Loan Approval case study by

Keerthi Gayam

Aniket

Problem Statements:

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

- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Approach used to understand the factors behind loan defaults

1. Read the data in to a python notebook from the csv file
2. Understand the structure of the data by looking at the number of rows, columns and data types.
3. Finding out the number of null values present in each column of our data set.
4. Dropping columns that have more than 13% of null values from the application_data data frame
5. Checking for null values percentage in rows of the application data set.
6. Checking the data type of our application data frame.
7. Checking for outliers in the data frame and tackling those outliers
8. Splitting Application data DF into two sub data frames based on Target field. 1 = Defaulters, 0 = All others
9. Finding out Data Imbalance
10. Univariate analysis of application data frame categorical variables
11. Univariate analysis of application data frame numerical variables with continuous data
12. Bivariate analysis of application data frame
13. Correlation maps to compare co- relation level between different variables of Target =1 and 0

The above approach is common for both application data set and previous data set.

Some of the factors that are effecting the loan repayment capability of the customers are:

	SK_ID_CURR
NAME_CONTRACT_TYPE	
Cash loans	93.54
Revolving loans	6.46

We can see that 93% of the loans with repayment difficulties are Cash loans. But this could also be because the cash loans are taken more often than revolving loans. We need to analyze more before we concolude whether contract type has any impact on repayment difficulties.

		SK_ID_CURR
TARGET	NAME_CONTRACT_TYPE	
0	Cash loans	91.654087
	Revolving loans	94.521671
1	Cash loans	8.345913
	Revolving loans	5.478329

As we can see out of the total cash loans taken 8.3% ended up with payment difficulties where as 5.4% of the revolving loans end up with payment difficulties. So cash loans has a sligtly higher risk comapred to revolving loans.

Some of the factors that are effecting the loan repayment capability of the customers are:

	SK_ID_CURR
CODE_GENDER	
F	57.079557
M	42.920443

We can see that 57% of the customers with repayment difficulties are female. Its is more risky to approve loans to female customers than male. But again this could be because the number of loans taken by men could be more than female.

		SK_ID_CURR
TARGET	CODE_GENDER	
0	F	93.000672
	M	89.857983
	XNA	100.000000
1	F	6.999328
	M	10.142017

We can conclude that providing loans to men has slightly higher risk compared to women.

Some of the factors that are effecting the loan repayment capability of the customers are:

	SK_ID_CURR
FLAG_OWN_CAR	
N	69.482377
Y	30.517623

We can conclude that people without a own vehicle are more likely to default on loans.

		SK_ID_CURR
TARGET	**FLAG_OWN_CAR**	
0	**N**	91.499773
	Y	92.756201
1	**N**	8.500227
	Y	7.243799

The conclusion that people without a own vehicle are more likely to default on loans holds true even when we compare it against the number of people without cars that applied for loan and defaults without cars.

Some of the factors that are effecting the loan repayment capability of the customers are:

	SK_ID_CURR
FLAG_OWN_REALTY	
N	31.589124
Y	68.410876

We can see that people with a own house are more likely to default on loans but this also could be people without reality are more likely to apply for loans when compared to people with reality, it is not conclusive without doing further analysis

		SK_ID_CURR
TARGET	FLAG_OWN_REALTY	
0	N	91.674982
	Y	92.038423
1	N	8.325018
	Y	7.961577

We can see that the number of people who apply for loan are people without own reality , so the previous analysis showed us that the people more likely to default on loan their loans are people with own reality but when we compared the same with the number of people who applied for loan vs defaluters in that category it is claer that people with no reality are more likely to default.

Some of the factors that are effecting the loan repayment capability of the customers are:

TARGET	NAME_INCOME_TYPE	SK_ID_CURR
0	Maternity leave	60.000000
	Unemployed	63.636364
	Working	90.411528
	Commercial associate	92.515743
	State servant	94.245035
	Pensioner	94.613634
	Businessman	100.000000
	Student	100.000000
1	Pensioner	5.386366
	State servant	5.754965
	Commercial associate	7.484257
	Working	9.588472
	Unemployed	36.363636
	Maternity leave	40.000000

People who are on Maternity leave or unemployed are very high risk customer as they defaulters in this category are very high.

Some of the factors that are effecting the loan repayment capability of the customers are:

NAME_EDUCATION_TYPE	SK_ID_CURR	TARGET
Academic degree	0.053331	0.053331
Lower secondary	1.240931	1.240931
Incomplete higher	3.341994	3.341994
Higher education	24.344820	24.344820
Secondary / secondary special	71.018923	71.018923

TARGET	NAME_EDUCATION_TYPE	SK_ID_CURR
0	Lower secondary	89.072327
	Secondary / secondary special	91.060071
	Incomplete higher	91.515034
	Higher education	94.644885
	Academic degree	98.170732
1	Academic degree	1.829268
	Higher education	5.355115
	Incomplete higher	8.484966
	Secondary / secondary special	8.939929
	Lower secondary	10.927673

People who are not having higher education and academic degree are risky customer when compared to others.

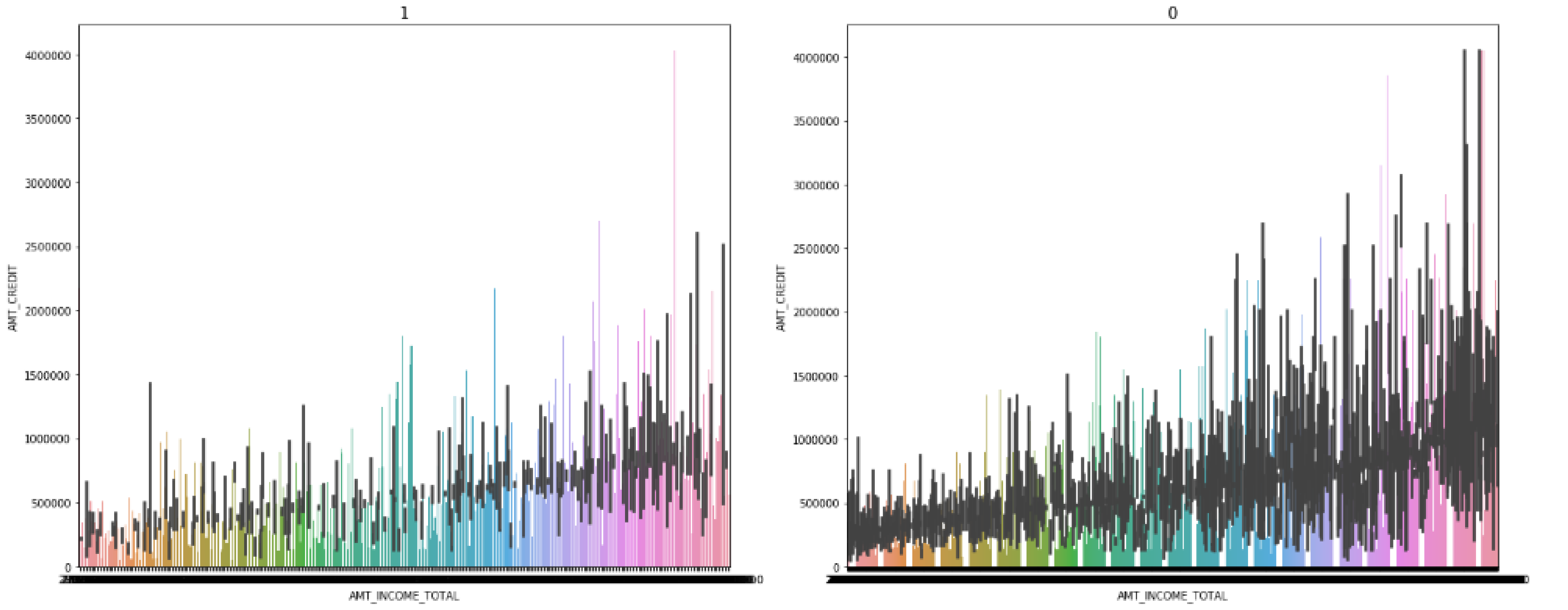
Some of the factors that are effecting the loan repayment capability of the customers are:

TARGET	NAME_HOUSING_TYPE	SK_ID_CURR
0	Rented apartment	87.686949
	With parents	88.301887
	Municipal apartment	91.460252
	Co-op apartment	92.067736
	House / apartment	92.204289
	Office apartment	93.427589
1	Office apartment	6.572411
	House / apartment	7.795711
	Co-op apartment	7.932264
	Municipal apartment	8.539748
	With parents	11.698113
	Rented apartment	12.313051

I see that people living in rented apartment or with parents are high risk customers compared to others

Some of the factors that are effecting the loan repayment capability of the customers are:

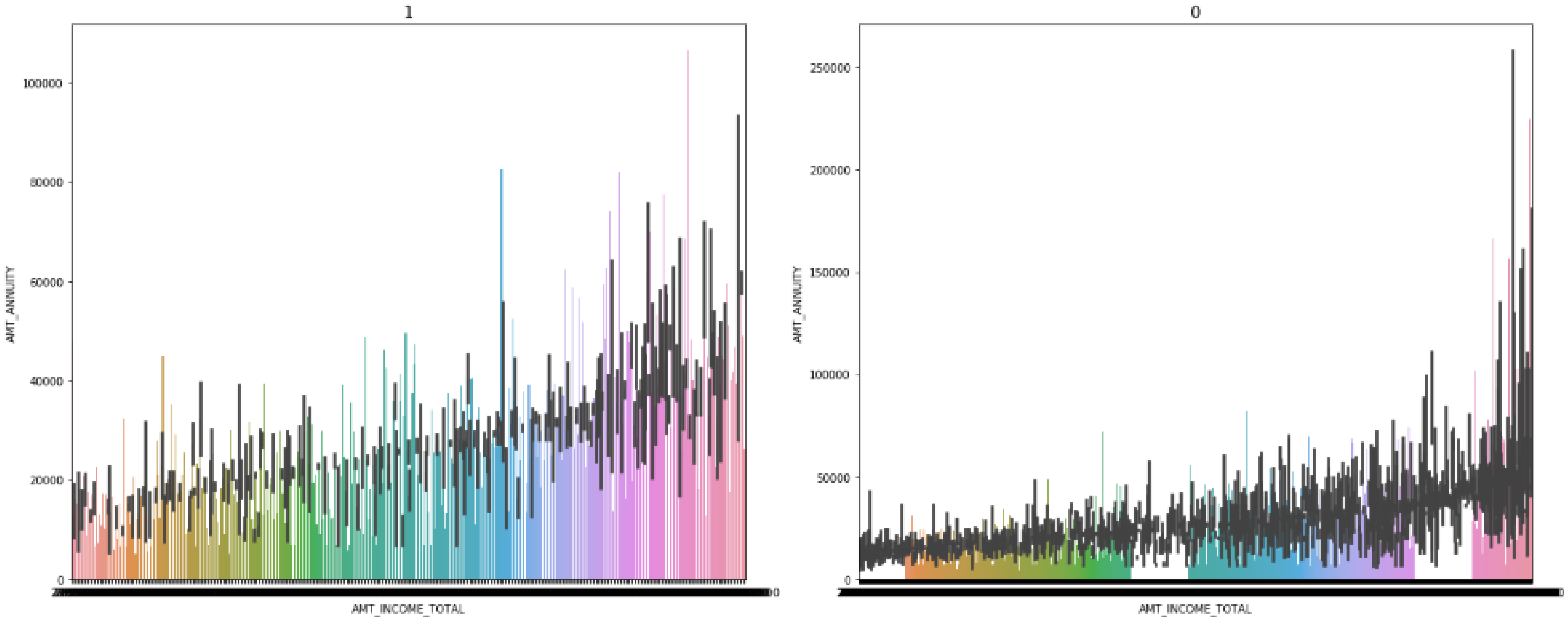
Amount of Income Vs Amount of Credit



In Case of of defaulters we can see that the Amount of Credit is much higher than the Amount of income. This can lead to repayment difficulties.

Some of the factors that are effecting the loan repayment capability of the customers are:

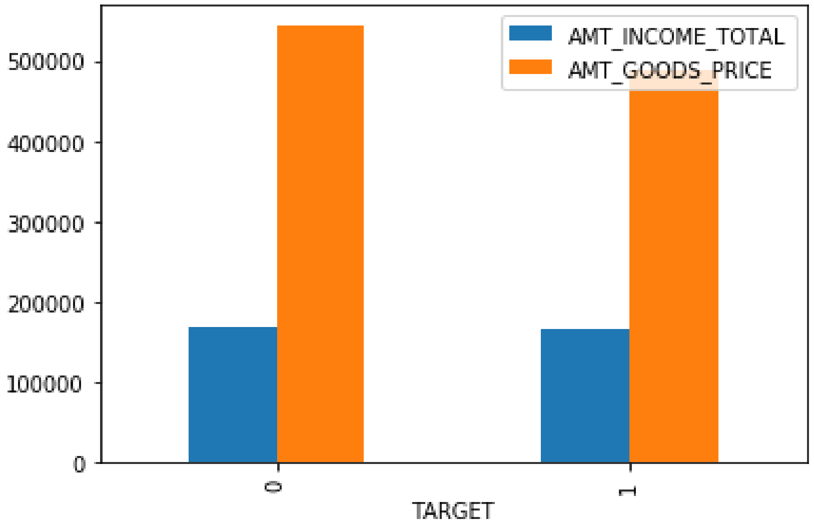
Amount of Income Vs Amount of Annuity



The amount of Income of the client Vs Loan Annuity shows the clients with good income to annuity ratio has better replayment capacity than others.

Some of the factors that are effecting the loan repayment capability of the customers are:

Amount of Income Vs Amount of Annuity



The conclusion that average of client which are having high salary tend to have good repayment with consumer loans by goods prices

Some of the factors that are effecting the loan repayment capability of the customers are:

Previous Application data analysis

	SK_ID_CURR
NAME_CONTRACT_TYPE	
Cash loans	44.757917
Consumer loans	43.656142
Revolving loans	11.565225
XNA	0.020716

44.75% contract are of Cash and 43.65 are consumer loans which is most people prefer because it may get approved but we can't tell now

Some of the factors that are effecting the loan repayment capability of the customers are:

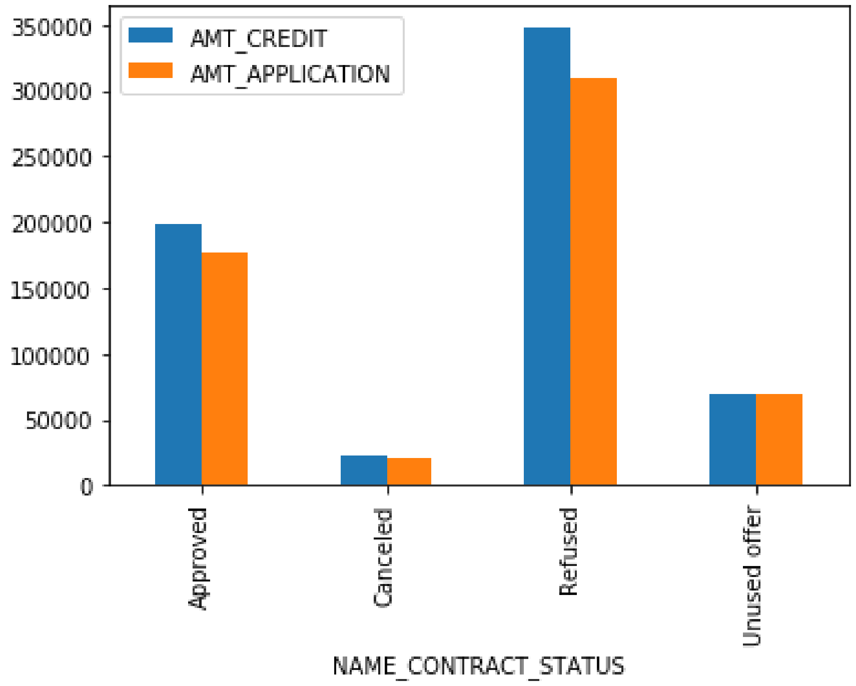
Previous Application data analysis

NAME_CONTRACT_STATUS	NAME_CONTRACT_TYPE	SK_ID_CURR
Approved	Cash loans	18.712572
	Consumer loans	37.508367
	Revolving loans	5.853801
Canceled	Cash loans	16.081233
	Consumer loans	0.093341
	Revolving loans	2.745397
	XNA	0.018860
Refused	Cash loans	9.934535
	Consumer loans	4.501519
	Revolving loans	2.965728
	XNA	0.001856
Unused offer	Cash loans	0.029577
	Consumer loans	1.552915
	Revolving loans	0.000299

44.75% contract are of Cash and 43.65 are consumer loans which is most people prefer but approval rate is higher for Consumer Loans and loans which are canceled by client are mostly client with cash loans

Some of the factors that are effecting the loan repayment capability of the customers are:

Previous Application data analysis



Refused rate is greater for greater final loan amount but approval rate is moderate for average Loan Amount

Conclusions

- Before approving loans the bank should consider the Amount of income and give loans that are in proportion to the income earned by the client.
- Before approving loans to customers it is good to cross check the employment stability and in case of women the possibility of going on long Maternity leave could effect the repayment capacity as we have clearly seen while analysing the application data.
- In case of newly employed clients it is better to give out smaller loans with some kind of repayment or recovery guaranty.
- Cash loans and male clients have slightly higher risk, so better analyse these types of loans on case to case basis and approve.
- These are some of the parameters that are to be looked out for before approving loans.
- Before approving higher loan amounts we need to analyse all risk factors and if possible it is better to reduce the loan amount than the requested amount where applicable.