EDA CREDIT

Assignment – DS C45-Jun'22

Name-Karishma Singh

Purpose of the study

The company wants to understand the driving factors behind loan default deriving the variables which are strong indicators of defaulter.

The company will then utilize this knowledge for its portfolio and risk assessment.

Steps taken into consideration

- Data understanding – Data dictionary
- Analysed Data quality
  - Checked the missing or NAN values
  - Imputed them on the basis of their missing category – MAR,MCAR,MNAR
  - Checked for Outliers and imputed it with Mean/Median based on values in the variable
  - Treated Irrelevant values in categorical variables (for e.g. Purpose of loan)
- Master Data Preparation by merging two Input datasets
- Data Analysis based on different segments and categories

Data Analysis

i) Applicant Age - Distribution of Applications and default cases using variable- Applicant Age

Plot 1

Response on age group

Target 0- Non Default
Target 1- Default
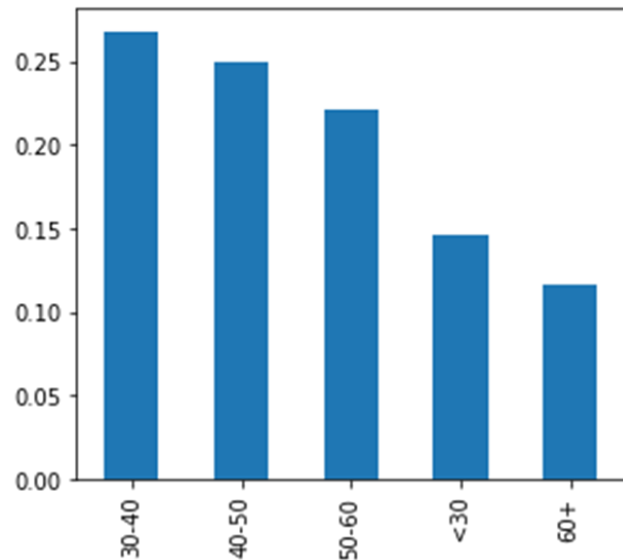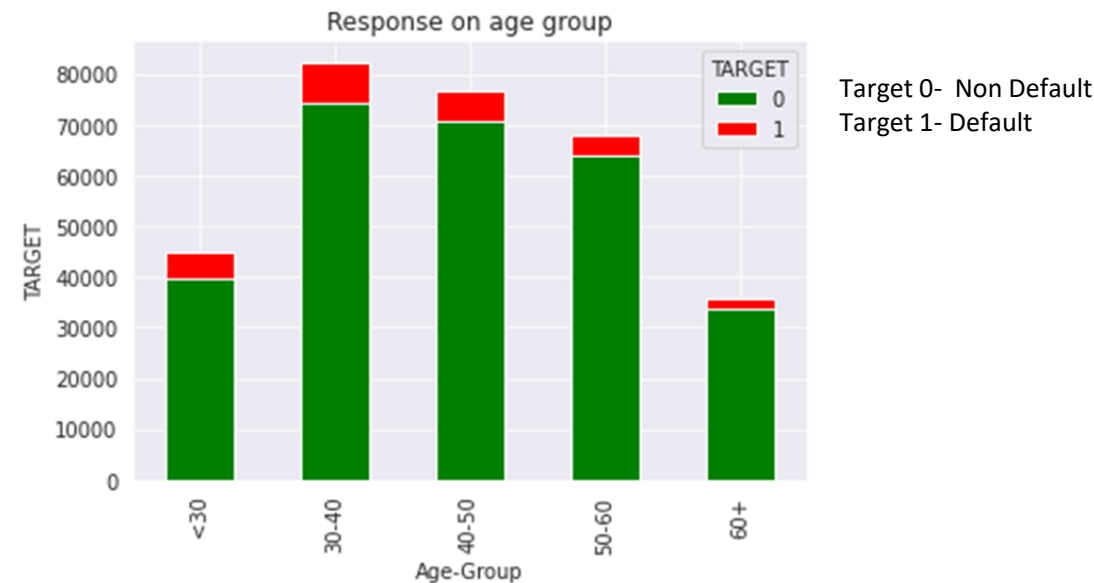
Plot 2

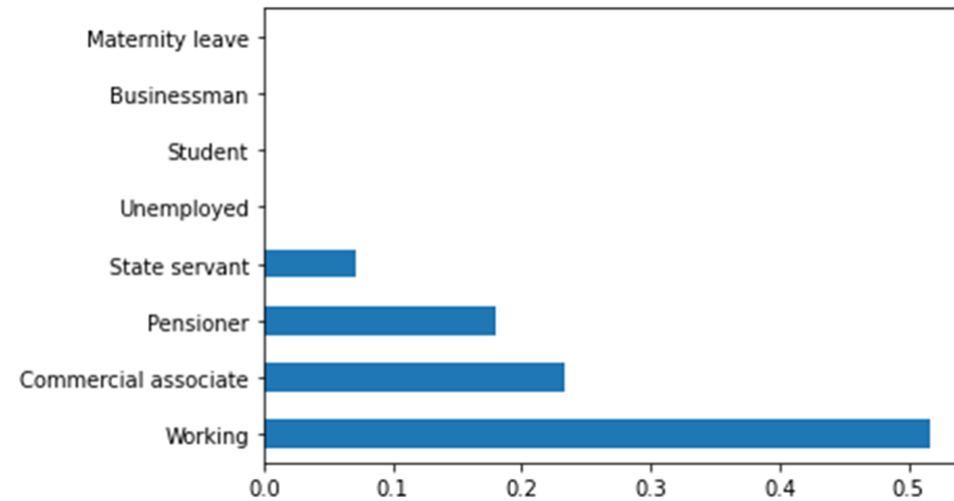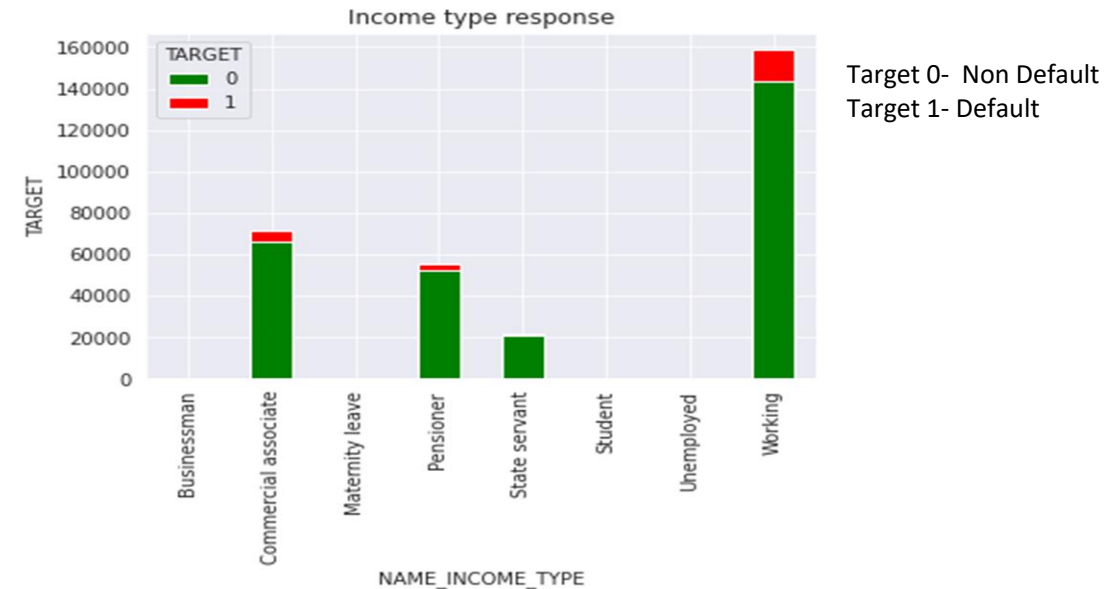| Applicant Age	# Default cases	# Total Applications	% Default cases
<30	5158	45021	11%
30-40	7888	82308	10%
40-50	5856	76541	8%
50-60	4172	68062	6%
60+	1751	35579	5%

Table 1

Insights-

- Loan applications are **highest in 30-40 age group** as exhibit in plot 1.

- % Default cases is highest in **age group less than 30**. Please refer the table-1 which shows comparison of volume and % default cases between different age groups as exhibit in plot 2

ii) Applicant Income Type -
Distribution of Applications and default cases using variable- Applicant Income type

Plot 1

Target 0- Non Default
Target 1- Default

Plot 2

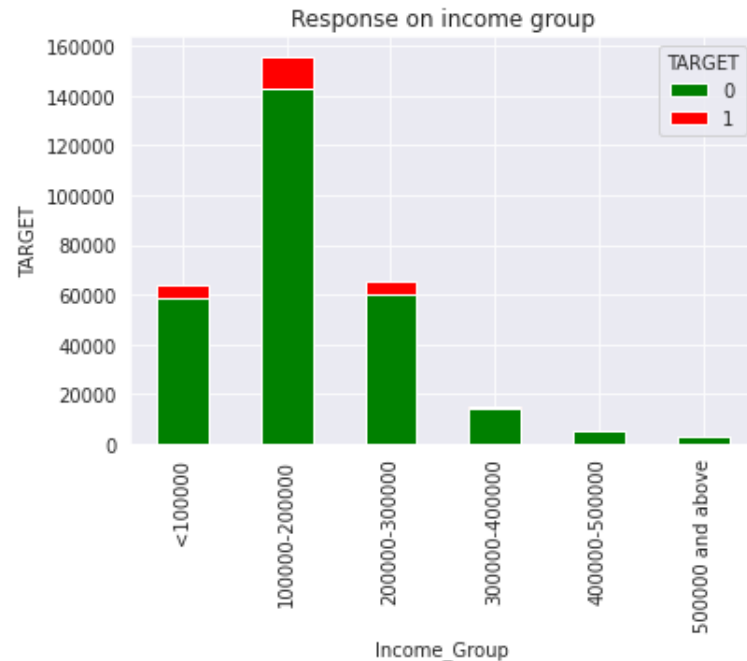| Income Type	# Default cases	# Total Applications	% Default cases
Working	15224	158774	10%
Commercial associate	5360	71617	7%
Pensioner	2982	55362	5%
State servant	1249	21703	6%
Unemployed	8	22	36%
Student	0	18	0%
Businessman	0	10	0%
Maternity leave	2	5	40%

Table 2

Insights-

- **Working class** category applied more applications compared to other income types as exhibit in plot 1

- % Default cases is highest in income type- *Unemployed and Maternity leave* but max. volume of default cases highest in- **Working and Commercial Associate**. Please refer the table-2 which shows comparison of volume and % default cases between income types as exhibit in plot 2

iii) Applicant Income Group -
Distribution of Applications and default cases using variable- Applicant Income Group

Response on income group

Plot 1

Target 0- Non Default
Target 1- Default

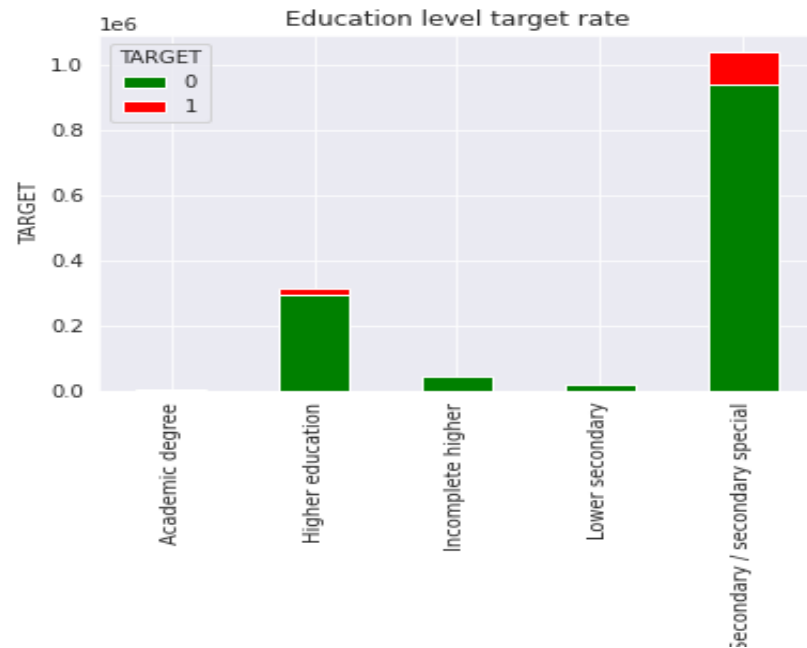| Income Group	# Default cases	# Total Applications	% Default cases
<100000	5225	63698	8.2%
100000-200000	13326	155898	8.5%
200000-300000	4921	65176	7.5%
300000-400000	885	14676	6%
400000-500000	322	5361	6%
500000 and above	146	2702	5%

Table 3

Insights-

- **Income Group -100000-200000 a**pplied more applications compared to other income groups as exhibit in plot 1

- % Default cases is highest in income group- **<100000, 100000-200000 and 200000-300000** but max. volume of default cases highest in- **100000-200000**. Please refer the table-3 which shows comparison of volume and % default cases between income types as exhibit in plot 2

iv) Applicant Education Level -

Distribution of Applications and default cases using variable- Applicant Education Level

Plot 1

Target 0- Non Default
Target 1- Default

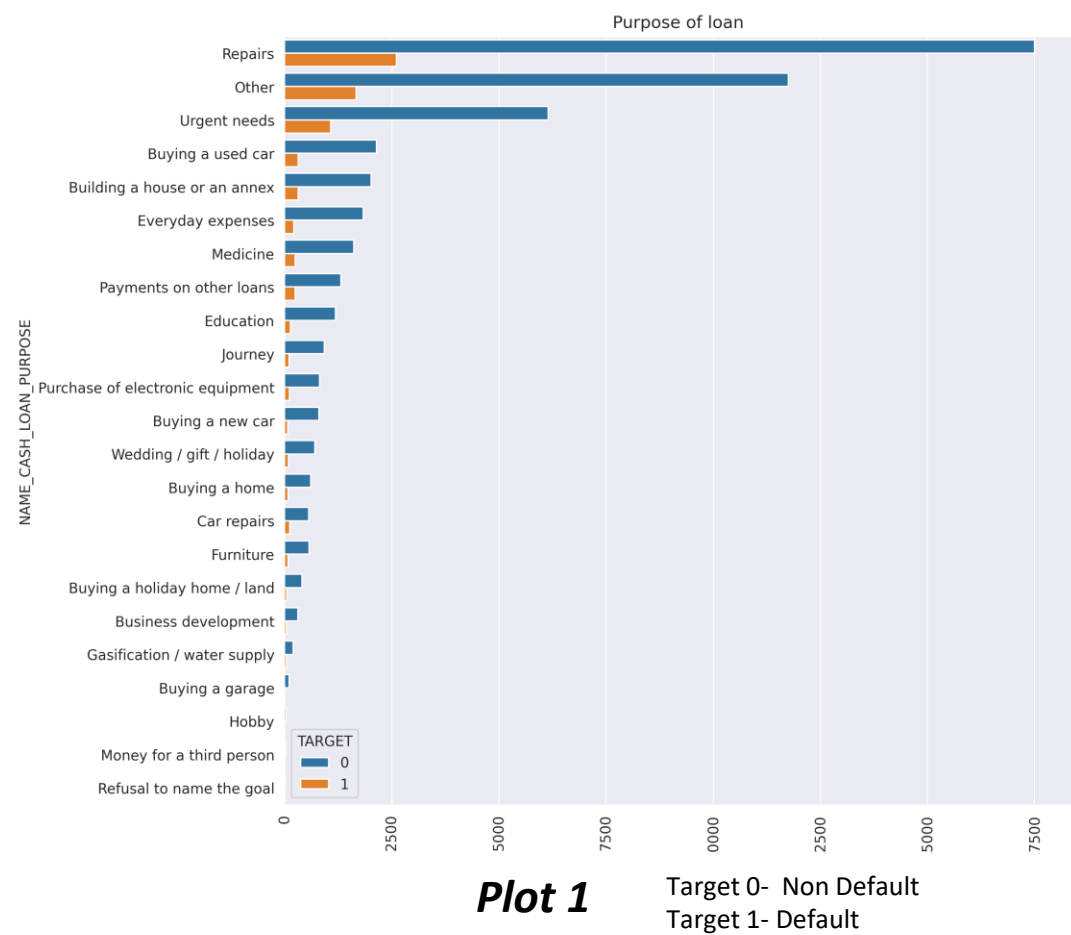| Education Level	# Default cases	# Total Applications	% Default cases
Academic degree	0	103	0%
Higher education	3184	53325	6%
Incomplete higher	673	7794	9%
Lower secondary	312	2893	11%
Secondary / secondary special	16445	176383	9%

Table 4

Insights-

- **Education Level- Secondary/Secondary Special a**pplied more applications compared to other Education level as exhibit in plot 1
- % Default cases is highest in income group- **Lower Secondary, Secondary/Secondary Special and Incomplete Higher** and max. volume of default cases highest in- **Secondary/Secondary Special**. Please refer the table-4 which shows comparison of volume and % default cases between Education level

v) Applicant Loan Purpose -
Distribution of Applications and default cases using variable- Applicant Loan Purpose

Purpose of loan

Plot 1

Target 0- Non Default
Target 1- Default

Top 6 Categories contributing to 80% applications

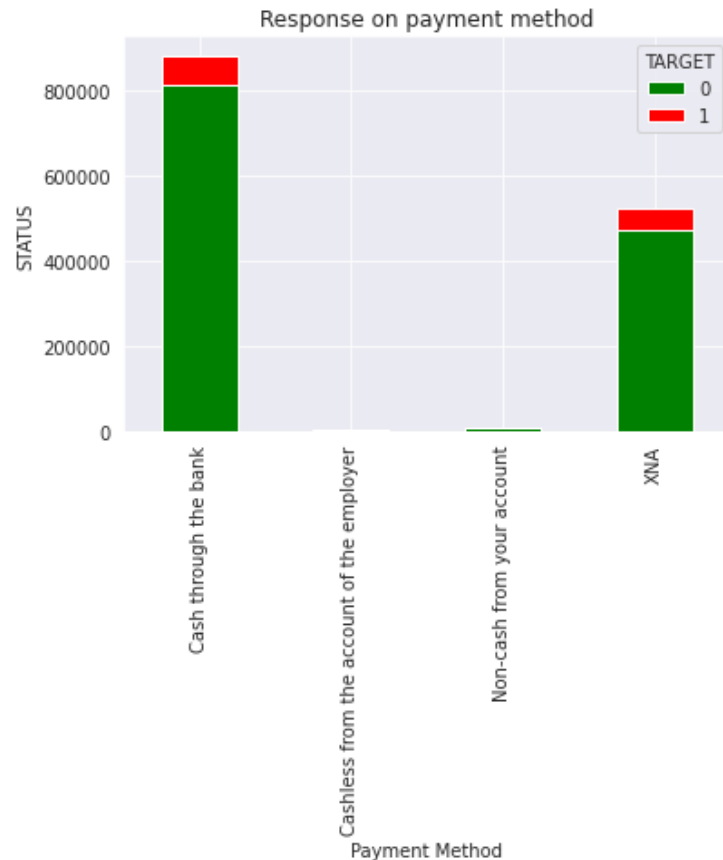| Education Level	# Default cases	# Total Applications	% Default cases
Repairs	423	3325	13%
Other	264	2192	12%
Urgent needs	190	1229	15%
Buying a used car	56	416	13%
Building a house or an annex	50	374	13%
Everyday expenses	34	338	10%

Table 5

Insights-

• With **Loan Purpose- 'Repairs'** bank received more applications compared to other loan purposes as exhibit in plot 1

• % Default cases is high in Loan Purpose- **'Urgent needs'**. Please refer the table-5 which shows comparison of volume and % default cases between Top 6 Loan Purpose which contributed to more than 80% of applications and default cases

v) Applicant Payment Method -

Distribution of Applications and default cases using variable- Applicant Payment Method

Response on payment method

Plot 1

Target 0- Non Default
Target 1- Default

Education Level	# Default cases	# Total Applications	% Default cases
Cash through the bank	12143	152033	8%
Cashless from the account of the employer	12	173	7%
Non-cash from your account	116	1291	9%
XNA	8343	87000	10%

Table 6

Insights-

- **'Cash through the bank'** payment method observed max applications as exhibit in plot 1
- % Default and Volume of default cases high for payment method- **'Cash through the bank'**. Please refer the table-6 which shows comparison of volume and % default cases between different Payment Methods.

Actionable Insights

High income and age group more than 40 have comparatively less number of unsuccessful payments.

Pensioners and state servants default less when compared with Working and Commercial Associate.

Maximum no. of defaulters and loan rejections are found in Repair purpose of loan.

'Country wide' and 'Stone' channel types got more approvals when compared with other Channels like Credit and cash offices.
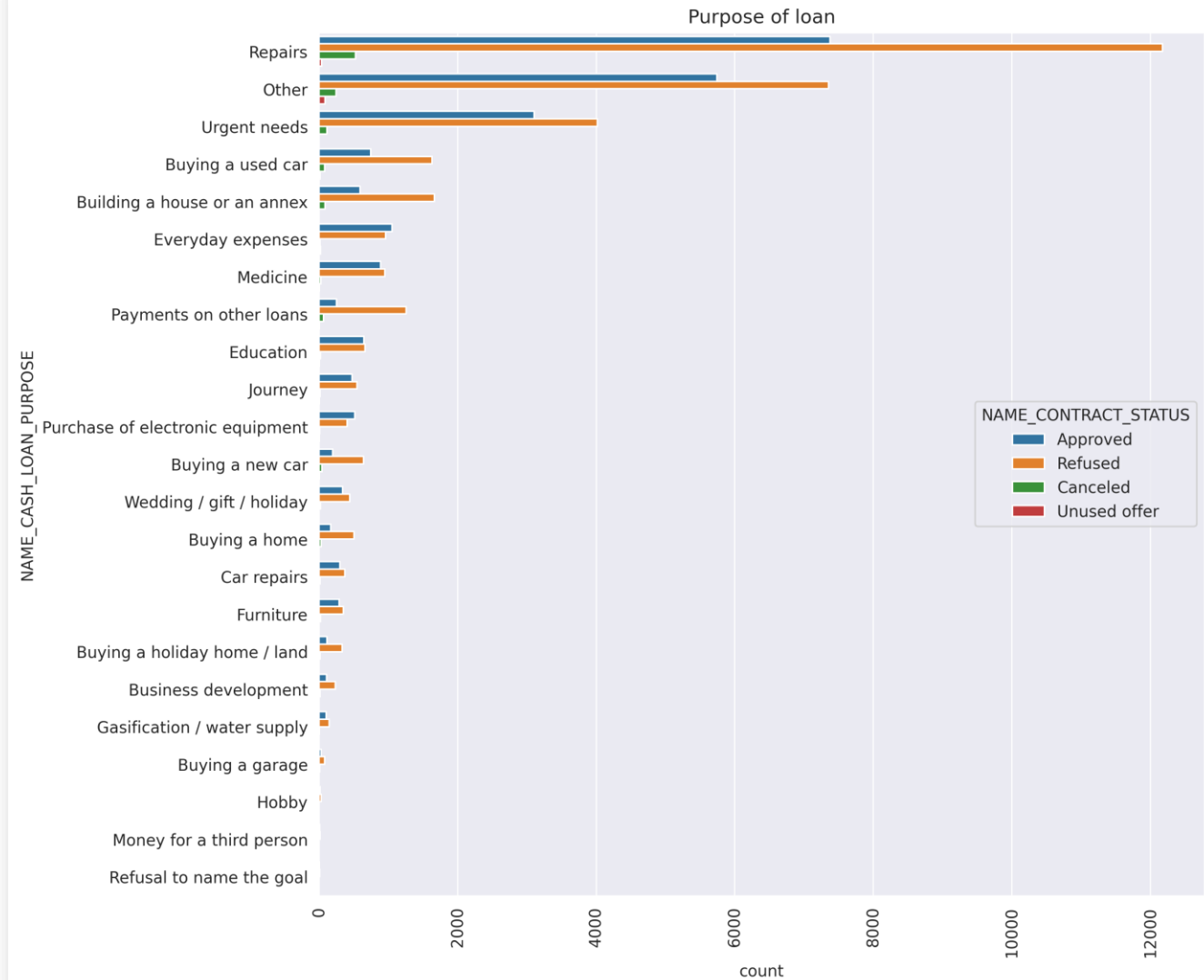
Loan Exposure is more in consumer electronics and connectivity industry. Bank can diversify the loan distribution.

Education Level- Secondary/Secondary Special has more proportion of default cases when compared with other education levels of the applicants.
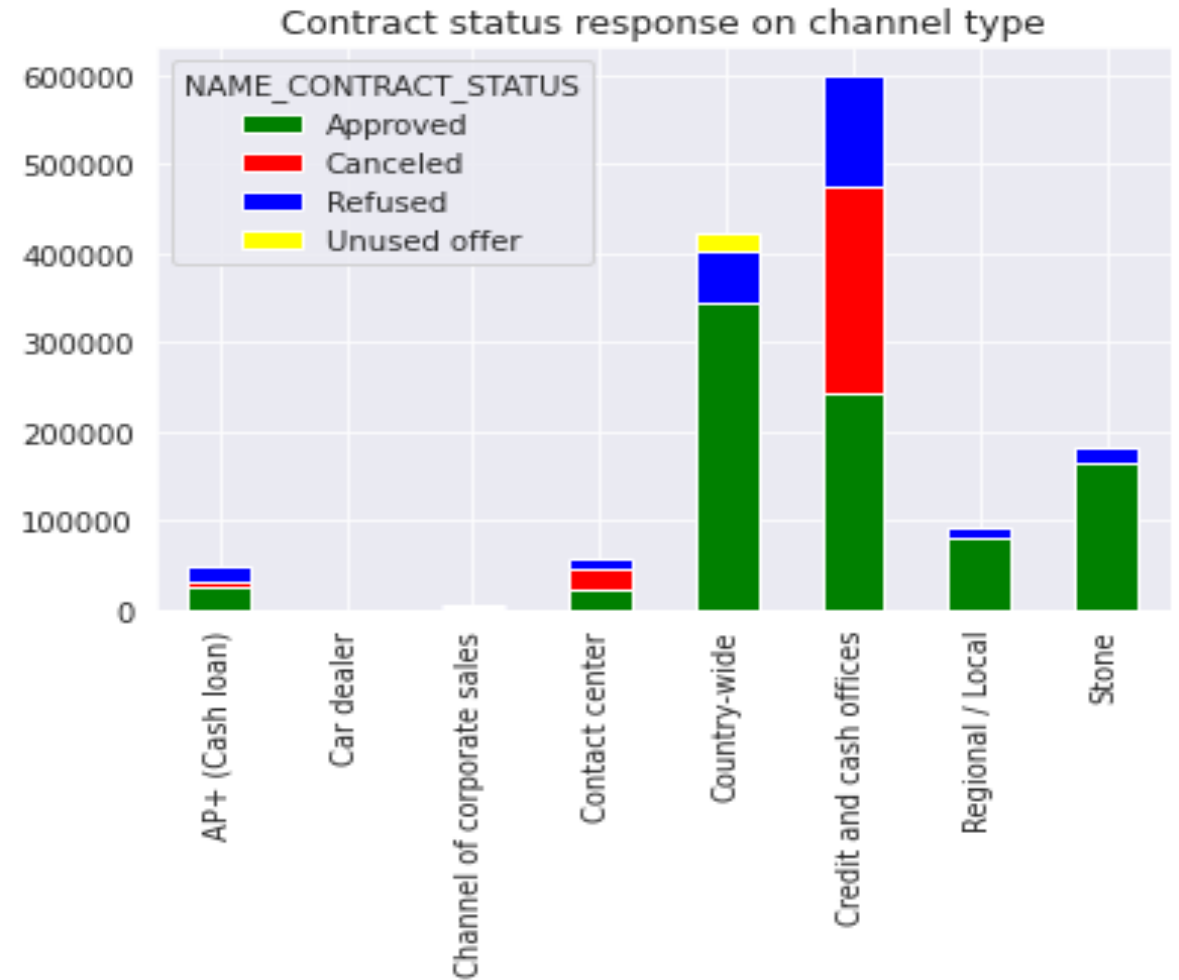
Appendix

- **Distribution of loan purpose and contract status-**

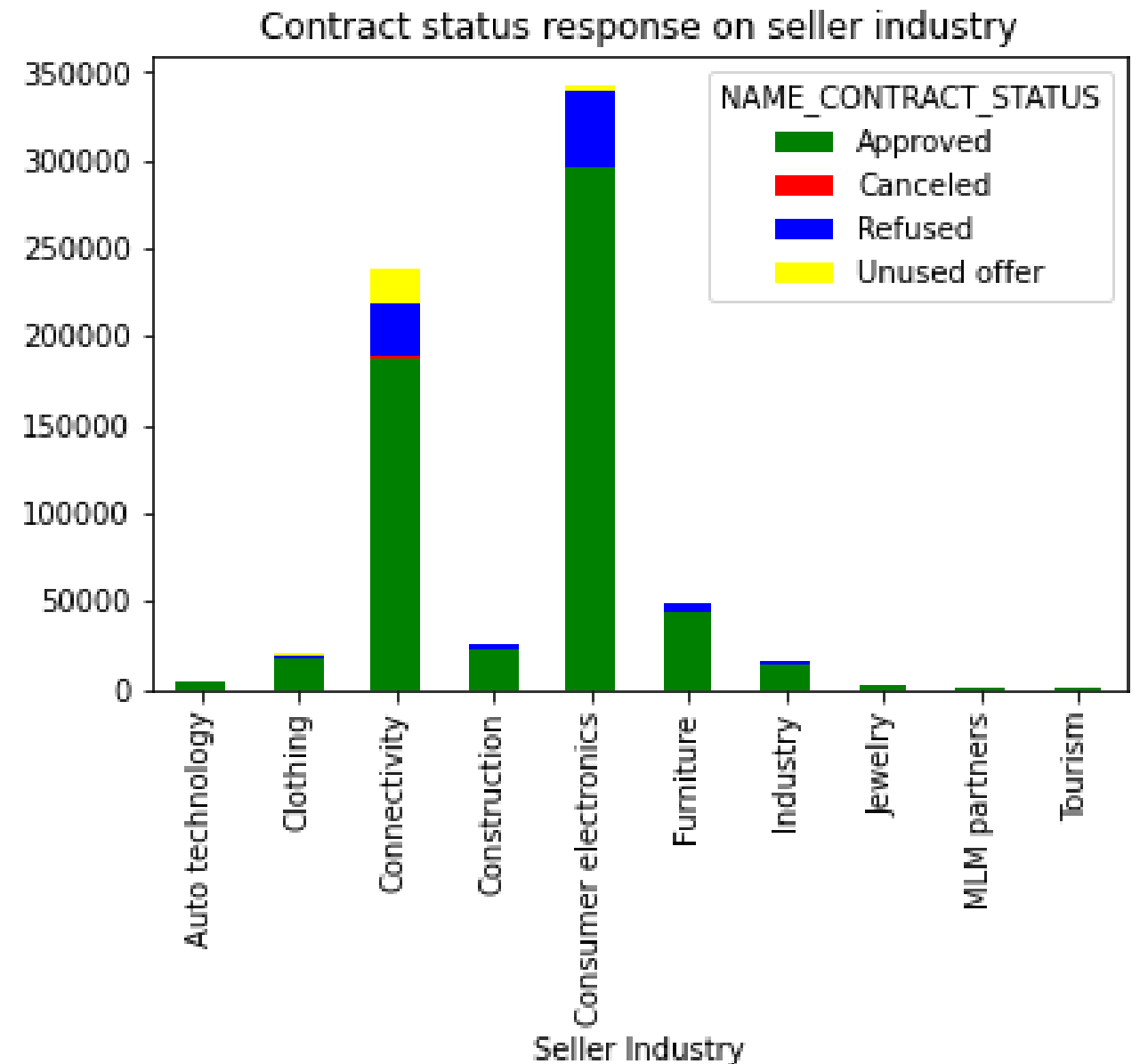- Most of the loan rejection and approval both were from Repairs.

Comparison in channel type and contract status

- Clients from country wide channel got more approvals and cancelled cases are more from credit and cash offices.

Contract status response on channel type

Contract status response on seller industry

Loan exposure is more in consumer electronics and connectivity industry. Bank can diversify the loan distribution.

Thank You!