



Credit EDA Case Study

Risk Analysis

BY- KRUNAL SONI

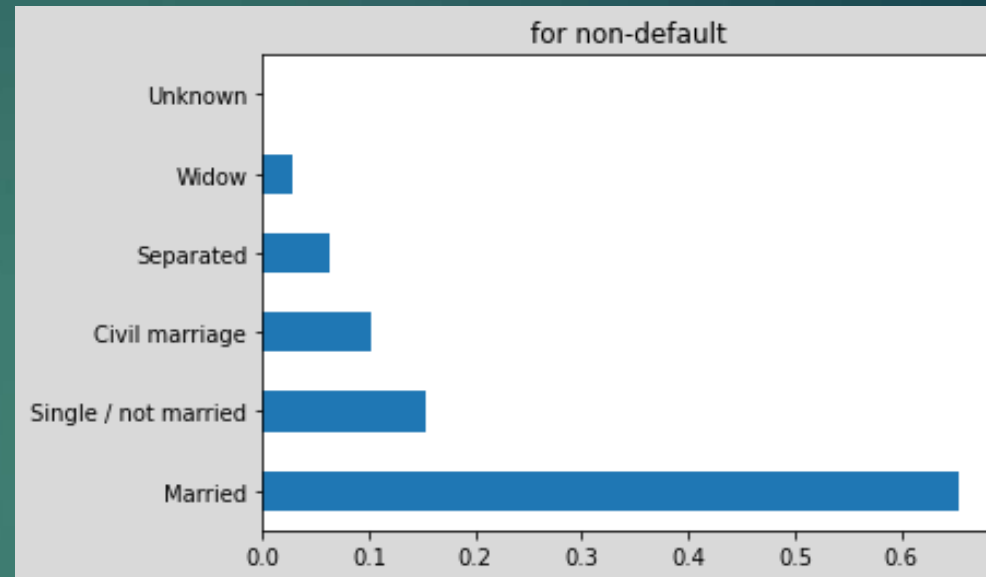
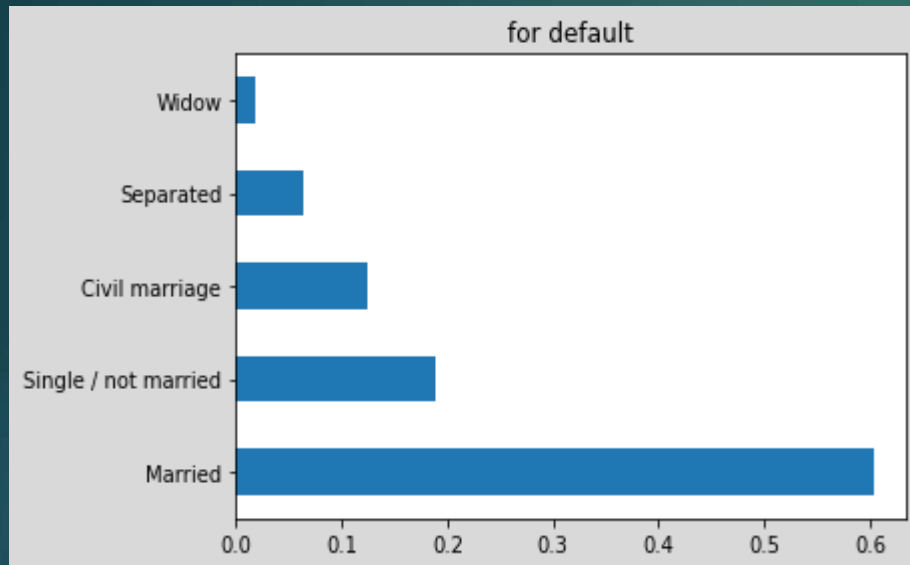
Problem statement-

credit risk analysis will help company to aware of risk of giving loan who have not good financial profile .
Save huge losses of company
and also aware of those good clients who have good financial background.

Approach-

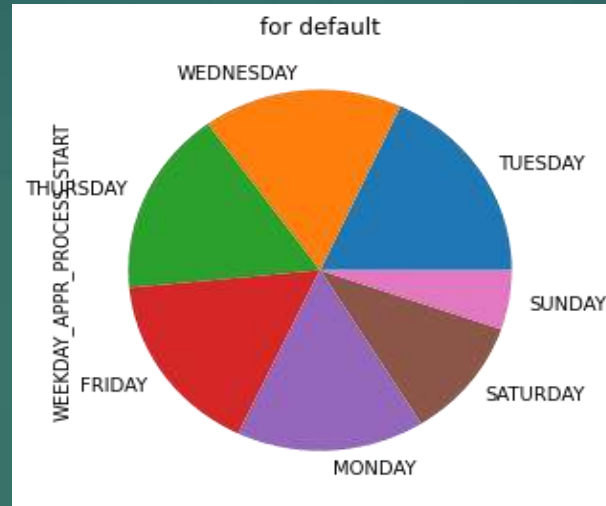
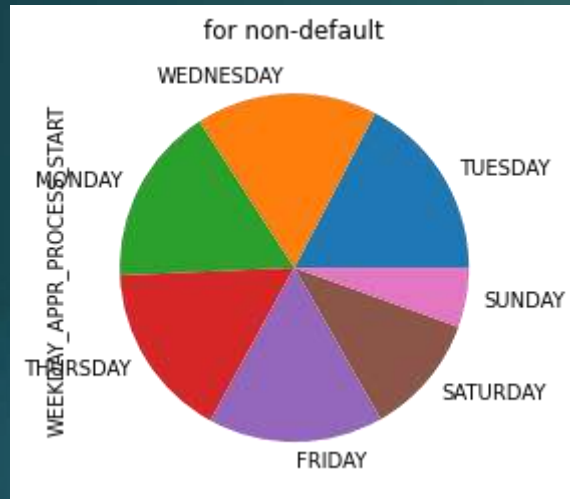
1. import required libraries
2. Data loading
3. Data cleaning
4. Finding missing values, based on percentage of missing we need to decide whether this column keep it or drop it , imputing missing values , finding outliers
5. Removing/imputing rows and column based on missing values and outliers
6. find data type and change them based on requirement
7. Check Data imbalance percentage
8. univariate, bivariate, segmented analysis, correlation
- 9 .merging the data sets
10. data visualization
11. conclusion

Univariate Analysis



Who are married most likely took loan compared to married those are also be defaulter , same as non – defaulter category

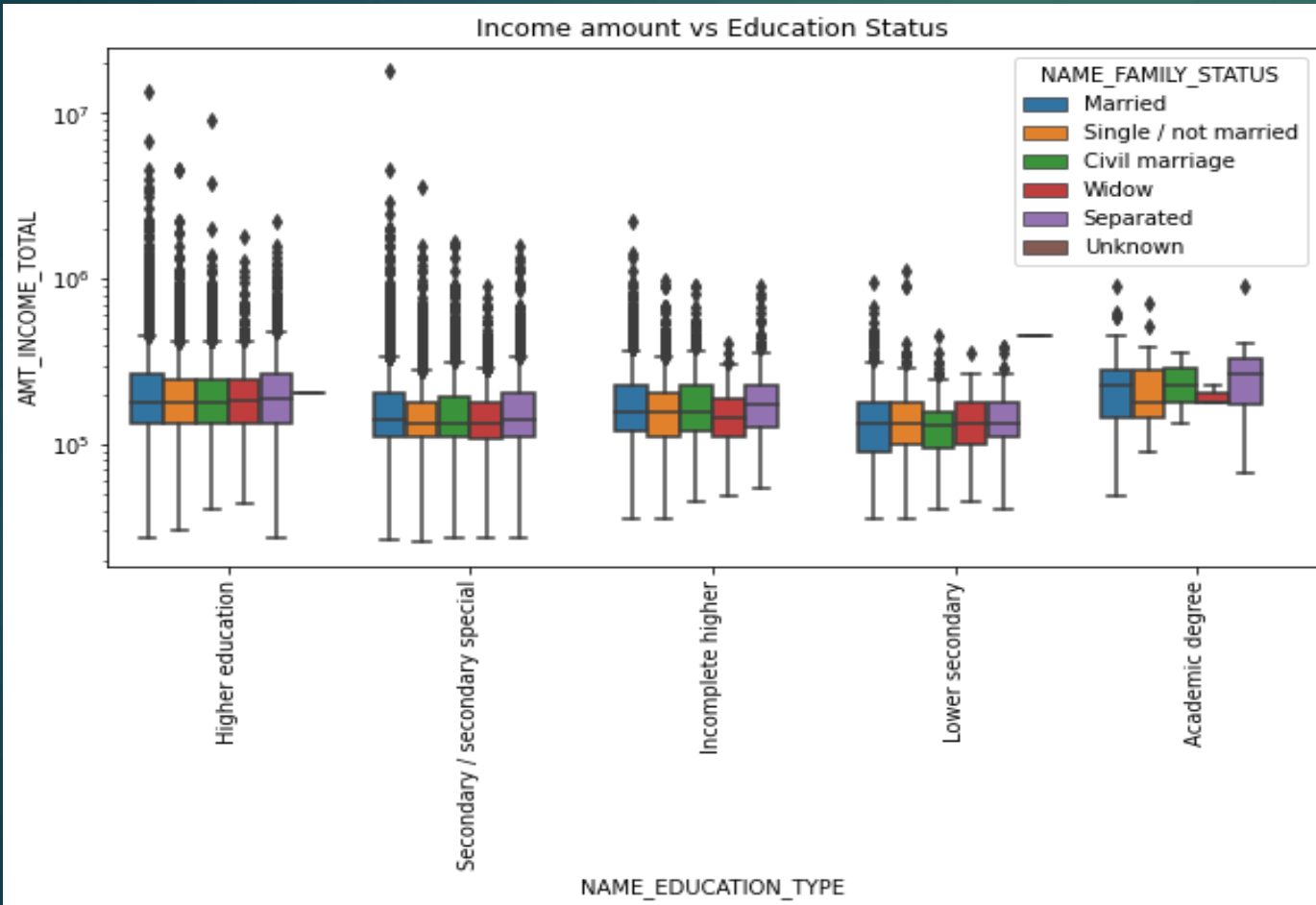
Those who likely come in weekdays
for defaulter or non defaulter category



mostly Wednesday, Tuesday, Thursday, more loan
inquiry came

Segmented analysis on(application data)

income vs education status



Conclusion –
Target=0(non-
defaulter),target=1 (defaulter)

we get insights from defaulter ,
non defaulter data sets those
who are in target=0 category
they have almost similar
income ,highest salary income
are also married ,taking more
loan and non defaulter
category while in target 1
category have varied income
but 50% of the people lies are
in similar income range (IQR)
similarity in target1 & target2
married have high income ,
outliers are also high

Top/least 10 Correlation between column top 10

	Column1	Column2	Correlation	Abs_Correlation
2978	AGE	DAYS_BIRTH	0.999591	0.999591
1734	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998491	0.998491
334	AMT_GOODS_PRICE	AMT_CREDIT	0.986726	0.986726
1119	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.949504	0.949504
992	CNT_FAM_MEMBERS	CNT_CHILDREN	0.893276	0.893276
1790	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.861454	0.861454
1343	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.860421	0.860421
1511	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.820828	0.820828
335	AMT_GOODS_PRICE	AMT_ANNUITY	0.766945	0.766945
279	AMT_ANNUITY	AMT_CREDIT	0.762117	0.762117

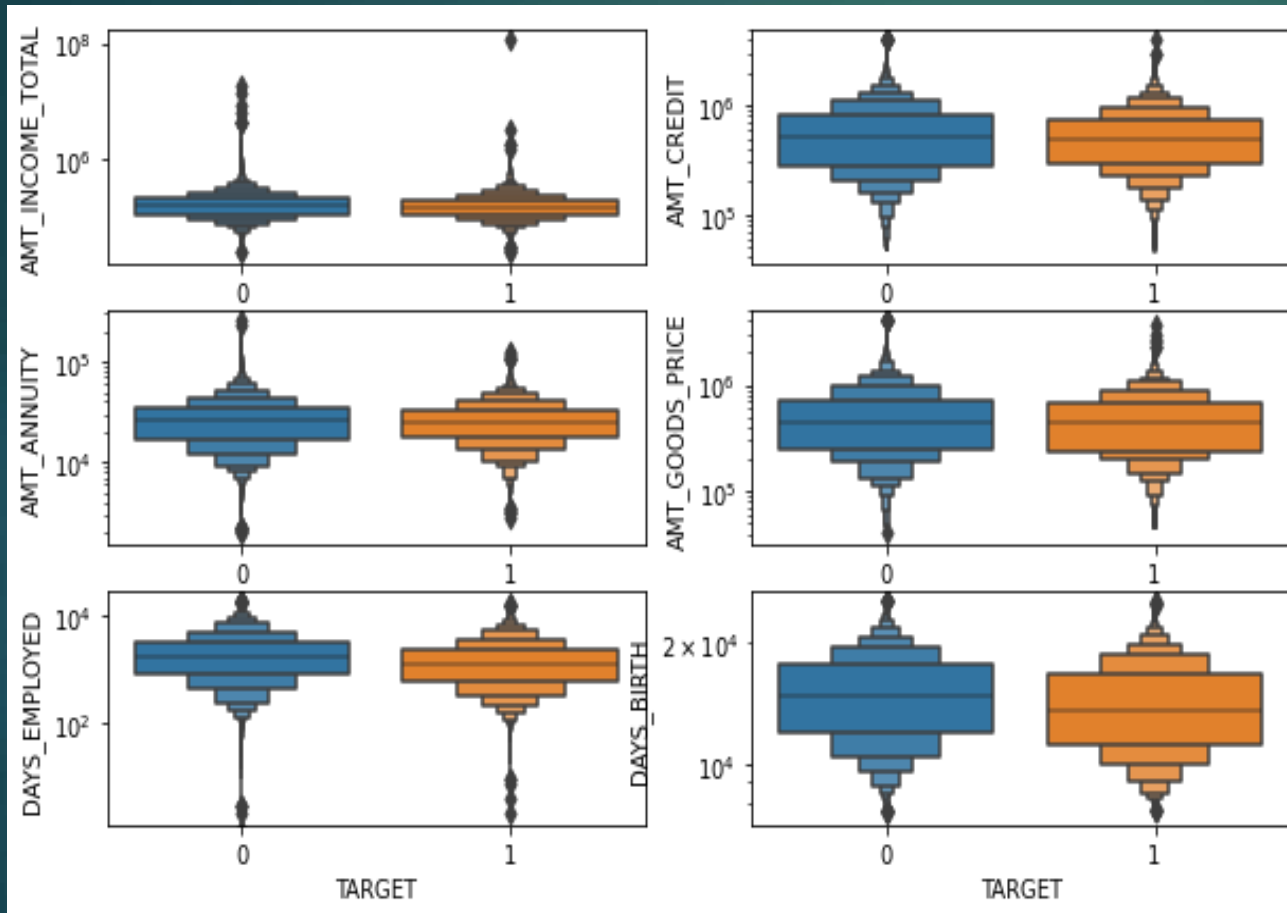
Conclusion-

Price of good amount and credit
Strongly correlated
Fam members and children have
Correlation

Least 10

	Column1	Column2	Correlation	Abs_Correlation
1992	FLAG_DOCUMENT_4	FLAG_MOBIL	0.000019	0.000019
2454	FLAG_DOCUMENT_12	FLAG_DOCUMENT_2	-0.000018	0.000018
2462	FLAG_DOCUMENT_12	FLAG_DOCUMENT_10	-0.000015	0.000015
727	FLAG_EMP_PHONE	FLAG_MOBIL	-0.000014	0.000014
1882	FLAG_DOCUMENT_2	FLAG_MOBIL	0.000013	0.000013
2422	FLAG_DOCUMENT_12	CNT_CHILDREN	0.000013	0.000013
2322	FLAG_DOCUMENT_10	FLAG_MOBIL	0.000011	0.000011
2432	FLAG_DOCUMENT_12	FLAG_MOBIL	0.000006	0.000006
495	DAYS_EMPLOYED	SK_ID_CURR	0.000005	0.000005
2931	FLAG_DOCUMENT_21	FLAG_PHONE	-0.000005	0.000005

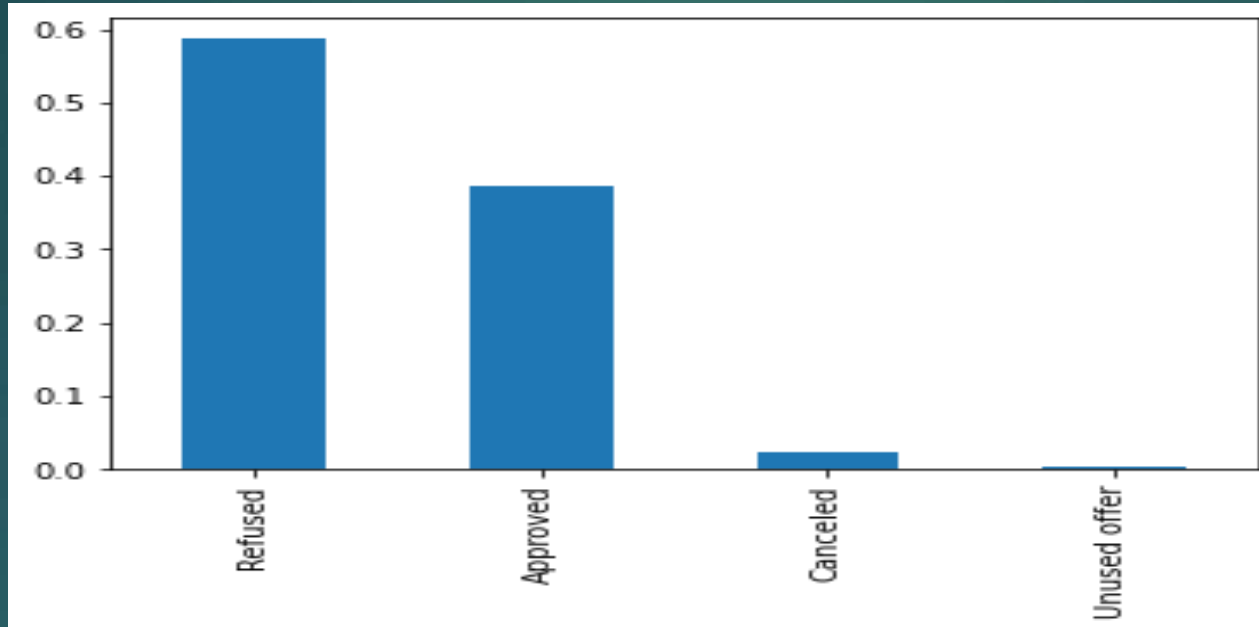
Bivariate analysis with respect to multiple column and target column



Insights-

people with higher no of employment days are less likely to default. Majority of defaulting people are having less total income. People with greater number of days born count are less likely to default Whose credit amount is greater than 50000 tends to be less default

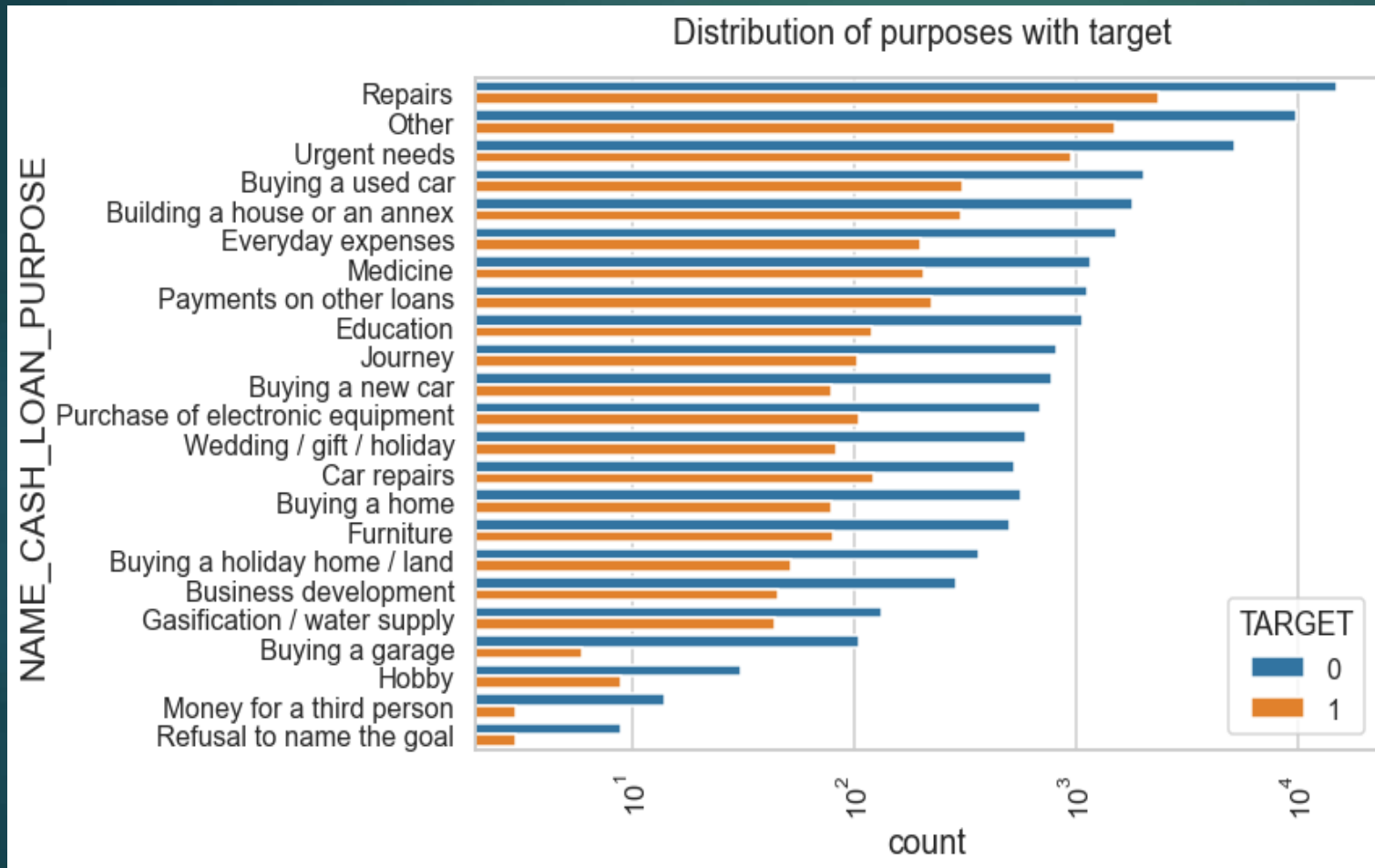
Insights from previous data loan contract status



Refused loan contract status high in previous , then approved rate are high , then canceled status

Merging the two data sets

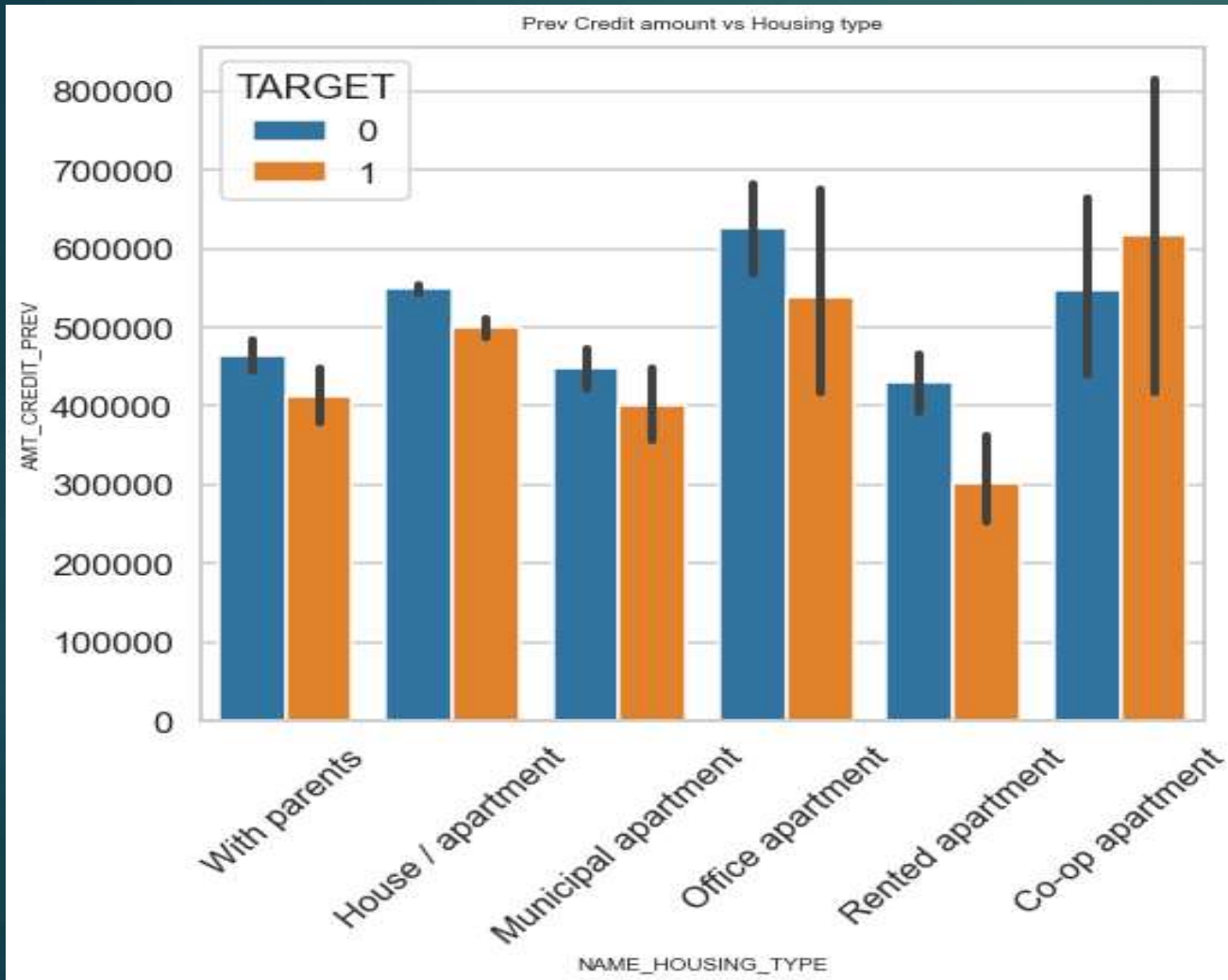
Distribution of purposes with target



Insights-

we can focus on client has minimum payment difficulties through data
buying garage , holiday home, building house has more loan required , more payment difficulties so we can thought how can we reduce our risk to give loan those people and not lossing the customer we can give better opotuninty

Bivariate analysis of merged data (credit amount vs housing type)



Conclusion-

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 or municipal apartment for successful payments.

Less chance of defaulter--

working as state servant

who have high income

who are married

old female client

widow who unused previous loan status

old , retired people

client have higher education level (mostly female)

student

businessmen

those who have unused loan status previously in

required

Conclusion

bank should focus on more like student , businessmen ,retired people to give loan

buying car old or new have , repair has more defaulter must need to work on this type

should also clients like with parents they less defaulter category

less cancelation also in water supply

Majority of defaulting people are having less total income. People with greater number of days born count are less likely to default Whose credit amount is greater than 50000 tends to be less default

secondary education have more defaulter compare to higher education and have more credit amount of bank , those are who lies in civil marriage category and have more credit history

higher education people have more defaulter compared to non defaulter data

92% are non defaulter category while 8% are in defaulter category