ADMISSION PREDICTOR

A DATA DRIVEN APPROACH

INTRODUCTION

- This presentation summarizes the analysis models and statistical modeling associated with admission predictor dataset which was created for prediction of Graduate Admissions from an Indian perspective. Primary goal of this presentation is to see to what extent we can predict the admission rate in universities.
- The presentation is as follows :

  Data Description

  Exploratory Data Analysis

  Inferential Data Analysis

  Final Model

  Conclusion

ABOUT THE DATA

The dataset is collected from Kaggle which is inspired by the UCLA Graduate Dataset and is owned by Mohan S Acharya. The dataset contains 9 quantitative variables and 500 records. This dataset was built with the purpose of helping students in shortlisting universities with their profiles.

Variable	Variable description
GRE Scores	GRE Scores (out of 340)
TOEFL Scores	TOEFL Scores (out of 120)
University Rating	University rating (out of 5)
SOP	Strength of Statement of Purpose (out of 5)
LOR	Strength of letter of recommendation (out of 5)
CGPA	Undergraduate GPA (out of 10)
Research	Research experience (0 or 1)
Chance of Admit	Chances of getting Admission (from 0 to 1)

EXPLORATORY DATA ANALYSIS

The table below explains the five-umber summary of the data(i.e., Min, Q1, Median, Q3, Max, IQR).

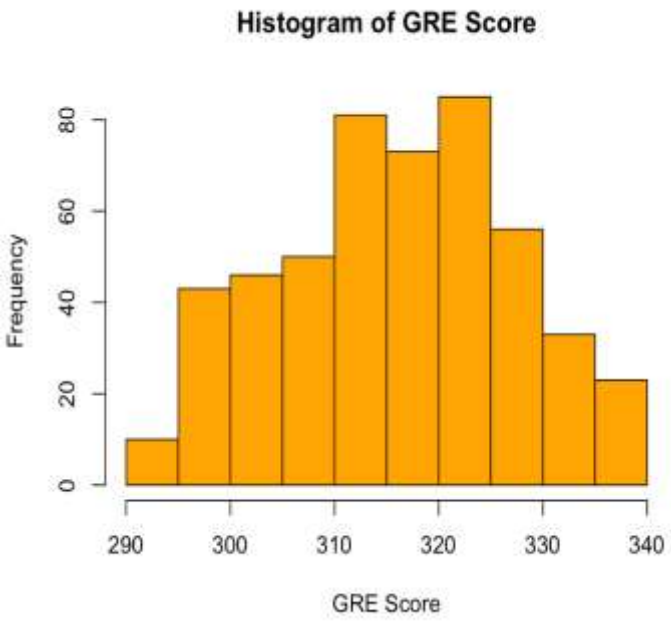
Variable Names -->	GRE Scores	TOEFL Scores	University Rating	SOP	LOR	CGPA	Chance of Admit
Min	290.0	92.0	1.00	1.00	1.00	6.80	0.34
Q1	308.0	103.0	2.00	2.50	3.00	8.13	0.63
Median	317.0	107.0	3.00	3.50	3.50	8.56	0.72
Q3	325.0	112.0	4.00	4.00	4.00	9.04	0.82
Max	340.0	120.0	5.00	5.00	5.00	9.92	0.97
IQR	17.0	9.0	2.00	1.50	1.00	0.91	0.19

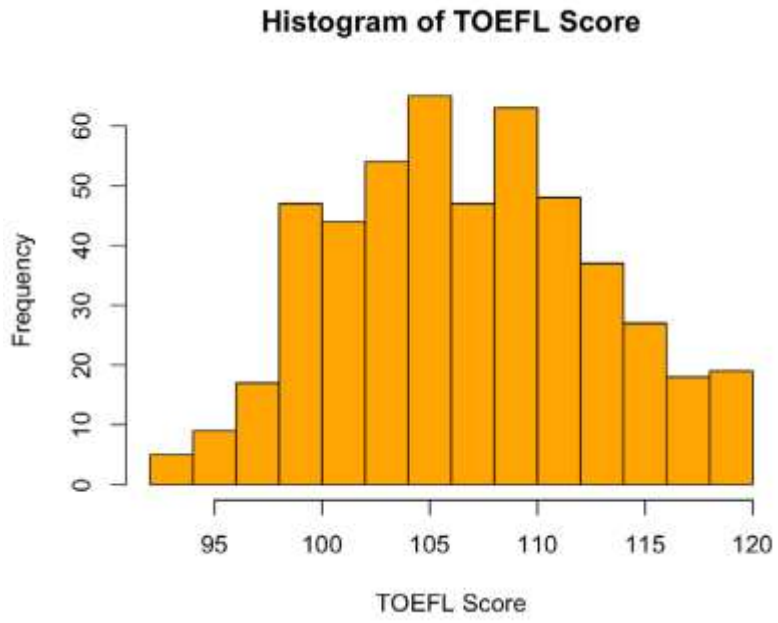
The table below shows the descriptive analysis i.e., Mean, Median, Standard Deviation, Mode ad Range of the variables in the dataset.

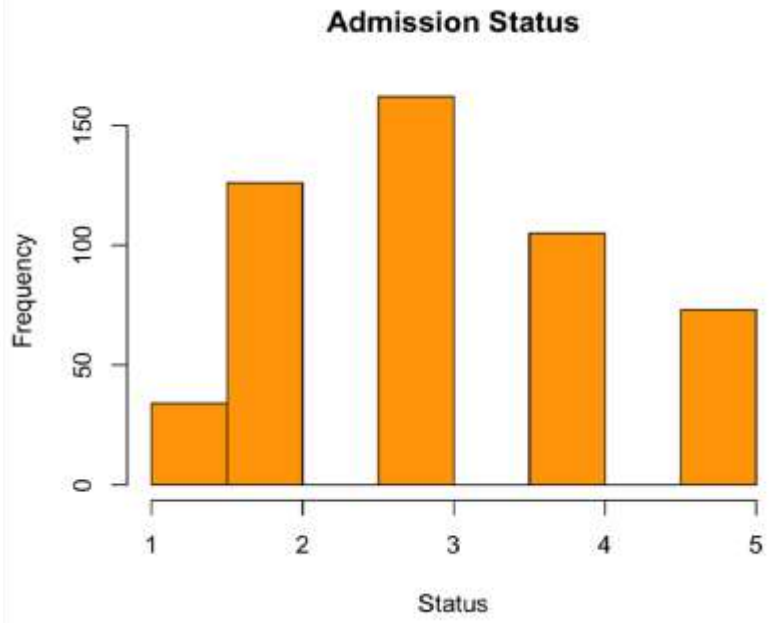
Variable Name -->	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Chance of Admit
Mean	316.5	107.2	3.114	3.374	3.484	8.6	0.72
Median	317.0	107.0	3.00	3.50	3.50	8.6	0.72
Standard Deviation	11.295	6.082	1.144	0.991	0.925	0.6	0.14
Mode	312	110	3	4	3	8.0	0.71
Range	50	28	4	4	4	3.1	0.63

HISTOGRAMS OF QUANTITATIVE VARIABLES

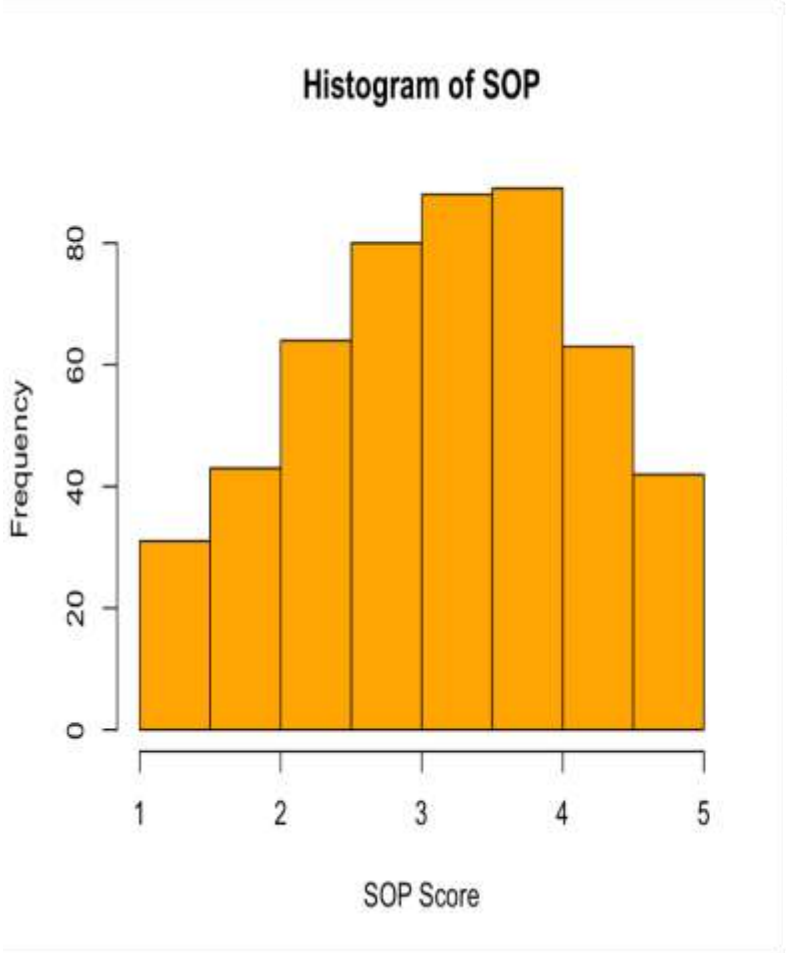
- GRE Score- The histogram of GRE Score looks unimodal, that is, normally distributed.
- TOEFL Score- The histogram for this is also unimodal with no skewness.
- University Rating- The histogram of the University Rating has values at certain datapoints and it doesn't look continuous.

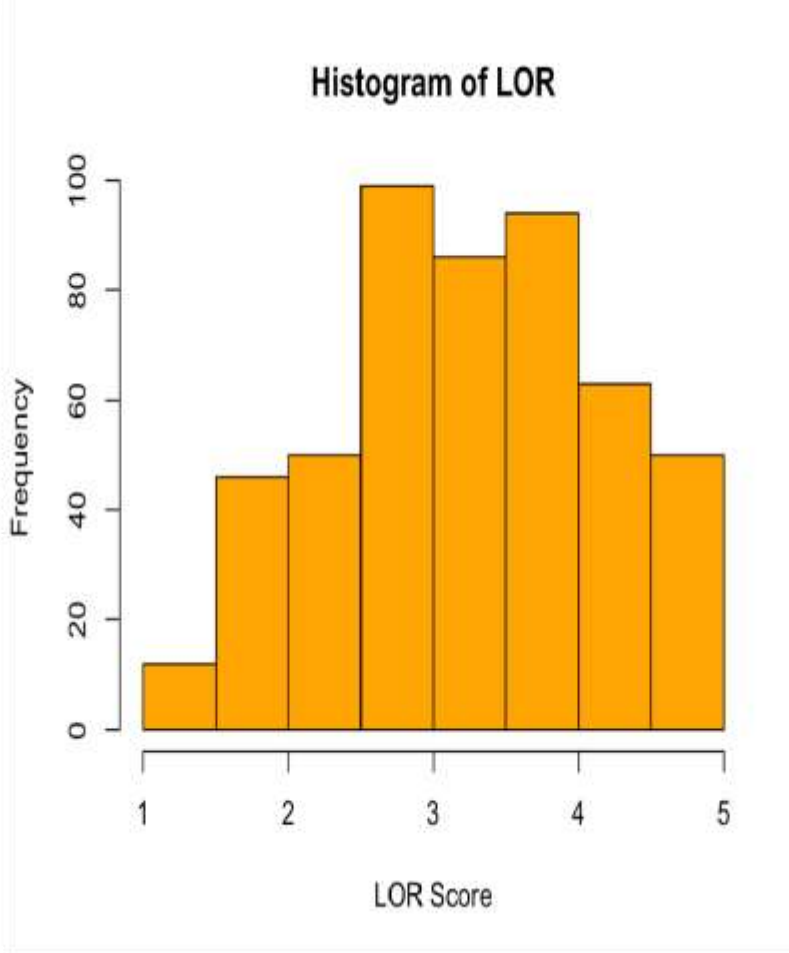


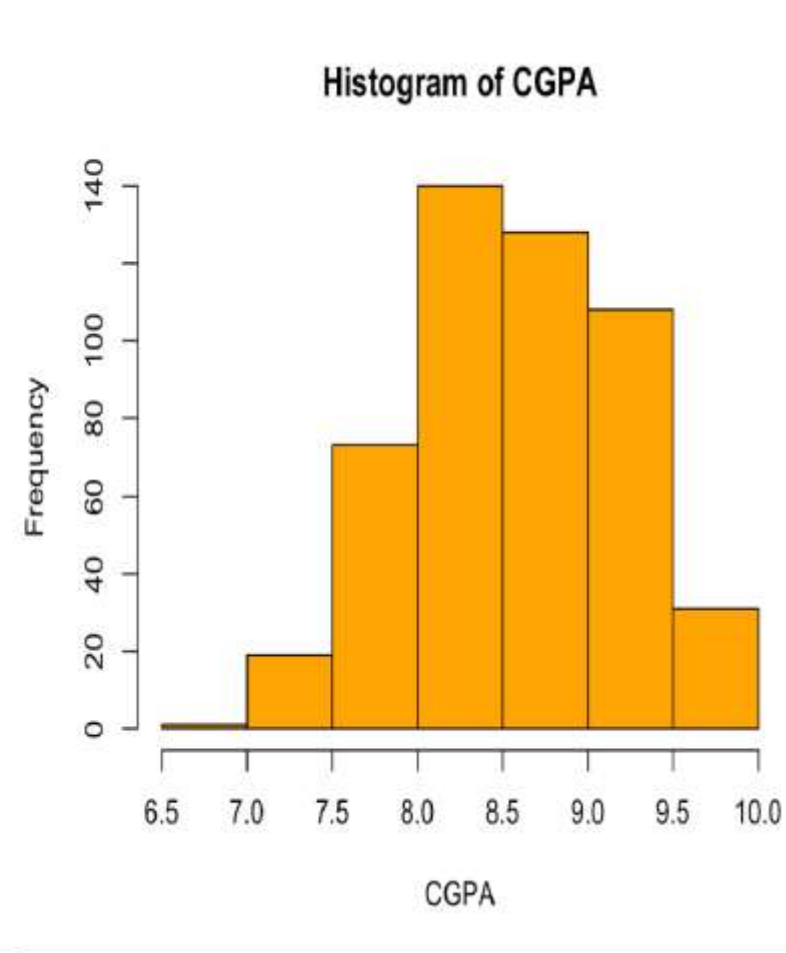




The predictor variables CGPA SOP, LOR, all appears normally distributed and are unimodal in shape.

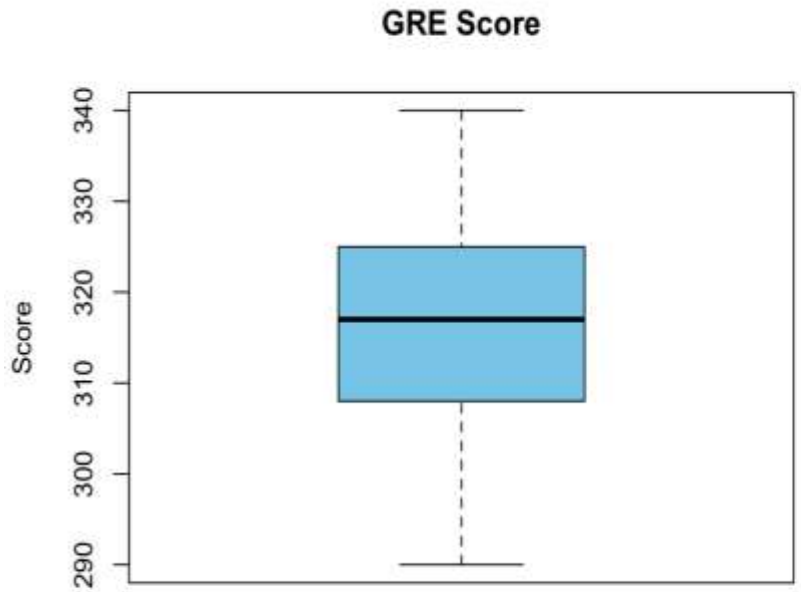


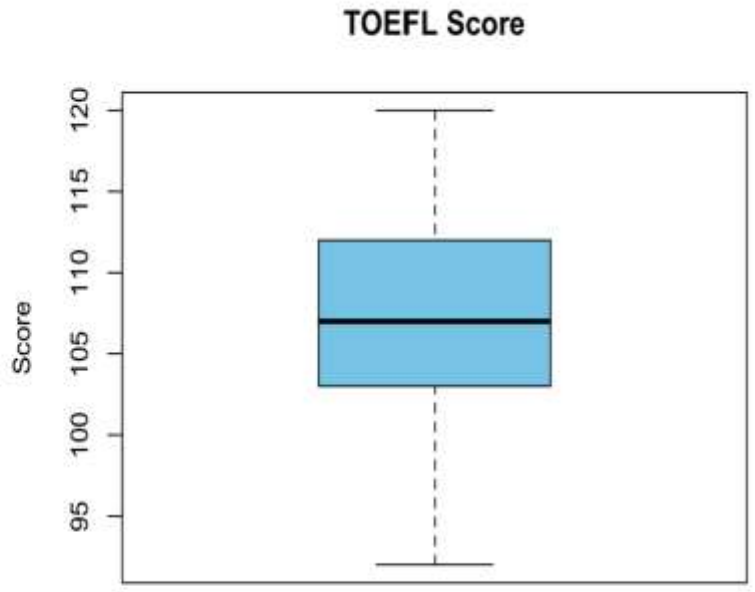


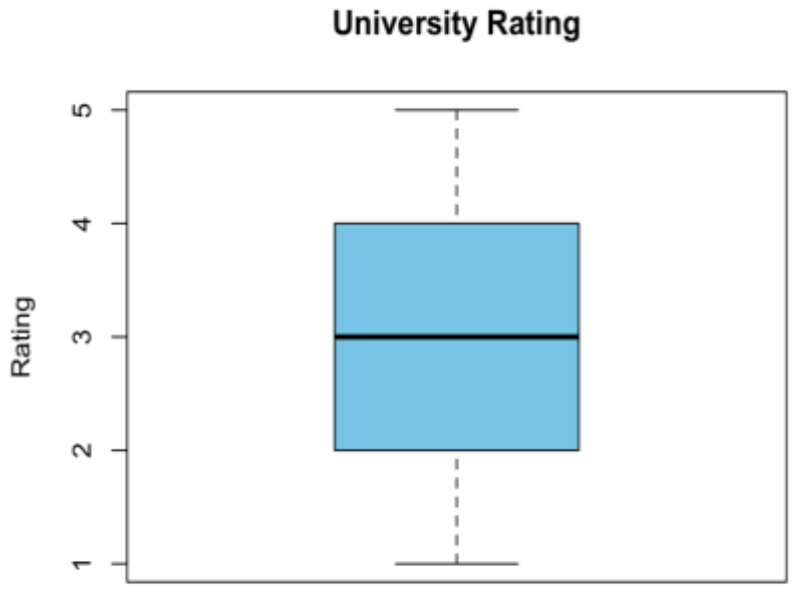


BOXPLOTS OF QUANTITATIVE VARIABLES

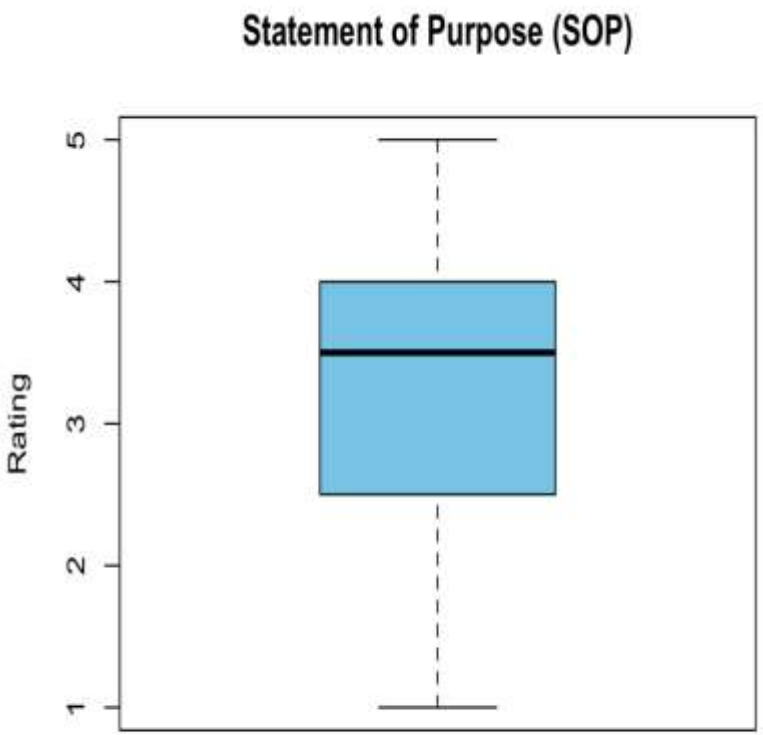
- The boxplot for GRE Score has no outliers, the boxplot of TOEFL Score doesn't have any outliers.
- However, with no outliers present in the plot University Rating the shape is symmetrical where the median is dividing the box in half.

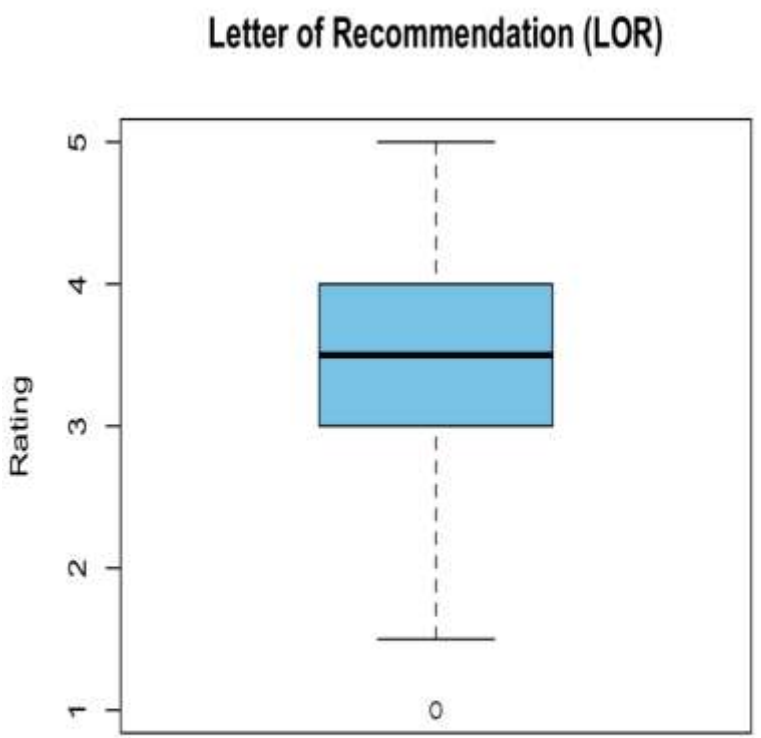


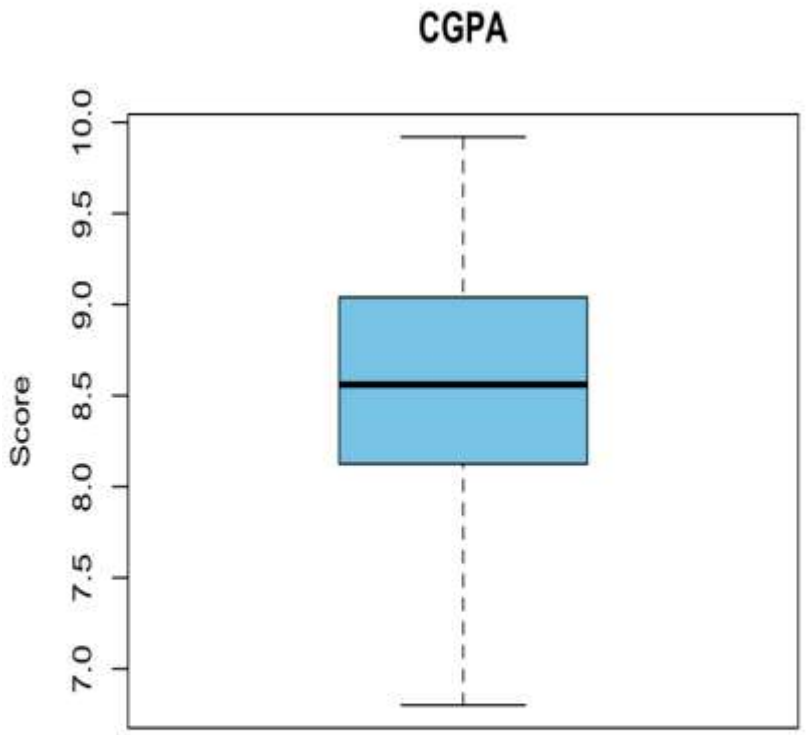




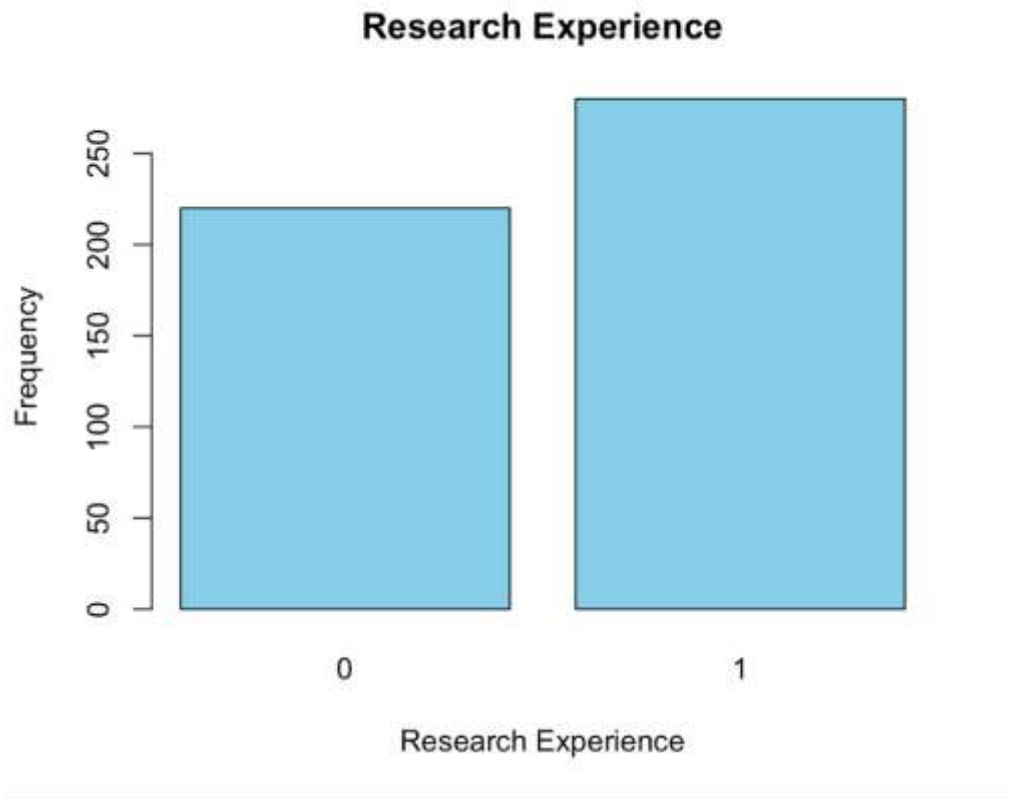
- Statement of Purpose(SOP) has no outliers.
- The boxplot of LOR is suggesting a normal distribution no outliers.
- Also, the boxplot of CGPA is suggesting a normal distribution no outliers.







BAR PLOT OF THE CATEGORICAL VARIABLE- RESEARCH EXPERIENCE

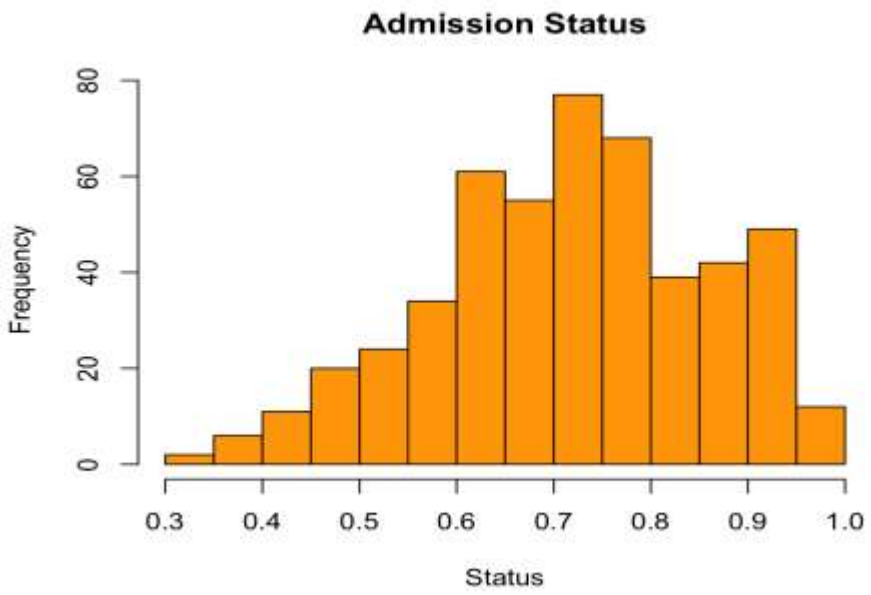


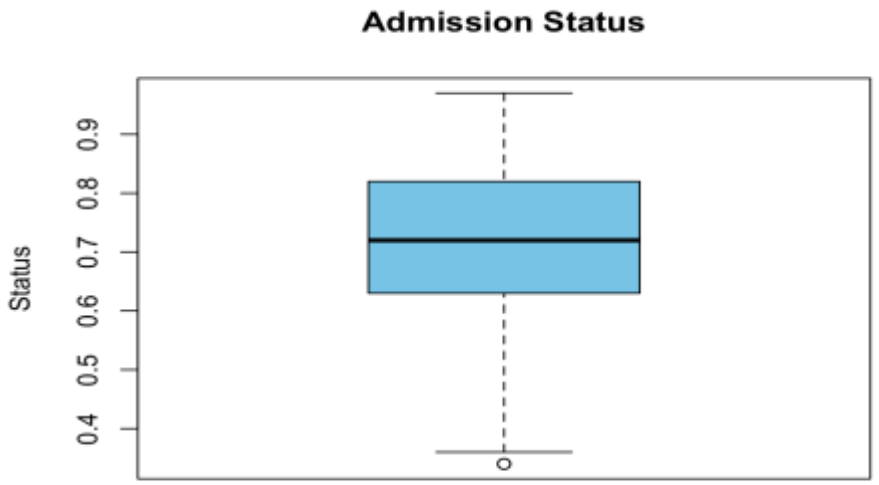

```
> table((adm$Research))

  0   1
220 280
>
```

OUTPUT VARIABLE ANALYSIS

Histogram & Boxplot of Admission Status

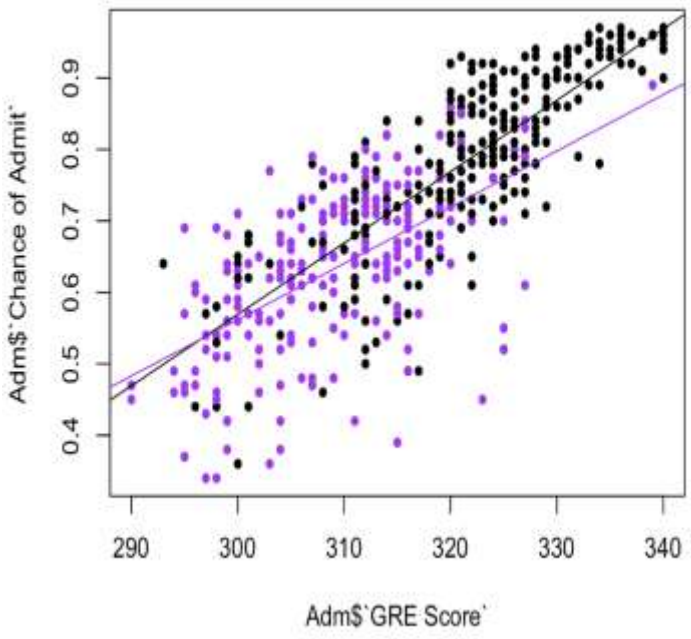


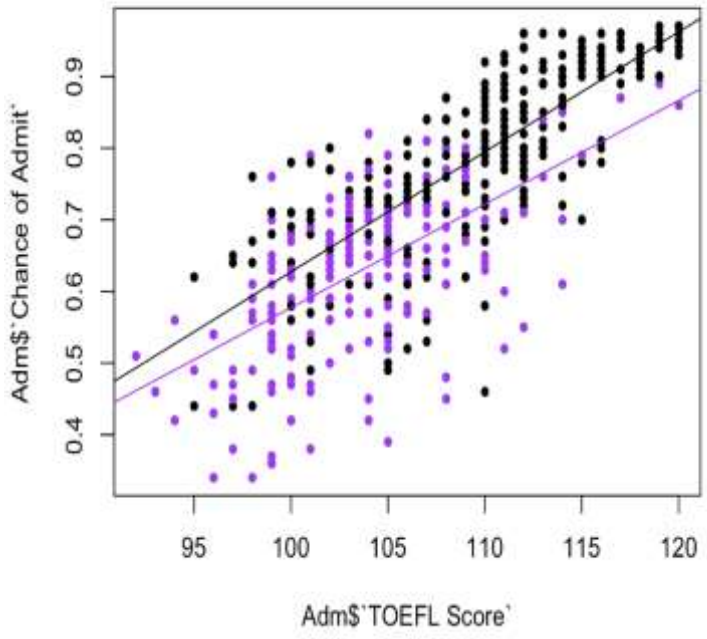


The histogram looks unimodal and skewed positively towards left as the mean is slightly greater than the median, while the boxplot for the response variable shows only one outlier.

SCATTERPLOT ANALYSIS

ADMIT CHANCE vs. GRE SCORE & TOEFL SCORE

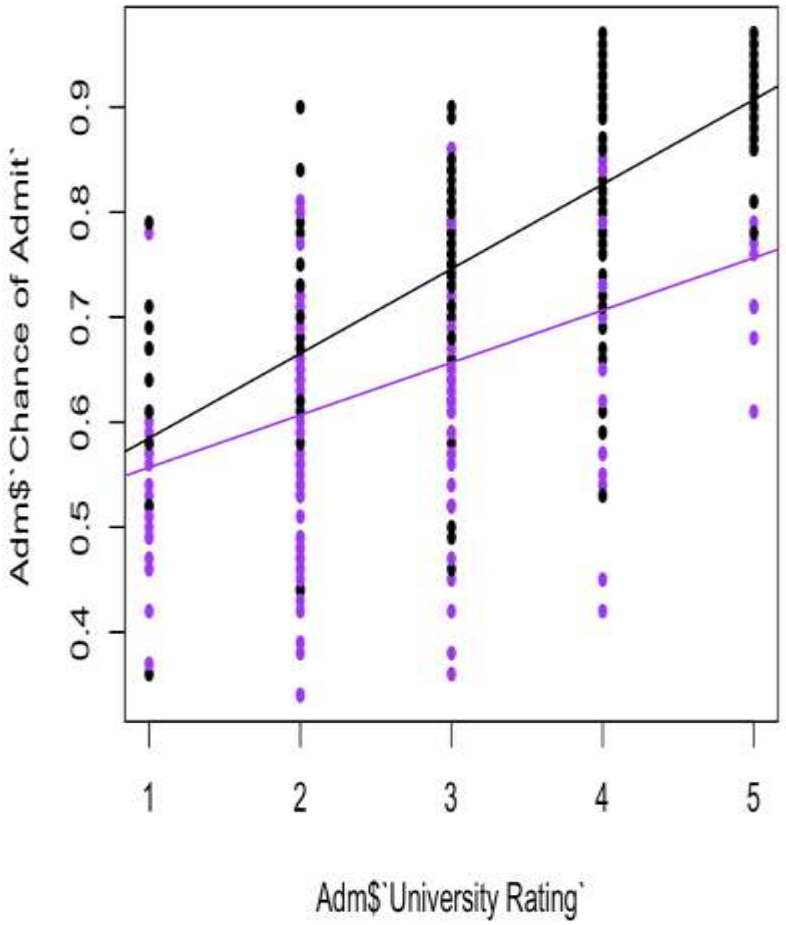


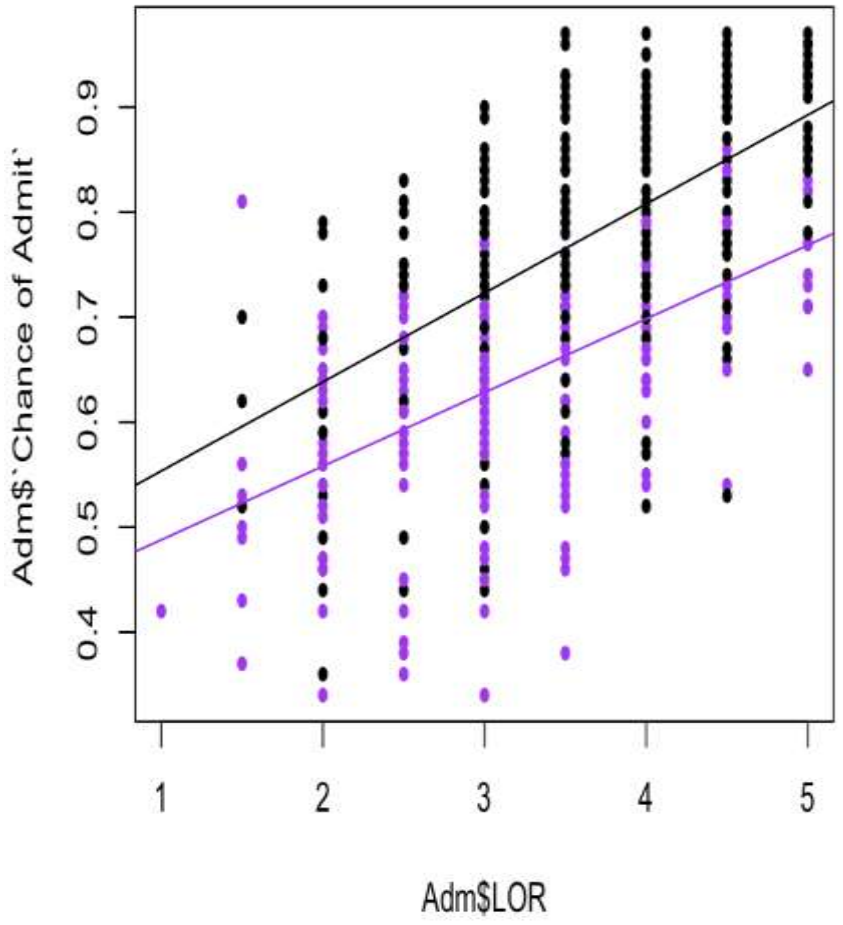


In the plot against GRE Scores and Chance of Admit, intersecting lines could suggest that there is a possible interaction between these two variables.

The variable TOEFL Score regression line is parallel with outcome variable(Chance of Admit), this can cannot have a potential interaction term including in the model.

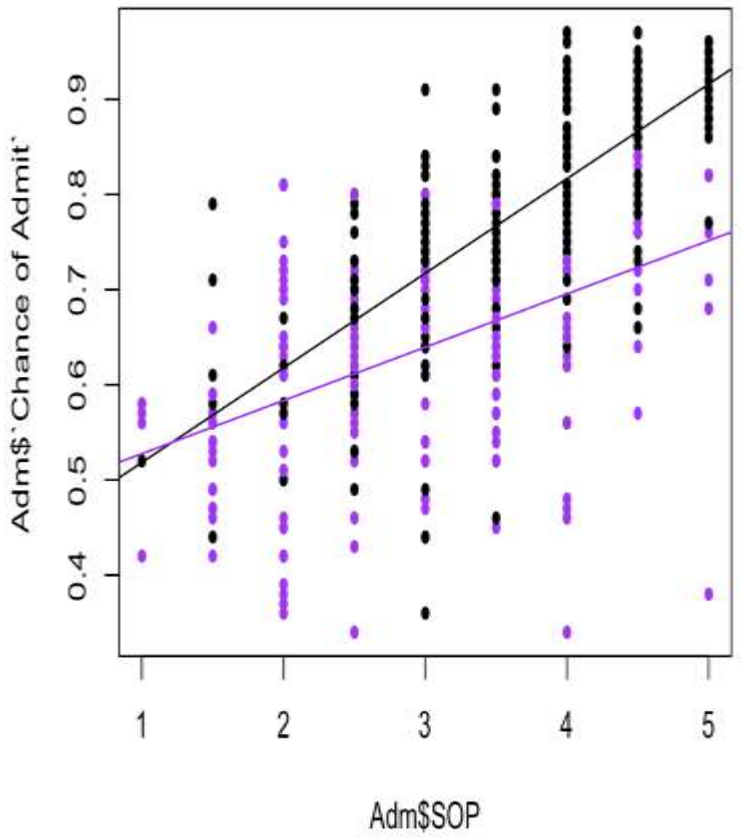
ADMIT CHANCE vs. UNIVERSITY RATING& LOR

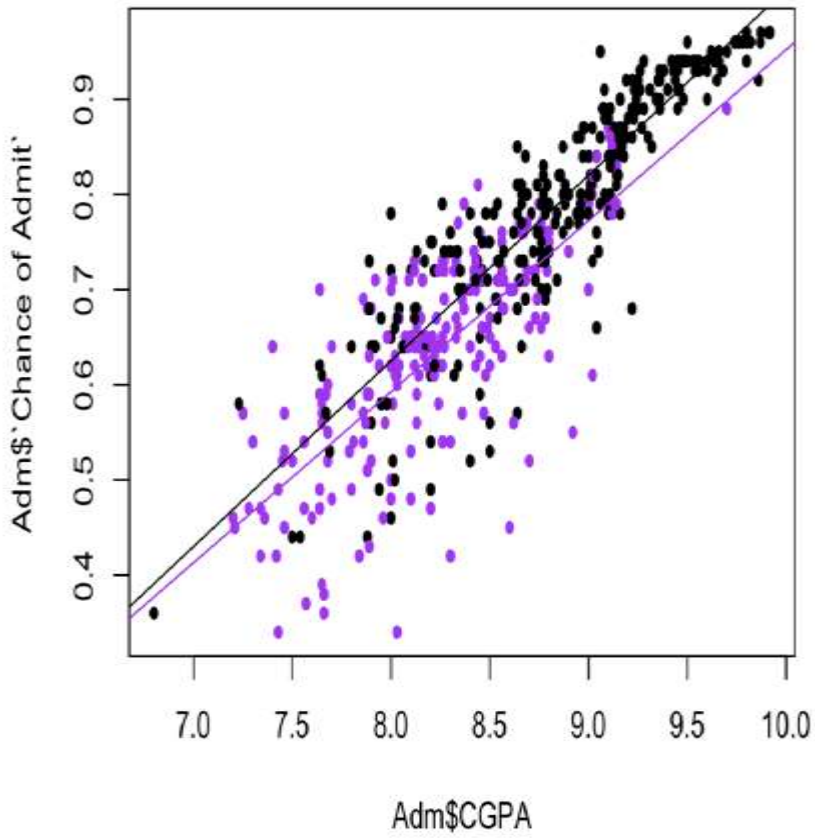




In both plots the predictor variables and outcome variable has no sign for interaction. As the regression lines are almost parallel to each other and are not intersecting.

ADMIT CHANCE vs. SOP & CGPA



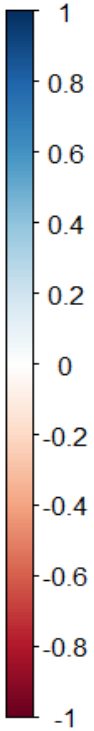


In the plot against SOP , Intersecting regression lines could suggest that there is a possible interaction effect between those two variables.

The variable CGPA line is almost parallel with outcome variable(Admit Chance),at every point this can cannot have a potential interaction term including in the model.

CORRELATION MATRIX

	GRE.Score	TOEFL.Score	University.Rating	SOP	LOR	CGPA	Research	Chance.of.Admit
GRE.Score	1.00	0.83	0.64	0.61	0.52	0.83	0.56	0.81
TOEFL.Score	0.83	1.00	0.65	0.64	0.54	0.81	0.47	0.79
University.Rating	0.64	0.65	1.00	0.73	0.61	0.71	0.43	0.69
SOP	0.61	0.64	0.73	1.00	0.66	0.71	0.41	0.68
LOR	0.52	0.54	0.61	0.66	1.00	0.64	0.37	0.65
CGPA	0.83	0.81	0.71	0.71	0.64	1.00	0.50	0.88
Research	0.56	0.47	0.43	0.41	0.37	0.50	1.00	0.55
Chance.of.Admit	0.81	0.79	0.69	0.68	0.65	0.88	0.55	1.00



The correlation matrix shows positive correlation between all the quantitative variables in the data set. We can see that some of the variables are highly correlated with one another. This suggests that there might be multicollinearity between them. The pairing to note are: {GRE, TOEFL}, {CGPA, TOEFL}.The variables GRE and TOEFL, are so highly correlated, that we have to remove one from the model. With the other pairing, one of the variables should be removed from the model and the effect on the Adjusted R should be noted. The dependent variable, Chance of Admit , is relatively highly correlated with the variable CGPA. One model must be tested that only uses CGPA to predict Admit Chance.

MODEL BUILDING USING LINEAR REGRESSION

```
> options(scipen=999)
> lm_model <- lm(`Chance of Admit` ~ `GRE Score` + `TOEFL Score` + `University Rating` + SOP + LOR + CGPA
+ Research, data = Adm)
> summary(lm_model)

Call:
lm(formula = `Chance of Admit` ~ `GRE Score` + `TOEFL Score` +
    `University Rating` + SOP + LOR + CGPA + Research, data = Adm)

Residuals:
      Min        1Q    Median        3Q       Max
-0.266657 -0.023327  0.009191  0.033714  0.156818

Coefficients:
                   Estimate Std. Error t value             Pr(>|t|)
(Intercept)       -1.2757251  0.1042962 -12.232 < 0.0000000000000002 ***
`GRE Score`        0.0018585  0.0005023   3.700             0.000240 ***
`TOEFL Score`      0.0027780  0.0008724   3.184             0.001544 **
`University Rating` 0.0059414  0.0038019   1.563             0.118753
SOP                0.0015861  0.0045627   0.348             0.728263
LOR                0.0168587  0.0041379   4.074            0.0000538 ***
CGPA               0.1183851  0.0097051  12.198 < 0.0000000000000002 ***
Research           0.0243075  0.0066057   3.680             0.000259 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05999 on 492 degrees of freedom
Multiple R-squared:  0.8219,    Adjusted R-squared:  0.8194
F-statistic: 324.4 on 7 and 492 DF,  p-value: < 0.00000000000000022
```

INITIAL MODEL

The Initial model is done by using all the variables in the dataset to predict Chance of Admit. The summary is shown on the left.

As we can see, the Adjusted-R-squared value is 81.94%. This means that approximately 82% of the variance seen in Chance of Admit can be explained by all the predictor variables.

Significant variables from this model are {GRE Score, TOEFL Score, LOR, CGPA, Research}.

```
> lm_model1 <- lm(`Chance of Admit` ~ `GRE Score` + `TOEFL Score`+ LOR + CGPA + Research, data = Adm)
> summary(lm_model1)

Call:
lm(formula = `Chance of Admit` ~ `GRE Score` + `TOEFL Score` +
    LOR + CGPA + Research, data = Adm)

Residuals:
      Min        1Q    Median        3Q       Max
-0.265965 -0.023835  0.008003  0.035543  0.158379

Coefficients:
               Estimate Std. Error t value            Pr(>|t|)
(Intercept)   -1.3357018  0.0990753 -13.482 < 0.0000000000000002 ***
`GRE Score`    0.0018892  0.0005024   3.760             0.000190 ***
`TOEFL Score`  0.0030174  0.0008619   3.501             0.000506 ***
LOR            0.0193203  0.0037939   5.092          0.000000504 ***
CGPA           0.1229798  0.0093018  13.221 < 0.0000000000000002 ***
Research       0.0251649  0.0065988   3.814             0.000154 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06007 on 494 degrees of freedom
Multiple R-squared:  0.8207,    Adjusted R-squared:  0.8188
F-statistic: 452.1 on 5 and 494 DF,  p-value: < 0.00000000000000022
```

MODEL – 1

In this model we conducted regression, only from the significant variables from the initial model. The summary is shown on the left.

As we can see, the Adjusted-R-squared value is 81.88%. This means that approximately 82% of the variance seen in Chance of Admit can be explained by the predictor/ independent variables.

The Adjusted R-squared value for this model is lower than the previous model which suggests that this model is not that good of an ideal model.

```
> lm_model2 <- lm(`Chance of Admit` ~ . - `GRE Score`, data = Adm)
> summary(lm_model2)

Call:
lm(formula = `Chance of Admit` ~ . - `GRE Score`, data = Adm)

Residuals:
     Min       1Q   Median       3Q      Max
-0.25240 -0.02560  0.01122  0.03571  0.16531

Coefficients:
                    Estimate Std. Error t value             Pr(>|t|)
(Intercept)       -0.9699361  0.0644277 -15.055 < 0.0000000000000002 ***
`TOEFL Score`      0.0042619  0.0007847   5.431          0.0000000879 ***
`University Rating` 0.0066299  0.0038459   1.724             0.085351 .
SOP                0.0010379  0.0046186   0.225             0.822286
LOR                0.0160269  0.0041846   3.830             0.000145 ***
CGPA               0.1326103  0.0090250  14.694 < 0.0000000000000002 ***
Research           0.0312936  0.0064111   4.881          0.0000014265 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06075 on 493 degrees of freedom
Multiple R-squared:  0.8169,    Adjusted R-squared:  0.8147
F-statistic: 366.7 on 6 and 493 DF,  p-value: < 0.00000000000000022
```

MODEL – 2

As from the correlation matrix, the GRE Score variable has a high positive correlation with the output variable. This might cause a multicollinearity issue.

In this model, we removed {GRE Score}. The summary is shown on the left.

The Adjusted-R-squared value is 81.47%. This means that around 81% of the variation seen in Admit Chance can be explained by the independent variables.

```
> lm_model3 <- lm(`Chance of Admit` ~ . - LOR, data = Adm)
> summary(lm_model3)

Call:
lm(formula = `Chance of Admit` ~ . - LOR, data = Adm)

Residuals:
      Min        1Q    Median        3Q       Max
-0.273891 -0.021844  0.008069  0.036529  0.143371

Coefficients:
                    Estimate Std. Error t value             Pr(>|t|)
(Intercept)        -1.2777149  0.1059323 -12.062 < 0.0000000000000002 ***
`GRE Score`         0.0017473  0.0005095   3.430             0.000655 ***
`TOEFL Score`       0.0027667  0.0008861   3.122             0.001900 **
`University Rating` 0.0081341  0.0038227   2.128             0.033843 *
SOP                 0.0071241  0.0044239   1.610             0.107954
CGPA                0.1266378  0.0096404  13.136 < 0.0000000000000002 ***
Research            0.0257946  0.0066992   3.850             0.000133 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06093 on 493 degrees of freedom
Multiple R-squared:  0.8159,    Adjusted R-squared:  0.8137
F-statistic: 364.1 on 6 and 493 DF,  p-value: < 0.00000000000000022
```

MODEL – 3

As from the plotted correlation matrix the LOR variable also has a high positive correlation with the output variable. This might cause a multicollinearity issue.

In this model that we removed the {LOR}. The summary is shown on the left.

The Adjusted-R-squared value is 81.37%. This means that around 81% of the variation seen in Admit Chance can be explained by the independent variables.

```
> lm_model4 <- lm(`Chance of Admit` ~ . - `GRE Score` - CGPA, data = Adm)
> summary(lm_model4)

Call:
lm(formula = `Chance of Admit` ~ . - `GRE Score` - CGPA, data = Adm)

Residuals:
      Min        1Q    Median        3Q       Max
-0.298349 -0.032011  0.009895  0.047825  0.209937

Coefficients:
                    Estimate Std. Error t value             Pr(>|t|)
(Intercept)        -0.661019   0.072955  -9.061 < 0.0000000000000002 ***
`TOEFL Score`       0.010821   0.000773  13.998 < 0.0000000000000002 ***
`University Rating` 0.016617   0.004535   3.664              0.000275 ***
SOP                 0.013553   0.005438   2.492              0.013018 *
LOR                 0.028580   0.004907   5.824        0.00000001035 ***
Research            0.046006   0.007586   6.065        0.00000000263 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07278 on 494 degrees of freedom
Multiple R-squared:  0.7368,    Adjusted R-squared:  0.7341
F-statistic: 276.6 on 5 and 494 DF,  p-value: < 0.00000000000000022
```

MODEL – 4

In this model that we removed both the highly correlated variables {GRE Score, CGPA}. As from the plotted correlation matrix these variables have a high positive correlation with the output variable. This might cause a multicollinearity.

The summary is shown on the left.

The Adjusted-R-squared value is 73.41%. This means that around 73% of the variation seen in Admit Chance can be explained by the predictor variables, the adjusted R squared value is the lowest of all the previous models.

INTERACTION TERMS

```
> lm_model5 <- lm(`Chance of Admit` ~ . +`GRE Score`: CGPA, data = Adm)
> summary(lm_model5)

Call:
lm(formula = `Chance of Admit` ~ . + `GRE Score`:CGPA, data = Adm)

Residuals:
      Min        1Q    Median        3Q       Max
-0.268664 -0.022666  0.008829  0.032882  0.160801

Coefficients:
                    Estimate Std. Error t value  Pr(>|t|)
(Intercept)       -2.4629057  1.0129346  -2.431  0.015395 *
`GRE Score`        0.0055754  0.0031942   1.745  0.081530 .
`TOEFL Score`      0.0028909  0.0008773   3.295  0.001055 **
`University Rating` 0.0063399  0.0038154   1.662  0.097215 .
SOP                0.0013899  0.0045639   0.305  0.760837
LOR                0.0166499  0.0041400   4.022 0.0000669 ***
CGPA               0.2577994  0.1187174   2.172  0.030369 *
Research           0.0245834  0.0066073   3.721  0.000222 ***
`GRE Score`:CGPA  -0.0004401  0.0003735  -1.178  0.239256
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05996 on 491 degrees of freedom
Multiple R-squared:  0.8224,    Adjusted R-squared:  0.8195
F-statistic: 284.2 on 8 and 491 DF,  p-value: < 0.00000000000000022
```

MODEL – 5

In this model, we included interaction terms which we found from the scatterplots. {GRE Score*CGPA} is one of the interaction term.

The summary is shown on the left.

The Adjusted-R-squared value is 81.95%. This means that approximately 82% of the variation seen in Admit Chance can be explained by the predictor variables, the value is higher for the first interaction term compared to the previous models.

```
> lm_model6 <- lm(`Chance of Admit` ~ . +`GRE Score`: SOP, data = Adm)
> summary(lm_model6)

Call:
lm(formula = `Chance of Admit` ~ . + `GRE Score`:SOP, data = Adm)

Residuals:
      Min        1Q    Median        3Q       Max
-0.265900 -0.023364  0.008881  0.033959  0.154983

Coefficients:
                    Estimate Std. Error t value             Pr(>|t|)
(Intercept)       -1.0945756  0.2803411  -3.904             0.000108 ***
`GRE Score`        0.0013052  0.0009404   1.388             0.165783
`TOEFL Score`      0.0027369  0.0008749   3.128             0.001862 **
`University Rating` 0.0056385  0.0038286   1.473             0.141465
SOP               -0.0516635  0.0766219  -0.674             0.500459
LOR                0.0169237  0.0041411   4.087            0.0000511 ***
CGPA               0.1179120  0.0097340  12.113 < 0.0000000000000002 ***
Research           0.0241341  0.0066139   3.649             0.000291 ***
`GRE Score`:SOP    0.0001702  0.0002445   0.696             0.486631
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06002 on 491 degrees of freedom
Multiple R-squared:  0.8221,    Adjusted R-squared:  0.8192
F-statistic: 283.6 on 8 and 491 DF,  p-value: < 0.00000000000000022
```

```
> lm_model7 <- lm(`Chance of Admit` ~ . + `GRE Score` * CGPA + `GRE Score` * SOP, data = Adm)
> summary(lm_model7)

Call:
lm(formula = `Chance of Admit` ~ . + `GRE Score` * CGPA + `GRE Score` *
    SOP, data = Adm)

Residuals:
      Min        1Q    Median        3Q       Max
-0.269193 -0.023627  0.008658  0.033306  0.159586

Coefficients:
                    Estimate Std. Error t value  Pr(>|t|)
(Intercept)       -4.0528595  1.2243484  -3.310  0.001001 **
`GRE Score`        0.0106196  0.0038685   2.745  0.006270 **
`TOEFL Score`      0.0029280  0.0008737   3.351  0.000867 ***
`University Rating` 0.0057139  0.0038088   1.500  0.134213
SOP               -0.2542289  0.1116857  -2.276  0.023259 *
LOR                0.0165287  0.0041226   4.009 0.0000704 ***
CGPA               0.5442015  0.1720638   3.163  0.001660 **
Research           0.0243235  0.0065799   3.697  0.000243 ***
`GRE Score`:CGPA  -0.0013513  0.0005445  -2.481  0.013420 *
`GRE Score`:SOP    0.0008157  0.0003561   2.291  0.022409 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0597 on 490 degrees of freedom
Multiple R-squared:  0.8243,    Adjusted R-squared:  0.8211
F-statistic: 255.4 on 9 and 490 DF,  p-value: < 0.00000000000000022
```

MODEL – 6

In this model, we included the interaction term found from the scatterplots which is {GRE*SOP}.

The Adjusted-R-squared value is 81.92%. This means that approximately 82% of the variation seen in Admit Chance can be explained from the predictor/independent variables, the value of Adjusted R squared has a minute difference from the previous interaction model(5).

MODEL – 7

In this model we included both the interaction terms of {GRE*SOP}& {GRE*CGPA}.

The Adjusted-R-squared value is 82.11%. This means that above 82% of the variation seen in Admit Chance can be explained from the independent/predictor variables, the adjusted R-squared of this model has the highest value compared to all previous models. So, we take this as the ideal model.

LOGISTIC REGRESSION

```
> Adm$Research <- as.factor(Adm$Research)
> logit_model <- glm(Research ~ ., data = Adm, family = binomial)
> summary(logit_model)

Call:
glm(formula = Research ~ ., family = binomial, data = Adm)

Coefficients:
                    Estimate Std. Error z value       Pr(>|z|)
(Intercept)        -30.09522    4.96333  -6.064 0.00000000133 ***
`GRE Score`          0.10756    0.02091   5.144 0.00000026883 ***
`TOEFL Score`       -0.04442    0.03643  -1.219       0.222759
`University Rating`  0.14981    0.15932   0.940       0.347043
SOP                  0.04753    0.18990   0.250       0.802361
LOR                  0.10735    0.17060   0.629       0.529191
CGPA                -0.49256    0.44927  -1.096       0.272924
`Chance of Admit`    6.10557    1.82023   3.354       0.000796 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 685.93  on 499  degrees of freedom
Residual deviance: 480.51  on 492  degrees of freedom
AIC: 496.51

Number of Fisher Scoring iterations: 5
```

```
> cbind(odds_ratios,conf_int)
                                odds_ratios                          2.5 %                 97.5 %
(Intercept)          0.0000000000008507739  0.00000000000000003981327     0.00000000117953
`GRE Score`          1.1135575415352143569  1.06971672704620490712557     1.16129525285279
`TOEFL Score`        0.9565531431475525758  0.88980552975942672766507     1.02676271845636
`University Rating`  1.1616172110499334291  0.84828253915824669562084     1.58656488131676
SOP                  1.0486777263865540987  0.72214977271230962330150     1.52289166332310
LOR                  1.1133257103880604965  0.79621507063512397817106     1.55639857067505
CGPA                 0.6110582803542766727  0.25108177258858221936321     1.46738744915907
`Chance of Admit`  448.3472628941013340409 13.44649718705661811668505 17430.63383887201417
```

- Intercept: The intercept is -30.09522.When all other predictors are zero, the odds of having a research experience is approximately -30.09522.

- GRE Score: The OR for GRE Score is 1.11355. For each one-unit increase in GRE Score, the odds of having research experience increases by approximately 135%, holding all other predictors constant.

- TOEFL Score: The OR for TOEFL Score is 0.95655. For each one-unit increase in TOEFL Score, the odds of having research experience decreases by approximately 4%, holding all other predictors constant. However, this coefficient is not statistically significant at the 0.05 level.

- University Rating: The OR for University Rating is 1.16161.For each one-unit increase in University Rating, the odds of having research experience increases by approximately 161%, holding all other predictors constant. However, this coefficient is not statistically significant at the 0.05 level.

- SOP (Statement of Purpose): The OR for SOP is 1.04867.For each one-unit increase in SOP, the odds of having research experience increases by approximately 48%, holding all other predictors constant.

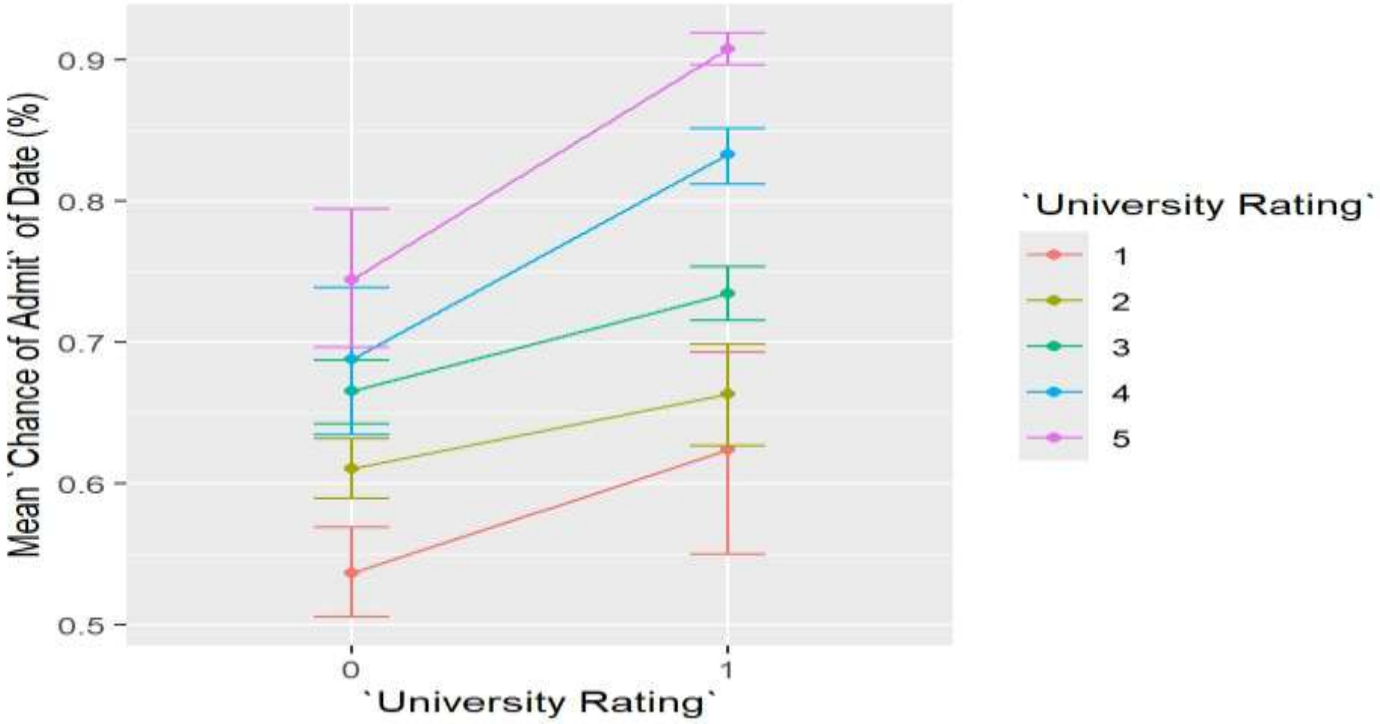
- LOR (Letter of Recommendation): The OR for LOR is 1.11332..For each one-unit increase in LOR, the odds of having research experience increases by approximately 113%, holding all other predictors constant.

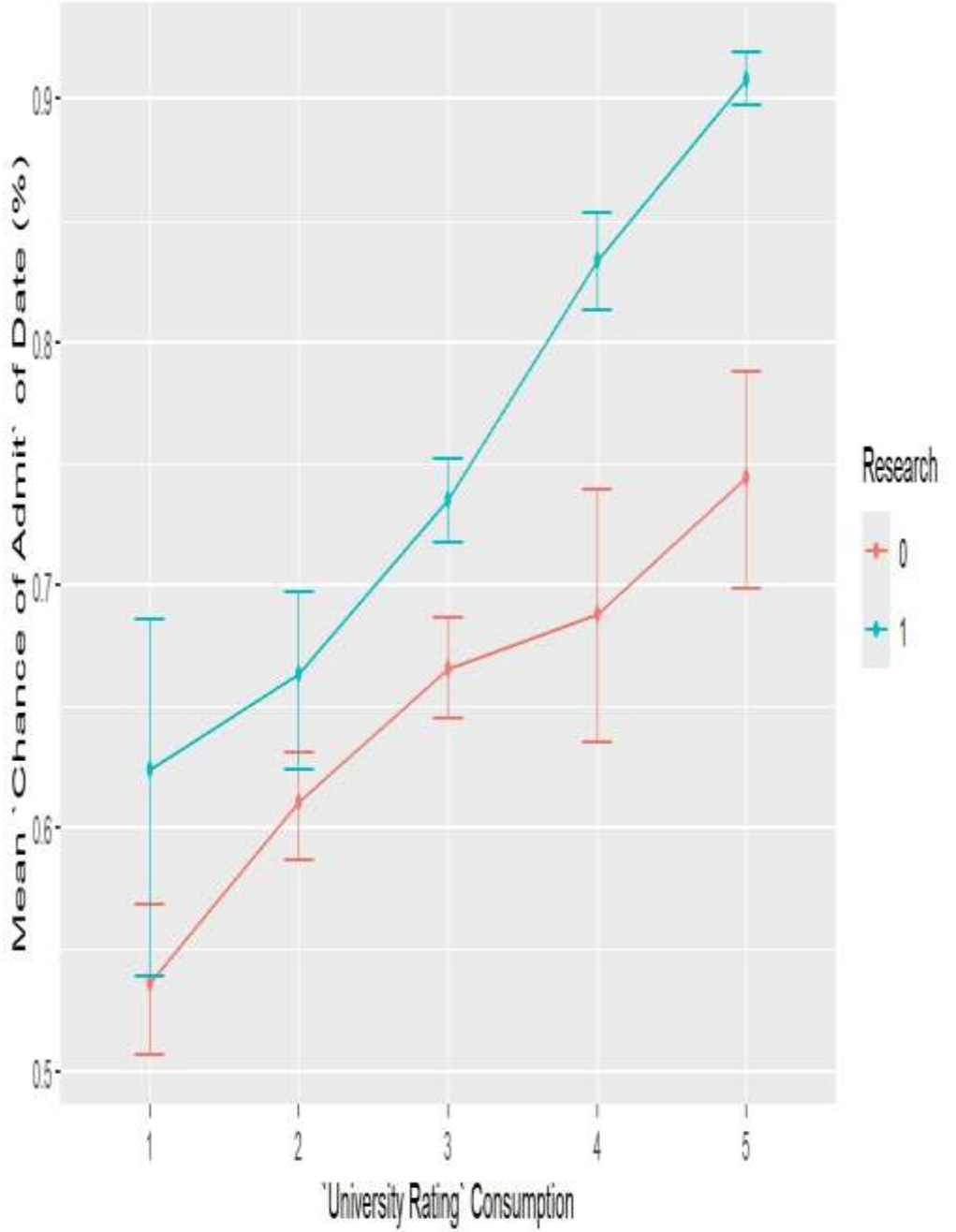
- CGPA: The OR for CGPA is 0.61105. For each one-unit increase in CGPA, the log-odds of having research experience decreases by approximately 39%, holding all other predictors constant.

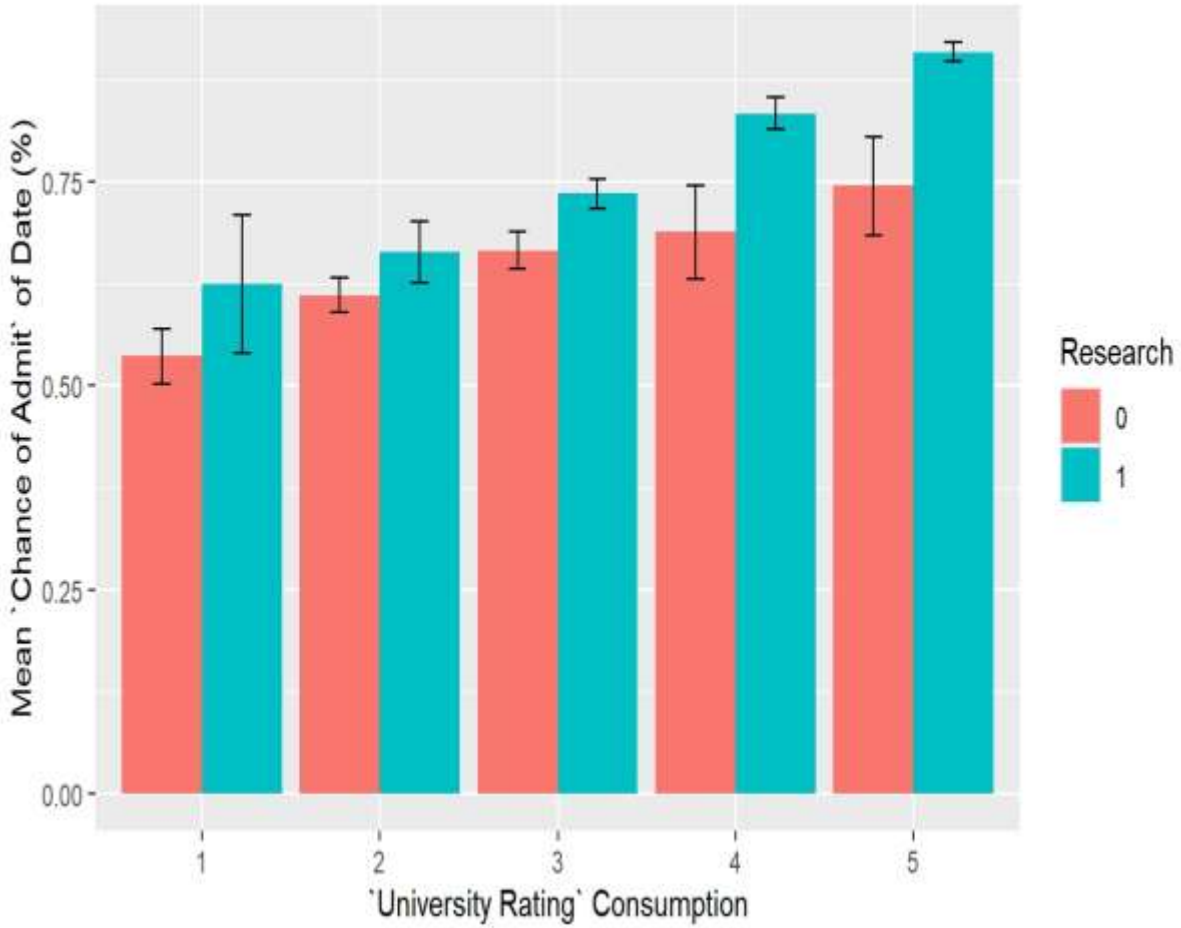
- Chance of Admit: The OR for Chance of Admit is 448.34726. For each one-unit increase in the Chance of Admit, the log-odds of having research experience increases by approximately 448%, holding all other predictors constant.

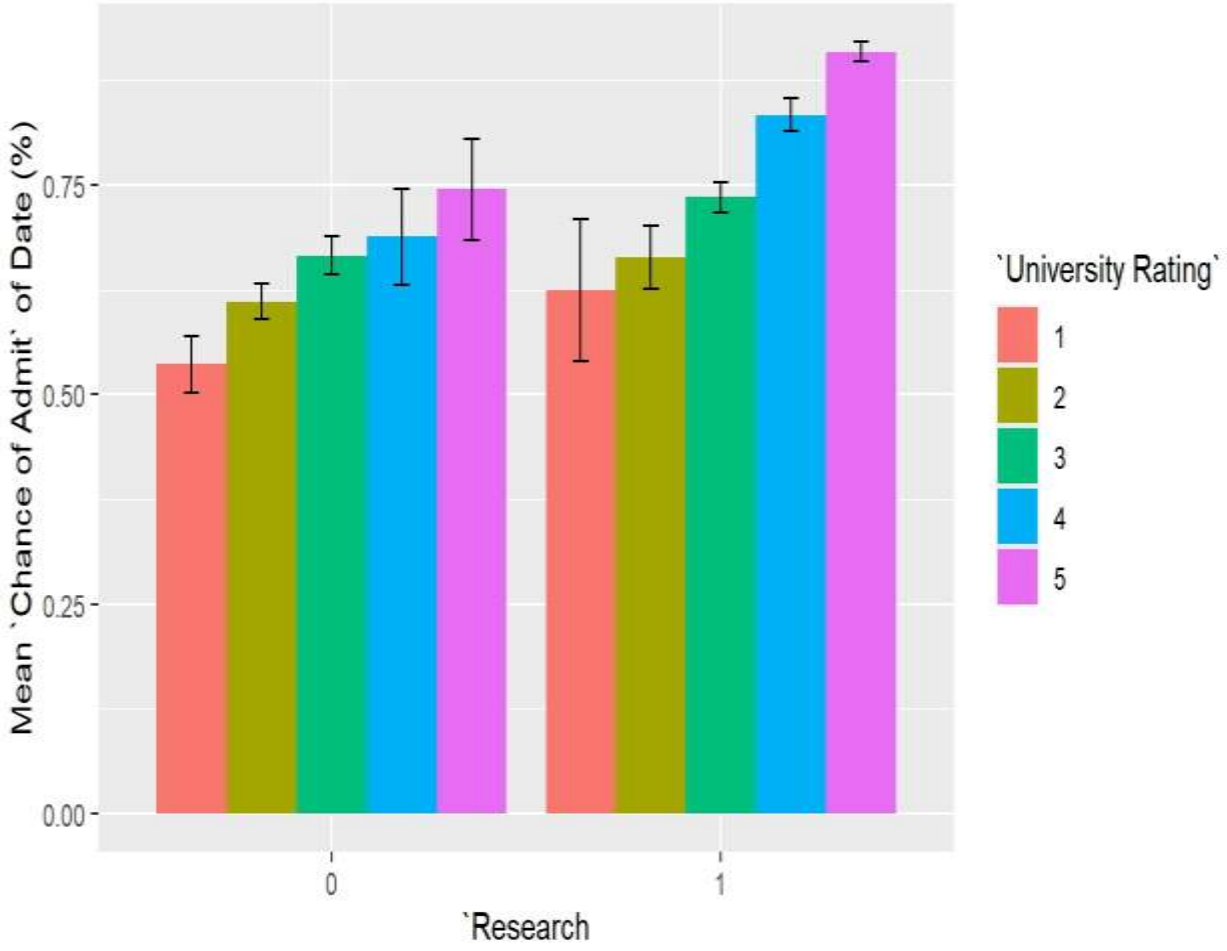
ANOVA

The main purpose of factorial ANOVA is to analyze the effects of two or more independent variables (factors) on a continuous dependent variable. It allows researchers to investigate how different combinations of levels of the independent variables influence the outcome variable









This output suggests that both the "Research" variable and the "University Rating" variable have significant independent effects on the dependent variable, and there is also a significant interaction effect between these two variables.

After analyzing bar graph, we can see that who has research experience chances increases significantly for admission in high rating university(4&5). Also, Chance for admit variable is influenced by university rating variable.

FINAL MODEL

Final model - 0.011*'GRE Score'+ '0.003*TOEFL Score' + 0.006*'University Rating'+ 0.544* CGPA- 0.254*SOP+0.017*LOR+ 0.024*Research –0.001 'GRE Score' : CGPA +0.001 'GRE Score' : SOP

The above model is the ideal model to predict the admission rate from all the predictor variables along with some interaction terms.The factors {GRE Score, TOEFL Score, University Rating, CGPA, SOP, LOR, and Research} are significant predictors of admission probability.

The inclusion of interaction terms suggests that the impact of one variable may depend on the value of another variable. For example, the effect of GRE Score on the outcome may vary depending on the applicant's SOP or CGPA.

Overall, this model aims to provide a quantitative way to predict the outcome (admission chances, perhaps) based on various factors commonly associated with the application process.

CONCLUSION

In this project, we aimed to develop a predictive model to estimate admission chances for prospective students based on various factors commonly considered during the application process. We employed a dataset containing information such as {GRE scores, TOEFL scores, University Ratings, CGPA, statements of purpose (SOP), letters of recommendation (LOR), and Research Experience}.

After thorough exploration we proceeded to build a predictive model using linear regression.Our final model revealed several key insights:

Significant Predictors: Among the predictors, CGPA emerged as the most influential factor in determining admission chances. Higher CGPA was strongly associated with higher predicted admission probabilities. Additionally, research experience and university ratings also played significant roles in predicting admission outcomes.

Variable Impact: While CGPA had the highest positive impact, SOP had a negative impact on admission chances, indicating the importance of a well-crafted statement of purpose.

Model Performance: The final model demonstrated satisfactory performance in predicting admission probabilities, as evidenced by metrics such as adjusted R-squared.

Practical Implications: Our findings suggest that applicants should focus on maintaining a high CGPA and obtaining research experience to enhance their admission prospects. Furthermore, attention to crafting strong SOPs is crucial, as they significantly influence admission decisions.

Limitations and Future Directions: Despite the model's predictive capability, it's essential to acknowledge its limitations, such as potential biases in the dataset and the inherent complexity of human decision-making in admissions. Future research could explore the inclusion of additional factors or employ more advanced modeling techniques to further improve predictive accuracy.