Project Ideas and Techniques:

I have loaded the data, and did wrangling techniques:

- we renamed the column names which have inconsistent names to be consistent

- Detecting Outliers

I have used box plots and scatter plots to find the outliers graphically. But no outliers detected

I have used z-score technique by taking 3rd standard deviation and found no outliers

Later I have done Exploratory Data Analysis:

- Bar plot GRE Score vs Chance of Admit

- Scatter plot Chance of Admit vs GRE Score

- Bar plot University Rating vs Chance of Admit

- Calculated highest probability for chance of admission and lowest probability for chance of admission

- Displaying those students records who has highest chance of getting admission and lowest chance of getting admission

- Did count plot for number of students has the high chance of admission and number of low chance of admission

- Did histograms to see continuous variables

- Did pairplots to check various relations

Conclusion:

I will create a Machine learning model by taking the cleaned dataset using Logistic Regression.

I will create a classification report and confusion matrix

Later I wanted to do some hyper parameter tuning to increase the model accuracy, I have to use the regularization parameter C to fine tune the model

I will also use alternate model to fit the data using random forest classifier

Apart from this model, we have some interesting things that we can observe in data analysis, like we can get the number of students approximately who got admitted into top universities, and the behaviour of various features towards 'Chance of Admit'