DATA-DRIVEN STRATEGIES

Credit Score

CREDIT CARD LOAN REPAYMENT DETECTION

Trends

Done By: Prisha Kundu

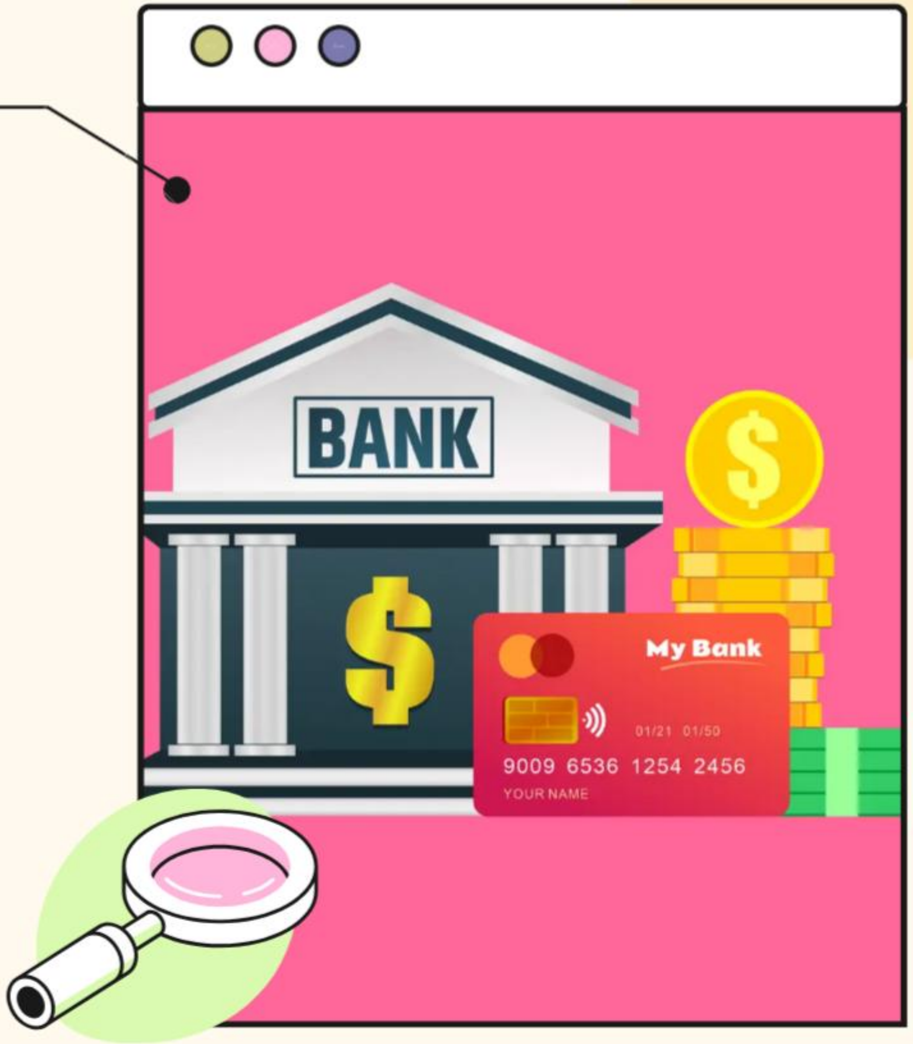
Debt Repayment

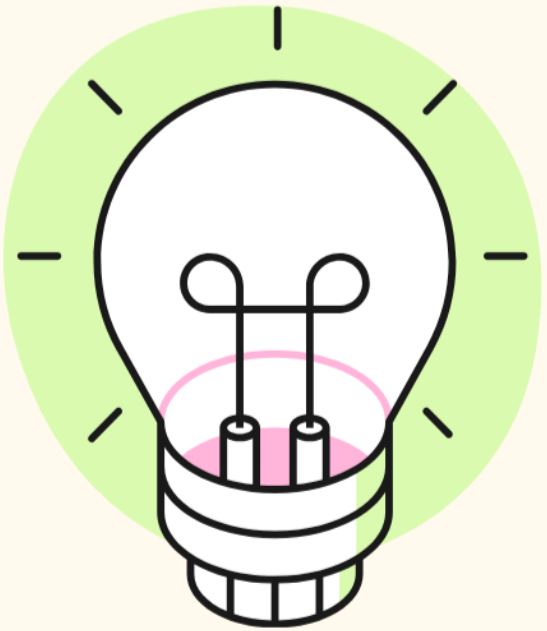
Feedback

INTRODUCTION

The loan providing companies find it hard to give loans to people due to their inadequate or missing credit history. Some consumers use this to their advantage by becoming a defaulter.

By using Exploratory Data analysis, patterns present in the particular dataset can be analyzed which will make sure that the loans are not rejected for the applicants capable of repaying.

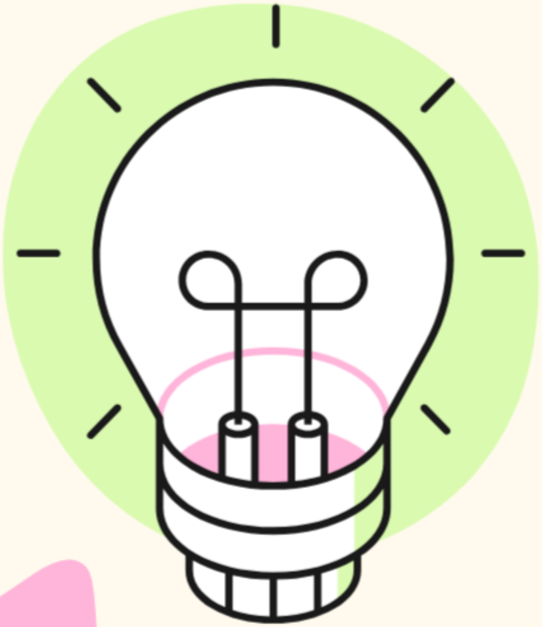




BUSINESS PROBLEM UNDERSTANDING

When the company receives a loan application, the company has to rights for loan approval based on the applicant's profile. Two types of risks are associated with the bank's or company's decision:

- If the aspirant is likely to repay the loan, then not approving the loan tends in a business loss to the company
- If the a is aspirant not likely to repay the loan, i.e. he/she is likely to default/fraud, then approving the loan may lead to a financial loss for the company.



The data contains information about the loan application. When a client applies for a loan, there are four types of decisions that could be taken by the bank/company:

1. **Approved**
2. **Cancelled**
3. **Refused**
4. **Unused offer: The loan has been cancelled by the applicant but at different stages of the process.**

In this project ,stakeholders will be financial institution which issues the loan .

OUR GOAL

Our project aims to identify patterns which indicate whether an applicant would be able to repay their installments which may be used for taking further actions such as denying the loan, reducing the amount of loan, lending at a higher interest rate, etc.

This will make sure that the applicants capable of repaying the loan are not rejected.

The use of EDA techniques using python forms the main base of this project

DATA PREPROCESSING

VARIOUS STEP FOR PREPROCESSING

Reading the file and accessing

```
df1 = pd.read_csv("application_data.csv")
df1.head()
```

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_(
0	100002	1	Cash loans	M	N	Y	0	202500.0	406597.5	24700.5	
1	100003	0	Cash loans	F	N	N	0	270000.0	1293502.5	35698.5	
2	100004	0	Revolving loans	M	Y	Y	0	67500.0	135000.0	6750.0	
3	100006	0	Cash loans	F	N	Y	0	135000.0	312682.5	29686.5	
4	100007	0	Cash loans	M	N	Y	0	121500.0	513000.0	21865.5	

The number of rows and columns

```
df1.shape
```

(307511, 122)

Statistical analysis

```
df1.describe()
```

	SK_ID_CURR	TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAY!
count	307511.000000	307511.000000	307511.000000	3.075110e+05	3.075110e+05	307499.000000	3.072330e+05	307511.000000	307511.000000	3(
mean	278180.518577	0.080729	0.417052	1.687979e+05	5.990260e+05	27108.573909	5.383962e+05	0.020868	-16036.995067	(
std	102790.175348	0.272419	0.722121	2.371231e+05	4.024908e+05	14493.737315	3.694465e+05	0.013831	4363.988632	1
min	100002.000000	0.000000	0.000000	2.565000e+04	4.500000e+04	1615.500000	4.050000e+04	0.000290	-25229.000000	-
25%	189145.500000	0.000000	0.000000	1.125000e+05	2.700000e+05	16524.000000	2.385000e+05	0.010006	-19682.000000	
50%	278202.000000	0.000000	0.000000	1.471500e+05	5.135310e+05	24903.000000	4.500000e+05	0.018850	-15750.000000	
75%	367142.500000	0.000000	1.000000	2.025000e+05	8.086500e+05	34596.000000	6.795000e+05	0.028663	-12413.000000	
max	456255.000000	1.000000	19.000000	1.170000e+08	4.050000e+06	258025.500000	4.050000e+06	0.072508	-7489.000000	3(

Identifying null values

```
(df1.isnull().sum()/len(df1)*100).sort_values(ascending = False).head(50)
```

```
COMMONAREA_MEDI                 69.872297
COMMONAREA_AVG                  69.872297
COMMONAREA_MODE                 69.872297
NONLIVINGAPARTMENTS_MODE        69.432963
NONLIVINGAPARTMENTS_MEDI        69.432963
NONLIVINGAPARTMENTS_AVG         69.432963
FONDKAPREMONT_MODE              68.386172
LIVINGAPARTMENTS_MEDI           68.354953
LIVINGAPARTMENTS_MODE           68.354953
LIVINGAPARTMENTS_AVG            68.354953
FLOORSMIN_MEDI                  67.848630
FLOORSMIN_MODE                  67.848630
FLOORSMIN_AVG                   67.848630
YEARS_BUILD_MEDI                66.497784
YEARS_BUILD_AVG                 66.497784
YEARS_BUILD_MODE                66.497784
OWN_CAR_AGE                     65.990810
LANDAREA_MODE                   59.376738
LANDAREA_AVG                    59.376738
LANDAREA_MEDI                   59.376738
BASEMENTAREA_MEDI               58.515956
BASEMENTAREA_AVG                58.515956
BASEMENTAREA_MODE               58.515956
EXT_SOURCE_1                    56.381073
NONLIVINGAREA_MEDI              55.179164
NONLIVINGAREA_AVG               55.179164
NONLIVINGAREA_MODE              55.179164
ELEVATORS_MODE                  53.295980
ELEVATORS_AVG                   53.295980
ELEVATORS_MEDI                  53.295980
WALLSMATERIAL_MODE              50.840783
APARTMENTS_MODE                 50.749729
APARTMENTS_AVG                  50.749729
APARTMENTS_MEDI                 50.749729
ENTRANCES_MEDI                  50.348768
ENTRANCES_MODE                  50.348768
```

Dropping columns with null values

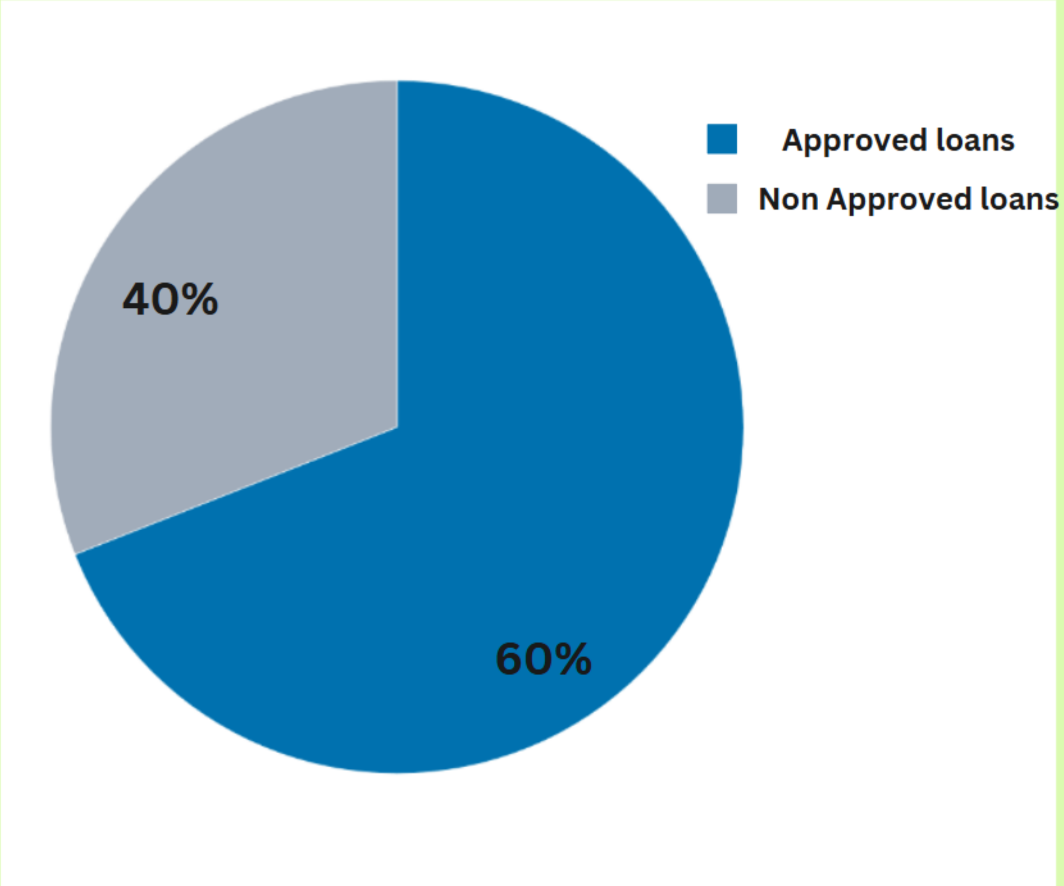
```
In [ ]: dfresult = df1.dropna(axis=1)
        print(dfresult)
```

```
df1.shape
```

```
(307511, 73)
```

DATA ANALYSIS

ANALYSIS OF APPROVED AND NON APPROVED LOANS



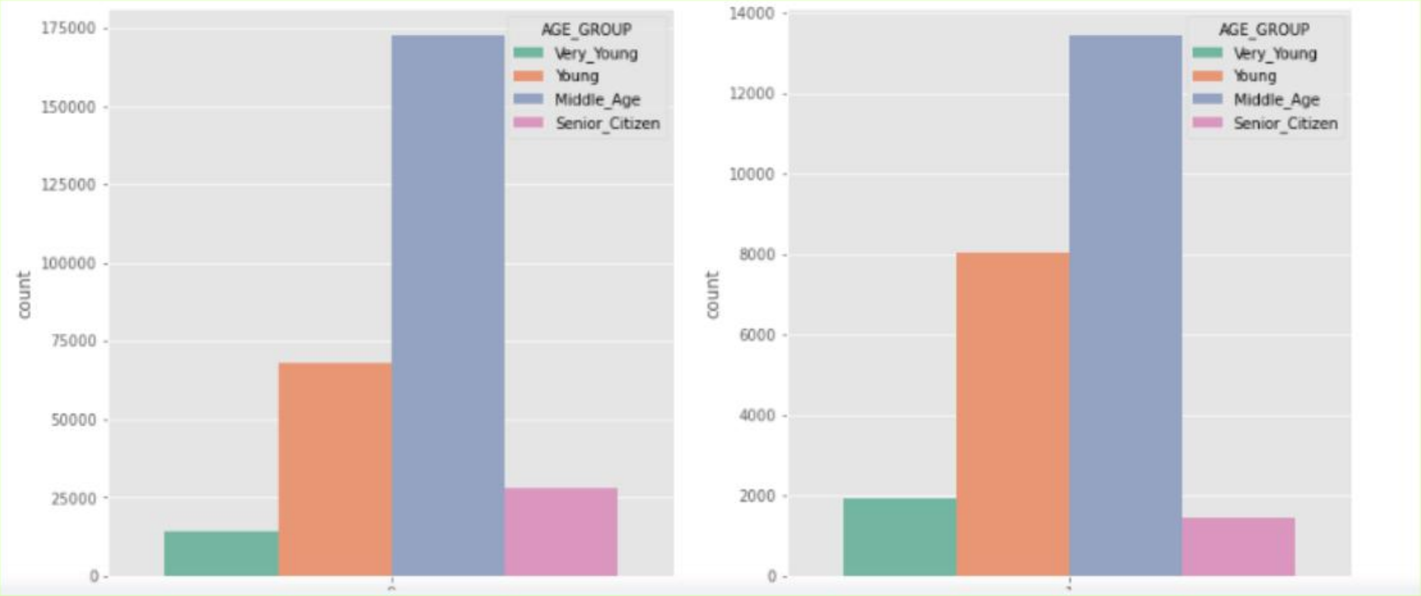

```
In [ ]: count1 = 0
        count0 = 0
        for i in df1['TARGET'].values:
            if i == 1:
                count1 += 1
            else:
                count0 += 1

        count1 = (count1/len(df1['TARGET']))*100
        count0 = (count0/len(df1['TARGET']))*100

        x = ['Approved(TARGET=1)','Non-Approved(TARGET=0)']
        y = [count1, count0]

        explode = (0.1, 0)  # only "explode" the 1st slice

        fig1, ax1 = plt.subplots()
        ax1.pie(y, explode=explode, labels=x, autopct='%1.1f%%',
                shadow=True, startangle=110)
        ax1.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
        plt.title('Data imbalance',fontsize=25)
        plt.show()
```

According to the pie chart, 40% percent of the clients have been rejected their credit loan, hence we take into their data to analyse the reasons why their loans should be not approved

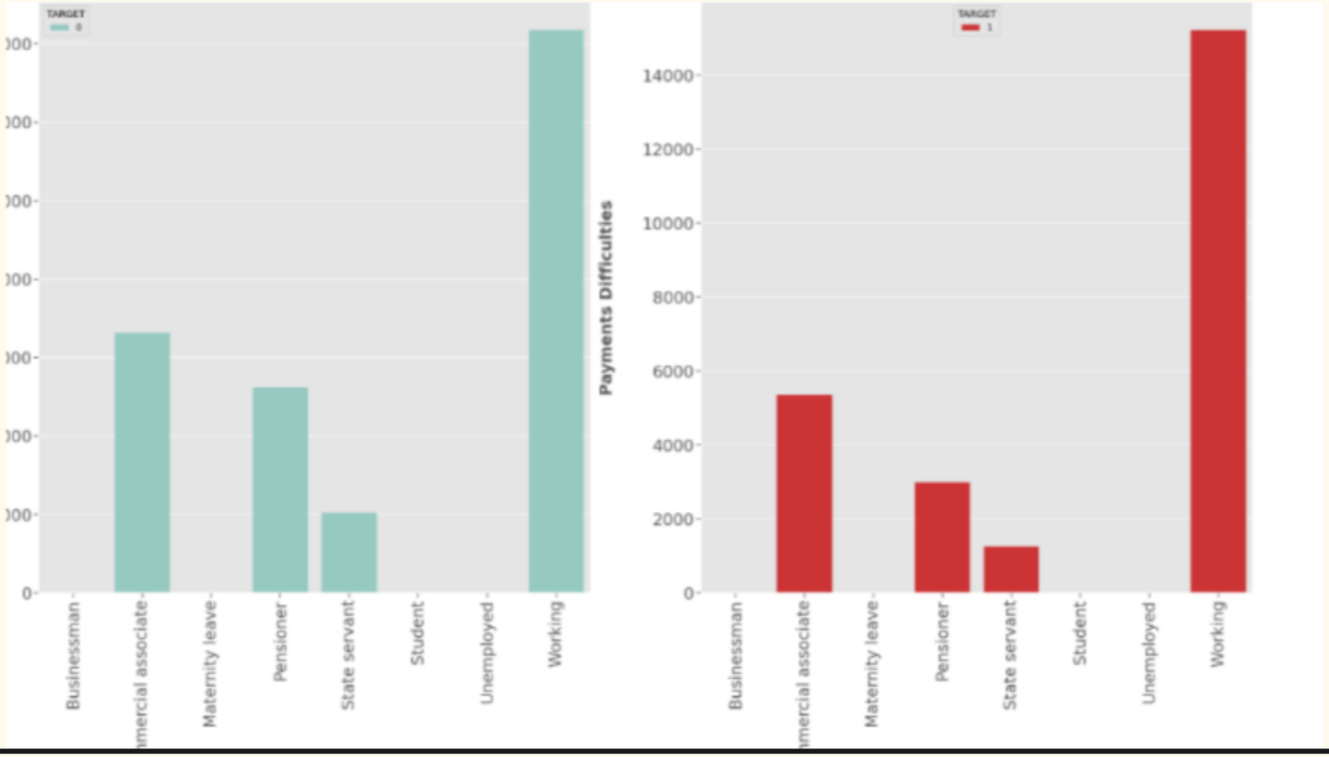
ANALYSIS BASED ON AGE

```
plt.figure(figsize=(15,7))
plt.subplot(121)
sns.countplot(x='TARGET',hue='AGE_GROUP',data=Target0,palette='Set2')
plt.subplot(122)
sns.countplot(x='TARGET',hue='AGE_GROUP',data=Target1,palette='Set2')
plt.show()
```

- Middle Age(35-60) the group seems to applied higher than any other age group for loans
- Also, Middle Age group facing paying difficulties the most.
- While Senior Citizens(60-100) and Very young(19-25) age group facing paying difficulties less as compared to other age groups.

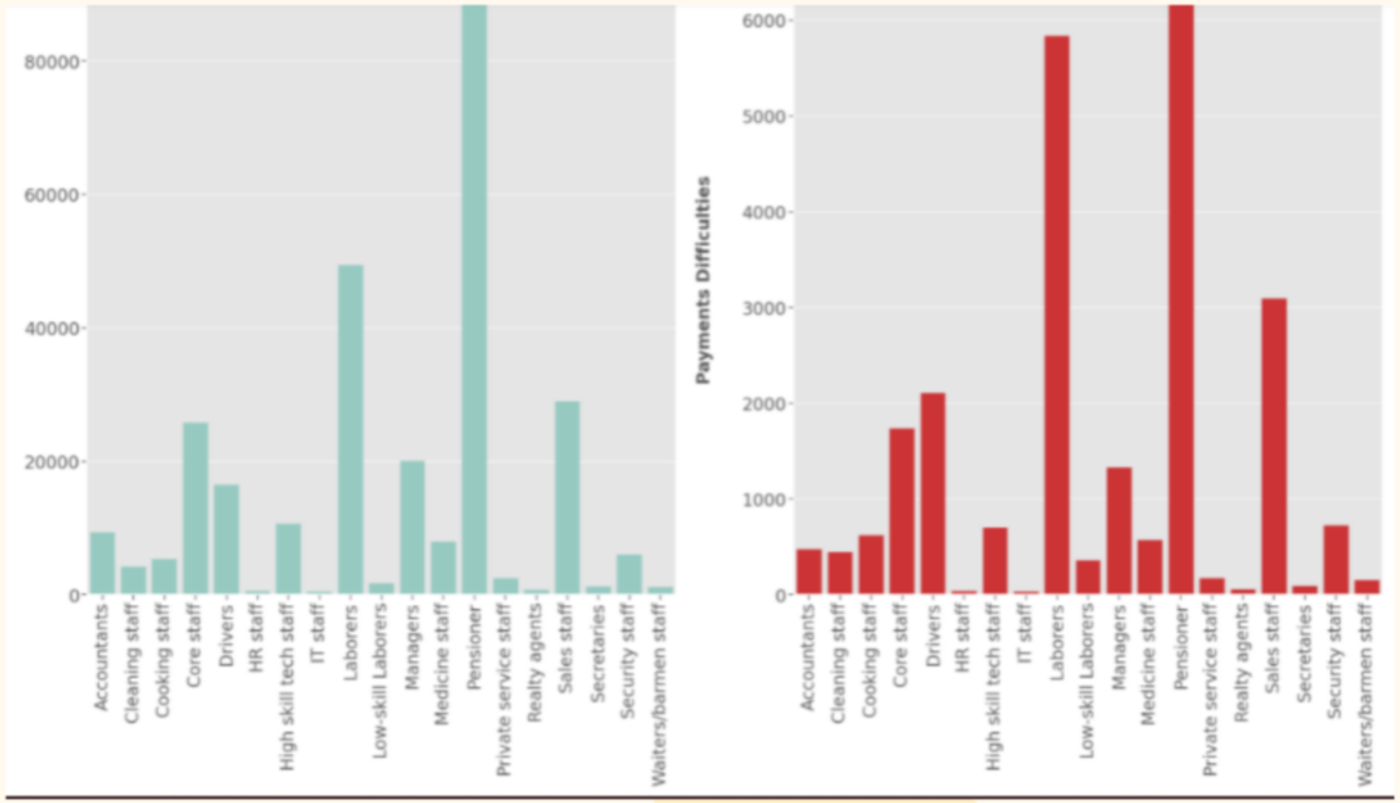


ANALYSIS BASED ON INCOME



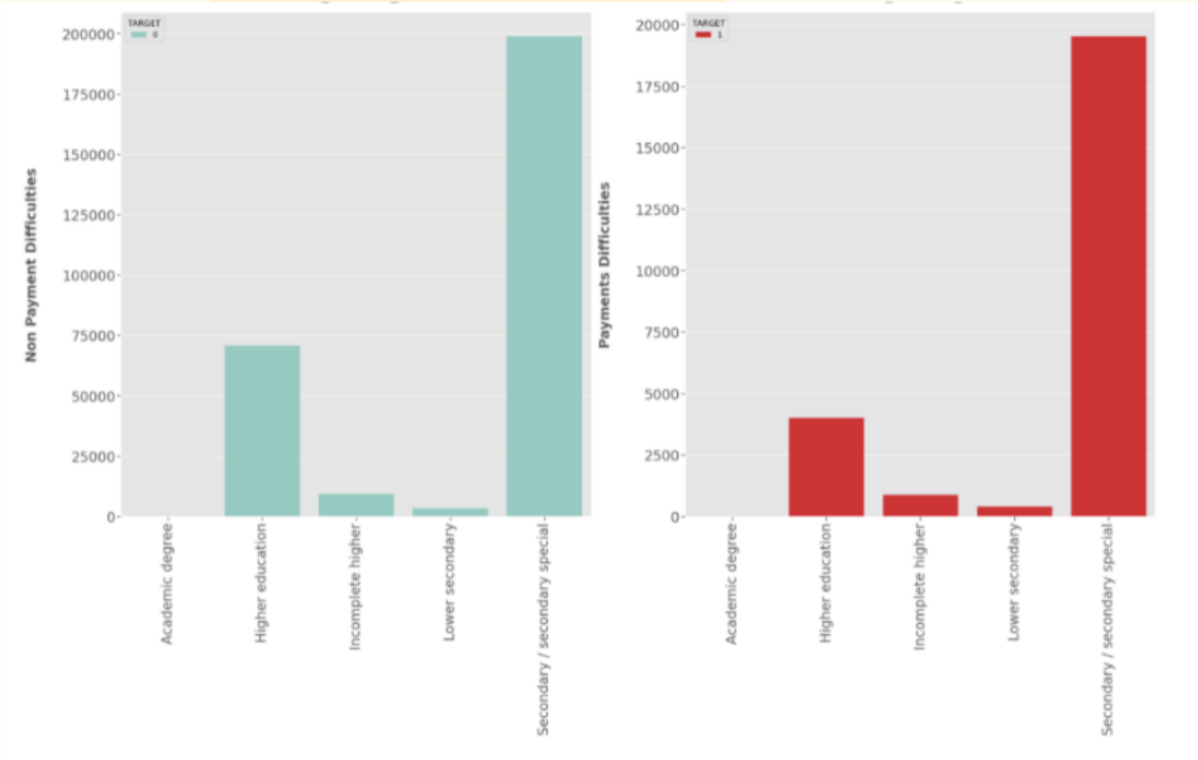
- Clients who applied for loans were getting income by Working,Commercial associate and Pensioner are more likely to apply for the loan, highest being the Working class category .
- Businessman, students and Unemployed less likely to apply for loan
- Working category have high risk of being unable to repay the loan
- State Servant is at a lower risk of being unable to repay as they are paid well according to the data

ANALYSIS BASED ON OCCUPUTION



- Pensioners have applied the most for the loan in this case
- Pensioners followed by Labourers are likely to be rejected and the working category as well and hence have low income.

ANALYSIS BASED ON EDUCATION

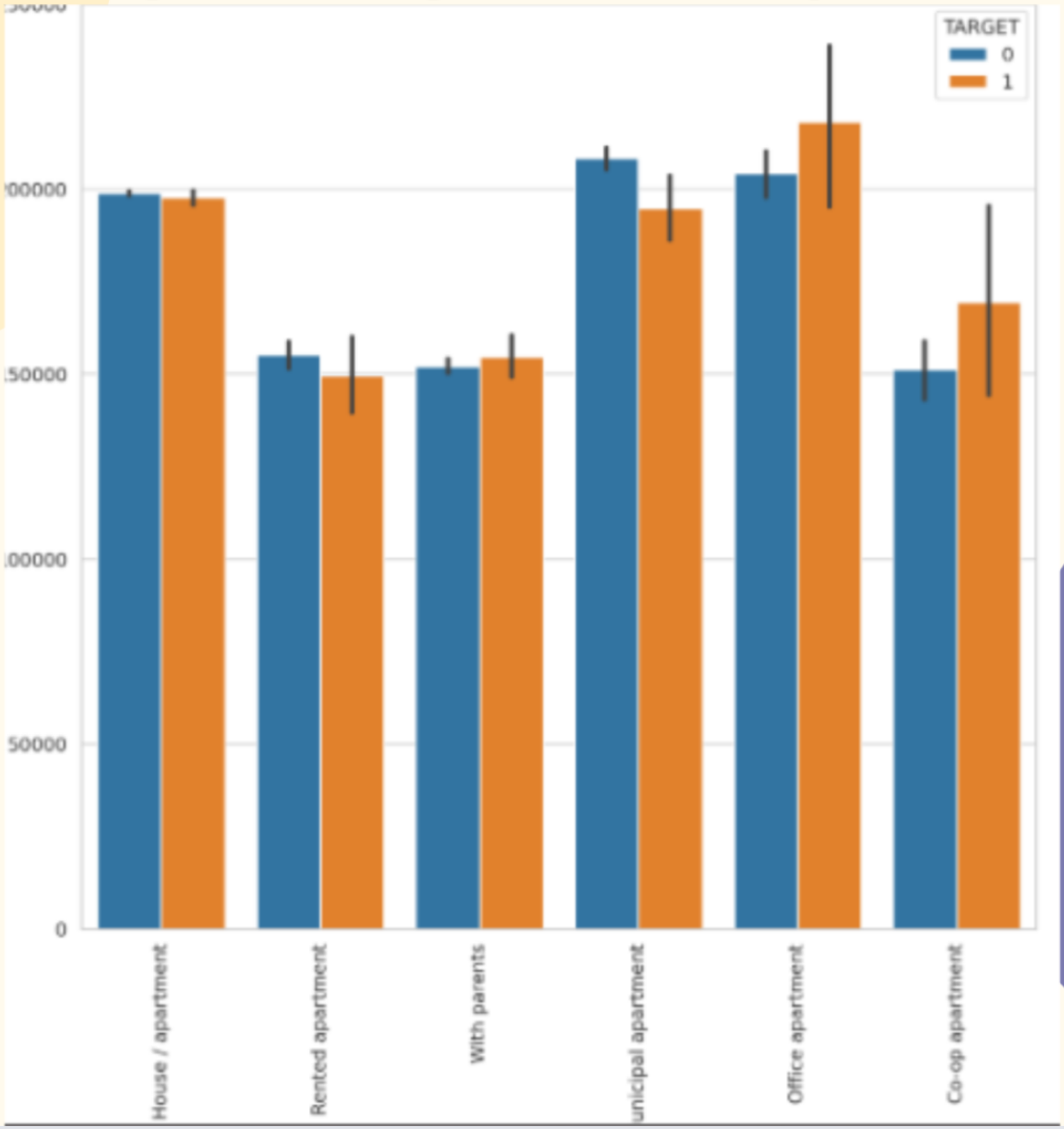


- Clients having education Secondary or Secondary Special are more likely to apply for the loan.
- Clients having education Secondary or Secondary Special have higher risk of not repaying.
- Other education types have minimal risk.

ANALYSIS BASED ON HOUSING

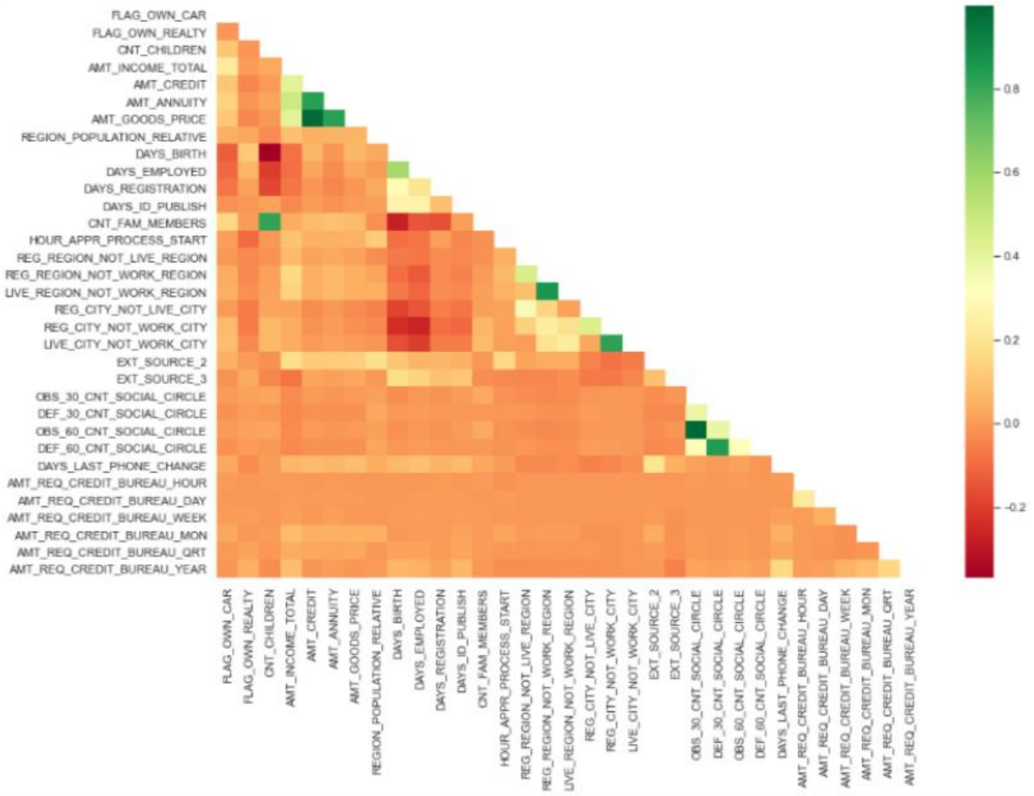
```
plt.figure(figsize=(15,15),dpi = 150)
plt.xticks(rotation=90)
sns.barplot(data =df_comb, y='AMT_CREDIT_PREV',hue='TA
RGET',x='NAME_HOUSING_TYPE')
plt.title('Prev Credit amount vs Housing type')
plt.show()
```

Here for Housing type, office appartment is having higher credit of target 0 and co-op apartment is having higher credit of target 1. So, we can conclude that bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment. Bank can focus mostly on housing type with parents or House\appartment or miuncipal appartment for successful payments.



USE OF CORRELATION

```
targets_corr(data=t0,title='Correlation for Target 0')
```




```
numerical_col = df1.select_dtypes(include='number').columns
numerical_col
```

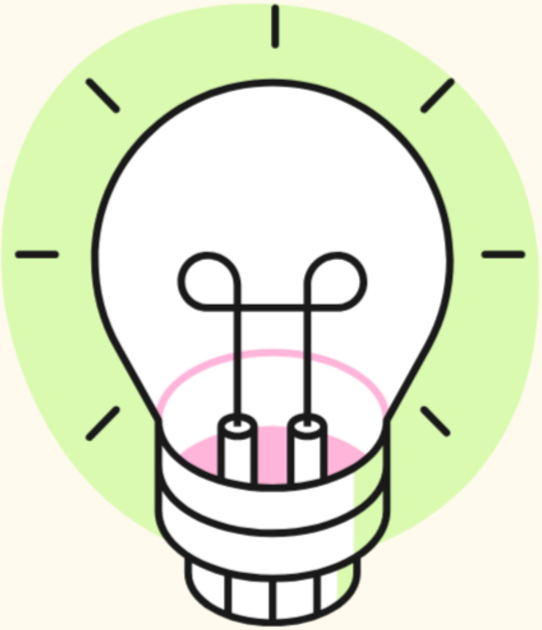
```
Index(['SK_ID_CURR', 'TARGET', 'FLAG_OWN_CAR', 'FLAG_OWN_REALTY',
       'CNT_CHILDREN', 'AMT_INCOME_TOTAL', 'AMT_CREDIT', 'AMT_ANNUITY',
       'AMT_GOODS_PRICE', 'REGION_POPULATION_RELATIVE', 'DAYS_BIRTH',
       'DAYS_EMPLOYED', 'DAYS_REGISTRATION', 'DAYS_ID_PUBLISH',
       'CNT_FAM_MEMBERS', 'HOUR_APPR_PROCESS_START',
       'REG_REGION_NOT_LIVE_REGION', 'REG_REGION_NOT_WORK_REGION',
       'LIVE_REGION_NOT_WORK_REGION', 'REG_CITY_NOT_LIVE_CITY',
       'REG_CITY_NOT_WORK_CITY', 'LIVE_CITY_NOT_WORK_CITY', 'EXT_SOURCE_2',
       'EXT_SOURCE_3', 'OBS_30_CNT_SOCIAL_CIRCLE', 'DEF_30_CNT_SOCIAL_CIRCLE',
       'OBS_60_CNT_SOCIAL_CIRCLE', 'DEF_60_CNT_SOCIAL_CIRCLE',
       'DAYS_LAST_PHONE_CHANGE', 'AMT_REQ_CREDIT_BUREAU_HOUR',
       'AMT_REQ_CREDIT_BUREAU_DAY', 'AMT_REQ_CREDIT_BUREAU_WEEK',
       'AMT_REQ_CREDIT_BUREAU_MON', 'AMT_REQ_CREDIT_BUREAU_QRT',
       'AMT_REQ_CREDIT_BUREAU_YEAR'],
      dtype='object')
```

```
len(numerical_col)
```

```
35
```

```
corr0=df1.iloc[0:,2:]
corr1=df1.iloc[0:,2:]

t0=corr0.corr(method='spearman')
t1=corr1.corr(method='spearman')
```

```
def targets_corr(data,title):
    plt.figure(figsize=(15, 10))

    mask= np.zeros_like(data)
    mask[np.triu_indices_from(mask)]=True
    with sns.axes_style("white"):
        ax= sns.heatmap(data, mask=mask,cmap='RdYlGn')
```



FINAL INFERENCE

From the analysis using the five parameters we have concluded the following:

- People of the middle age group (35-60),being at a crucial stage of life find it difficult to repay their loans .
- The working category and pensioners have a high risk of being unable to repay the loan .
- Clients having secondary or higher secondary education have higher chance of being unable to pay the loan.
- The clients owning a co-op apartment are likely to have difficulties in repayment.

THANK YOU