

Micro Credit Loan Use

Case

Submitted by:

Ram kumar

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

Thanks all.

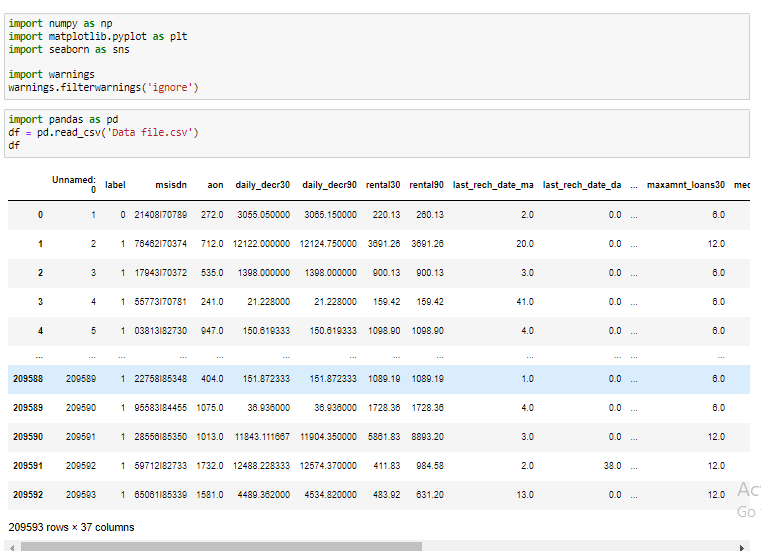
Ram kumar

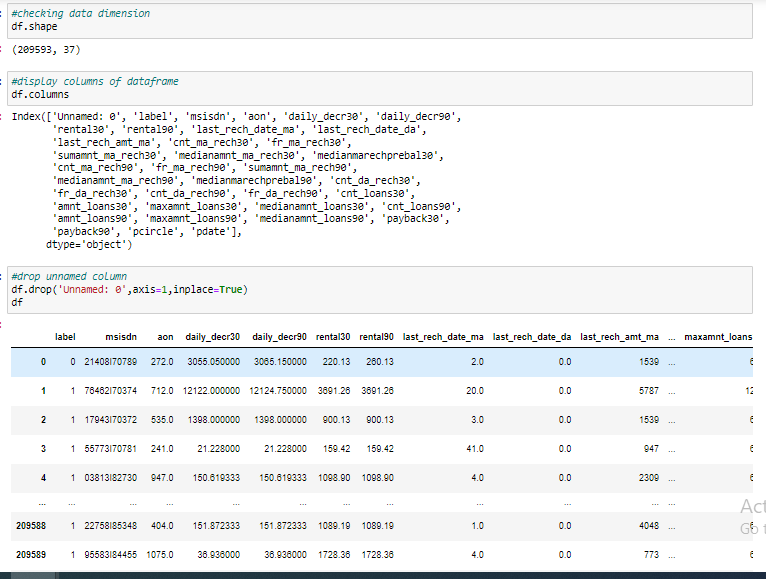
**INTRODUCTION**

* 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 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.
* 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.
* 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

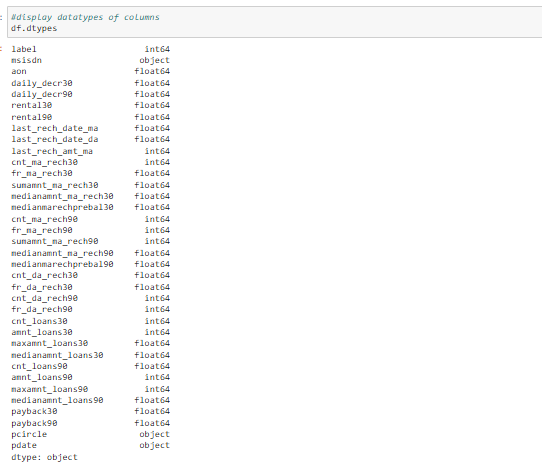
**Analytical Problem Framing**

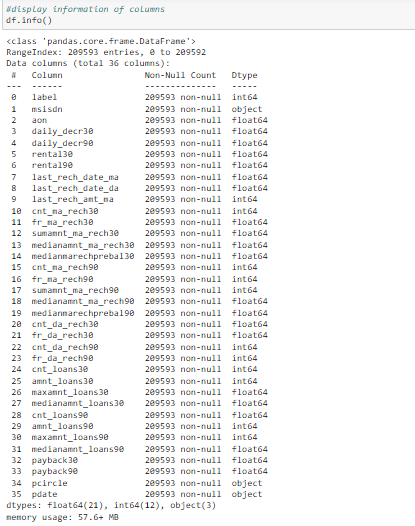
* Import library and load the dataset.



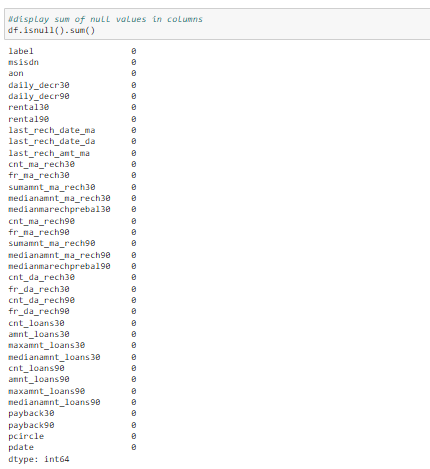


* Display all column name of dataset.

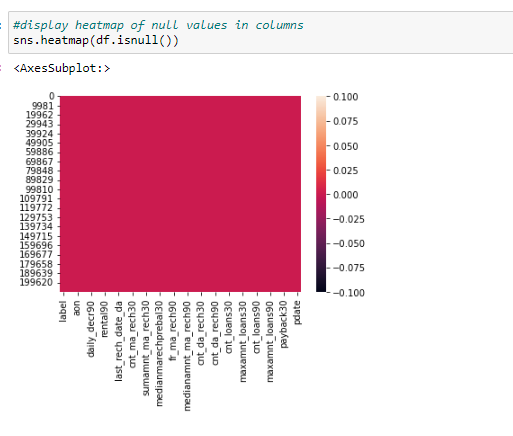




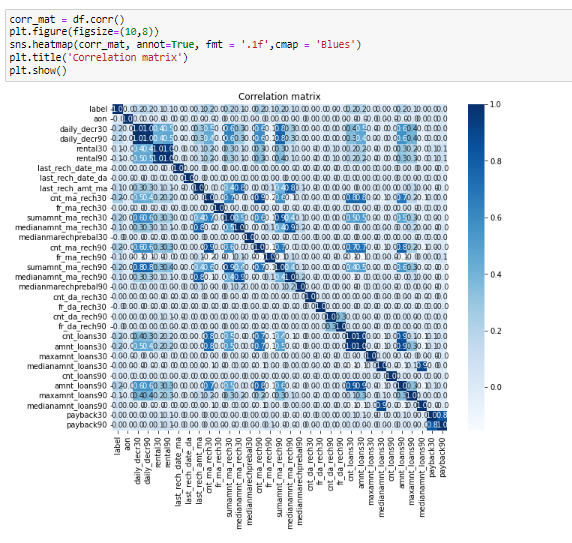
* Display datatypes and sum of null values.



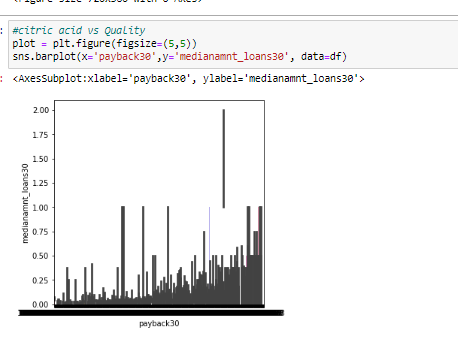
* Display null values of columns using heatmap.



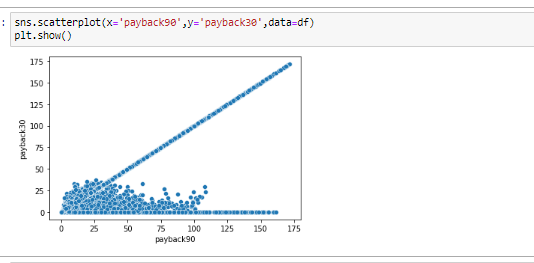
* Display correlation of columns using heatmap.



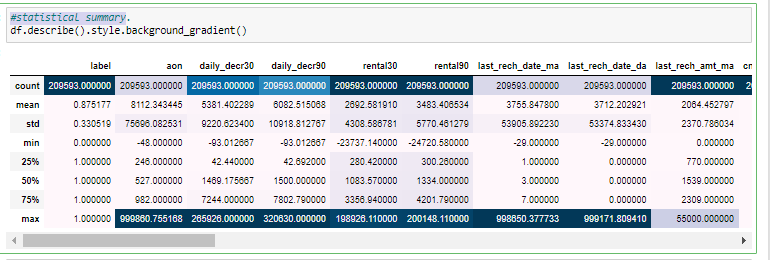
* Display barplot of all columns.



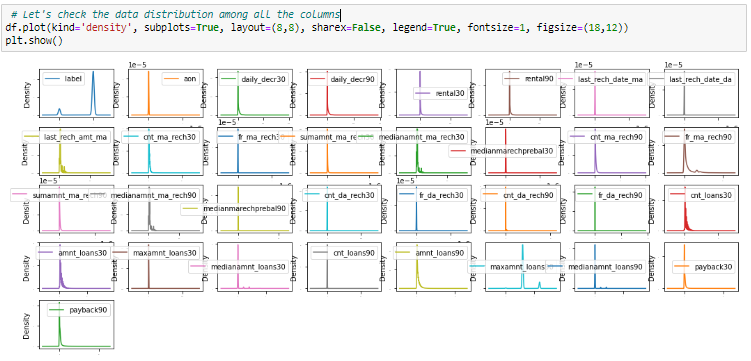
* Display Scatterplot of payback30column.



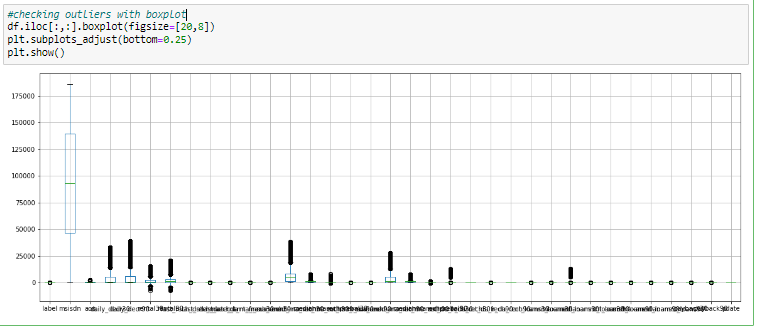
* Display statistical summary.



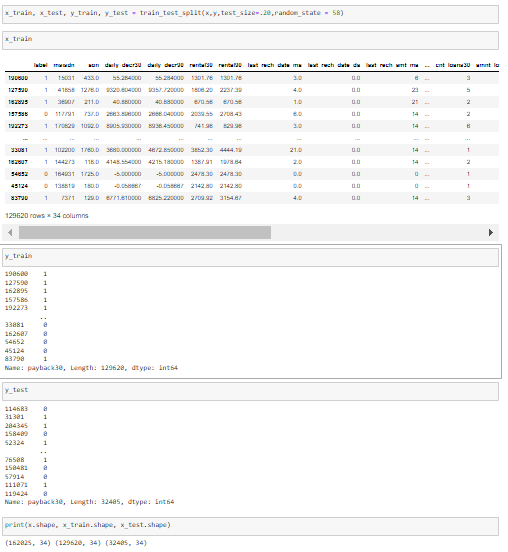
* Check the data distribution among all the columns



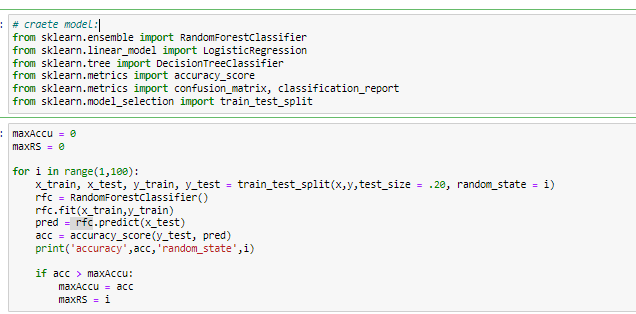
* **Checking outliers with boxplot.**

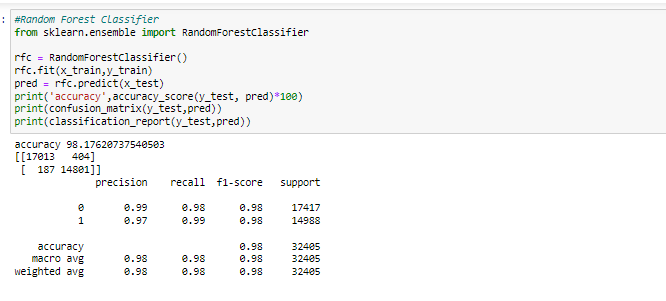


* Train test split Here:

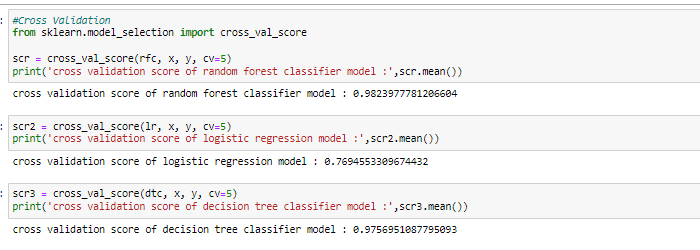


* Display creating model:





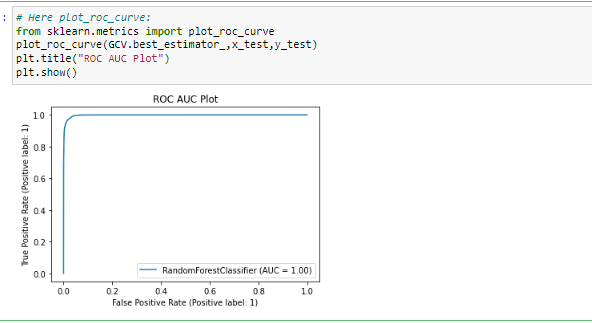
* Cross Validation Here:



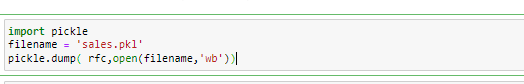
* Hyper parameter tuning here:



* Plot roc curve here:



* Save the mode here:



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

**CONCLUSION**

* The [Random Forest approach](https://www.analyticsvidhya.com/blog/2021/10/an-introduction-to-random-forest-algorithm-for-beginners/) is appropriate for classification and regression tasks on datasets with many entries and features that are likely to have missing values when we need a highly accurate result while avoiding overfitting.
* the random forest provides relative feature significance, enabling you to select the most important features. It is more interpretable than neural network models but less interpretable than decision trees.
* Predicting Loan Default is highly dependent on the demographics of the people, people with lower income are more likely to default on loans.