

Bank Loan Case Study

Advance Excel

For our Final Project, we are provided with 3 datasets of Bank loan details. The loan-providing banks find it hard to give loans to people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming defaulters. Suppose we work for a consumer finance company that specializes in lending various types of loans to urban customers. We have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected. When a client applies for a loan, there are four types of decisions that could be taken by the company:

1. Approved: The company has approved the loan application
2. Cancelled: The client canceled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk the client received worse pricing which he did not want.
3. Refused: The company had rejected the loan (because the client does not meet their requirements etc.)
4. Unused Offer: The loan has been canceled by the client but in different stages of the process.

Project Approach Used

This project is completely based on Microsoft Excel and its advance and statistical functions. Raw data was provided in 3 different Microsoft Excel workbooks itself.

Tech Stack Used

From data cleaning, to delivering the results Microsoft Excel was used. For sharing the insights Microsoft Word was used.

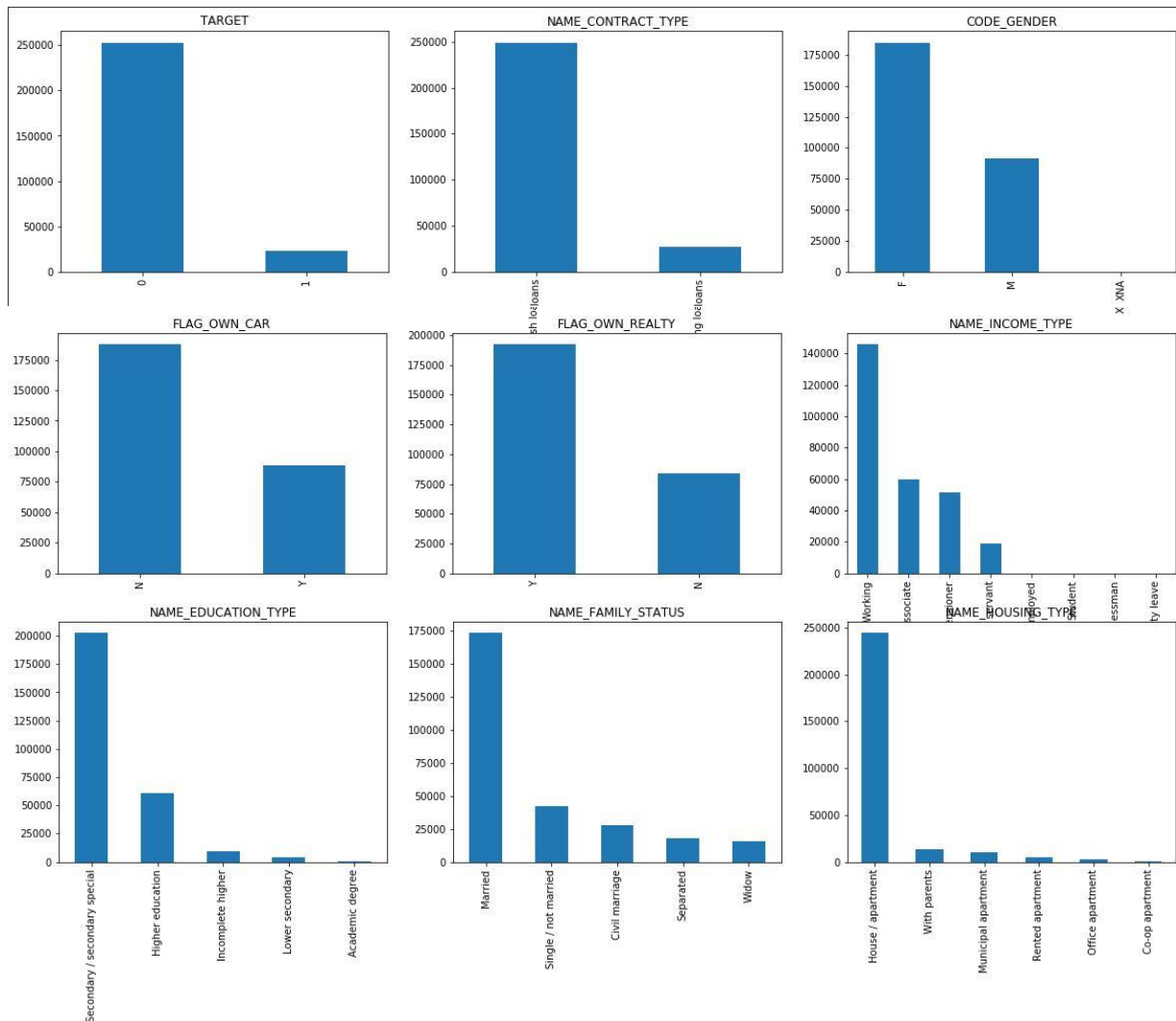
Project Insights

- **Current Applications(application_data.csv):**

Removing unnecessary columns:

We have removed all the unnecessary columns from the CSV file. And the column that has more than 30% null values, were removed too, the original data set has around 122 columns after removing the unnecessary columns we are left with only 26 columns which were useful for our analysis.

Data Imbalance-



We can see that there is a data imbalance in below columns:

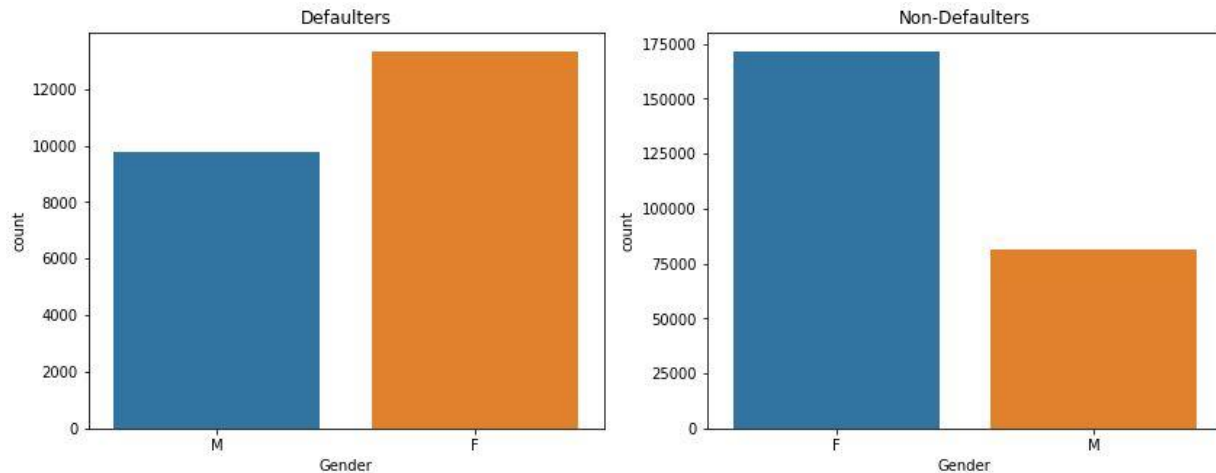
TARGET: There are very few defaulters (1) compared to non-defaulters (0)

NAME_CONTRACT_TYPE: There are very few Revolving loans than Cash loans

NAME_EDUCATION_TYPE: Most of the loans applied by Secondary/Secondary special educated people
NAME_FAMILY_STATUS: Most of the loans are applied by Married people.
NAME_HOUSING_TYPE: Most of the applications came from the Home/apartment owner

Defaulters and non-defaulters on basis of gender-

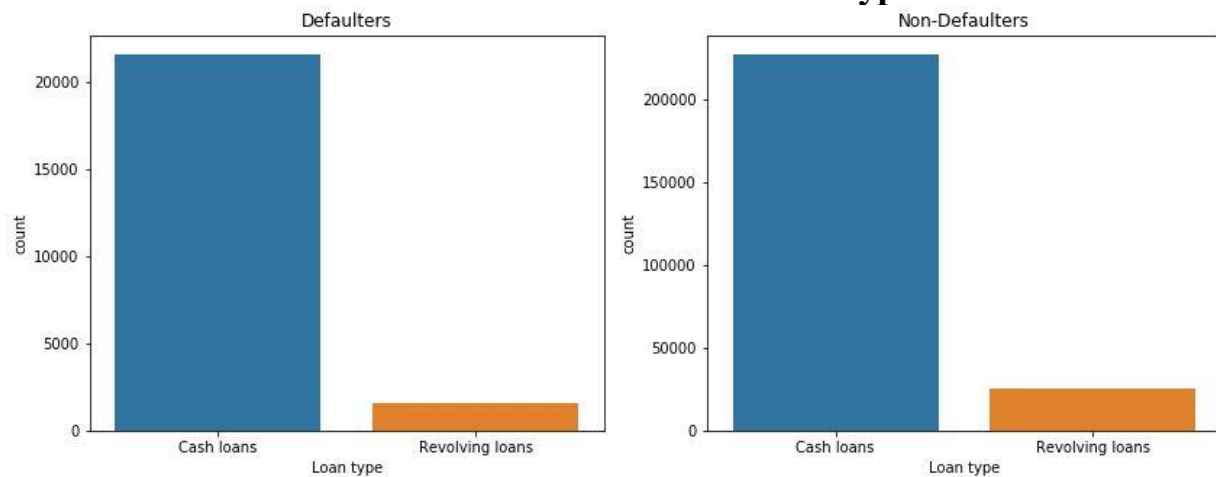
Data was divided into two different sheets one is for defaulters and another one is for non-defaulters.



Analysis

We can see that females are slightly more in the number of defaulters than males. The same pattern continues for non-defaulters as well. The females are more in number here than males.

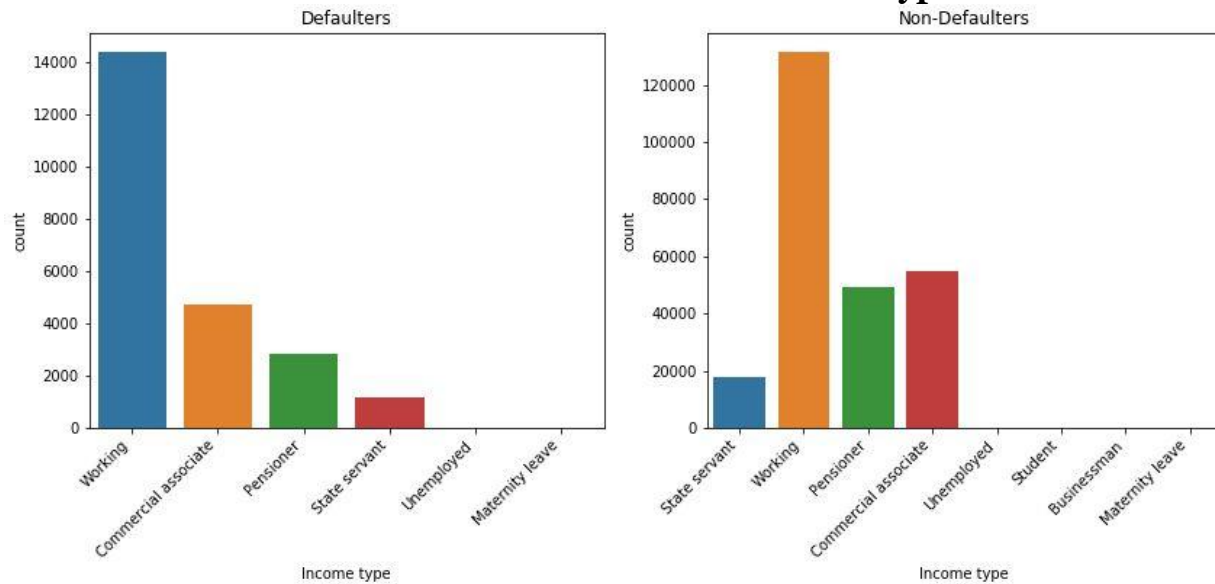
Defaulters and non-defaulters on the basis of Loan type



Analysis

We see in both cases that Revolving loans are very less in number compared to Cash loans.

Defaulters and non-defaulters on the basis of Income type

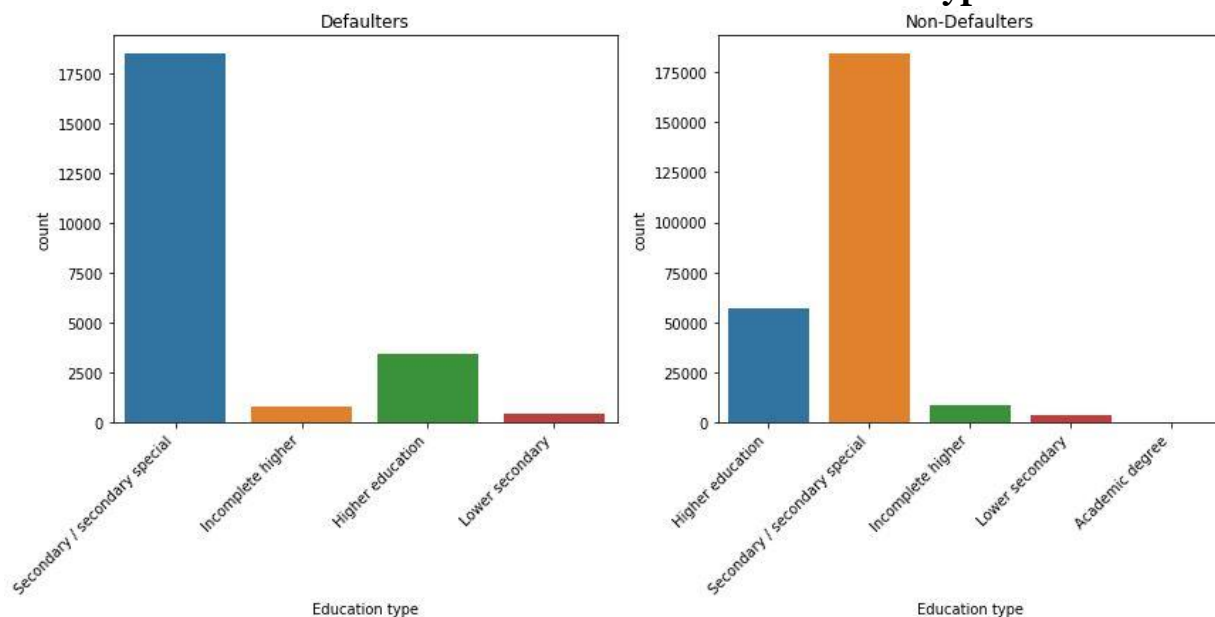


Analysis

Working people have mostly defaulted as their numbers are high with compare to other professions.

Similarly, here also working people are more in the number who do not default.

Defaulters and non-defaulters on the basis of Education type

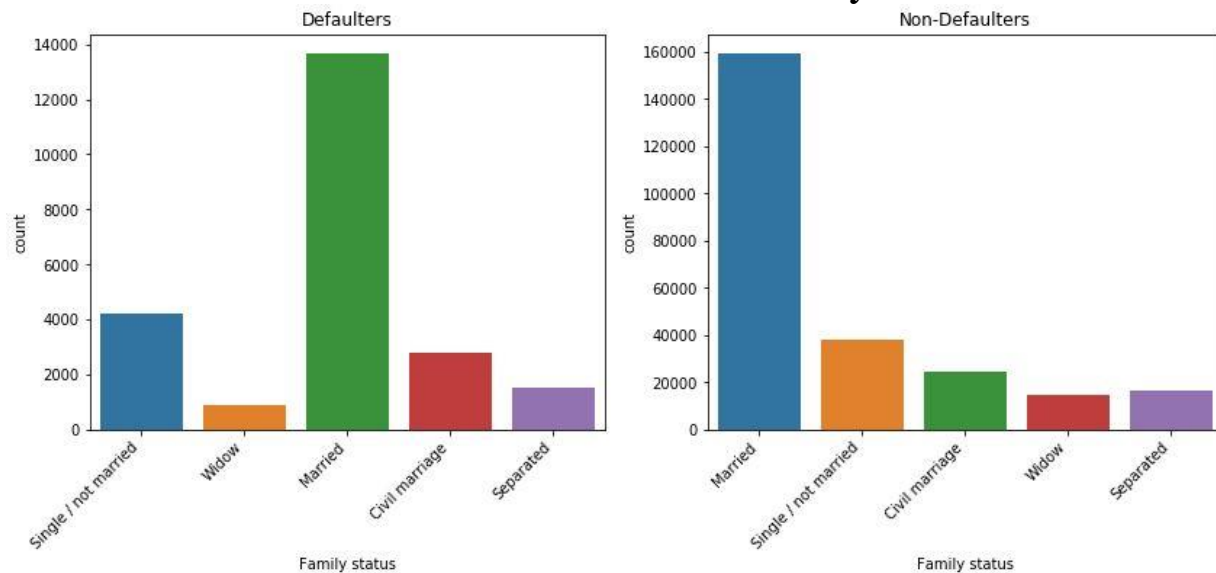


Analysis

Education with Secondary/Secondary special customers are more number in defaulters compared with other levels of educated people.

Here also Secondary/Secondary special are more in numbers.

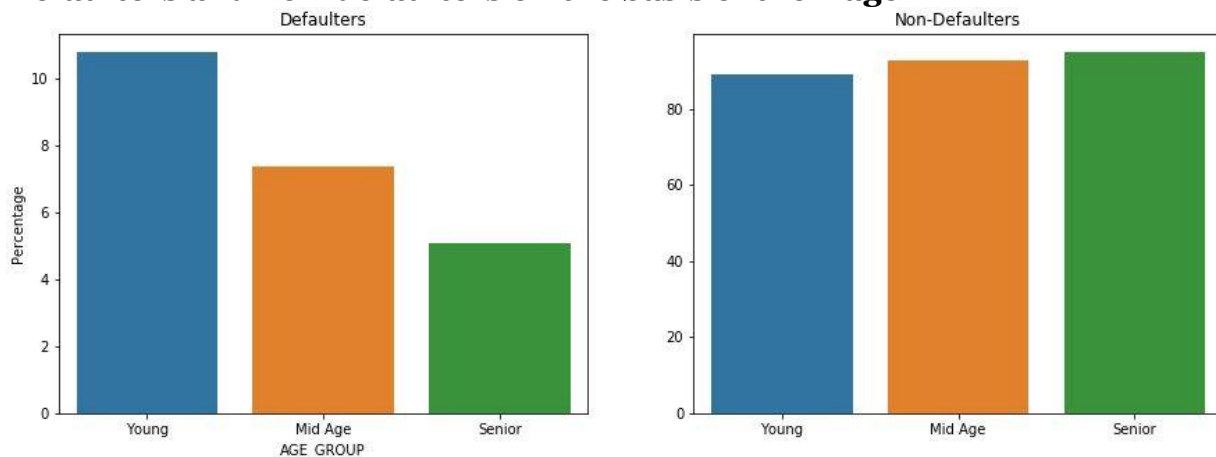
Defaulters and non-defaulters on the basis of Family status



Analysis

For both the customers (defaulters and non-defaulters) married people are more in number compared with single, separated, widows, etc.

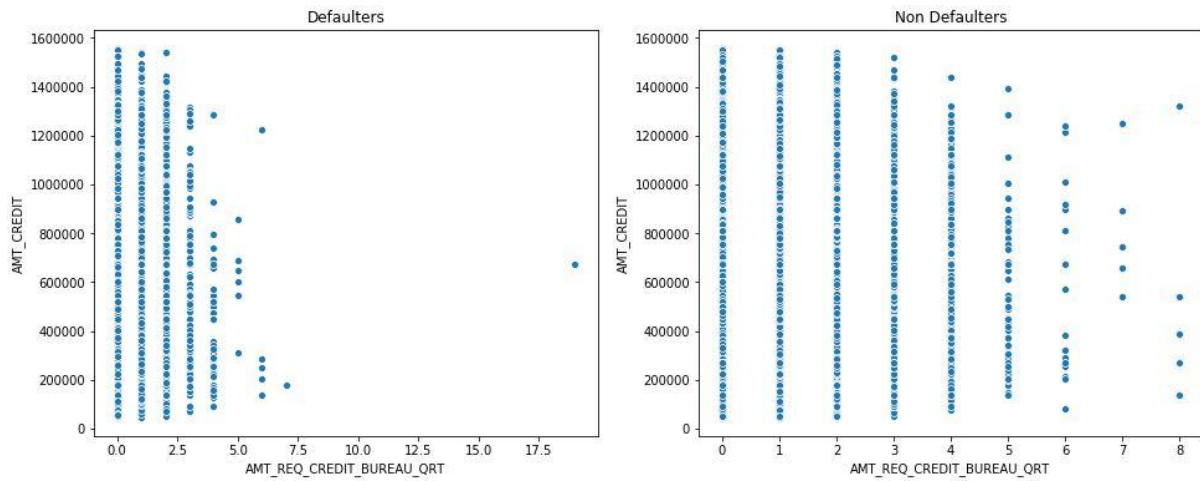
Defaulters and non-defaulters on the basis of their age



Analysis

We see that Young people are more likely to default than the other two age groups. Whereas, Senior citizens are less likely to default than others. There is not much difference in the likelihood of non-defaulters in the age groups.

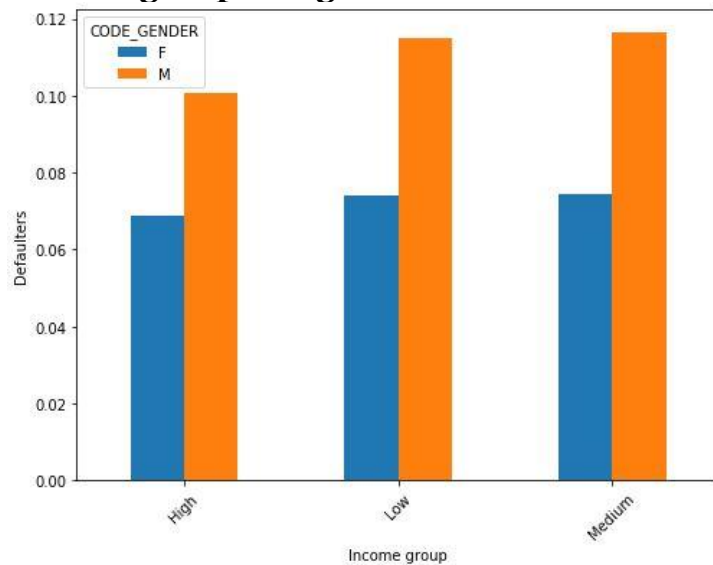
Credit amount of the loan according to the Credit Bureau



Analysis

We see that the greater number of inquiries the lesser the amount of loan credited for both defaulters and non-defaulters.

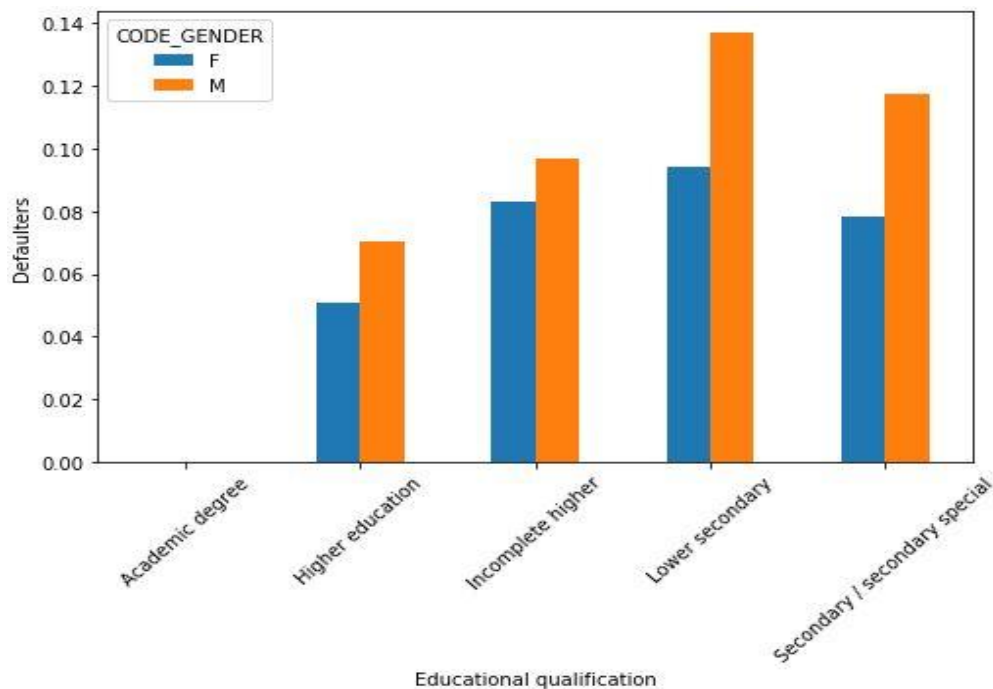
Income group and gender



Analysis

We can see that Males are more likely to default than Females across all income groups.

Education and gender



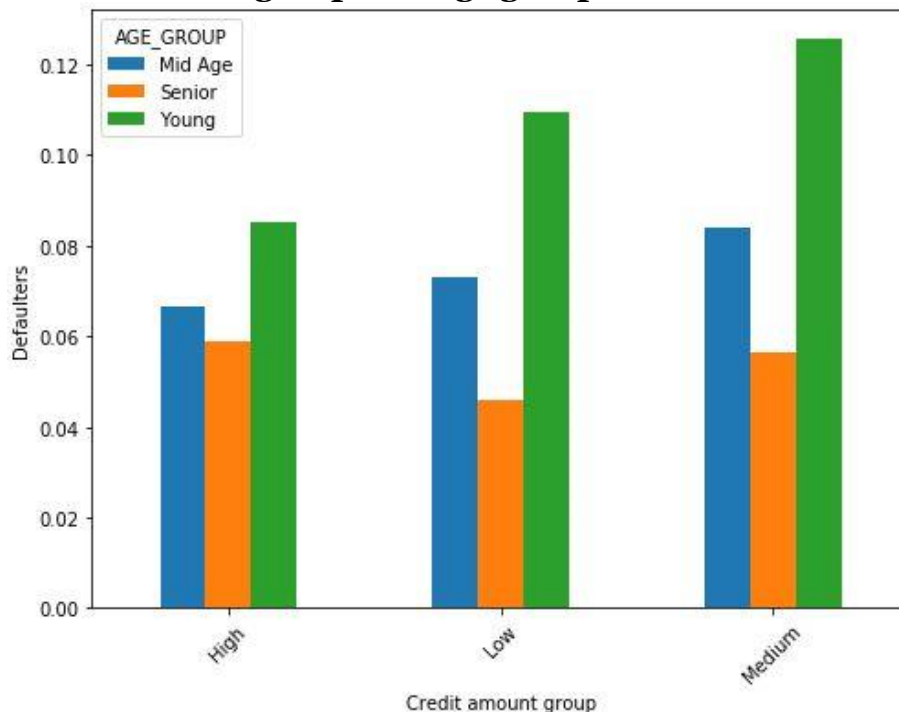
Analysis

Lower secondary educated clients have done more defaults followed by Secondary and Incomplete clients.

The Higher educated clients have done fewer defaults.

Across all educated levels, Females are less likely to do a default.

Credit amount group and Age group

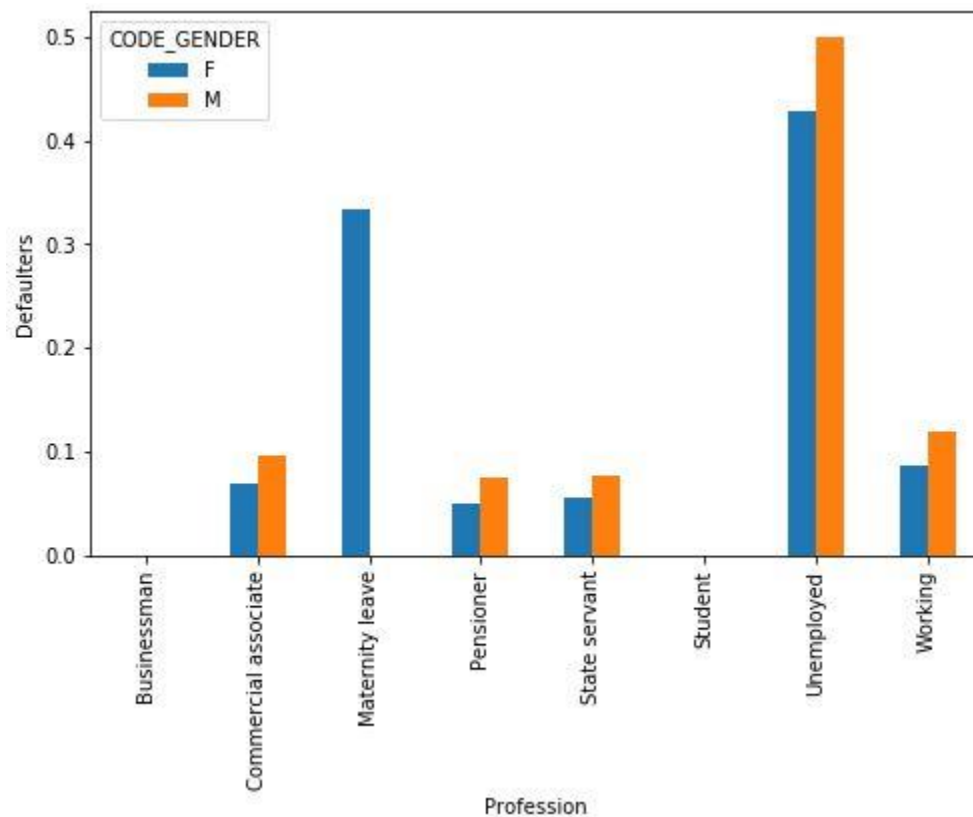


Analysis

Young clients with medium and low credit amount group are highly defaulted.

Senior citizens across all credit amount groups are less likely to default.

Profession and Gender



Analysis

No surprise that unemployed clients are more defaulted.

Clients with maternity leave are expected to default more.

The default rate is lesser in all other professions.

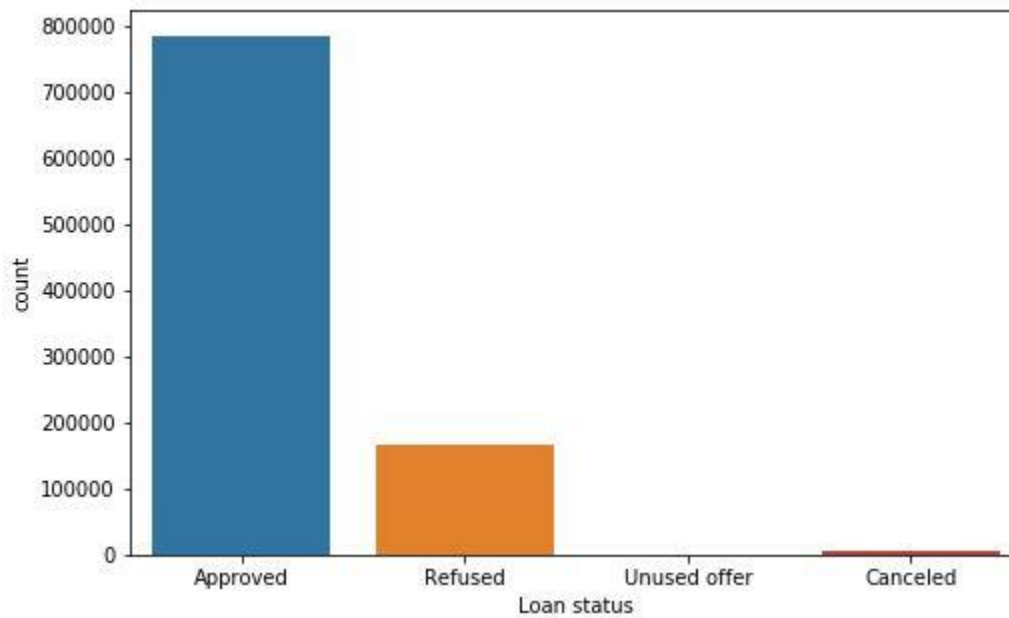
Males are more defaulted on their respective professions compared to females.

- **Previous Applications (pervious_application.csv)**

Removing unnecessary columns:

We have removed all the unnecessary columns from the CSV file. And the column that has more than 30% null values, was removed too.

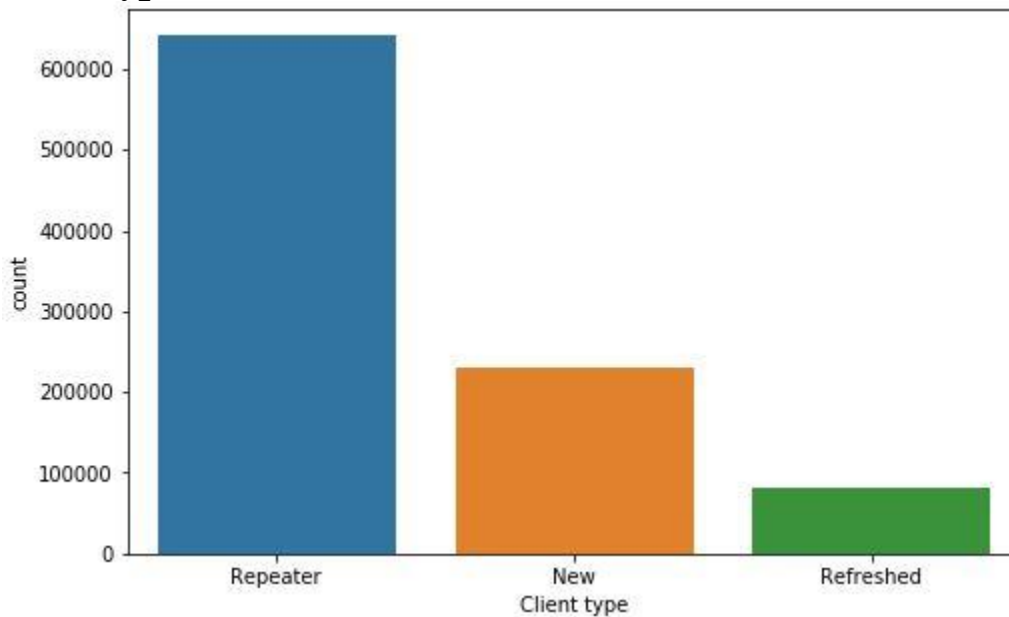
Previous Loan status



Analysis

There is a huge number of the Approved loan than Refused. Hardly, there is any Canceled or Unused offer loan.

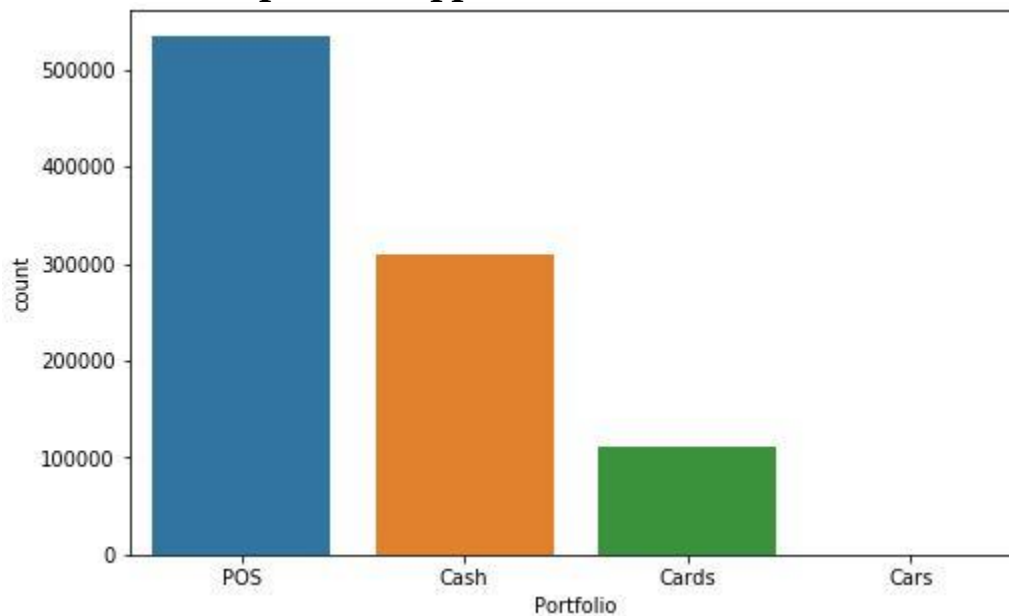
Client type



Analysis

Mostly the applicants were Repeaters.

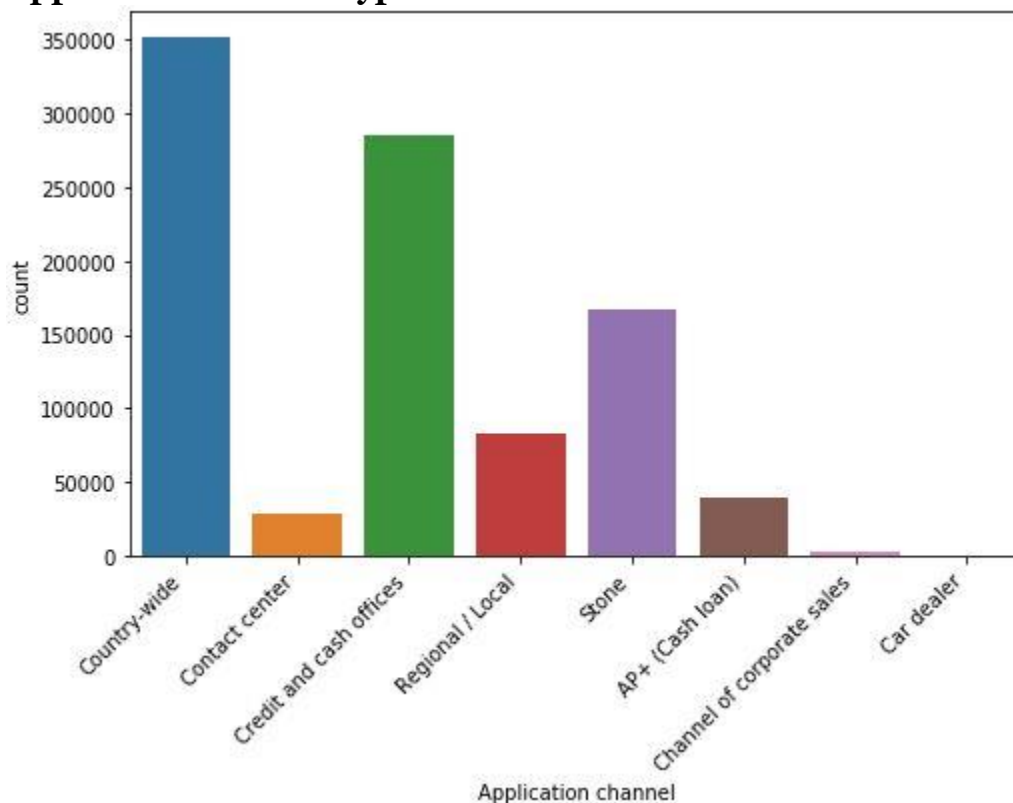
Portfolio of the previous applications



Analysis

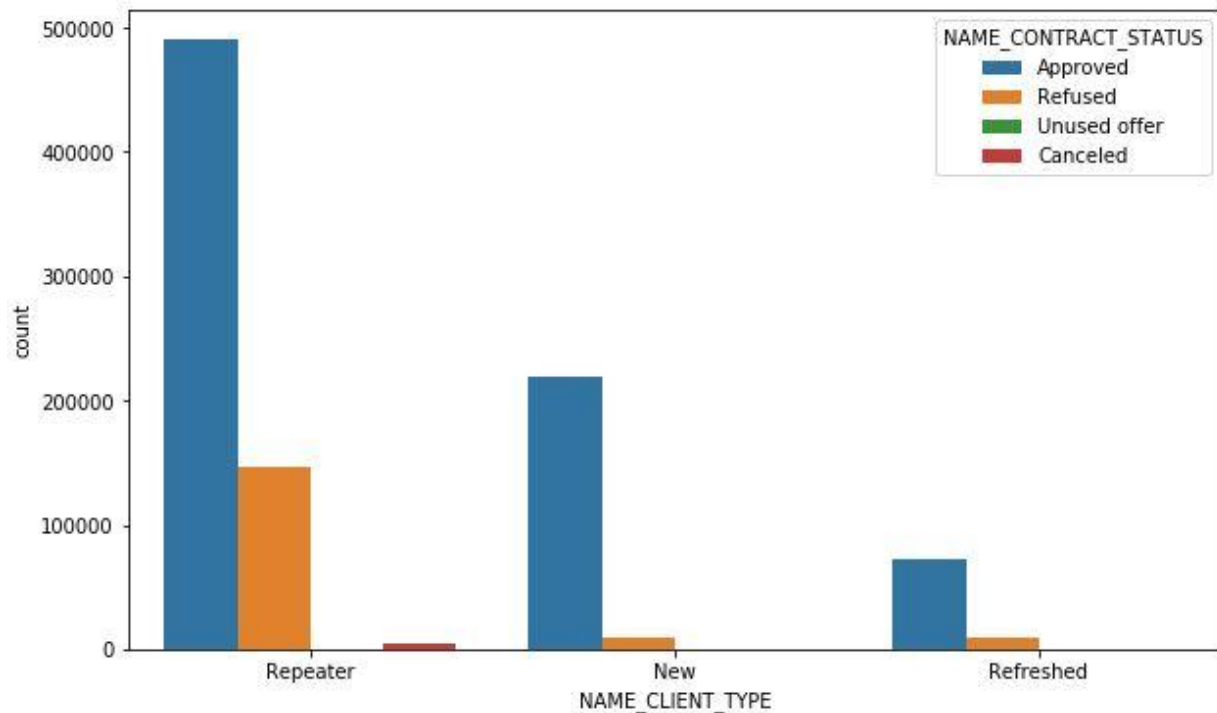
The highest number of the previous applications was for POS. Applications for Cash also have a good number. Applications for Cards were very few.

Application channel type



We see that Country-wide was heavily used for previous applications followed by Credit and Cash offices, Stone, and Regional. The rest of other channels are hardly used.

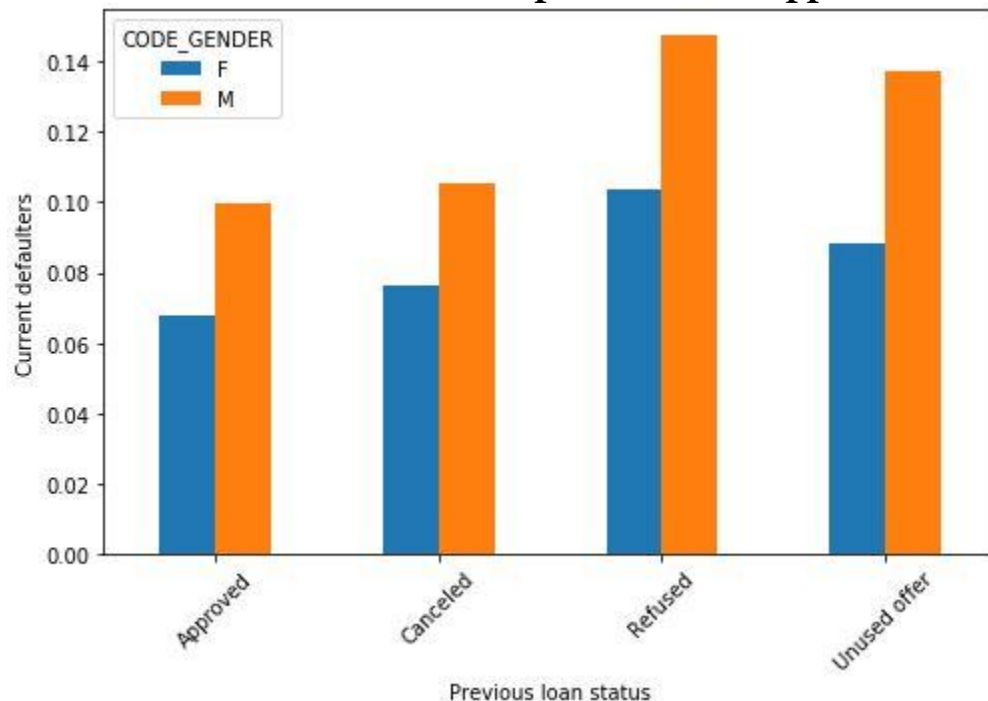
Status and Client type



Analysis

We see that the Repeater clients have more approved loans than New and Refreshed clients.

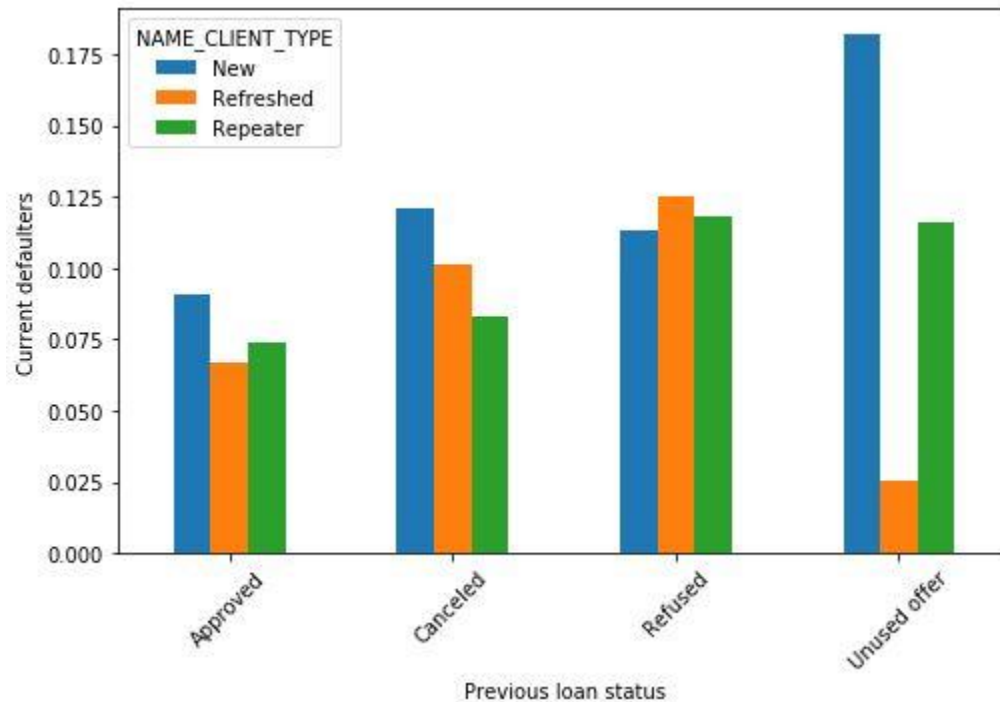
Current loan defaulter status to previous loan application status



Analysis

We see that previously Refused clients have done more defaults than previously Approved clients. Also, in all the cases Males have done more defaults than Females.

Current loan defaulter status to previous loan application status and client types



Analysis

We can see that the Defaulters are more for previously Unused offers loan status clients, who were New.

For previously Approved status the New clients were more defaulted followed by Repeater.

For previously Refused applicants the Defaulters are more Refreshed clients.

For previously Canceled applicants the Defaulters are more New clients.

RESULTS

After running all the formulas in Microsoft Excel and plotting the charts we analyzed all the data sets of this Project.

In the making of this report, we used our Microsoft Excel knowledge in a real-world example.

DRIVE LINK

<https://drive.google.com/drive/folders/1tk6jKHpyzKUfn86blXt3xBbRyn-oVZoI?usp=sharing>