Assessment on Statistical analysis for data science(C7081)

22302900

2022-11-23

GitHub link: [<https://github.com/kksam2705/my-project.git>]

# Background

The data set contains several parameters which are considered important during the application for Master’s Programs.

The important parameters are GRE score(out of 340), TOEFL score(out of 120), university rating(out of 5), SOP (Statement of Purpose)(out of 5), LOR (Letter of Recommendation)(out of 5), CGPA(out of 10), Research(0 or 1), and a chance of admission(out of 10).

In this data set, 400 entries are included with 8 variables.

# Objective

we are going to Analyze which are all factors that are important for getting Admission to the university.

# Source

This data set was downloaded from Kaggle. <https://www.kaggle.com/datasets/mohansacharya/graduate-admissions>

# Methods

# 1. Simple Linear Regression

Simple linear regression is used to estimate the relationship between two quantitative variables. How strong the relationship is between two variables.

# 2. Multiple Regression

Multiple regression is an extension of linear regression into a relationship between more than two variables. In simple linear relation, we have one predictor and one response variable, but in multiple regression, we have more than one predictor variable and one response variable.

# 3. Decision Tree

A Decision tree is a graph to represent choices and their results in form of a tree. The nodes in the graph represent an event or choice and the edges of the graph represent the decision rules or conditions.

# 4. Random Forest

In the random forest approach, a large number of decision trees are created. Every observation is fed into every decision tree. The most common outcome for each observation is used as the final output. A new observation is fed into all the trees and takes a majority vote for each classification model.

# 5. Logistic Regression

Logistic Regression is a regression model in which the response variable (dependent variable) has categorical values such as True/False or 0/1. It actually measures the probability of a binary response as the value of the response variable based on the mathematical equation relating to the predictor variables.

# Setup

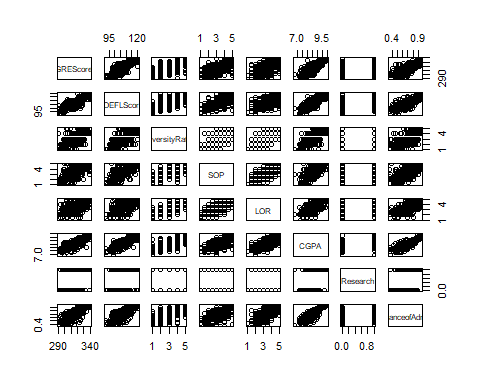
setwd("C:/Users/DELL/Downloads/admission into university")  
library(readr)  
adm\_data <- read\_csv("adm\_data.csv")

## Rows: 400 Columns: 8  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (8): GREScore, TOEFLScore, UniversityRating, SOP, LOR, CGPA, Research, C...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

names(adm\_data)

## [1] "GREScore" "TOEFLScore" "UniversityRating" "SOP"   
## [5] "LOR" "CGPA" "Research" "ChanceofAdmit"

plot(adm\_data)



GRE Score, TOEFL Score, and CGPA are important because they are highly correlated.

# Summary of all variables

summary(adm\_data)

## GREScore TOEFLScore UniversityRating SOP LOR   
## Min. :290.0 Min. : 92.0 Min. :1.000 Min. :1.0 Min. :1.000   
## 1st Qu.:308.0 1st Qu.:103.0 1st Qu.:2.000 1st Qu.:2.5 1st Qu.:3.000   
## Median :317.0 Median :107.0 Median :3.000 Median :3.5 Median :3.500   
## Mean :316.8 Mean :107.4 Mean :3.087 Mean :3.4 Mean :3.453   
## 3rd Qu.:325.0 3rd Qu.:112.0 3rd Qu.:4.000 3rd Qu.:4.0 3rd Qu.:4.000   
## Max. :340.0 Max. :120.0 Max. :5.000 Max. :5.0 Max. :5.000   
## CGPA Research ChanceofAdmit   
## Min. :6.800 Min. :0.0000 Min. :0.3400   
## 1st Qu.:8.170 1st Qu.:0.0000 1st Qu.:0.6400   
## Median :8.610 Median :1.0000 Median :0.7300   
## Mean :8.599 Mean :0.5475 Mean :0.7244   
## 3rd Qu.:9.062 3rd Qu.:1.0000 3rd Qu.:0.8300   
## Max. :9.920 Max. :1.0000 Max. :0.9700

Checking the Average score, minimum score, and maximum score of all variables.

## Linear Regression

adm.lm <- lm(ChanceofAdmit ~ CGPA, data = adm\_data)  
adm.lm

##   
## Call:  
## lm(formula = ChanceofAdmit ~ CGPA, data = adm\_data)  
##   
## Coefficients:  
## (Intercept) CGPA   
## -1.0715 0.2088

Make Liner regression with Chance of Admit as an independent variable and CGPA as a dependent variable.

# summary

summary(adm.lm)

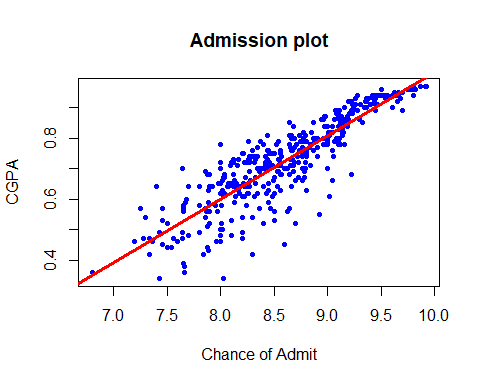
##   
## Call:  
## lm(formula = ChanceofAdmit ~ CGPA, data = adm\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.274575 -0.030084 0.009443 0.041954 0.180734   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.07151 0.05034 -21.29 <2e-16 \*\*\*  
## CGPA 0.20885 0.00584 35.76 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06957 on 398 degrees of freedom  
## Multiple R-squared: 0.7626, Adjusted R-squared: 0.762   
## F-statistic: 1279 on 1 and 398 DF, p-value: < 2.2e-16

The estimates (Estimate) for the model parameters – the value of the y-intercept (in this case -1.071) and the estimated effect of Chance of Admit on CGPA (0.208).

The final three lines are model diagnostics – the most important thing to note is the p-value (here it is 2.2e-16, or almost zero), which will indicate whether the model fits the data well.

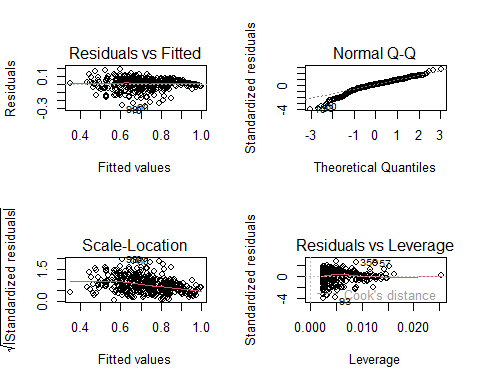
# Plot the data

plot(  
 ChanceofAdmit ~ CGPA,  
 data = adm\_data,  
 xlab = "Chance of Admit",  
 ylab = "CGPA",  
 main = "Admission plot",  
 col = "blue",  
 lwd = 1,  
 pch = 20  
)  
  
# Add regression line  
abline(reg = adm.lm, col = "red", lwd = 3)



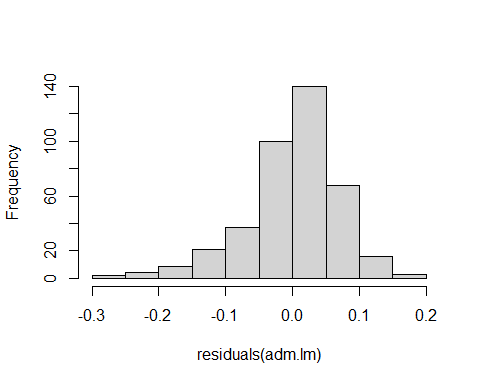
# Diagnostic plots

par(mfrow = c(2, 2))  
plot(adm.lm)



# Residual distribution

par(mfrow = c(1, 1))  
hist(residuals(adm.lm), main = "")



The most important thing to look for is that the red lines representing the mean of the residuals are all basically horizontal and centered around zero. This means there are no outliers.

In the Normal Q-Q plot in the top right, we can see that the real residuals from our model form an almost perfectly one-to-one line with the theoretical residuals from a perfect model.

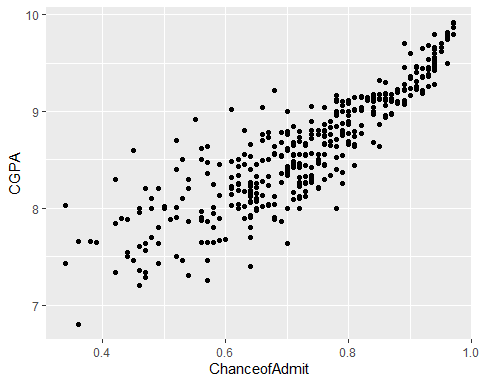
Based on these residuals, we can say that our model meets the assumption of homoscedasticity.

# Visualize the results with Graph

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.2

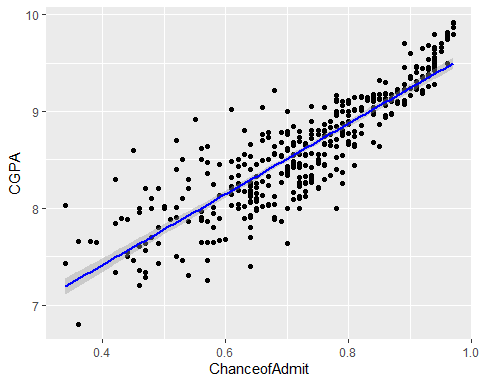
adm.graph <-  
 ggplot(adm\_data, aes(x = ChanceofAdmit, y = CGPA)) + geom\_point()  
adm.graph



# Add regression line

adm.graph <- adm.graph + geom\_smooth(method = "lm", col = "blue")  
adm.graph

## `geom\_smooth()` using formula = 'y ~ x'



# Conclusion

From these results, we can say that there is a significant positive relationship between the Chance of Admit and CGPA (p-value < 0.001), with a 0.208-unit (+/- 0.01) increase in Chance of Admit for every unit increase in CGPA.

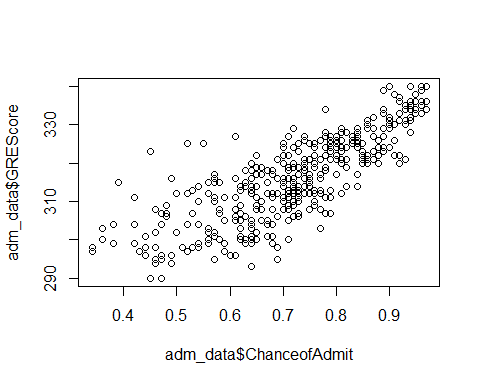
Since there is a significant positive relationship between the Chance of Admit and CGPA so, we can reject the null hypothesis.

## Multiple regression

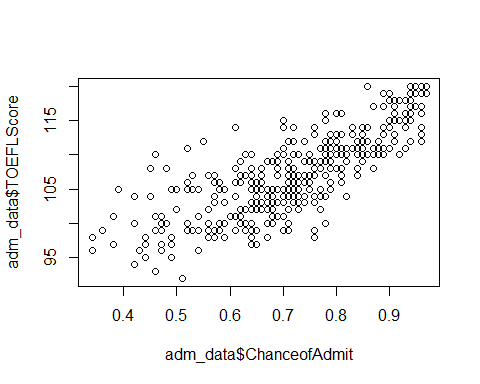
Check for linearity.

We can check this using two scatter plots: one for GRE Score and Chance of admission, and one for TOEFL Score and Chance of admission.

plot(adm\_data$ChanceofAdmit, adm\_data$GREScore)



plot(adm\_data$ChanceofAdmit, adm\_data$TOEFLScore)



adm.multi <-  
 lm(ChanceofAdmit ~ GREScore + TOEFLScore, data = adm\_data)  
adm.multi

##   
## Call:  
## lm(formula = ChanceofAdmit ~ GREScore + TOEFLScore, data = adm\_data)  
##   
## Coefficients:  
## (Intercept) GREScore TOEFLScore   
## -2.128397 0.005814 0.009412

# summary

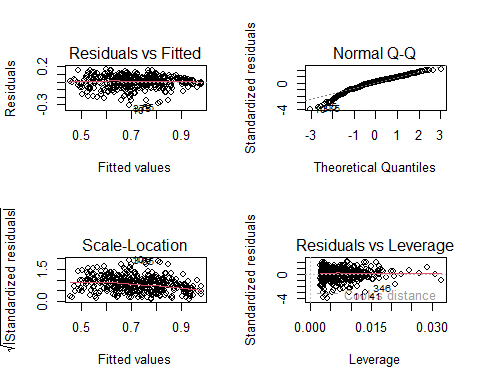
summary(adm.multi)

##   
## Call:  
## lm(formula = ChanceofAdmit ~ GREScore + TOEFLScore, data = adm\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.31590 -0.03886 0.01252 0.04963 0.16019   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.1283967 0.1163739 -18.289 < 2e-16 \*\*\*  
## GREScore 0.0058137 0.0006304 9.223 < 2e-16 \*\*\*  
## TOEFLScore 0.0094119 0.0011916 7.899 2.8e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.07928 on 397 degrees of freedom  
## Multiple R-squared: 0.6925, Adjusted R-squared: 0.691   
## F-statistic: 447 on 2 and 397 DF, p-value: < 2.2e-16

The estimated value of chance of admission on the GRE Score is 0.0058, while the estimated value of the TOEFL Score is 0.0094.

# Check for homoscedasticity

par(mfrow = c(2, 2))  
plot(adm.multi)



par(mfrow = c(1, 1))

# Create the Equation for a Regression model

# y = a + b1x1 + b2x2 +…bnxn

# Apply Equation for predicting New Values

we can predict the chance of Admission when we set new values to GRE Score and TOEFL Score.

Y = -2.128+(0.005814)\*325+(0.009412)\*105  
Y

## [1] 0.74981

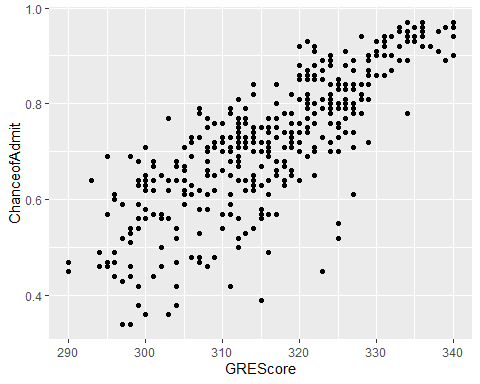
# Conclusion

If the person with GRE Score is 325 and TOEFL Score is 105 and his chance of Admit is 0.74%.

Since there is a positive relationship between the GRE Score, TOEFL Score, and the Chance of admitting so, we can reject the null hypothesis.

# Visualize the results with graph

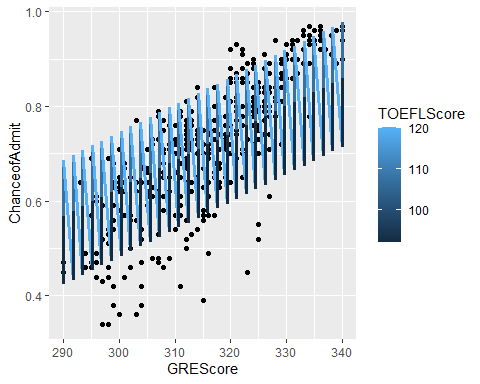
plotting.data <- expand.grid(  
 GREScore = seq(min(adm\_data$GREScore), max(adm\_data$GREScore), length.out =  
 30),  
 TOEFLScore = c(  
 min(adm\_data$TOEFLScore),  
 mean(adm\_data$TOEFLScore),  
 max(adm\_data$TOEFLScore)  
 )  
)  
plotting.data$predicted.y <-  
 predict.lm(adm.multi, newdata = plotting.data)  
  
plotting.data$GREScore <- round(plotting.data$GREScore, digits = 2)  
  
multi.plot <-  
 ggplot(adm\_data, aes(x = GREScore, y = ChanceofAdmit)) + geom\_point()  
multi.plot



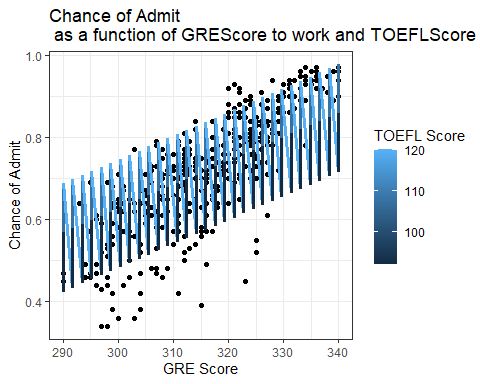
multi.plot <- multi.plot +  
 geom\_line(data = plotting.data,  
 aes(x = GREScore, y = predicted.y, color = TOEFLScore),  
 size = 1.25)

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.

multi.plot



multi.plot <- multi.plot + theme\_bw() +  
 labs(  
 title = "Chance of Admit \n as a function of GREScore to work and TOEFLScore",  
 x = "GRE Score",  
 y = "Chance of Admit",  
 color = "TOEFL Score" )  
multi.plot



## Decision tree

library(randomForest)

## Warning: package 'randomForest' was built under R version 4.2.1

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':  
##   
## margin

library(party)

## Warning: package 'party' was built under R version 4.2.2

## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Warning: package 'strucchange' was built under R version 4.2.1

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.2.1

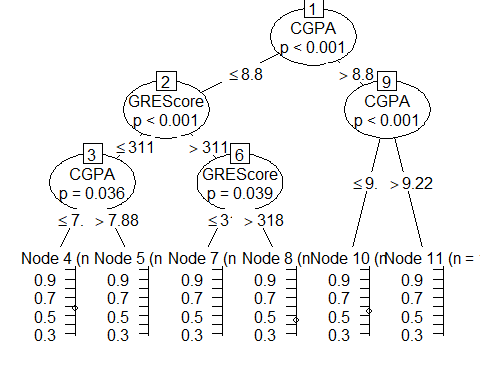
##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 4.2.1

input.data <- adm\_data[c(1:105),]  
output.tree <-  
 ctree(ChanceofAdmit ~ CGPA + GREScore + TOEFLScore, data = input.data)  
plot(output.tree)



# conclusion

From the decision tree, we can conclude that anyone whose CGPA is more than 8.8 and whose GRE Score is more than 311 will have a high chance of Admission.

## Random forest

library(randomForest)  
library(party)  
adm.reg <- randomForest(ChanceofAdmit ~ ., data = adm\_data)  
print(adm.reg)

##   
## Call:  
## randomForest(formula = ChanceofAdmit ~ ., data = adm\_data)   
## Type of random forest: regression  
## Number of trees: 500  
## No. of variables tried at each split: 2  
##   
## Mean of squared residuals: 0.004355389  
## % Var explained: 78.53

round(importance(adm.reg), 2)

## IncNodePurity  
## GREScore 1.75  
## TOEFLScore 1.37  
## UniversityRating 0.75  
## SOP 0.57  
## LOR 0.48  
## CGPA 2.56  
## Research 0.26

# Conclusion

From the random forest, we can conclude that the CGPA, GRE Score and TOEFL Score are the important factors for getting admission into the university.

## Logistic Regression

glm.adm <- glm(ChanceofAdmit ~ ., data = adm\_data)  
glm.adm

##   
## Call: glm(formula = ChanceofAdmit ~ ., data = adm\_data)  
##   
## Coefficients:  
## (Intercept) GREScore TOEFLScore UniversityRating   
## -1.259432 0.001737 0.002920 0.005717   
## SOP LOR CGPA Research   
## -0.003305 0.022353 0.118939 0.024525   
##   
## Degrees of Freedom: 399 Total (i.e. Null); 392 Residual  
## Null Deviance: 8.115   
## Residual Deviance: 1.595 AIC: -1057

summary(glm.adm)

##   
## Call:  
## glm(formula = ChanceofAdmit ~ ., data = adm\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.26259 -0.02103 0.01005 0.03628 0.15928   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.2594325 0.1247307 -10.097 < 2e-16 \*\*\*  
## GREScore 0.0017374 0.0005979 2.906 0.00387 \*\*   
## TOEFLScore 0.0029196 0.0010895 2.680 0.00768 \*\*   
## UniversityRating 0.0057167 0.0047704 1.198 0.23150   
## SOP -0.0033052 0.0055616 -0.594 0.55267   
## LOR 0.0223531 0.0055415 4.034 6.6e-05 \*\*\*  
## CGPA 0.1189395 0.0122194 9.734 < 2e-16 \*\*\*  
## Research 0.0245251 0.0079598 3.081 0.00221 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.004068258)  
##   
## Null deviance: 8.1146 on 399 degrees of freedom  
## Residual deviance: 1.5948 on 392 degrees of freedom  
## AIC: -1056.7  
##   
## Number of Fisher Scoring iterations: 2

# Coefficients

Intercept coefficient(-1.2594325)

GRE Score coefficient(0.0017374)

TOEFL Score Coefficient(0.0029196)

LOR Coefficient( 0.0223531)

CGPA Coefficient(0.1189395)

Research Coefficient(0.0245251)

These are all the factors that impact the Chance of Admit.

# p-value

The p-value of the University Rating and SOP are so high (more than 0.05) which implies that insignificant to the Chance of Admit.

while other factors such as GRE Score, TOEFL Score, CGPA, and LOR with low p-value(less than 0.05) are significant to the Chance of Admit.

# Predict()

glm.predict <- (predict(glm.adm, type = "response"))  
glm.predict

## 1 2 3 4 5 6 7 8   
## 0.9514586 0.8056367 0.6547367 0.7383624 0.6352064 0.8658537 0.7129861 0.6011195   
## 9 10 11 12 13 14 15 16   
## 0.5472143 0.7125863 0.7333123 0.8379894 0.8545403 0.6392162 0.6346553 0.6413325   
## 17 18 19 20 21 22 23 24   
## 0.7094836 0.6513064 0.7318737 0.6484267 0.6111222 0.6948068 0.9273824 0.9591770   
## 25 26 27 28 29 30 31 32   
## 0.9554978 0.9622796 0.7568611 0.5233641 0.4248012 0.4876430 0.6015018 0.7144819   
## 33 34 35 36 37 38 39 40   
## 0.9284189 0.9352382 0.9599034 0.8602838 0.6562453 0.5431920 0.4966729 0.5815495   
## 41 42 43 44 45 46 47 48   
## 0.6455257 0.6550270 0.6801592 0.8387336 0.8825544 0.8295116 0.9027980 0.9338149   
## 49 50 51 52 53 54 55 56   
## 0.8139169 0.7587543 0.6916946 0.6234559 0.7122800 0.6839449 0.6581535 0.5873833   
## 57 58 59 60 61 62 63 64   
## 0.5451374 0.4942130 0.4246499 0.6212652 0.6013721 0.6151223 0.6437020 0.7010294   
## 65 66 67 68 69 70 71 72   
## 0.7495430 0.7826933 0.7886606 0.7322475 0.8274395 0.8576062 0.9568227 0.9605277   
## 73 74 75 76 77 78 79 80   
## 0.8946757 0.7985727 0.7731511 0.7818900 0.7716647 0.5764729 0.4997538 0.4437467   
## 81 82 83 84 85 86 87 88   
## 0.6452140 0.9541930 0.8514861 0.8895155 0.9298408 0.7117632 0.6793102 0.6554650   
## 89 90 91 92 93 94 95 96   
## 0.6501089 0.7604875 0.6566903 0.5331888 0.5647905 0.5770725 0.5335400 0.5669753   
## 97 98 99 100 101 102 103 104   
## 0.5842660 0.8495644 0.8952733 0.8057129 0.7269795 0.6252136 0.6532890 0.6883527   
## 105 106 107 108 109 110 111 112   
## 0.8075247 0.7840749 0.8612207 0.9309831 0.9183219 0.6985639 0.6801261 0.7724885   
## 113 114 115 116 117 118 119 120   
## 0.6762211 0.7122630 0.6896628 0.7969605 0.6652738 0.5074803 0.4895124 0.7851907   
## 121 122 123 124 125 126 127 128   
## 0.9496003 0.9346629 0.6567658 0.6525047 0.6620838 0.6858885 0.8356931 0.7358755   
## 129 130 131 132 133 134 135 136   
## 0.8134717 0.9240677 0.9637077 0.7150316 0.6973561 0.7889842 0.8820961 0.7726838   
## 137 138 139 140 141 142 143 144   
## 0.6784314 0.6501533 0.8137494 0.7684106 0.8114225 0.8744926 0.8923142 0.9834470   
## 145 146 147 148 149 150 151 152   
## 0.8021953 0.7393192 0.6694368 0.8221528 0.9462346 0.6642673 0.8988774 0.9081654   
## 153 154 155 156 157 158 159 160   
## 0.8512089 0.7474116 0.7651201 0.7087515 0.6527852 0.6242094 0.6105636 0.5336233   
## 161 162 163 164 165 166 167 168   
## 0.5688978 0.5024609 0.6965775 0.6886455 0.8298244 0.7911769 0.6711758 0.6434025   
## 169 170 171 172 173 174 175 176   
## 0.5793164 0.5893072 0.6397218 0.8817111 0.8530214 0.8642349 0.8128201 0.7863596   
## 177 178 179 180 181 182 183 184   
## 0.8821985 0.7749858 0.6344768 0.6296728 0.6086142 0.6433962 0.5690079 0.7472772   
## 185 186 187 188 189 190 191 192   
## 0.6812239 0.8552593 0.7332825 0.9063700 0.8811465 0.8587999 0.8446949 0.8426346   
## 193 194 195 196 197 198 199 200   
## 0.8238122 0.9523416 0.7223743 0.7039906 0.6186115 0.6351539 0.6900639 0.7348743   
## 201 202 203 204 205 206 207 208   
## 0.7082579 0.7066831 0.9991509 0.9967980 0.6756091 0.5222586 0.5822472 0.6503759   
## 209 210 211 212 213 214 215 216   
## 0.6190763 0.6524723 0.8200630 0.8389770 0.9697485 0.9669547 0.9219350 0.9030292   
## 217 218 219 220 221 222 223 224   
## 0.8614935 0.8260113 0.8072414 0.6715639 0.7191558 0.7238591 0.7762847 0.6898929   
## 225 226 227 228 229 230 231 232   
## 0.6021294 0.5580165 0.6876850 0.6852715 0.7334273 0.8037420 0.7213580 0.6810324   
## 233 234 235 236 237 238 239 240   
## 0.6600703 0.6019461 0.8771469 0.8564955 0.8576538 0.8880620 0.6670999 0.5359087   
## 241 242 243 244 245 246 247 248   
## 0.5276867 0.6092766 0.7783162 0.8104371 0.7092141 0.7944797 0.7199570 0.6745514   
## 249 250 251 252 253 254 255 256   
## 0.7991548 0.7921045 0.7071287 0.7193125 0.7066965 0.9427932 0.8372478 0.7008646   
## 257 258 259 260 261 262 263 264   
## 0.6779293 0.7633035 0.7893017 0.9054299 0.8247511 0.6434905 0.6878485 0.7399815   
## 265 266 267 268 269 270 271 272   
## 0.7502151 0.6793385 0.6549397 0.6802403 0.8454789 0.7027186 0.6512082 0.5263789   
## 273 274 275 276 277 278 279 280   
## 0.4717123 0.5848493 0.5803676 0.7840314 0.9032376 0.6869043 0.6659564 0.6958092   
## 281 282 283 284 285 286 287 288   
## 0.7225506 0.8250623 0.7442832 0.7813823 0.9386901 0.8850012 0.8895491 0.8534625   
## 289 290 291 292 293 294 295 296   
## 0.7806910 0.7552315 0.5606250 0.5468231 0.5550858 0.6274162 0.6464448 0.6599964   
## 297 298 299 300 301 302 303 304   
## 0.7065827 0.8349452 0.8731365 0.7117673 0.5974717 0.7222555 0.7047103 0.7394214   
## 305 306 307 308 309 310 311 312   
## 0.6443887 0.7715206 0.8119443 0.8262575 0.6851490 0.6980810 0.7497016 0.8395480   
## 313 314 315 316 317 318 319 320   
## 0.7968599 0.5844006 0.6349417 0.6093974 0.5392394 0.5618284 0.7523475 0.7750802   
## 321 322 323 324 325 326 327 328   
## 0.7184778 0.7271034 0.6747217 0.6019054 0.6453711 0.8522607 0.5611516 0.5307219   
## 329 330 331 332 333 334 335 336   
## 0.7859875 0.5119905 0.7657953 0.6239956 0.6476481 0.7358546 0.7457919 0.8475979   
## 337 338 339 340 341 342 343 344   
## 0.7245503 0.9366030 0.7858968 0.7768521 0.7000814 0.7783697 0.6395204 0.6195786   
## 345 346 347 348 349 350 351 352   
## 0.4575807 0.5048934 0.5118188 0.4322740 0.4604275 0.6114742 0.6990839 0.7828210   
## 353 354 355 356 357 358 359 360   
## 0.6218918 0.5927879 0.5251927 0.6496494 0.7895535 0.6082230 0.5492444 0.6528730   
## 361 362 363 364 365 366 367 368   
## 0.7890244 0.9066233 0.9113759 0.6374955 0.7602548 0.8369977 0.7228263 0.5126037   
## 369 370 371 372 373 374 375 376   
## 0.5079727 0.5954216 0.6189958 0.8184305 0.9378960 0.7310724 0.5650003 0.5353999   
## 377 378 379 380 381 382 383 384   
## 0.4684549 0.4810220 0.5179977 0.6616384 0.7745943 0.7425580 0.8348693 0.6216590   
## 385 386 387 388 389 390 391 392   
## 0.9623014 0.9805245 0.5883075 0.6269769 0.5169499 0.7352220 0.6246785 0.7089648   
## 393 394 395 396 397 398 399 400   
## 0.8310911 0.7054498 0.8559911 0.8081979 0.8111549 0.9080171 0.7226391 0.9299496

## Result

By performing various statistical analyses CGPA, GRE Score, and TOEFL Score, are all crucial factors that obtain the Chance of Admit to the University.

## Literature citation

Mohan S. Acharya, Asfia Armaan, Aneeta S Antony: A Comparison of Regression Models for Prediction of Graduate Admissions, IEEE International Conference on Computational Intelligence in Data Science 2019.