Assignment 1

Instructions:

ABC bank has decided to select a model that helps in identifying whether the loan should be granted to an applicant or not based on the dataset available with the bank.

You are a data scientist. You are required to perform various activities pertaining to the data such as, preparing the dataset for analysis; investigating the relationships in the data set with visualization; creating a model; evaluating the performance of the classification model.

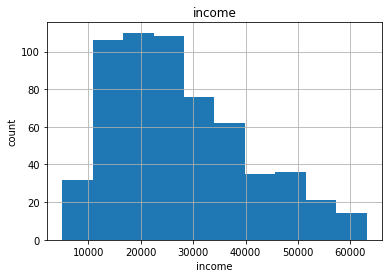
Demonstrate KDD process with following activities for your problem statement:

* Demonstrate the data
* Perform exploratory data analysis
* Select the features
* Select Training, test split
* Train the model
* Predictions and reporting
* Evaluating remarks

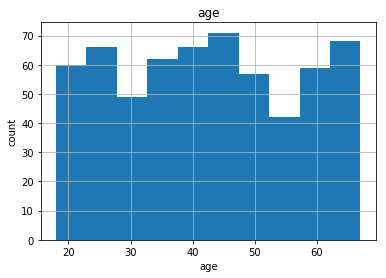
Using Bank Dataset, apply decision tree to predict whether loan can be granted or not.

**Solution:**

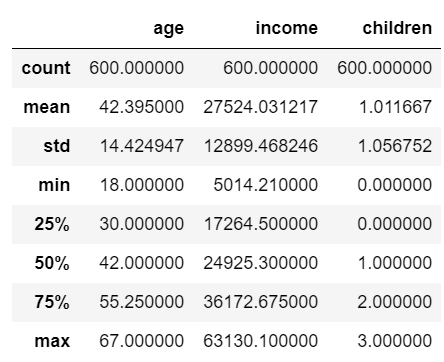
Distribution of data by income:



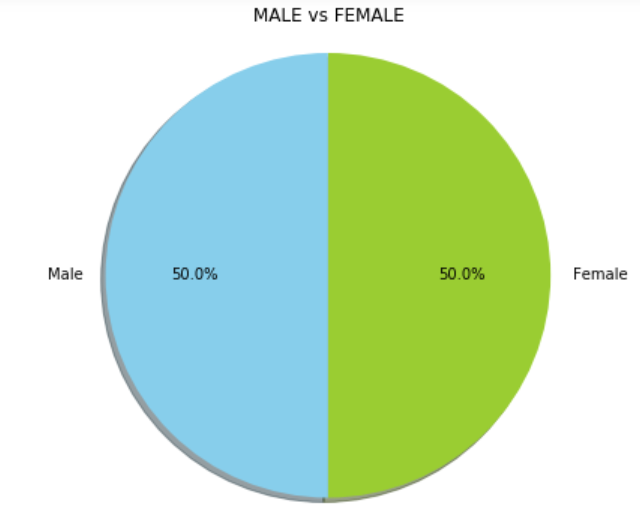
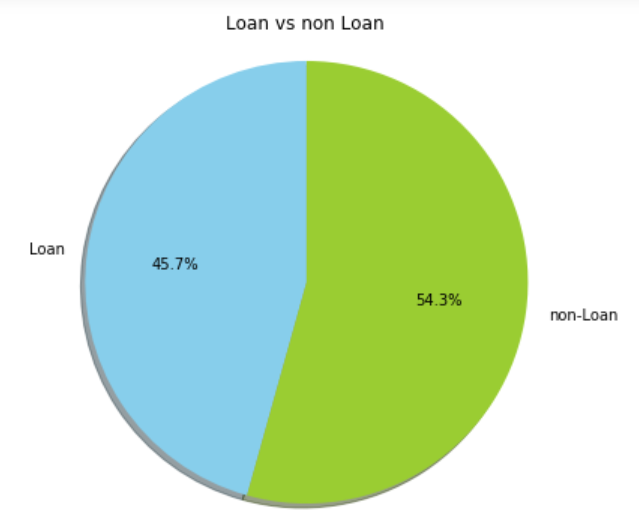
Distribution of data by age:

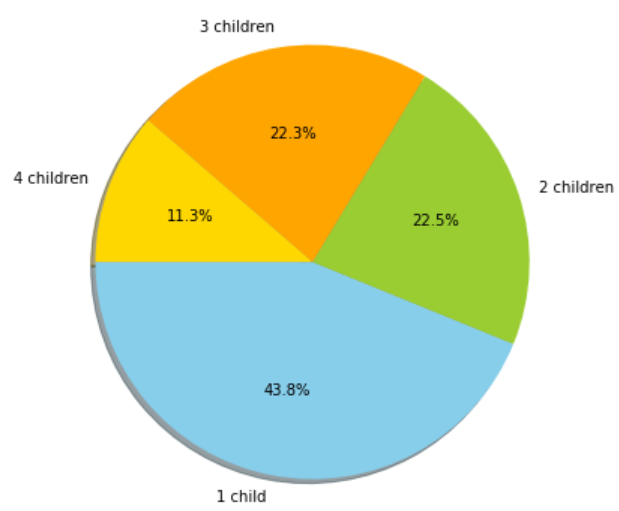
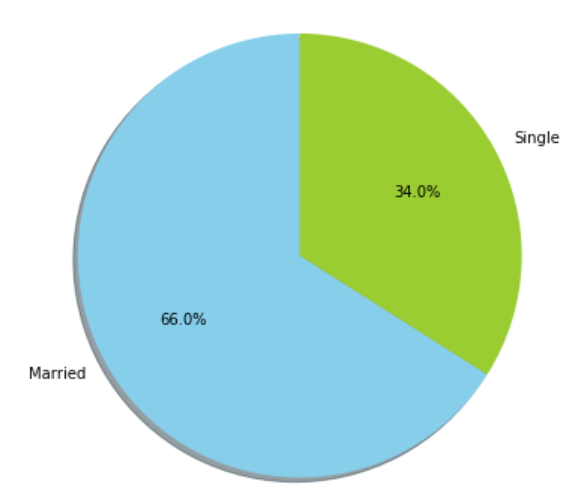


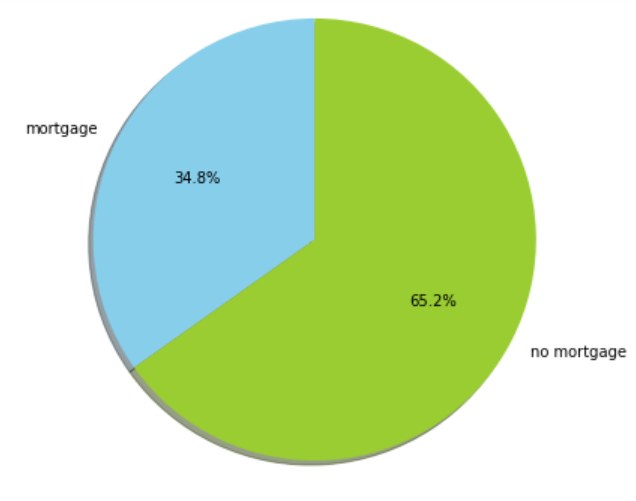
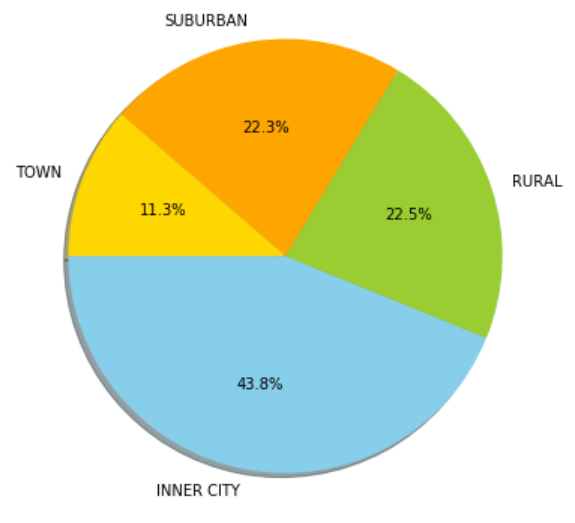
Exploratory data analysis:

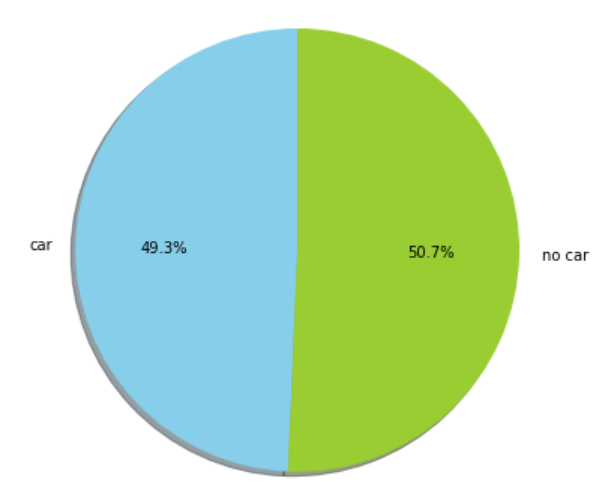
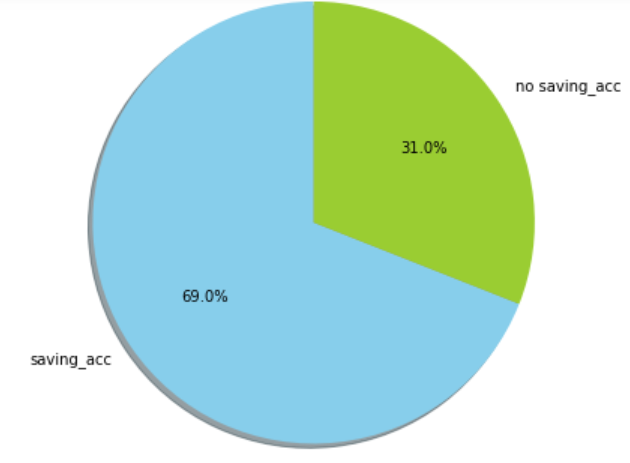


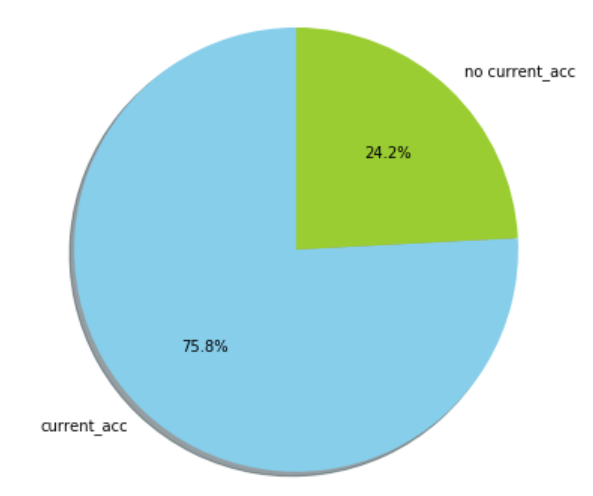
Plotting pie chart for each attribute:



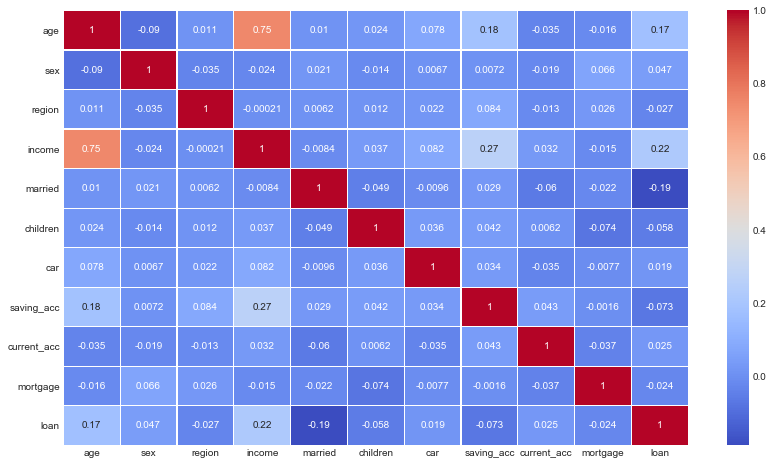
 



Correlation Matrix:





Training split = 80%

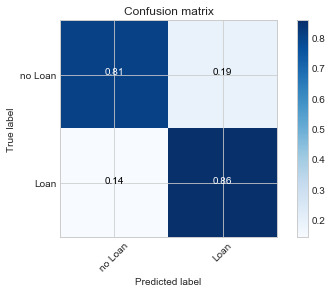
Test split = 20%

A decision tree classifier model was built based on tuned parameters and a training split of 80%.

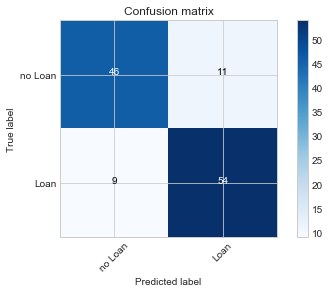
The Accuracy of the model was calculated to be 83.34%

Confusion Matrix:

Normalized



Not normalized



ROC Curve:

