**Data Set**

The data used in this project is based on loans disbursed from March to May 2021 by a lending institution. It contains information about the loans disbursed, repayment date, and demographic information about the customers obtaining these loans.

**Objective**

The objective of this analysis is to gain some insights about the customers who apply for these loans, their reasons and how these loans are actually performing by calculating the portfolio at risk using PAR7 AND PAR30 ratios.

**Tools**

**Python (Numpy, Pandas and Matplotlib Libraries), Jupyter Notebook** and **Excel** were used to conduct this analysis although Excel was only used briefly to visually access the data. Most of the analysis occurred in the Jupyter Notebook (link below) using Python.

**Additional Resources**

Below is a link to a pdf version of the notebook used for this analysis. The Pdf file contains all the code that was used to produce this report with additional explanations on the rationale behind the analysis. Also in case you require the live notebook where you can test the code, please refer to the second link.

[QuickCheck Case Study.Pdf](https://drive.google.com/file/d/1R9242g7I9q-rYWA-eU2GohKmR4KN2TkZ/view?usp=sharingn-G1QkpoT4A/view?usp=sharing)

[Live Notebook(IPYNB FILE)](https://drive.google.com/file/d/1X7Dj2rRMRQZ_U1L0G6pKD5UhUW9hH8fj/view?usp=sharingt2dQ8wpCaLW9KIuPM4u4q1FU/edit?usp=sharing)

**Data Gathering**

The data was obtained from the lending institution directly and permission was given to use the data for the purpose of this project.

**Data Cleaning**

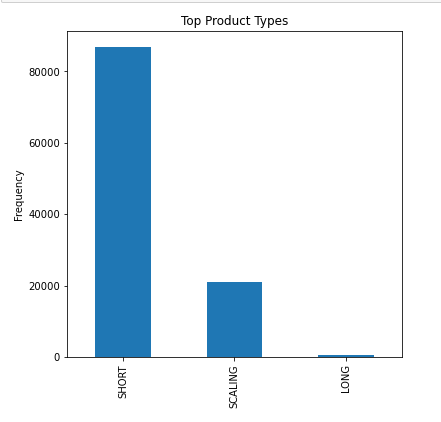
Since the original data set was a csv file, a lot of data type conversion had to be done to make the data ready for exploration using the pandas library. All cleaning steps are detailed in the pdf file attached above.

Most importantly, a lot of duplicated data had to be dropped. The data set contains different snapshots of the same loan(s) as at different dates. The most recent iteration of these loans(sorted by date) were used for the purpose of this analysis with all earlier iterations discarded. This was necessary to get the actual loan portfolio, balance yet to be repaid and other key metrics

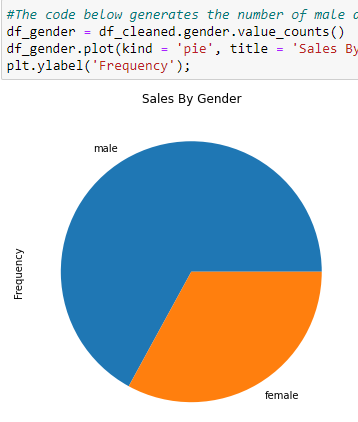
**Exploration and Insights**

I have included screenshots from the [QuickCheck Case Study.Pdf](https://drive.google.com/file/d/1tkKyPkzBg_CbCJPiBmNDin-G1QkpoT4A/view?usp=sharing) for easy reference and visualisation of the insights listed below:

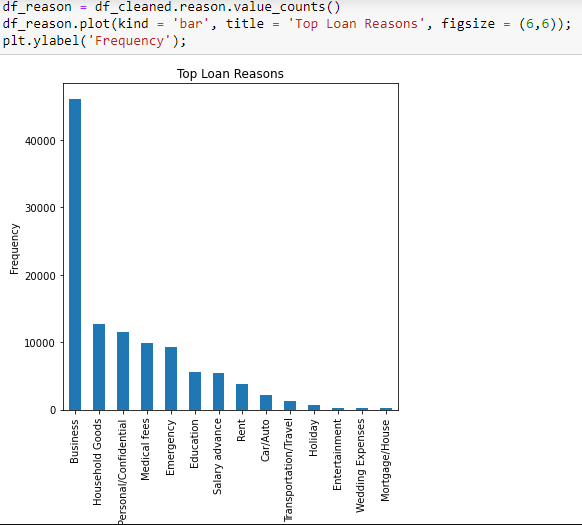
1. **Product Types**: Short loans are the most common type of loan approved by a wide margin. This is the major source of business for the company.



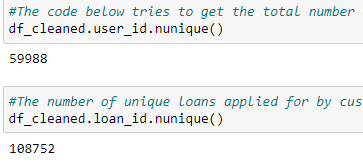
1. **Customer Gender :** Most of the company’s customers are male.



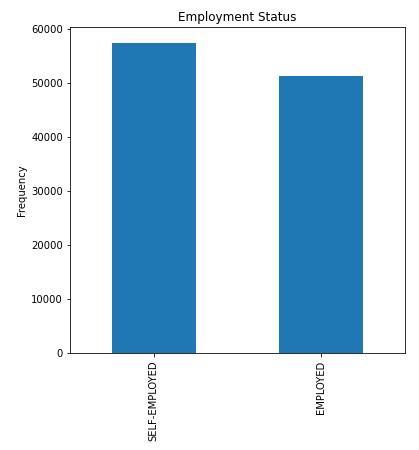
1. Top Loan Reasons: ‘Business’ was the most common reason provided by customers on loan applications.

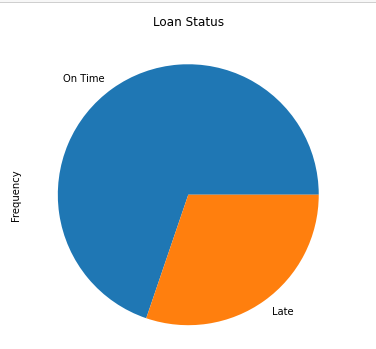


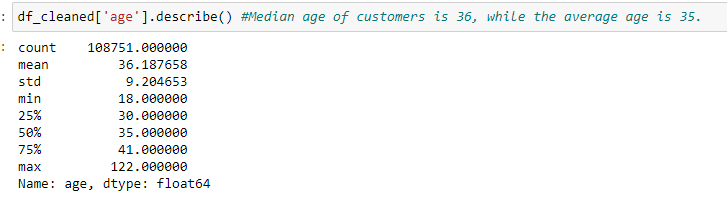
1. **Number of Users and Loans**: There are **59,988** customers captured in this dataset. Additionally about **108,752** unique loans were disbursed.



1. **Employed/Self Employed:** More customers run their own business.
2. **Loan Status:** There are more loans with repayment on schedule than late loans.
3. **Age:** The average age of our customers is about 36, while the median range is 35







1. **PAR7 and PAR30:** PAR was calculated as ‘portfolio at risk / total portfolio’. For PAR7, loans overdue by 7 days or more were considered at risk while only loans overdue by 30 days or more were considered at risk for PAR30. PAR 7 = 20.19%, while PAR30 is a little lower at 13.79%.



*Regards,*

*Olatunde Oyediran, FMVA*