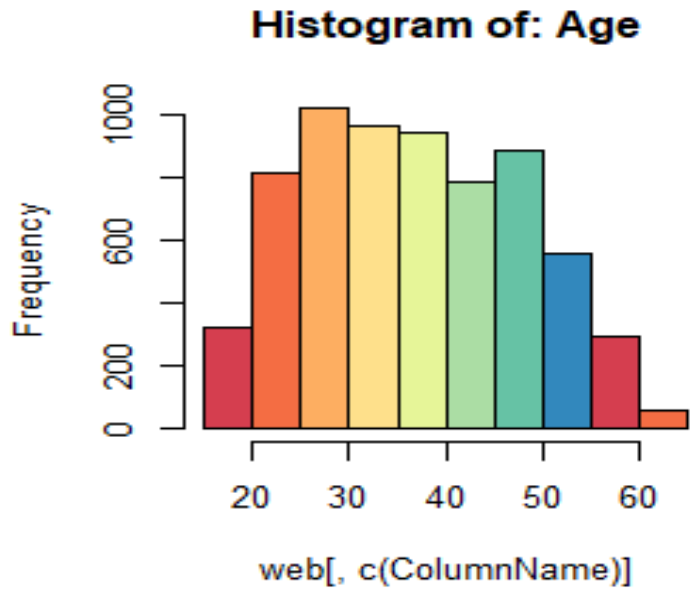
The objective here is to perform a Logistic Regression analysis to arrive at a model that can be used to predict who is likely going to click on the Advertisement so it can contribute to the more revenue generation to the organization.

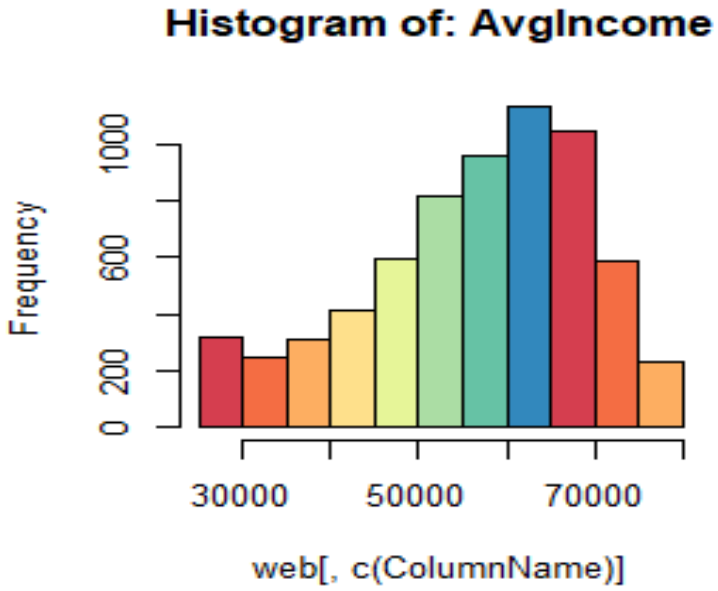
UNIVARIATE ANALYSIS

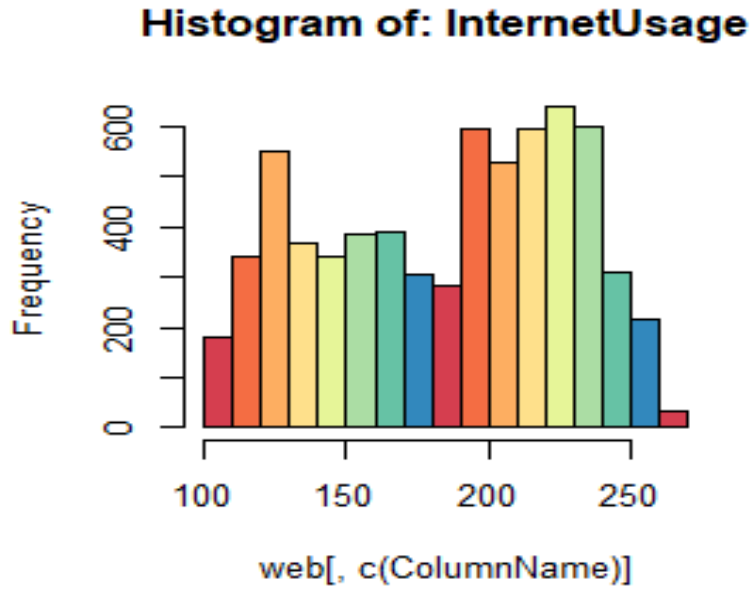
Continuous column:- Using Histogram

- TimeSpent is left skewed and the maximum time spent is between 70 to 90.
- AvgIncome is also left skewed and the ad clicks are maximum between 50000 to 70000.

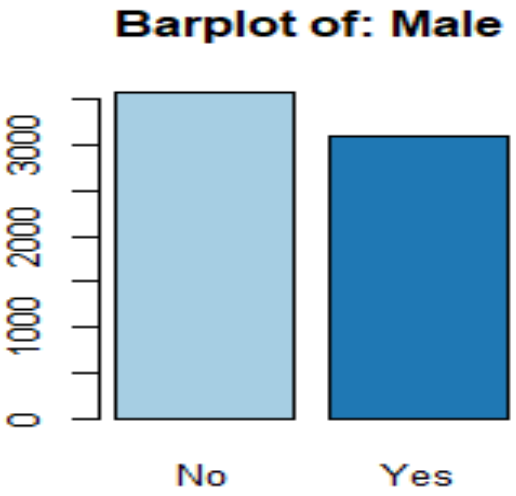


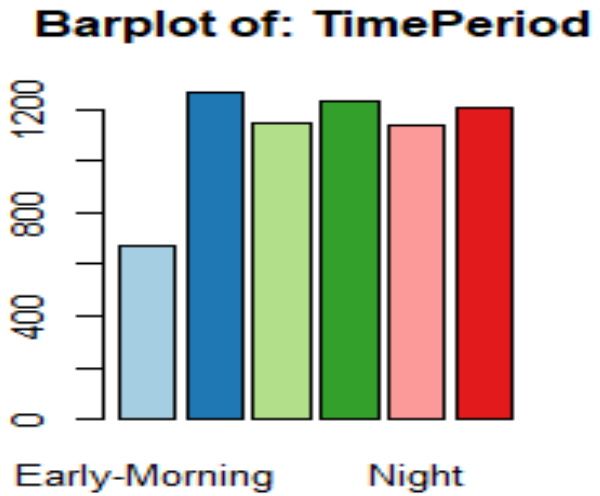




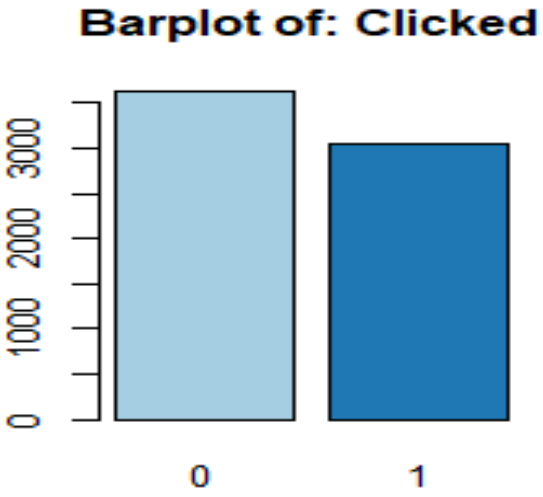


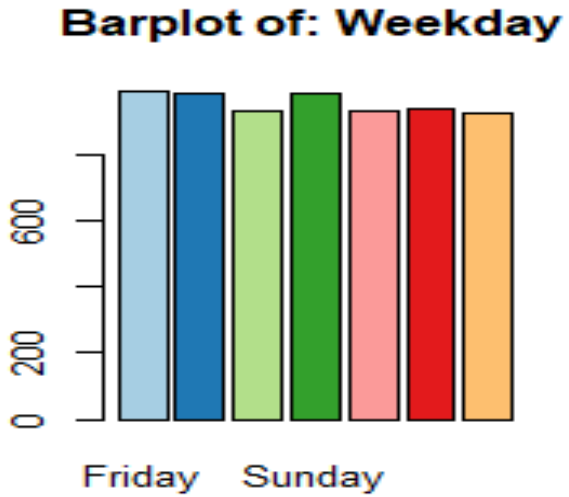
Categorical column:- Using Barplot



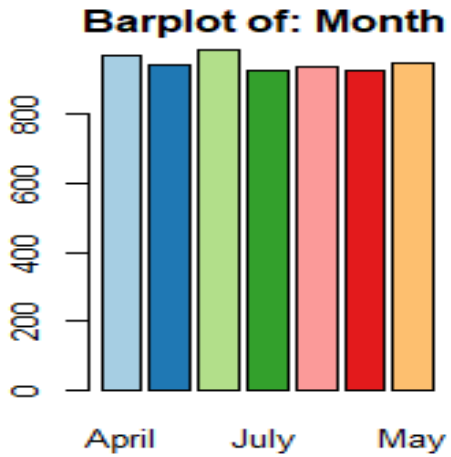


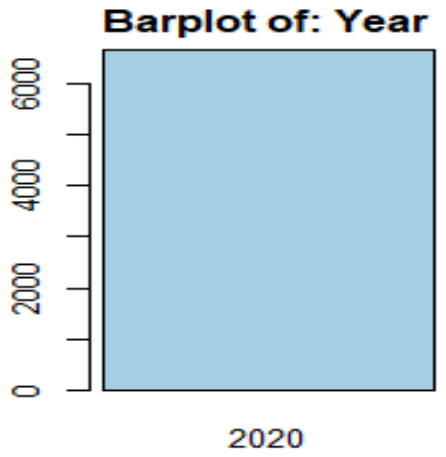
- Number of unclicked is more than clicked.
- Number of customers are less in early-morning as compared to other time.

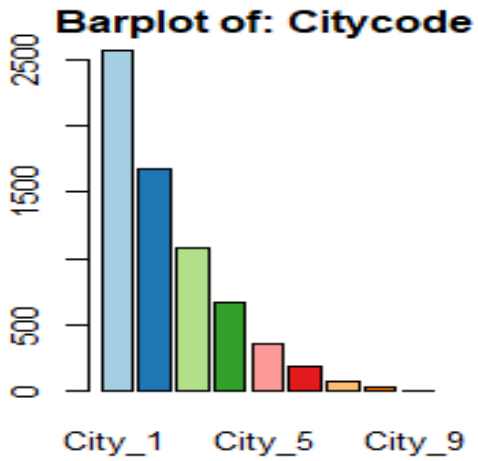




- City_1 has clicked more on ads as compared to the other cities.



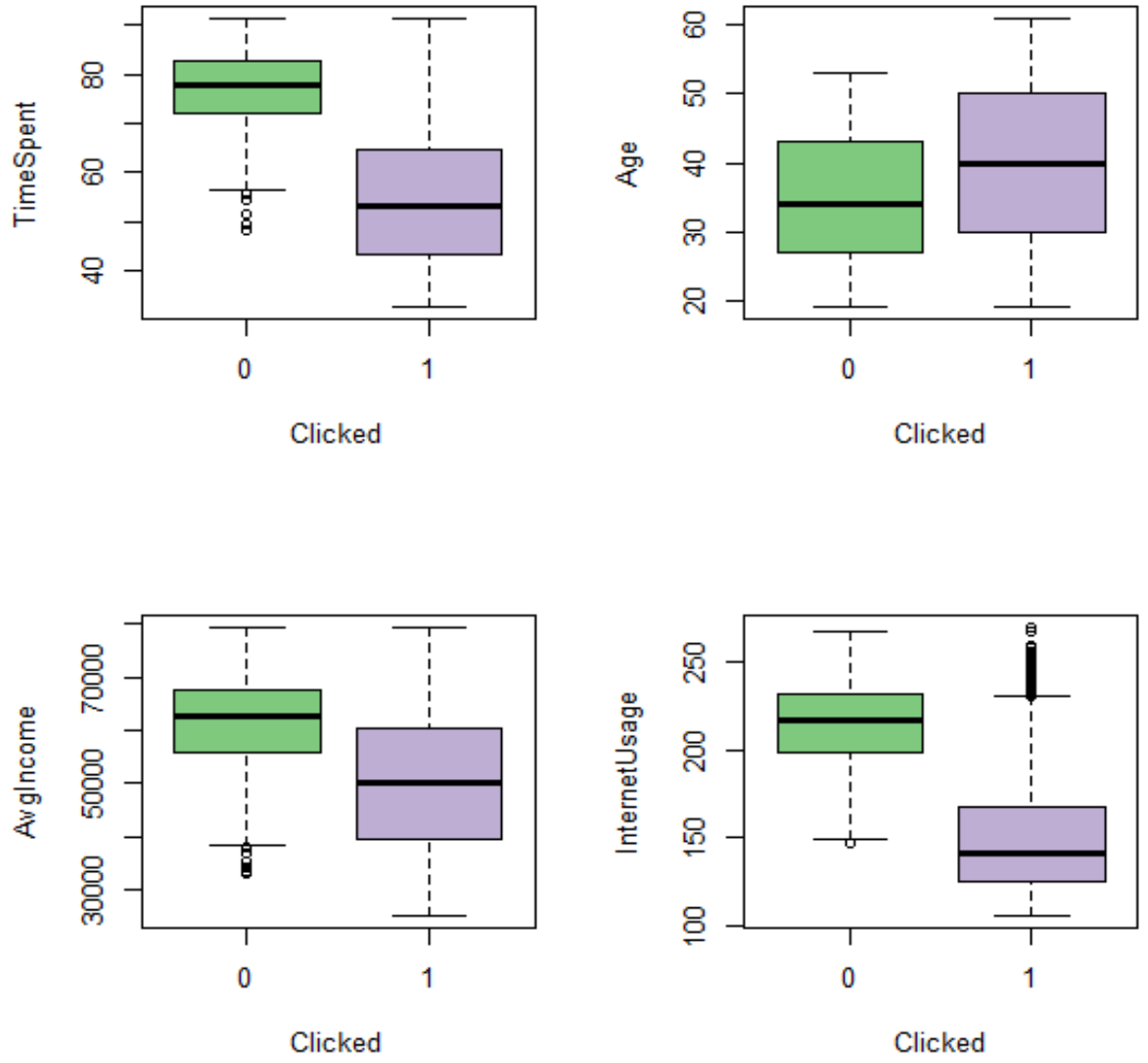




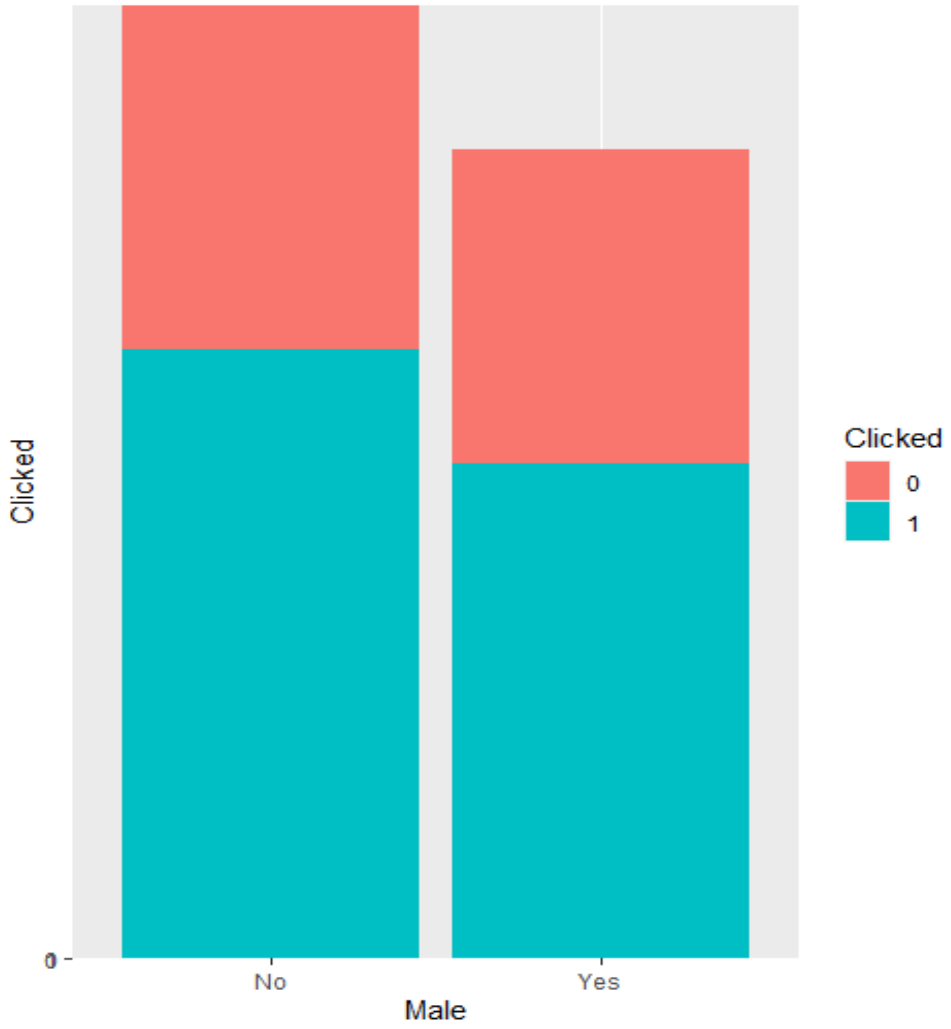
BIVARIATE ANALYSIS

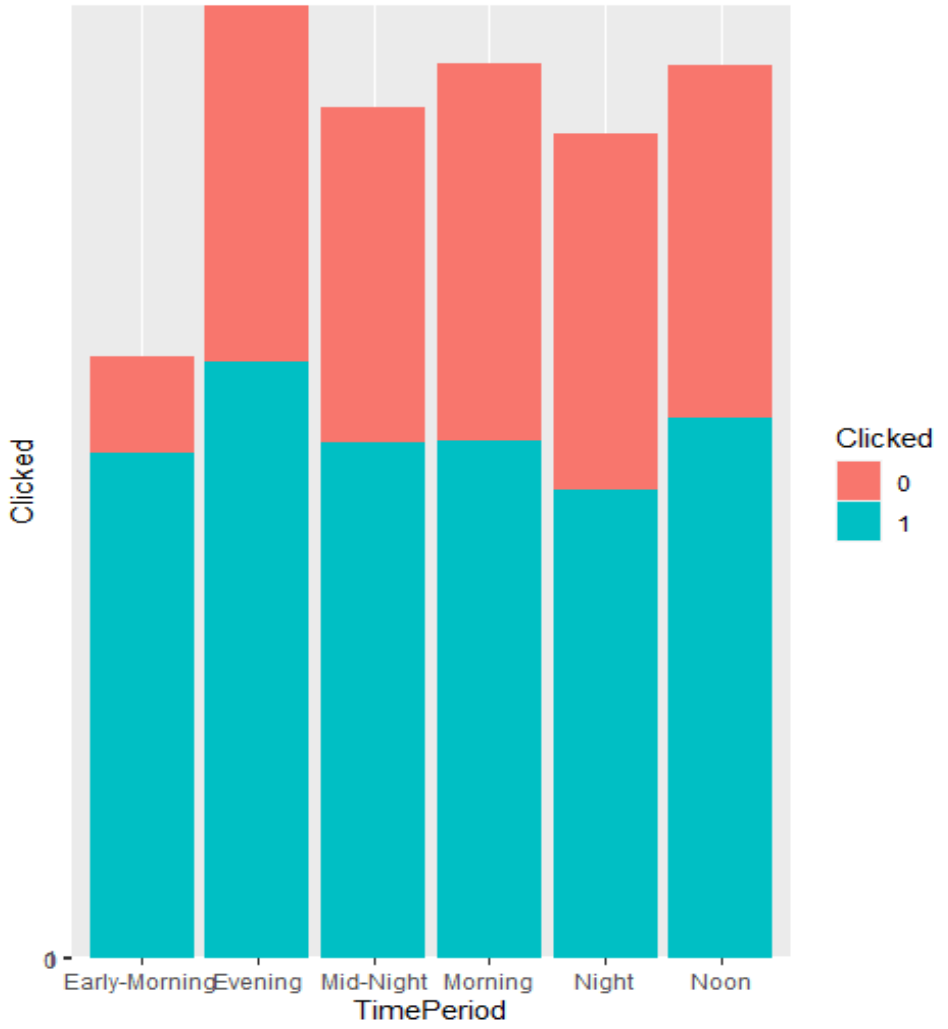
Using Boxplot

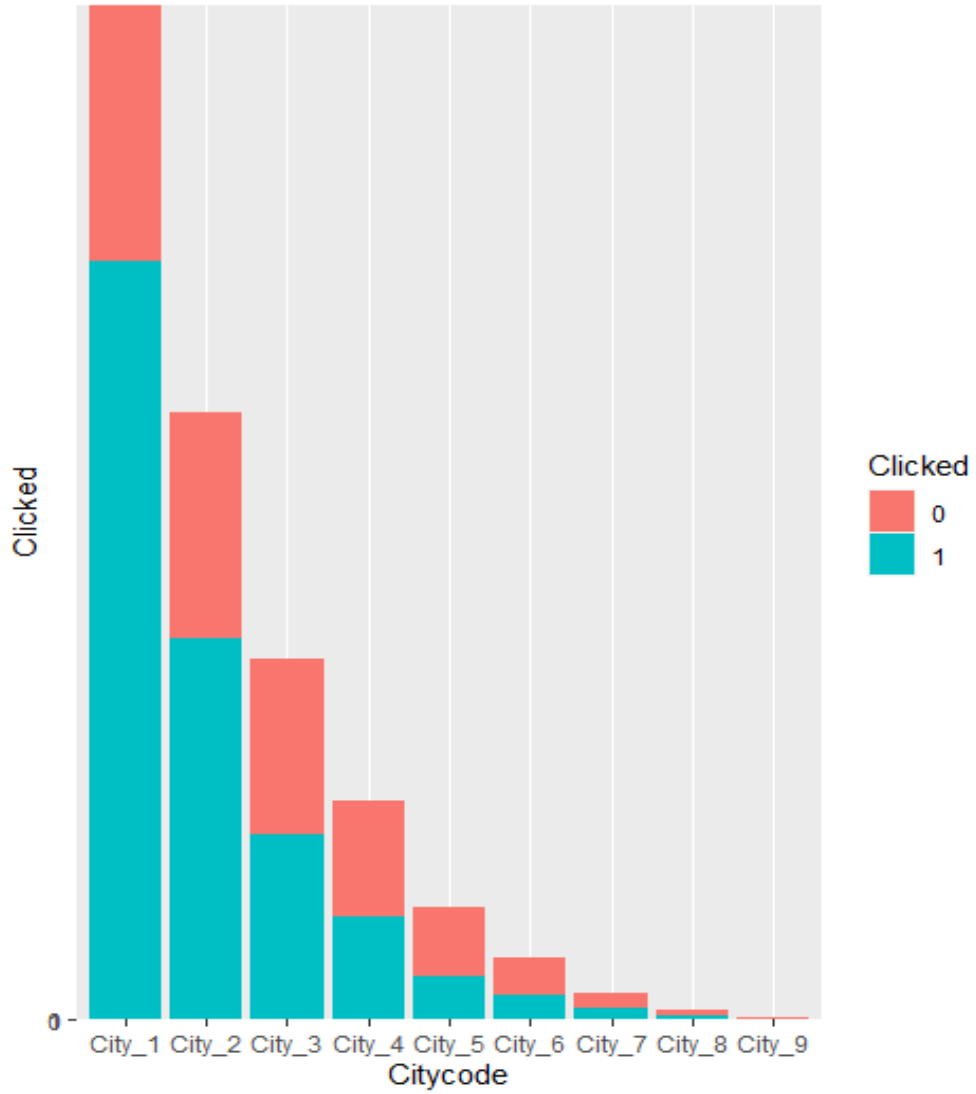
- The average age of the user is 36 years old, whilst the youngest is 19 and the oldest is 61.
- It is safe to conclude that the site's target audience is adults.
- The daily time spent by users on the website ranges between 32.6–91.4 minute

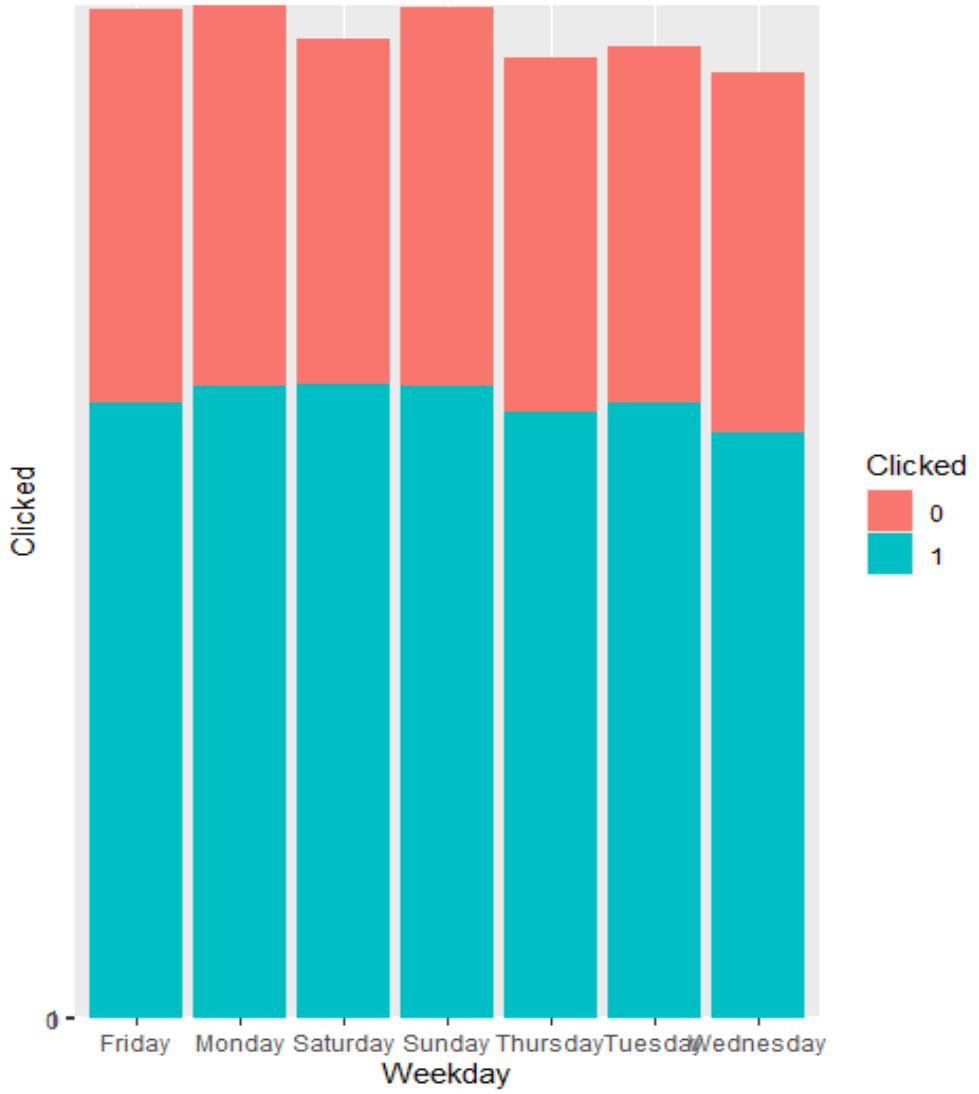


- Using Group bar chart






Clicked
0
City_1
City_2
City_3
City_4
City_5
City_6
City_7
City_8
City_9
Citycode
Clicked
0
1


Clicked
0
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Weekday
Clicked
0
1

ANOVA (Analysis of Variance)

It gives the relationship between continuous and categorical columns whether they are statistically significant or not.

H0- Null hypothesis means the variables are not correlated:

1. If Small P-Value < 5% (0.05) means the variables are correlated. Null hypothesis H0 is rejected
2. If Large P-Value > 5% (0.05) means the variables are not correlated and the Null hypothesis H0 is accepted

In this dataset, the variables (p-value < 0.05) that are found significant from ANOVA test:

- TimeSpent
- Age
- AvgIncome
- InternetUsage

Chi-Square Test

Chi-square is a quantitative measure used to determine whether relationship exists between two categorical variables.

H0- Null hypothesis means the variables are not correlated:

1. If Small P-Value < 5% (0.05) means the variables are correlated. Null hypothesis H0 is rejected

2. If Large P-Value > 5% (0.05) means the variables are not correlated and the Null hypothesis H0 is accepted

The variables (p-value < 0.05) that are found significant from Chi-square test:

- Male
- TimePeriod
- Citycode
- Year

The variables Week and Month found with large p-value i.e. p-value > 0.05. So, we can reject these variables.

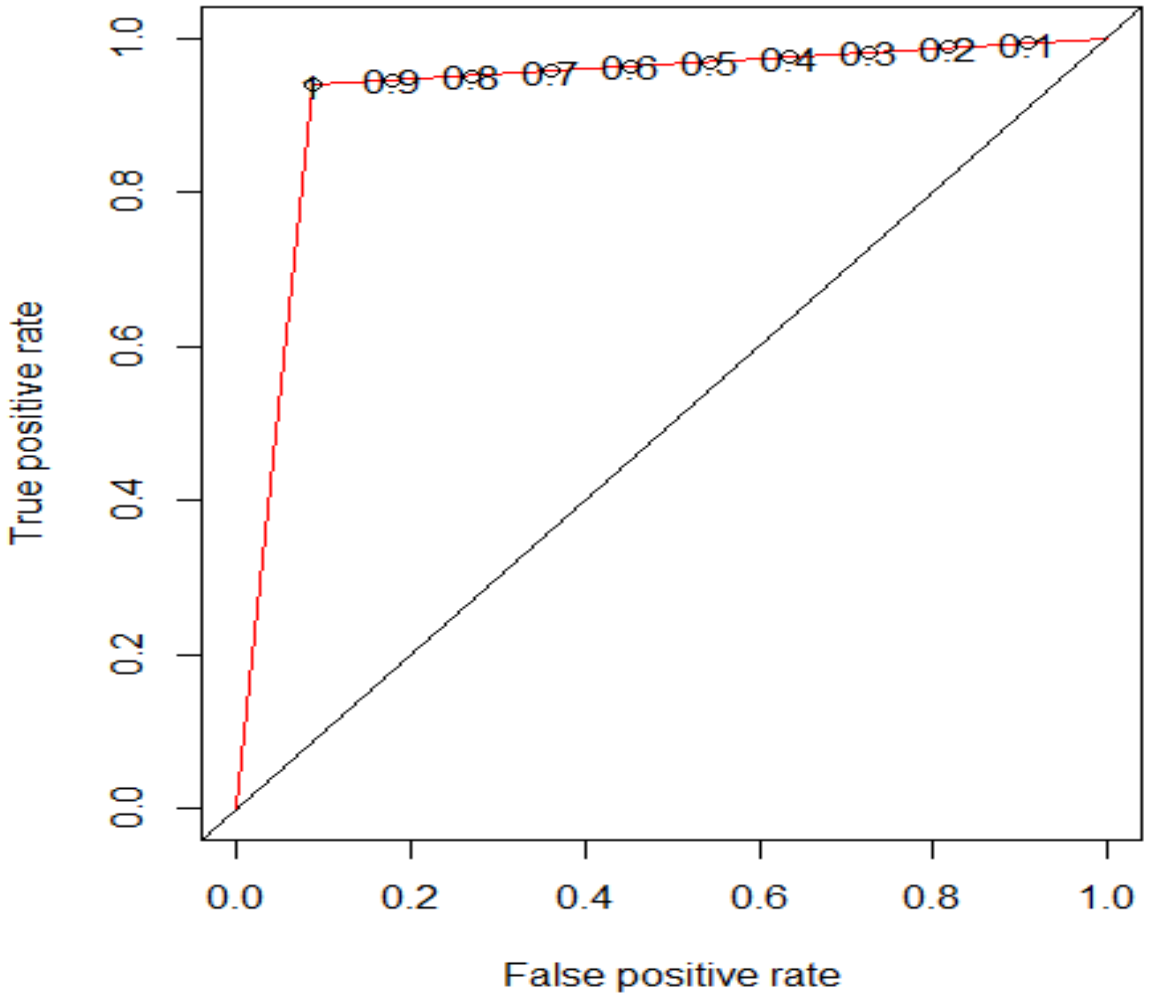
Model Building and Outcome

Significant Variables

- TimeSpent
- Age
- AvgIncome
- InternetUsage
- TimePeriod
- Citycode

Sigmoid Curve

ROC curve plots the true positive rate against false positive rate.



Observations

Accuracy of Logistic Regression model: 0.9237

Sensitivity or Recall : 0.9121

Specificity : 0.9390

Positive Pred Value or Precision : 0.9514

Negative Pred Value : 0.8908

Balanced Accuracy : 09255

F1-Measures or F1-Score : 0.9313356

Business Recommendations

- We can see that more people aged between 30 to 40 are spending more time on site daily. We can give more advertisement related to discounts to the people between 50 to 60 in order to generate more revenue for the organisation.
- User tends to click on a Ad later in a day or probably early in the morning. It is expected based on the age feature that most people are working so it seems appropriate as they either find time early or late in the day. Also Sunday seems to be effective for clicking on a ad from the bar chart.
- The percentage of males visiting the website is slightly lower than females who are clicking on ads. We can show more advertisements related to men products or sports in which men are interested like football ads. This will increase the clicking on ads and can elevate the revenue of the organisation.
- The area income of users ranges between $13,996.50 - $79,484.80. Quite a large distribution of incomes. This tells us that site visitors hail from various social classes.

THANK YOU