

Credit EDA Submission

Analysis by Suresh Babu R

Problem Statement

In the absence of a credit report of the customer making a loan application, it is difficult for a banking institution to take a decision on either approving or denying a new loan application.

The banking institution will have a loss if:

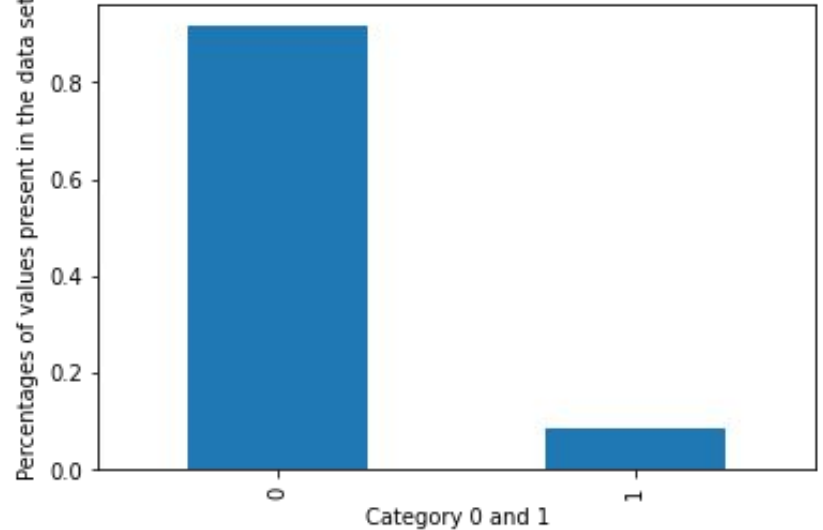
1. They deny loan to a customer who is capable of repaying the loan with interest; thus losing on profit earned.
2. Approving loan to a customer who defaults on repayments; thus losing out on money lent and the profit.

Can inferences be made of the behavior of customers who have previously defaulted and thereby a decision taken on either approving or denying a new loan application through the use of strong indicators of default?

Assumptions

The TARGET variable has data imbalance, and no corrective measures are taken to correct the imbalance.

Plot of non-defaulters(0) against defaulters(1); indicative of data imbalance



Approach

49 columns with more than 32% of null values present in them are dropped.

96391 rows in OCCUPATION_TYPE, 41857 rows in EXT_SOURCE_3, 752 rows in NAME_TYPE_SUITE, 513 rows in OBS_30_CNT_SOCIAL_CIRCLE, 255 rows in EXT_SOURCE_2, and 11 rows in AMT_ANNUITY columns having null values are dropped to keep data clean.

Outlier of 117 million in AMT_INCOME_TOTAL is removed to keep data clean. The rest of the outliers under 9M are retained.

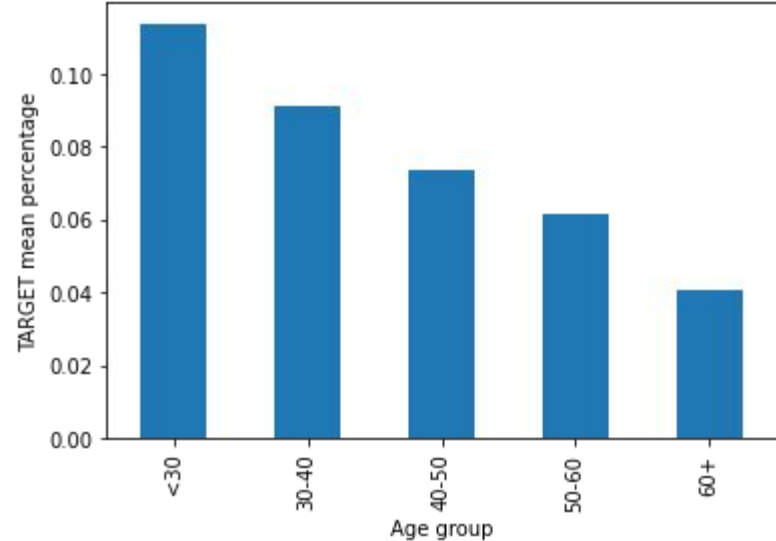
Outliers in the AMT_CREDIT column are retained.

Outliers in the AMT_GOODS_PRICE are retained.

1. Age as an indicator

- Customers less than 30 years have a tendency to default.
- Customers in the age group 30-40 too should be dealt with caution.

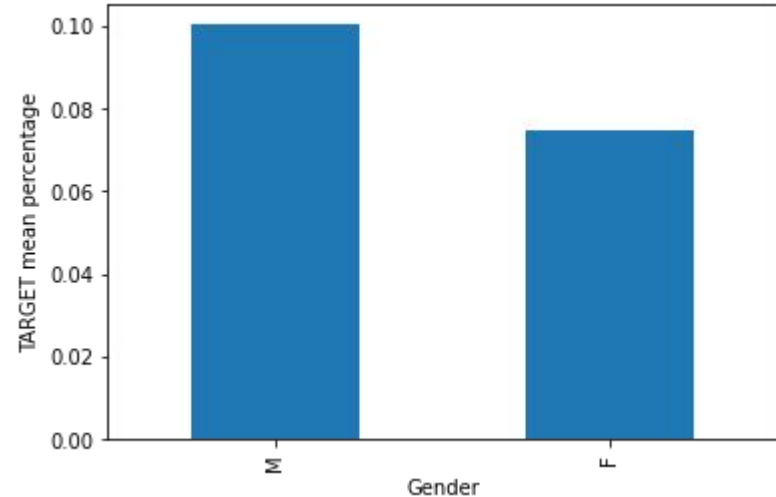
Plot of Age group against the TARGET mean percentages found in data set



2. Gender as an indicator

- Male customers are more prone to default than female customers.
- Female customers are a safer bet to lend after taking other indicators into account.

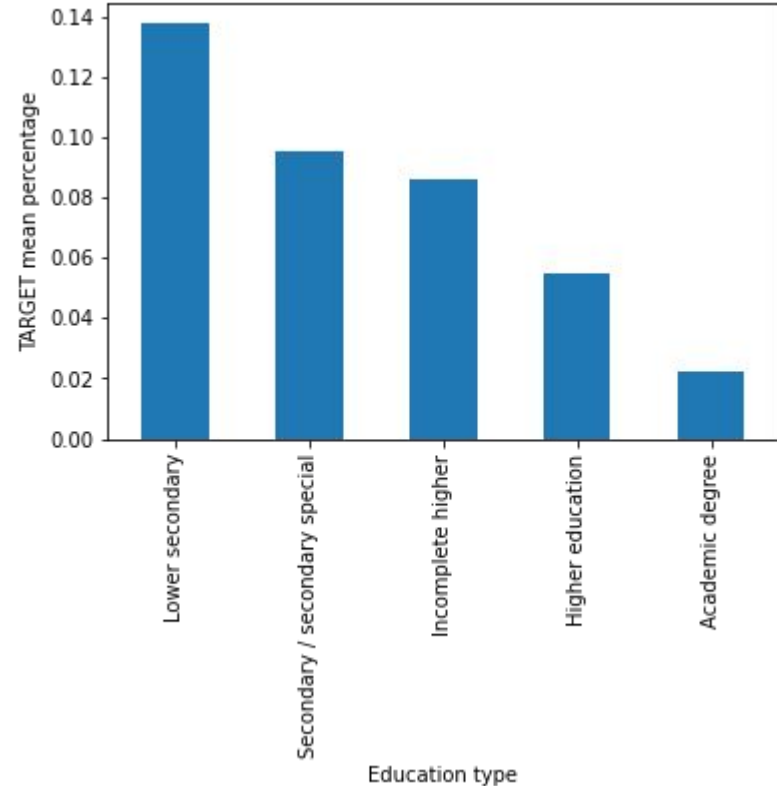
Plot of Gender against the TARGET mean percentages found in data set



3. Education type as an indicator

- Customers who are less educated have a tendency to default.

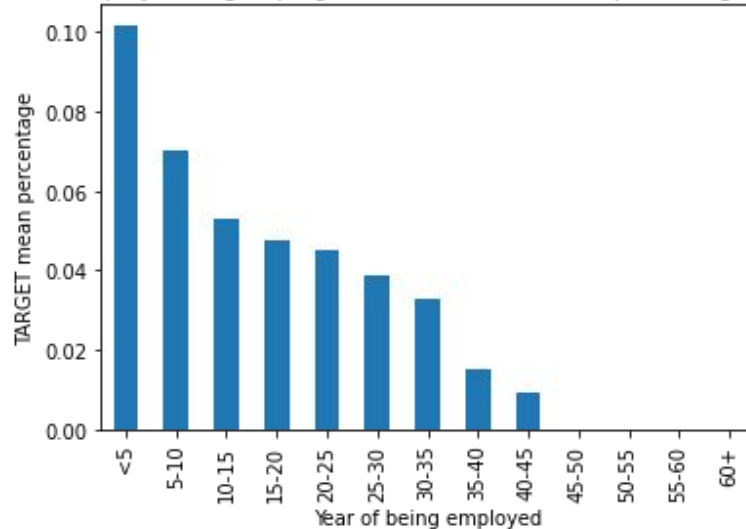
Plot of Education type against the TARGET mean percentages found in data set



4. Years of being employed as an indicator

- Customers with less than 5 years of employment have a tendency to default.

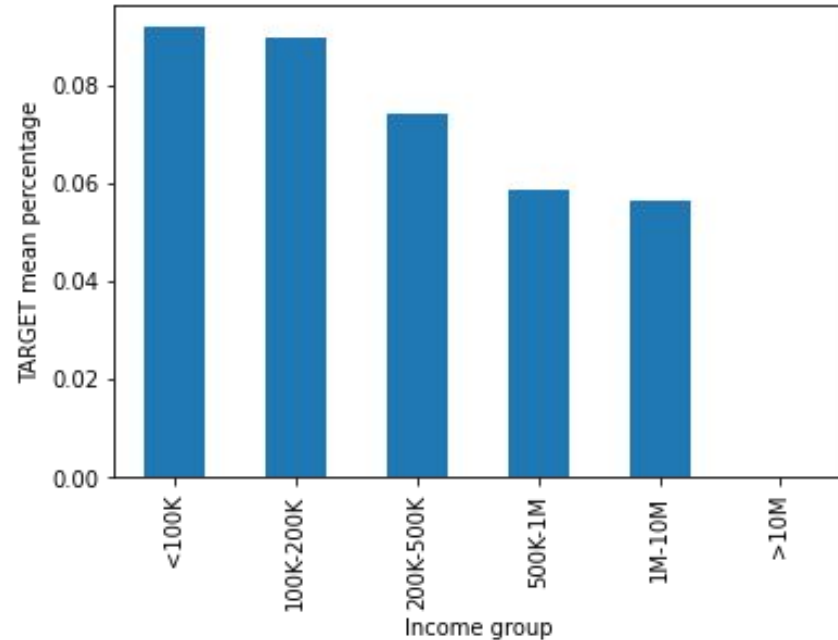
Plot of Years of employment group against the TARGET mean percentages found in data set



5. Income earned being employed as an indicator

- Customers whose income is less than 100K and so also customers whose income is between 100K-200K have a tendency to default.

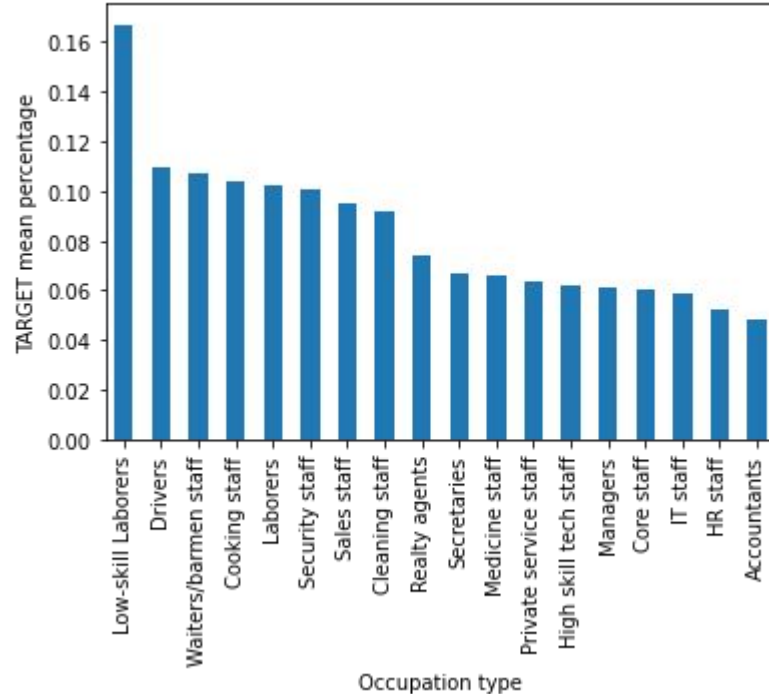
Plot of Income group against the TARGET mean percentages found in data set



6. Occupation type being employed as an indicator

- Customers who are employed in low-skill activities have a tendency to default.

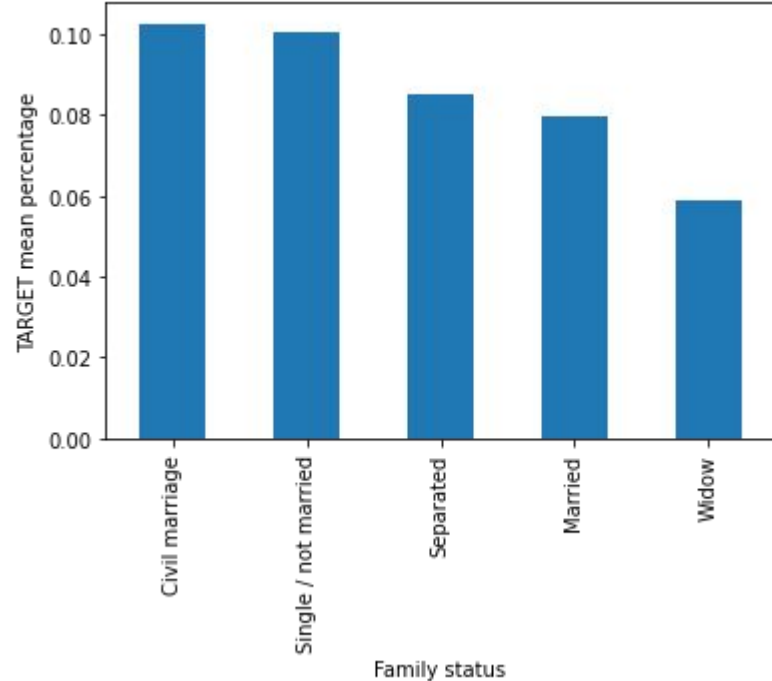
Plot of Occupation type against the TARGET mean percentages found in data set



7. Family status being employed as an indicator

- Customers who have had civil marriage and are either single or not married have more tendency to default.

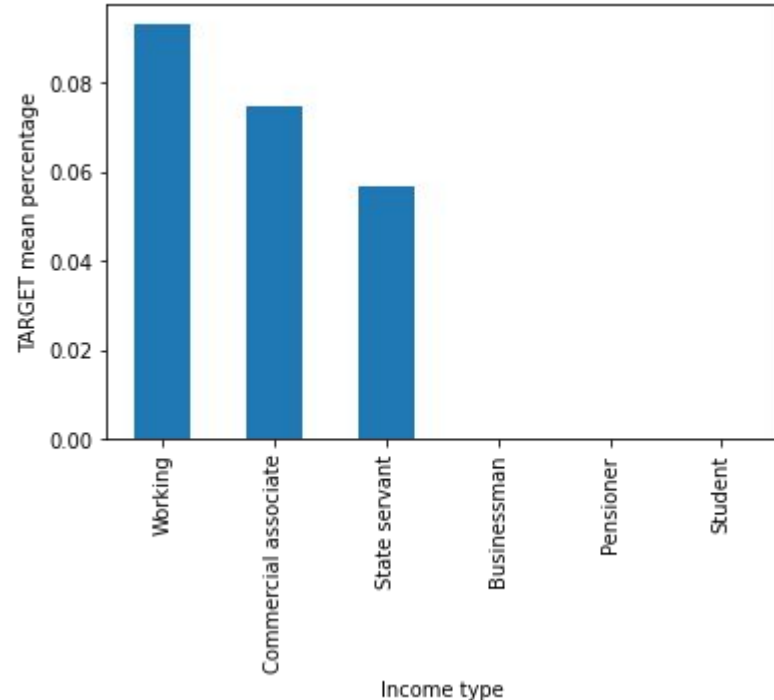
Plot of Family status against the TARGET mean percentages found in data set



8. Income type being employed as an indicator

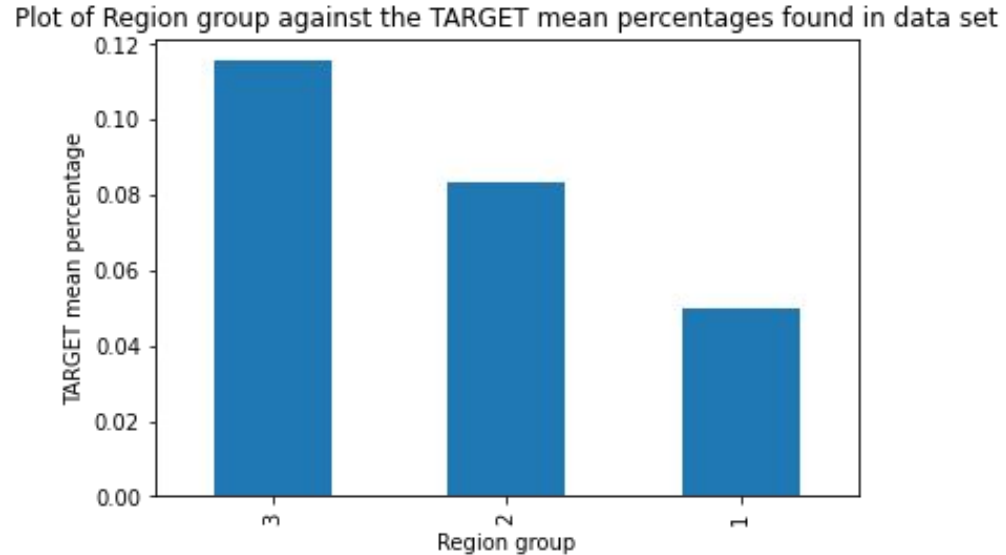
- Customers whose income type is 'working' have a major tendency to default.

Plot of Income type against the TARGET mean percentages found in data set



9. Region being employed as an indicator

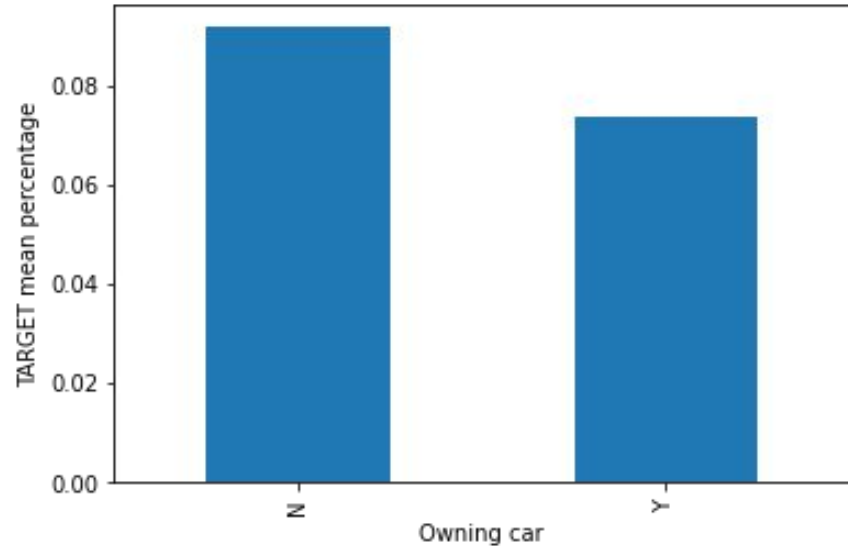
- Major defaults are seen by customers residing in the Region identified as 3.



10. Owning a car being employed as an indicator

- Customers who do not own a car have a greater tendency to default.

Plot of Owning car against the TARGET mean percentages found in data set



Conclusion

The analysis is as per the given data set.