

Micro Credit Defaulter

Submitted by:

Dipak Someshwar

**ACKNOWLEDGMENT**

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals. We would like to extend my sincere thanks to SME. Khushboo Garg.

We are highly indebted to Flip Robo technology for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I thank and appreciations also go to our colleague in developing the project and people who have willingly helped us out with their abilities.

Thanks all.

Dipak Someshwar

**INTRODUCTION**

A Microfinance Institution (MFI) is an organization that offers financial services to low income populations. MFS becomes very useful when targeting especially the unbanked poor families living in remote areas with not much sources of income. The Microfinance services (MFS) provided by MFI are Group Loans, Agricultural Loans, Individual Business Loans and so on.

Many microfinance institutions (MFI), experts and donors are supporting the idea of using mobile financial services (MFS) which they feel are more convenient and efficient, and cost saving, than the traditional high-touch model used since long for the purpose of delivering microfinance services. Though, the MFI industry is primarily focusing on low income families and are very useful in such areas, the implementation of MFS has been uneven with both significant challenges and successes.

Today, microfinance is widely accepted as a poverty-reduction tool, representing $70 billion in outstanding loans and a global outreach of 200 million clients.

We are working with one such client that is in Telecom Industry. They are a fixed wireless telecommunications network provider. They have launched various products and have developed its business and organization based on the budget operator model, offering better products at Lower Prices to all value conscious customers through a strategy of disruptive innovation that focuses on the subscriber.

They understand the importance of communication and how it affects a person’s life, thus, focusing on providing their services and products to low income families and poor customers that can help them in the need of hour.

They are collaborating with an MFI to provide micro-credit on mobile balances to be paid back in 5 days. The Consumer is believed to be defaulter if he deviates from the path of paying back the loaned amount within the time duration of 5 days. For the loan amount of 5 (in Indonesian Rupiah), payback amount should be 6 (in Indonesian Rupiah), while, for the loan amount of 10 (in Indonesian Rupiah), the payback amount should be 12 (in Indonesian Rupiah).

The sample data is provided to us from our client database. It is hereby given to you for this exercise. In order to improve the selection of customers for the credit, the client wants some predictions that could help them in further investment and improvement in selection of customers.

1. **Model Building Phase:**

Build a model which can be used to predict in terms of a probability for each loan transaction, whether the customer will be paying back the loaned amount within 5 days of insurance of loan. In this case, Label ‘1’ indicates that the loan has been paid i.e. Non- defaulter, while, Label ‘0’ indicates that the loan has not been paid i.e. defaulter.

Before model building do all data pre-processing steps.

Try different models with different hyper parameters and select the best model.

1. Data Cleaning

2. Exploratory Data Analysis

3. Data Pre-processing

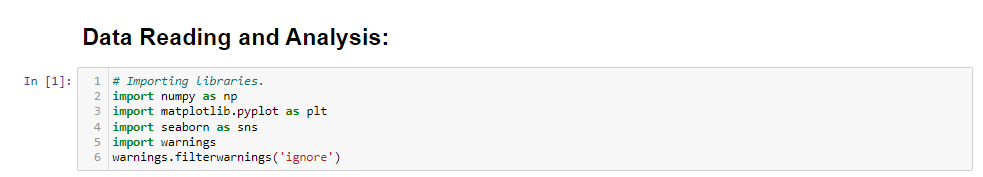
4. Model Building

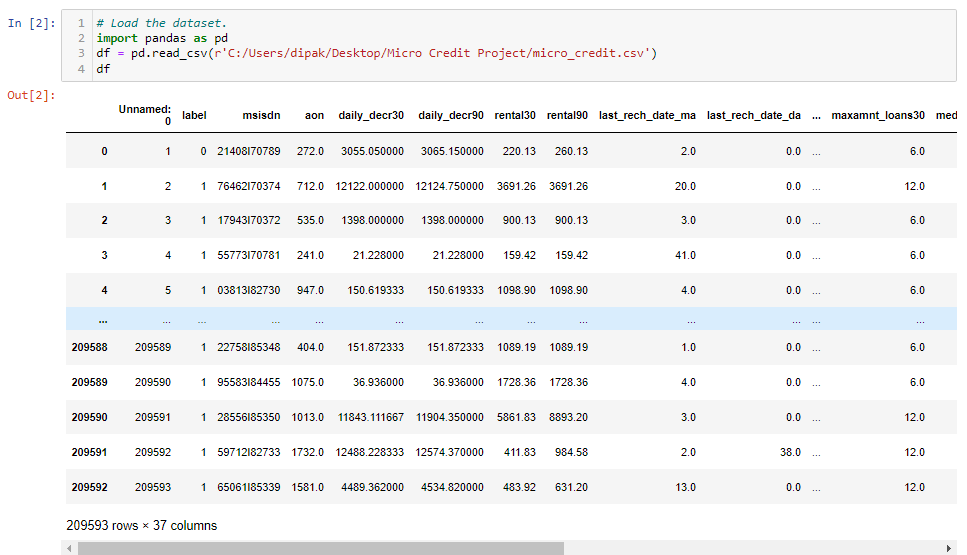
5. Model Evaluation

6. Selecting the best model

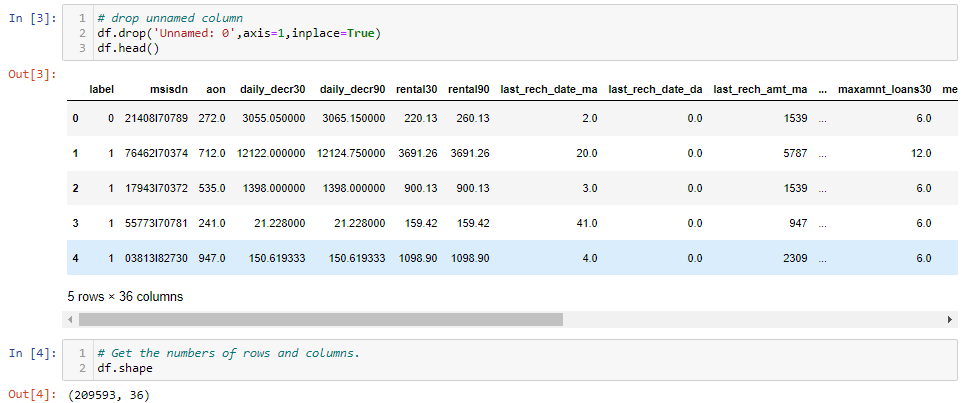
**Analytical Problem Framing**

* Import library and load the dataset.

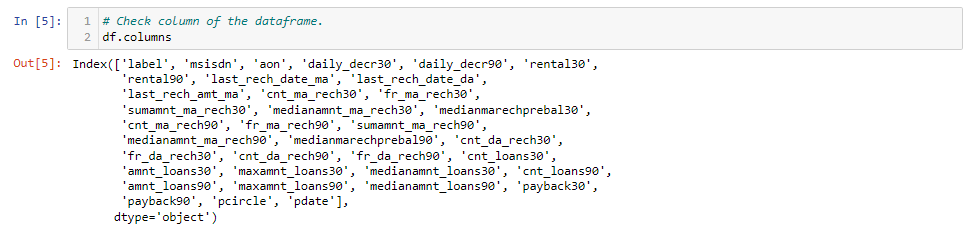




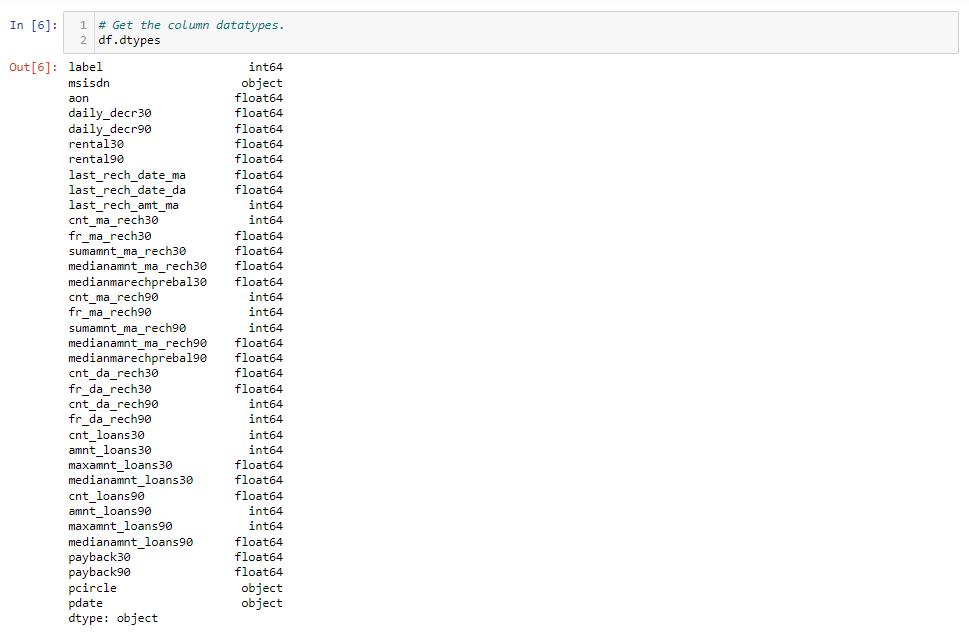
* Drop unnamed column and get the shape of data frame.



* Display all column name of dataset.

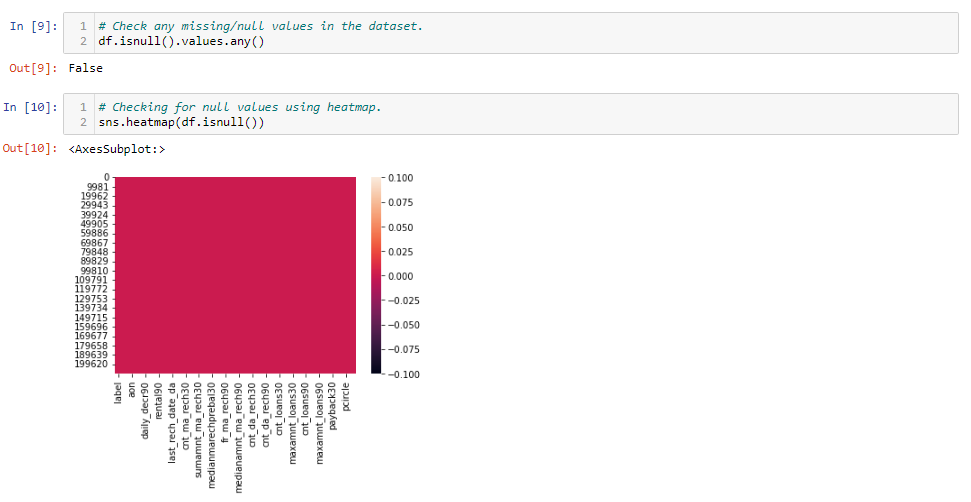
****

* Display datatypes and sum of null values.

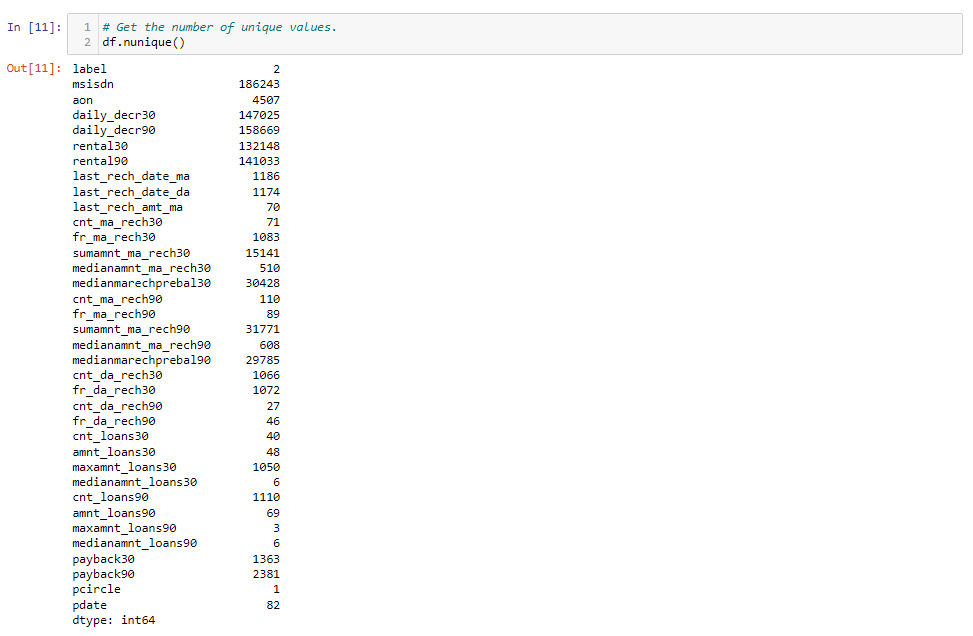
****



* Display null values of columns using heatmap.

****

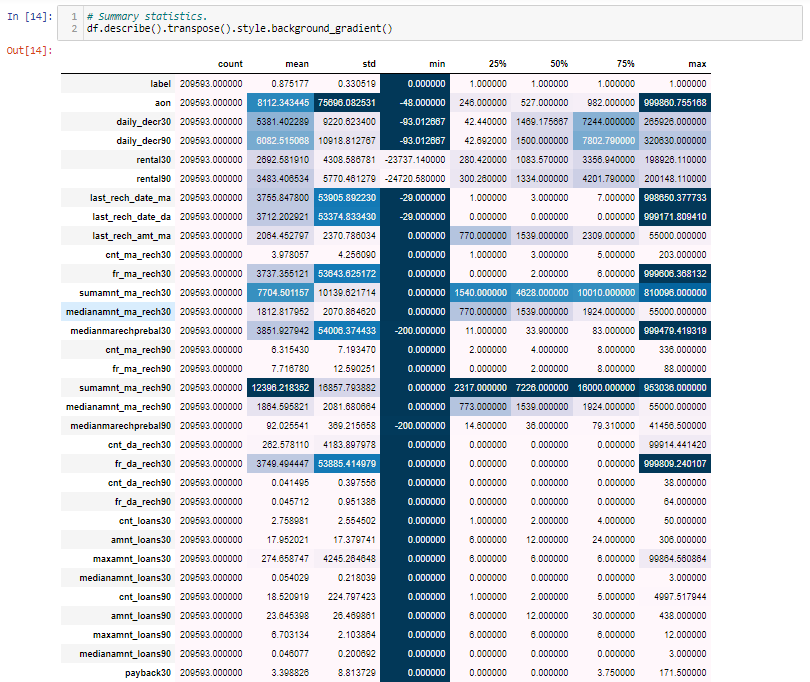
* Get the nunique values.

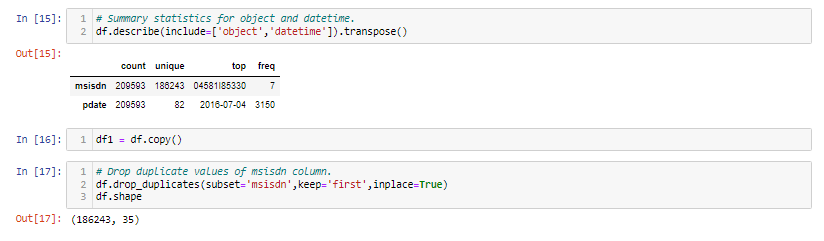


* Data Preprocessing

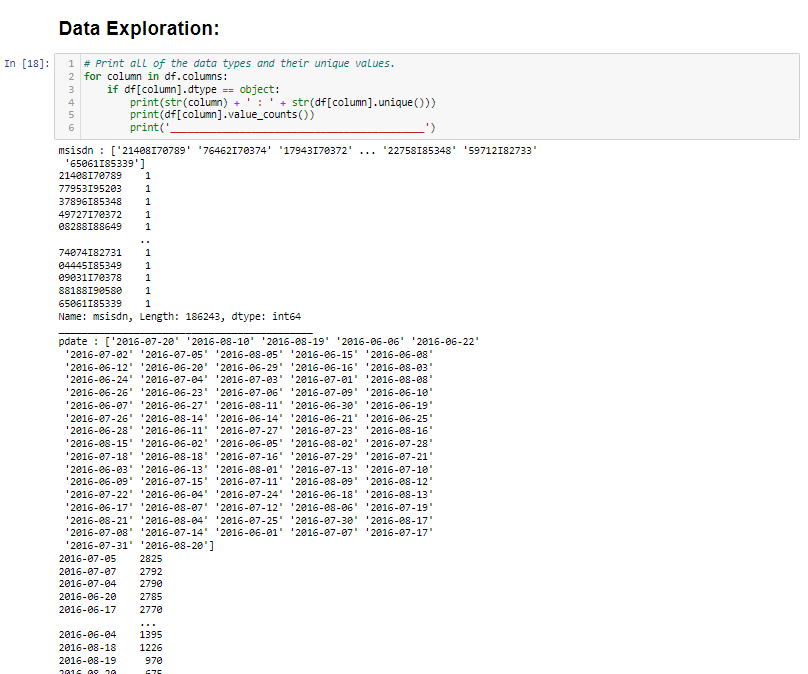


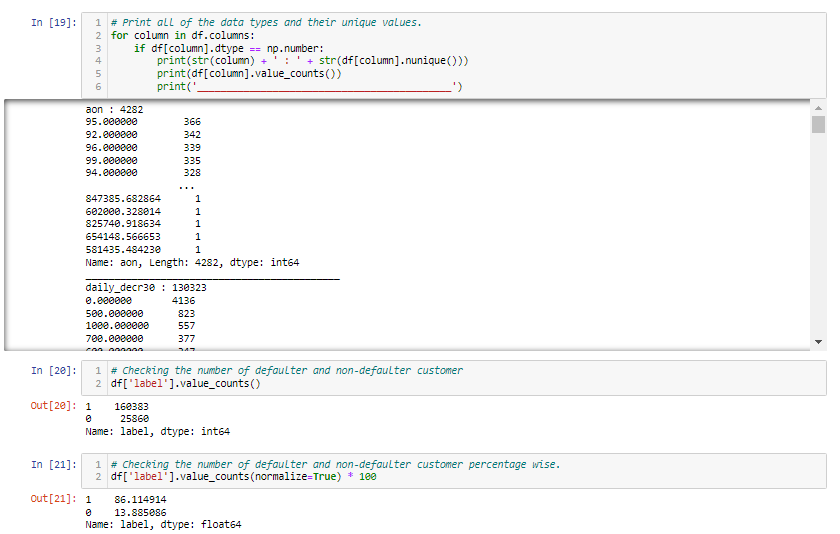
* Summary Statistics and Drop Duplicates value.

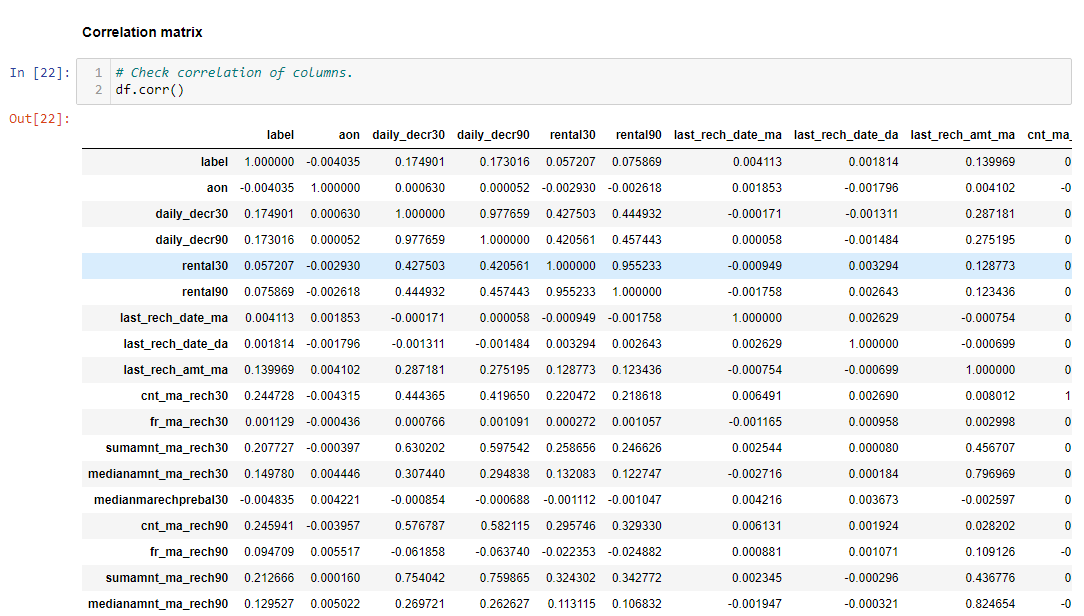


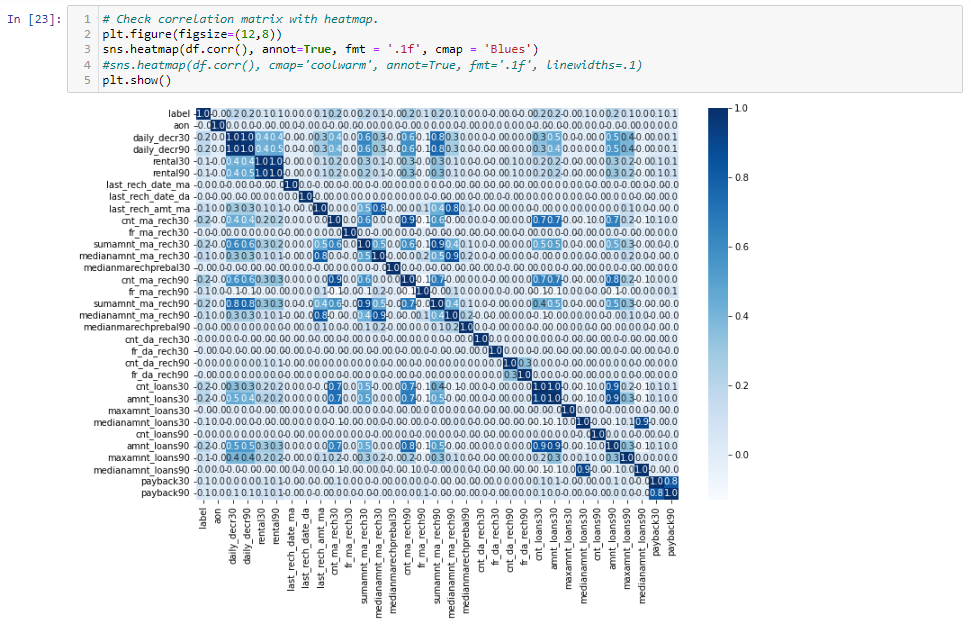


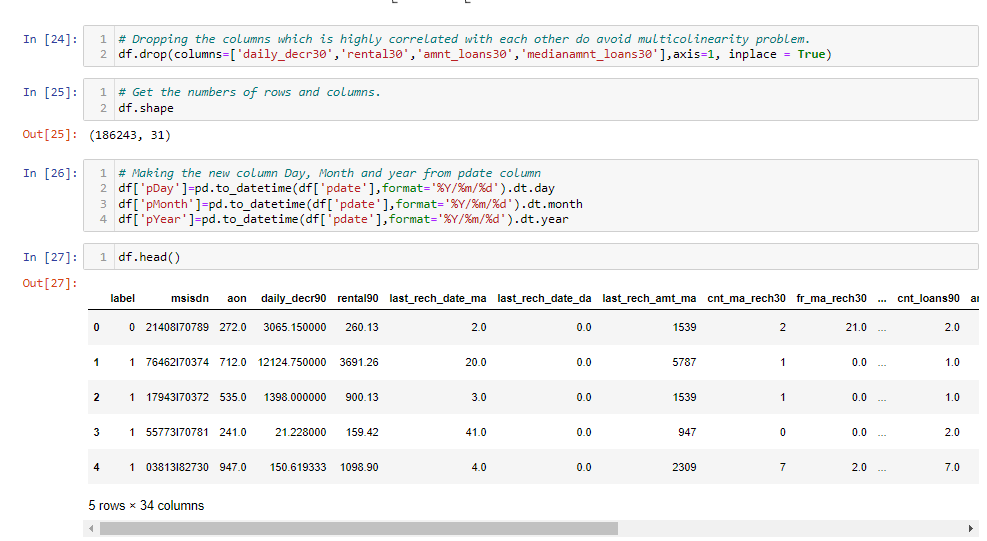
* Data Exploration.

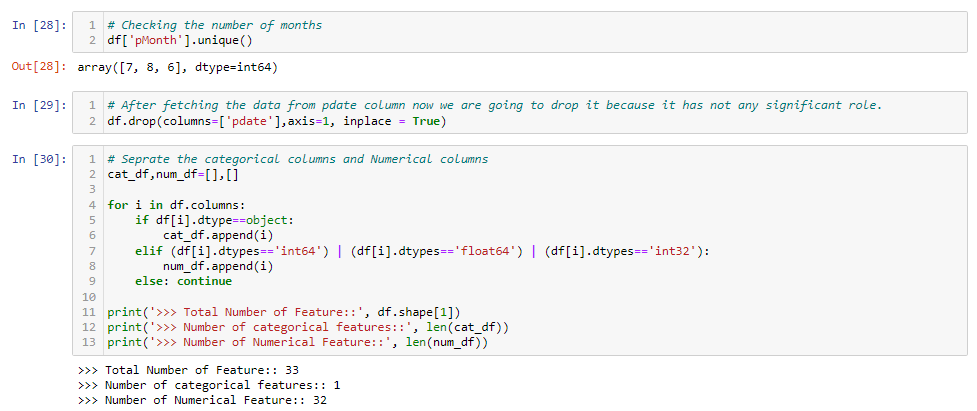




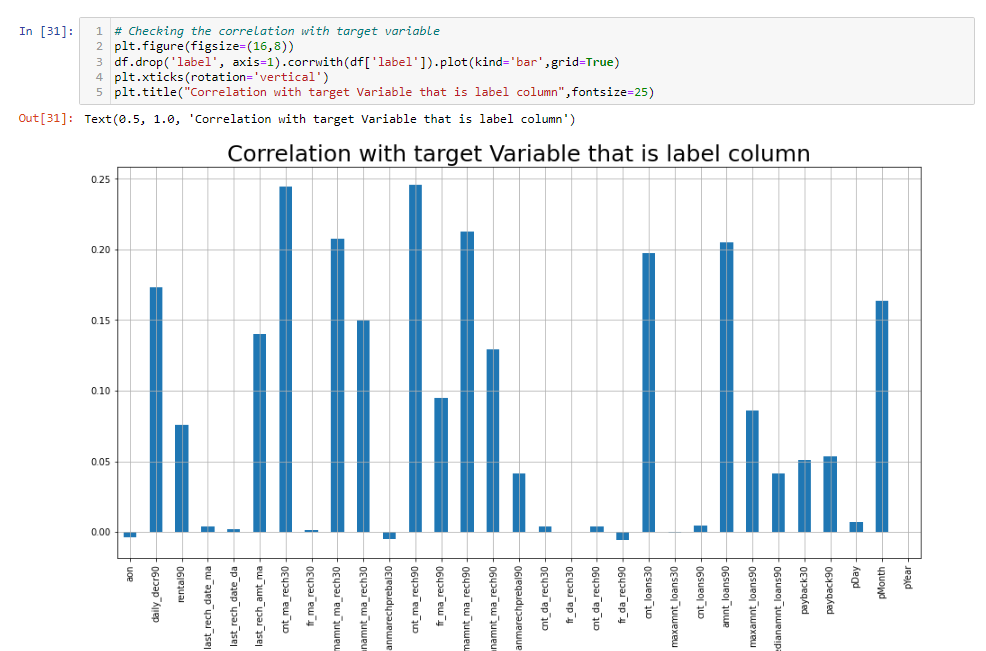


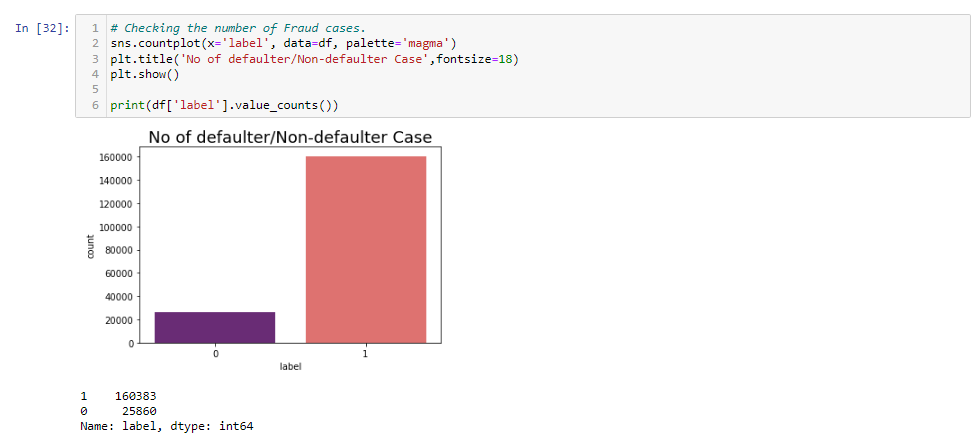




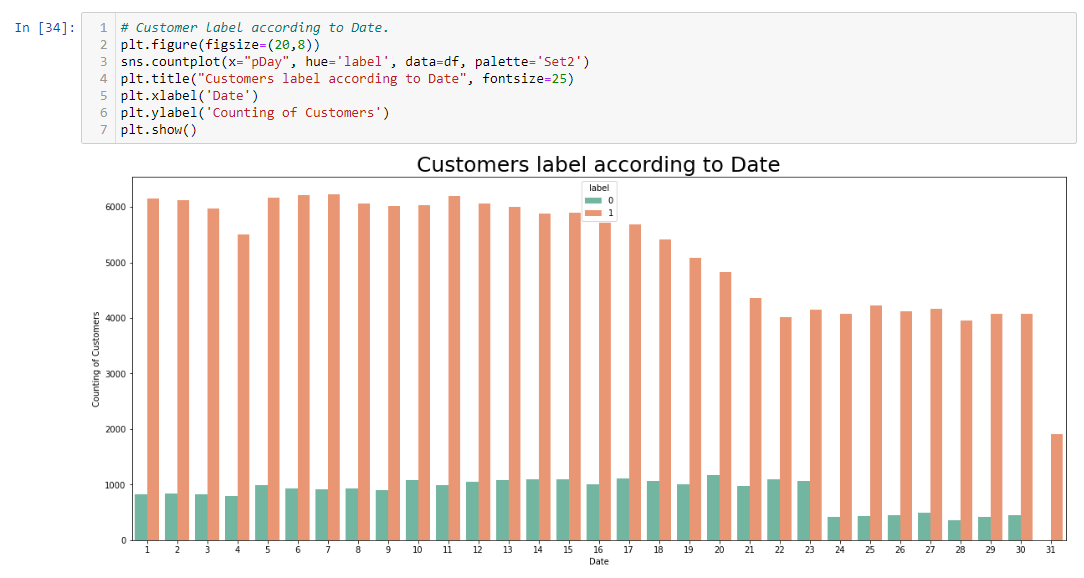


* Data Visualization.

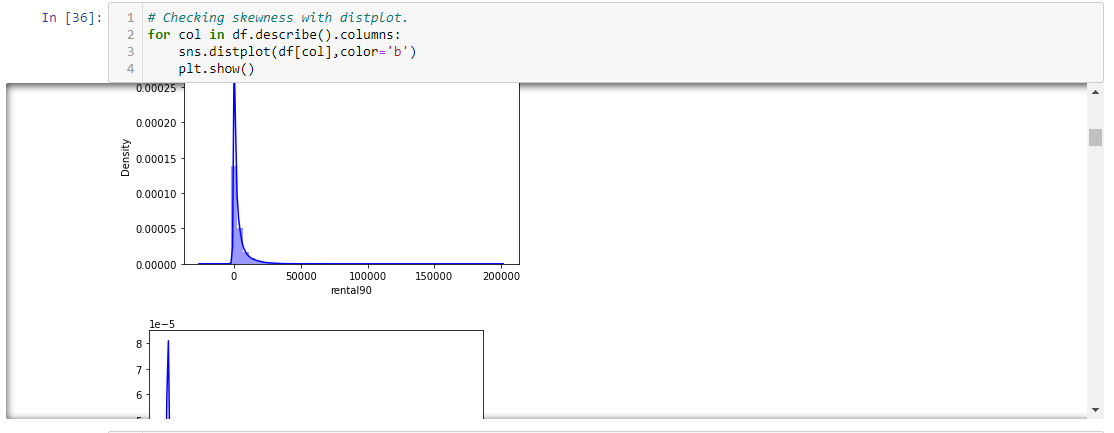




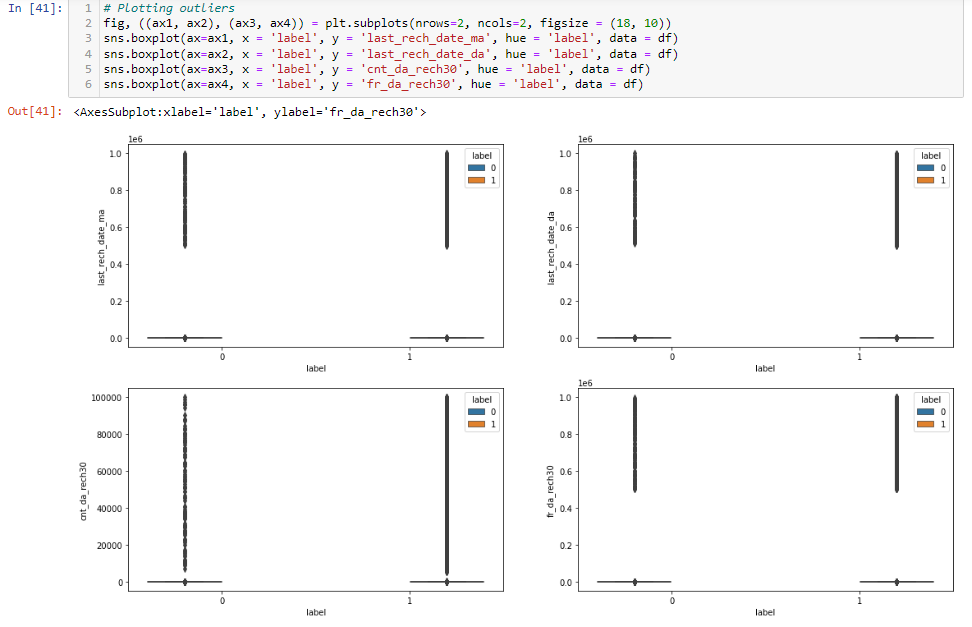




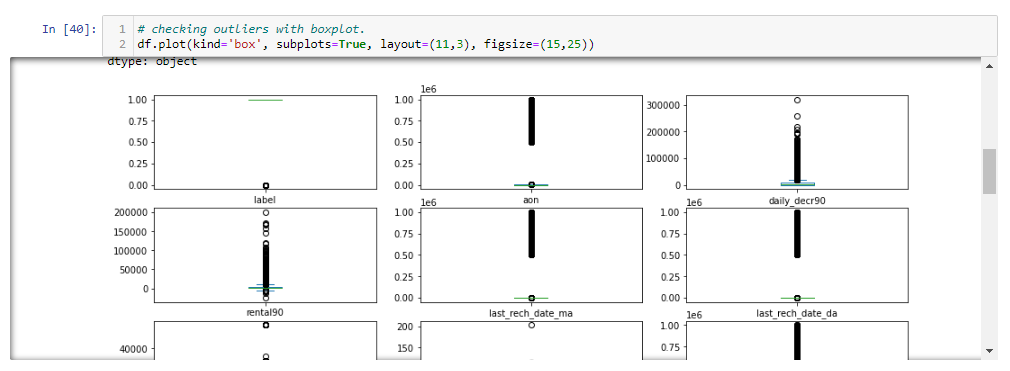




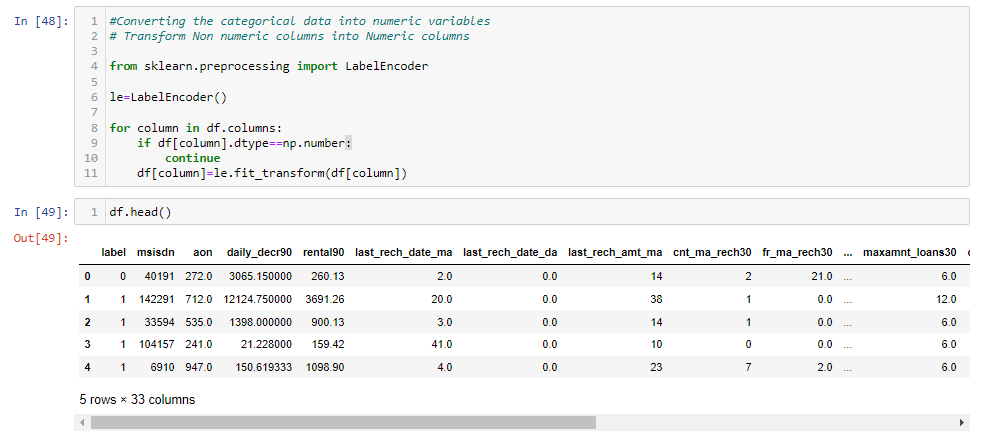
* Display boxplot of columns to compare with label.



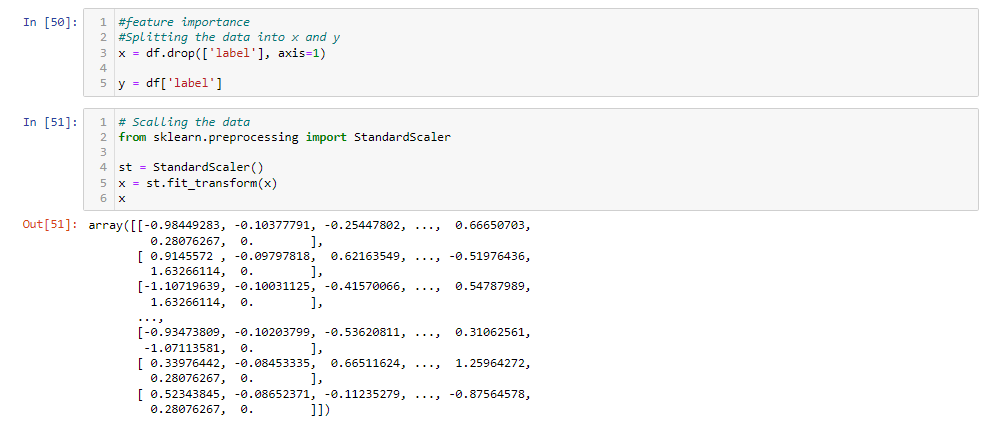
* Display outliers of all columns.



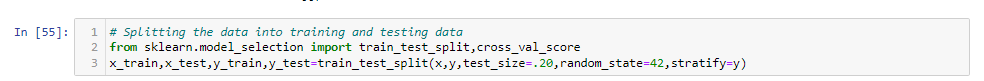
* Label Encoding.



* Data Pre-processing and Scalling the data.



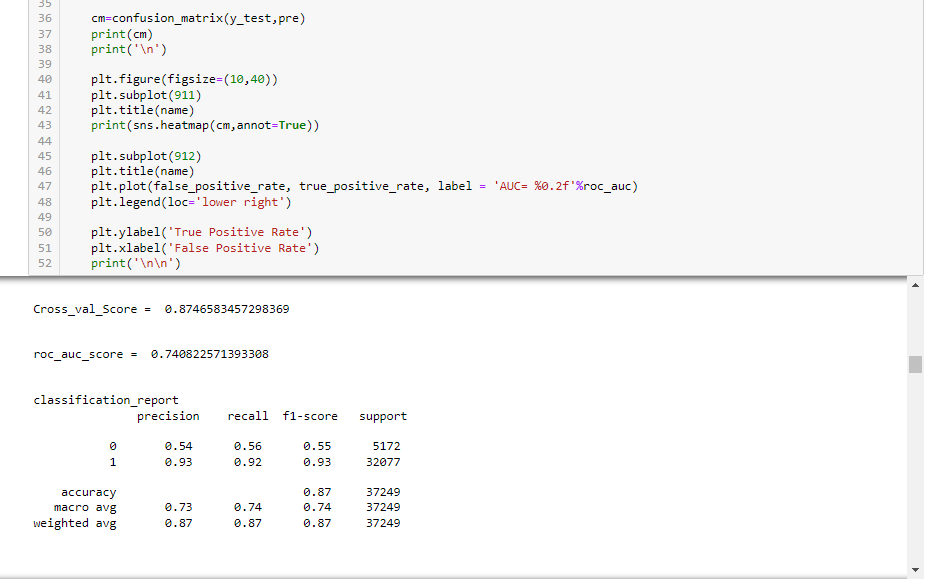
* Train-Test Splitting.



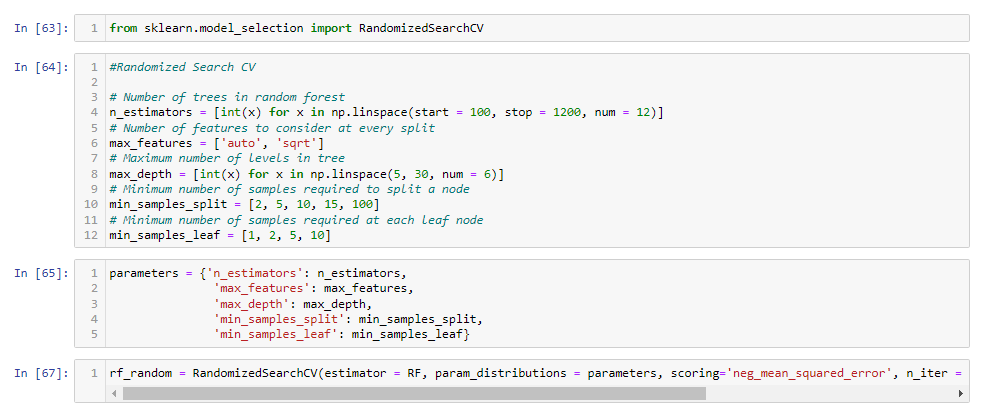
* Run and evaluate selected models.







* Hypertuning of the Model.



* Hardware and Software Requirements and Tools Used
* **Language :-** Python
* **Tool:-** Jupyter
* **OS:-** Windows 10
* **RAM:-** 8gb

**CONCLUSION:**

* This Kernel investigates different models for car price prediction.
* Different types of Machine Learning methods including LinearRegression, RandomForestRegressor, AdaBoostRegressor, GradientBoostingRegressor and DecisionTreeRegressor in machine learning are compared and analysed for optimal solutions.
* Even though all of those methods achieved desirable results, different models have their own pros and cons.
* The RandomForestRegressor is probably the best one and has been selected for this problem.
* Finally, the RandomForestRegressor is the best choice when parameterization is the top priority.