Bank Loan Case Study

(Final Project-2)

Project Description :

We are provided with couple of datasets of Bank loan details. The loan providing banks find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose we work for a consumer finance company which specialises in lending various types of loans to urban customers. We have to use EDA to analyse 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 loan application

2. Cancelled : The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he 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 : Loan has been cancelled by the client but on different stages of the process.

This case study aims to identify the patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending at a higher interest rate, etc. This will ensure that the customers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

For this case study, we are provided with 3 datasets as explained below:

1. "application_data.csv" contains all the information of the client at the time of application. The data is about whether a client has any payment difficulty.

2. "previous_application.csv" contains information about the client's loan history. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

3. "columns description.csv"

In this case study, we will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Approach :

Here we will try to identify patterns using EDA with Excel & Python (Dataset too large to load in Excel) on given databases to let company understand driving factors behind loan default.
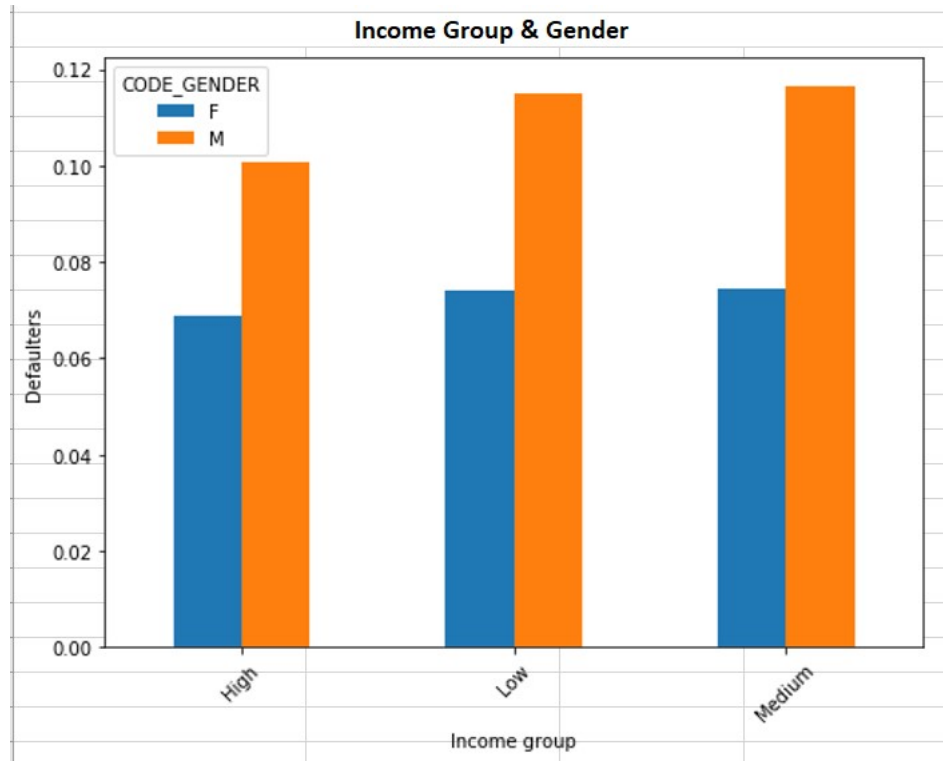
Tech-Stack Used :

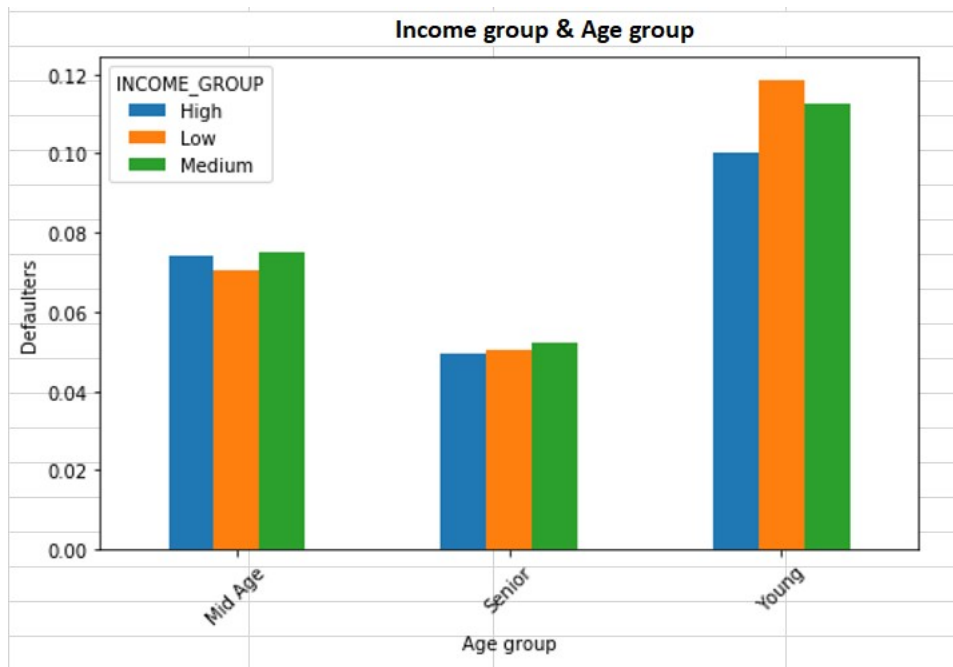Excel by Microsoft Corporation – For carrying out EDA on the datasets & Visualisation.

Jupyter Notebook By Google - For carrying out EDA on the datasets & Visualisation.

WordPad by Microsoft Corporation – For creating the project report.

Insights :

Current Applications (application_data.csv) :

Income group & Age group

Observations :

1.People with High Income Groups are less defaulters compared to Low Income Groups.

2.Mid age & Senior Age Groups including all income groups are less defaulters.

Recommended :

1.It is safe to grant loan to Mid Age & Senior Age Groups with higher income

2.It is risky to grant loan to young people with low income

Family status & Age group

Family status & Gender
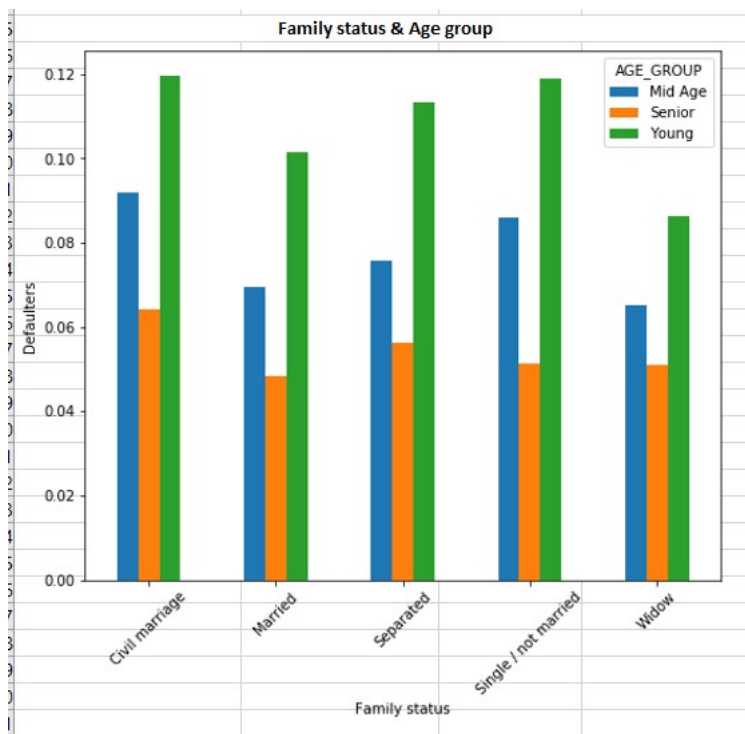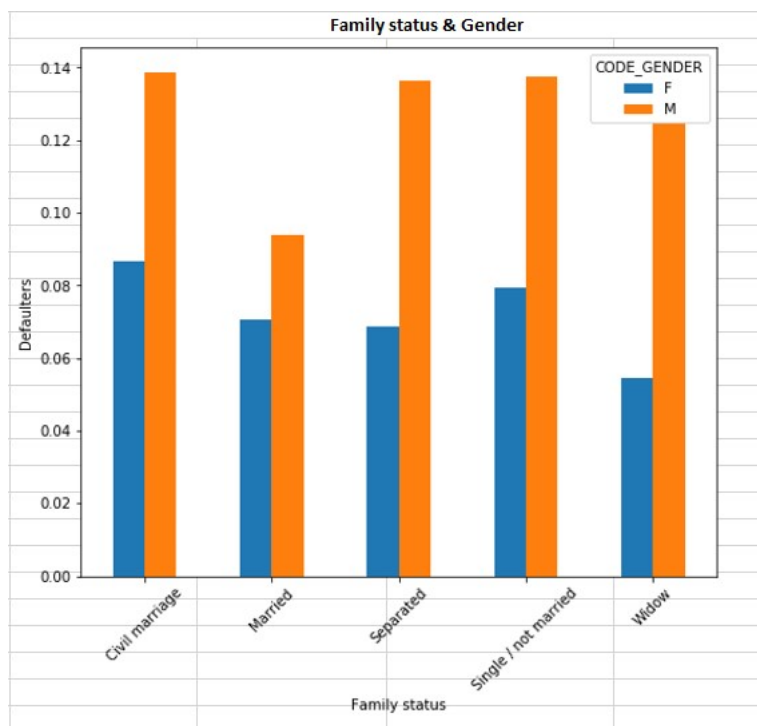
Observations:

1.Seniors irrespective or family status are less likely to be defaulted.

2.Young people are more likely to be defaulted in all Family Statuses.

3.Males are more likely to be defaulted than Females.

Recommended:

1.It is safe to grant loan to all Family status' Senior Age group.

2.It is risky to grant loan to Single, Separated & Civil Marriage of Young Age group.

Credit amount group & Income group

Credit amount group & Age group
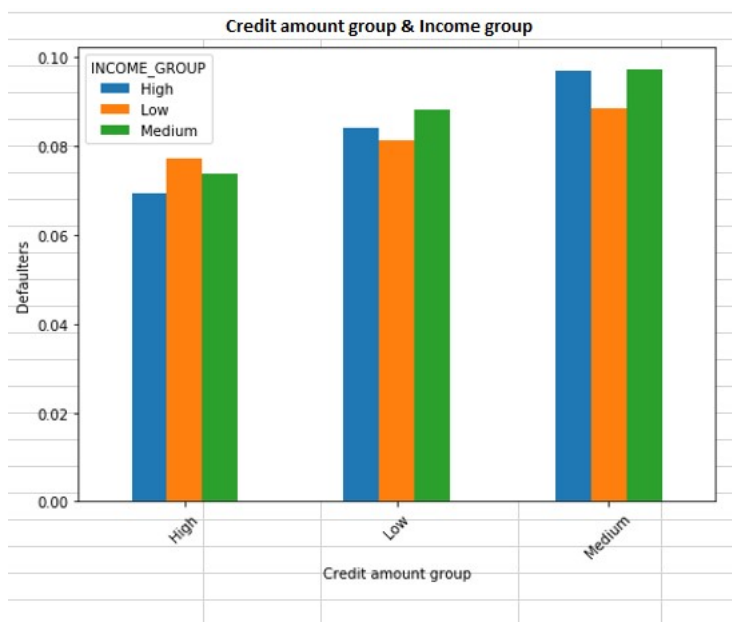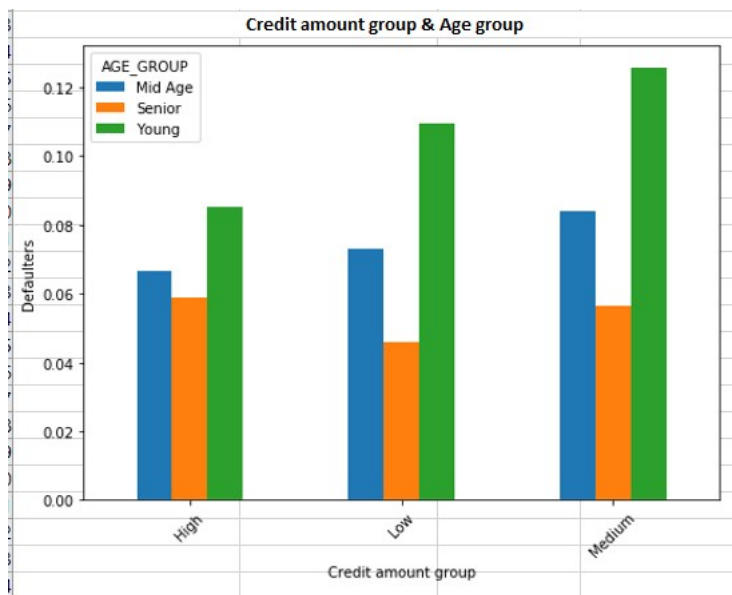
Observations :
1.Across all Income Group clients with Medium Credit Amount is highly defaulted, followed by Low & High Credit Amount
2.Young Age Groups with Medium & Low Credit Amount are most likely defaulted.

Recommended:
1.It is very risky to grant Medium & Low Credit Amount to Young Age Group.

Analysis of two segmented variables

Education & Gender
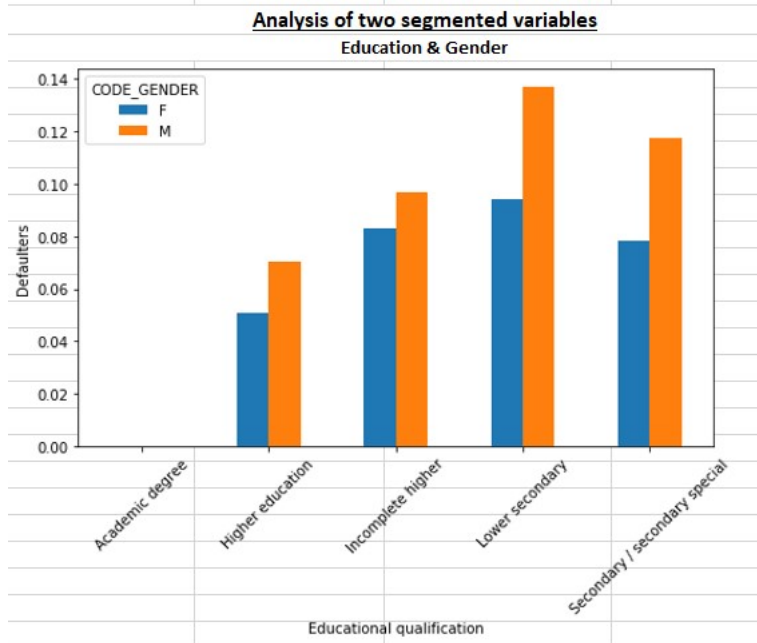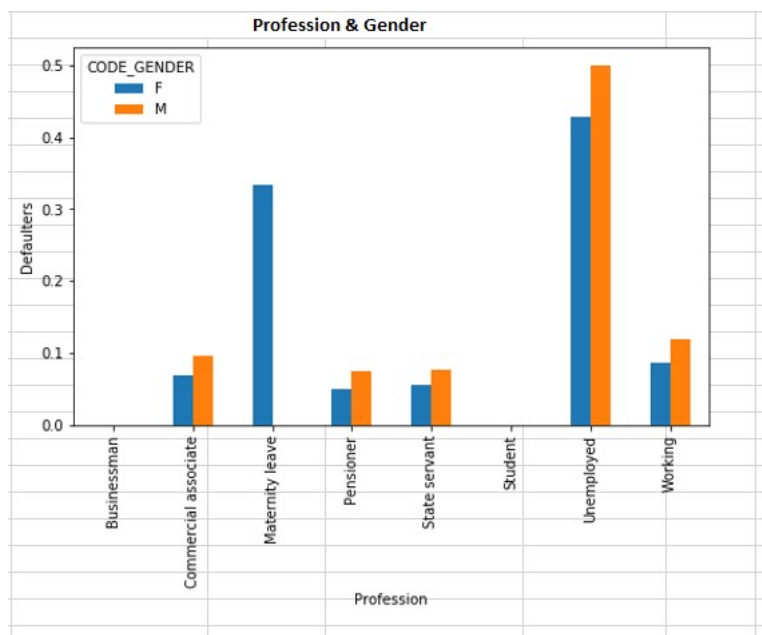
Observations:

1.People with Higher Education are less defaulted & People with Lower Secondary education are more defaulted

2.Unemployed & Maternity Leave clients are very much defaulted

Recommended:

1.It is safe to grant loan to people with Higher Education in all professions except Females & Females with Maternity Leave.

Loan Application Status Relations

Current & Previous (application_data.csv & previous_application.csv) :

Current loan defaulter status with respect to previous loan application status and client types
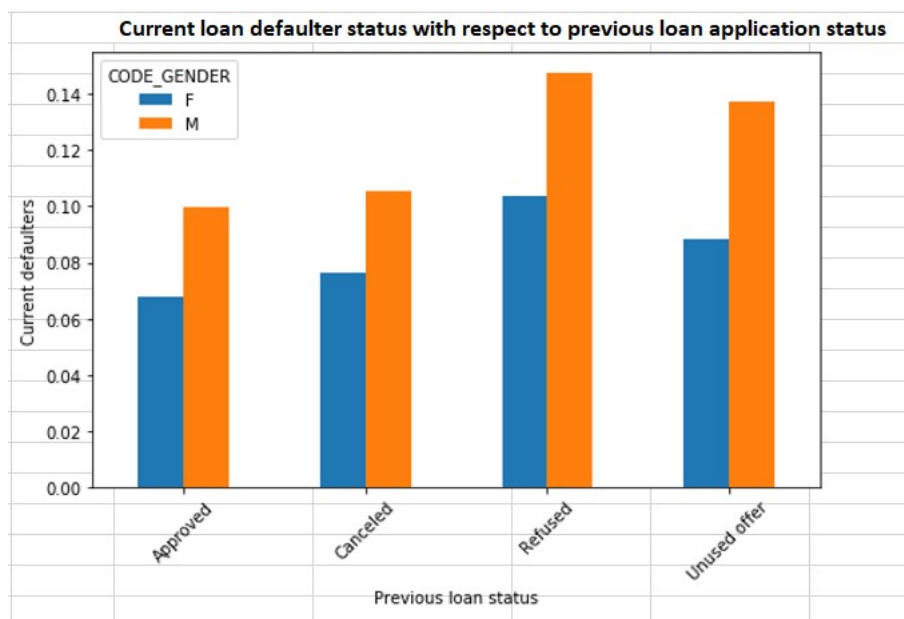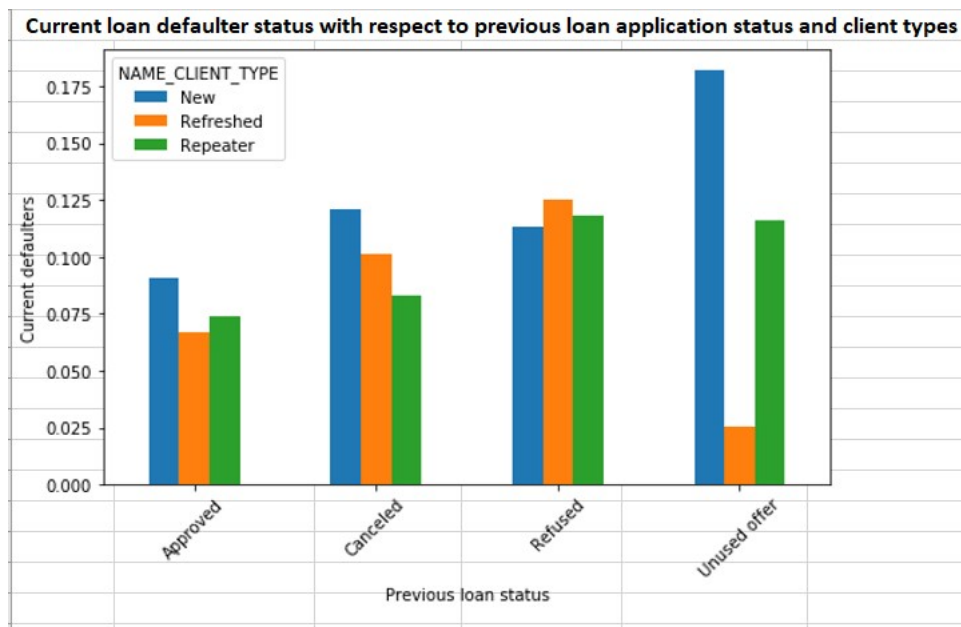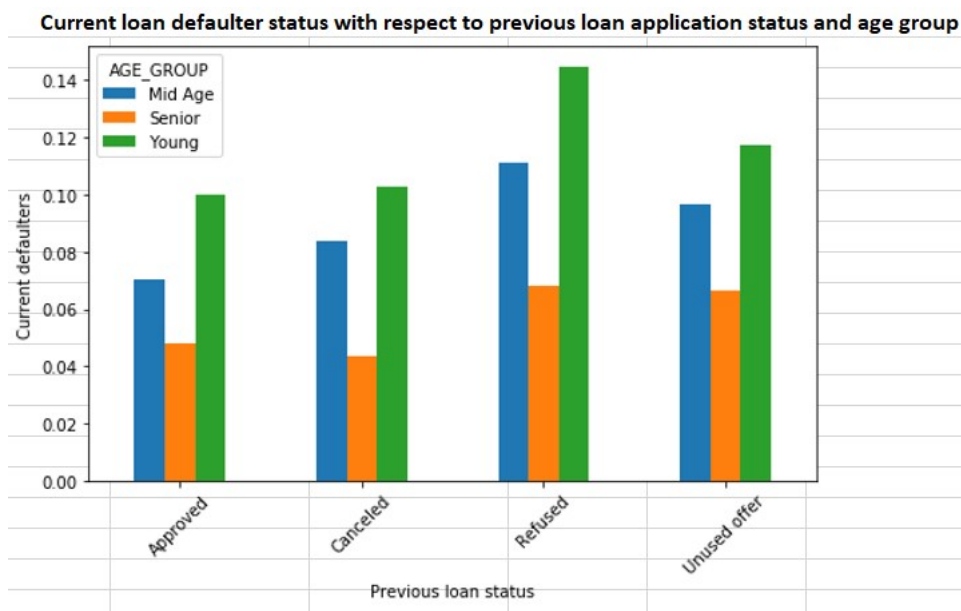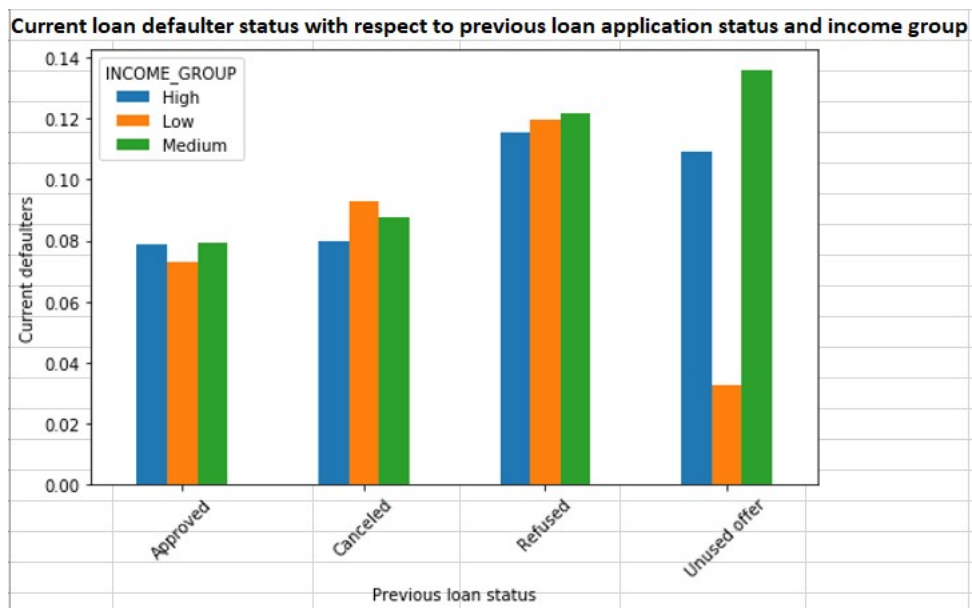
Observations :

1.Previously Refused & Unused offer applications were more defaulted in Males.

2.New clients with previously Unused offers are more defaulted.

Recommended :

1.It is safe to provide loans to previously approved Females.

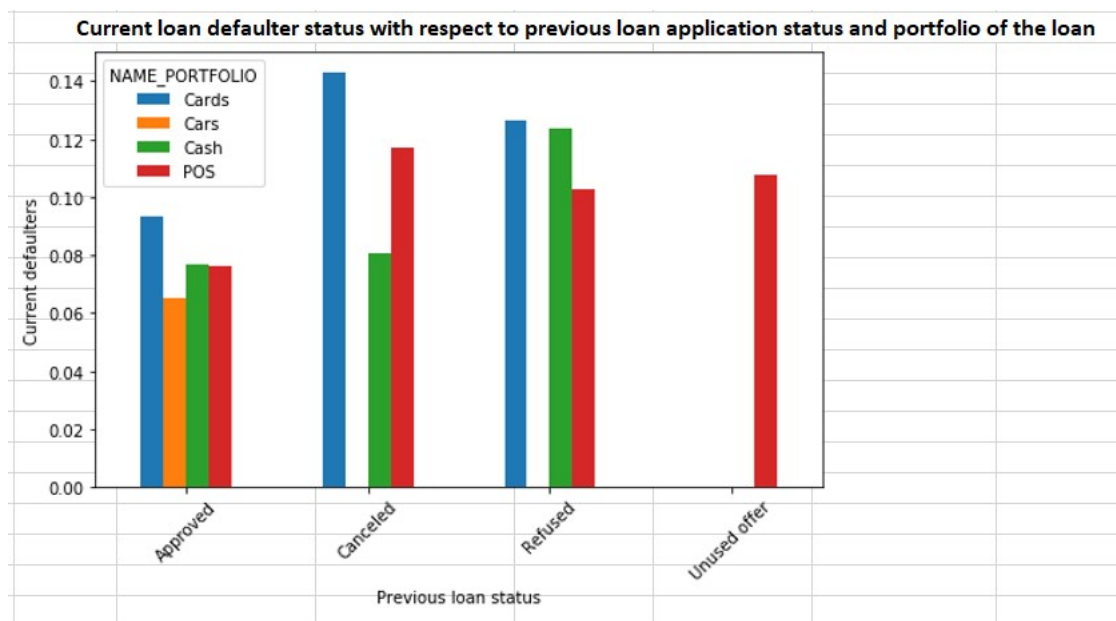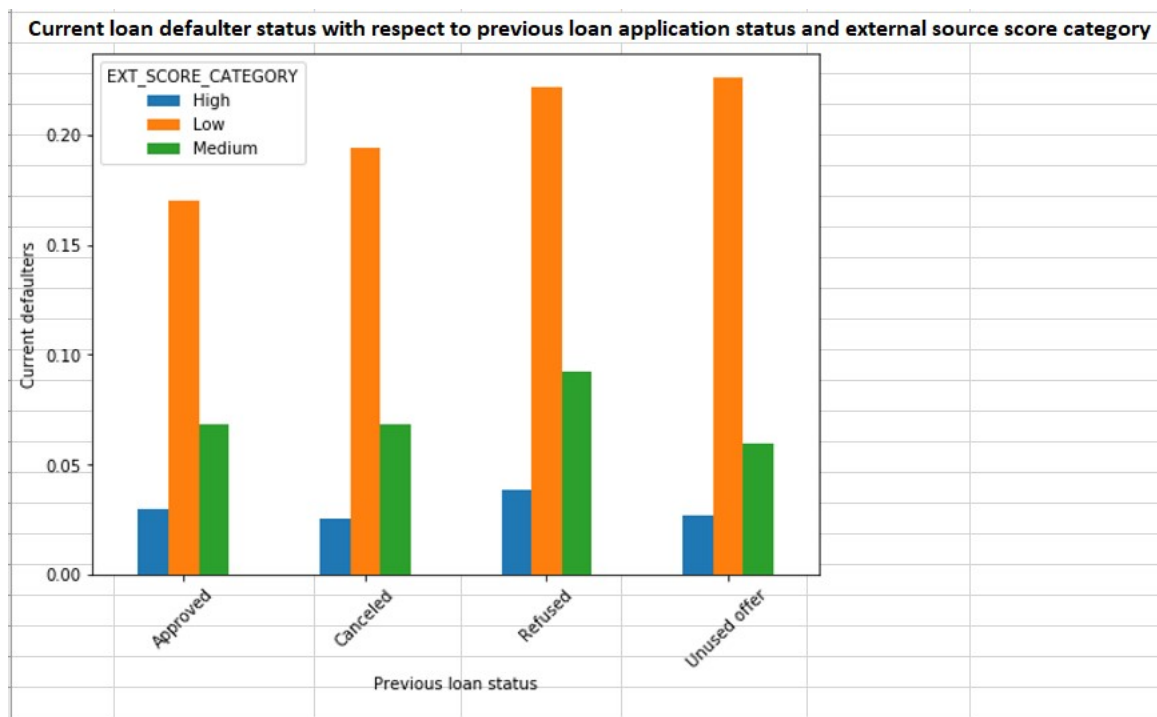2.It is risky to grant loans to clients whose applications were previously Refused & Unused offer.

Current loan defaulter status with respect to previous loan application status and age group

Current loan defaulter status with respect to previous loan application status and income group

Observations:

1.Young clients which are previously Refused are highly defaulted.

2.Senior clients are less defaulted irrespective of their previous loan status.

3.Previously refused application in all Income Groups are highly defaulted.

Recommended:

1.It is safe to grant loan to Senior clients.

2.It is less risky to grant loans for approved applicants in all Income Groups.

Current loan defaulter status with respect to previous loan application status and portfolio of the loan

Current loan defaulter status with respect to previous loan application status and external source score category

Observations:

1.The previous applications for Cards & POS are mostly defaulted.

2.Previously refused application for Cash is highly defaulted.

3.Low External Source Scorer are highly defaulted.

Recommended:

1.It is safe to grant loan to any portfolio for previously approved clients.

2.It is highly risky to grant loans to clients with poor external source score whose loan was previously Refused, Unused offer or Cancelled.

Result:

We carried out EDA on the given datasets using Microsoft Excel & Jupyter Notebook-Python (dataset is too large to work on Excel) and got answers as following which helped us to take data driven decisions.

Highly Recommended Groups :

1.Approved clients in previous applications

2.Clients with Higher Education with High income

3.Clients with Higher external source score

4.Senior clients in all categories

5.Married clients

6.Females

High Risk Groups :

1.Clients who are previously Refused, Cancelled or Unused offer.

2.ow income groups with previously refused status.

3.Clients who are Unemployed.

4.Clients with Poor external source score.

5.Clients who are Young.

6.Clients who are educated Lower Secondary & Secondary.

Please check application_data.xlsx, previous_application.xlsx & Bank_Loan_Case_Study.ipynb for all tables, charts and analysis in Drive folder.