Credit EDA Case Study

By Gunavardhan Sattu

Business Objective

- Use data analysis to assist a loan company in making better lending choices
- We aim to avoid lending to unreliable borrowers and ensure those who can repay aren't turned away
- By looking at data, we aim to make loan approvals better while also avoiding potential losses.
- Figure out the main reasons why people might not be able to pay back loans.
- We want to find clues in the data that can help us predict if someone might not pay on time.

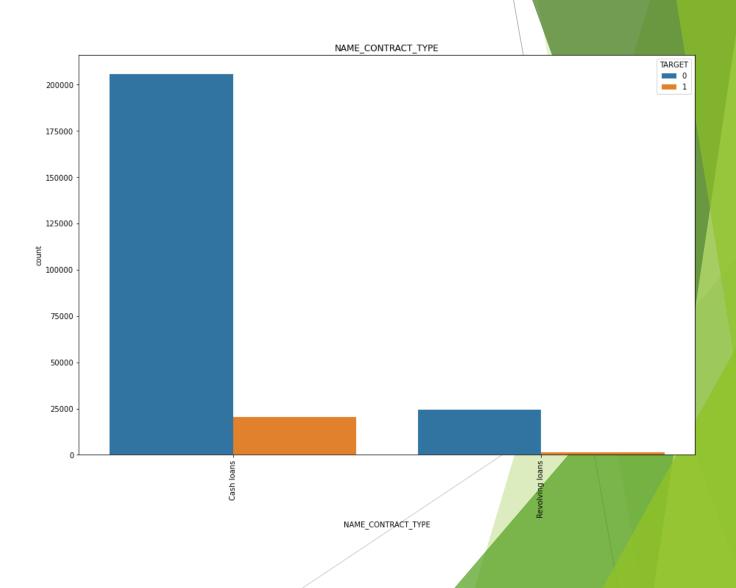
Approach

- Data Understanding
- Data Cleaning
 - ▶ Removing columns with Null values greater than 40
 - ► Impute/Remove missing values
 - ► Fixing invalid values and Filter Data
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

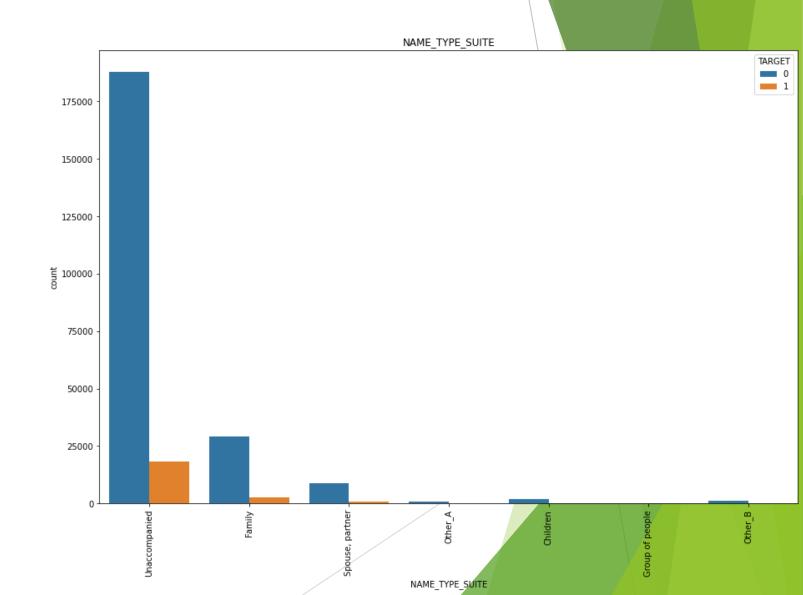
Assumptions

- ▶ Setting a threshold of 40% to eliminate columns with null values
- Filling missing values with median or mode
- In the dataset, the "TARGET" column has been divided into "Target 1" for clients with payment difficulties and "Target 0" for all other cases
- Columns containing values like "XPA," "XNA," or other non-relevant values are treated as missing data

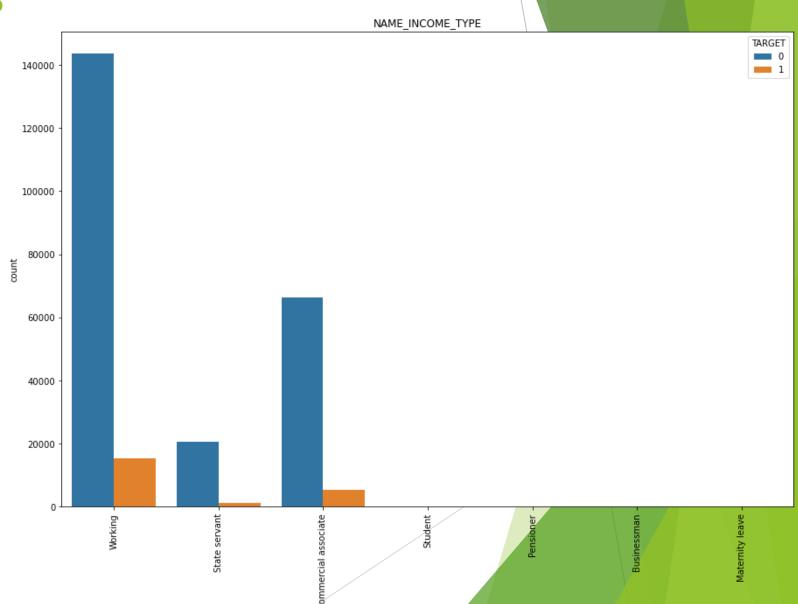
☐ The majority opted for cash loans, with a higher 9% default rate, compared to a 5.65% default rate for revolving loans.



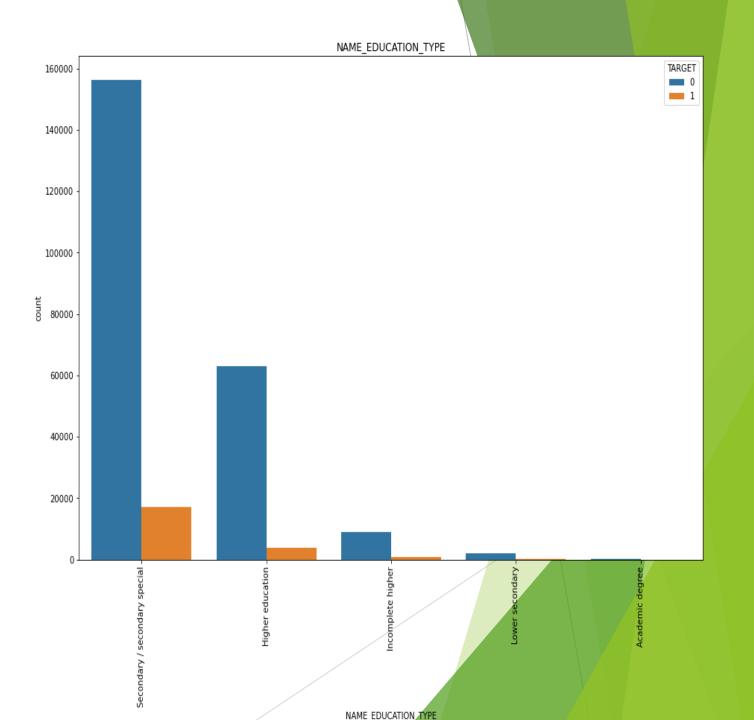
□ Accompanied customers take more loans, with an 8.5% default rate.
 ○ Other_B name type is prone to default, while clients with children are less likely to default.



■ Working professionals have a 9.5% default rate, seeming secure with higher loan count. Maternity leave individuals show higher default rates.

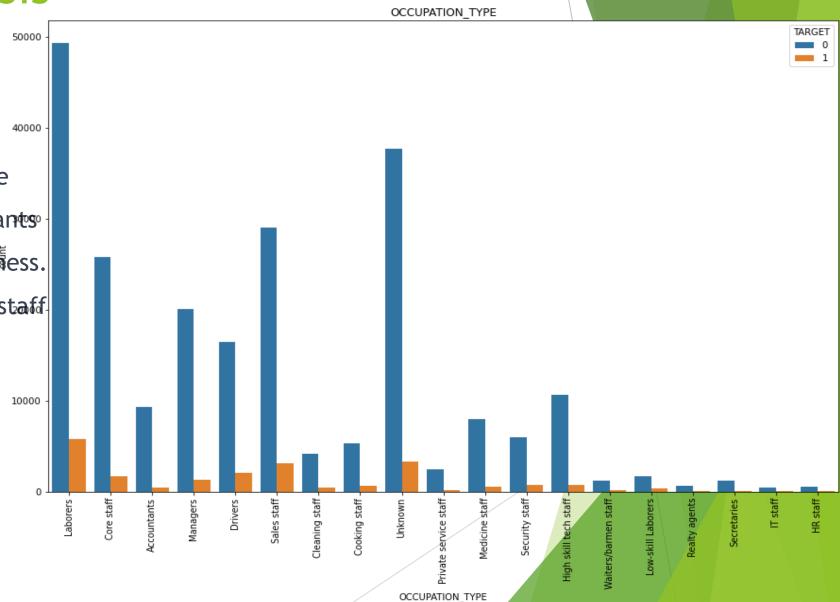


□ People who are low educated are more likely to default whereas people with higher education and academic degree are less likely to default.

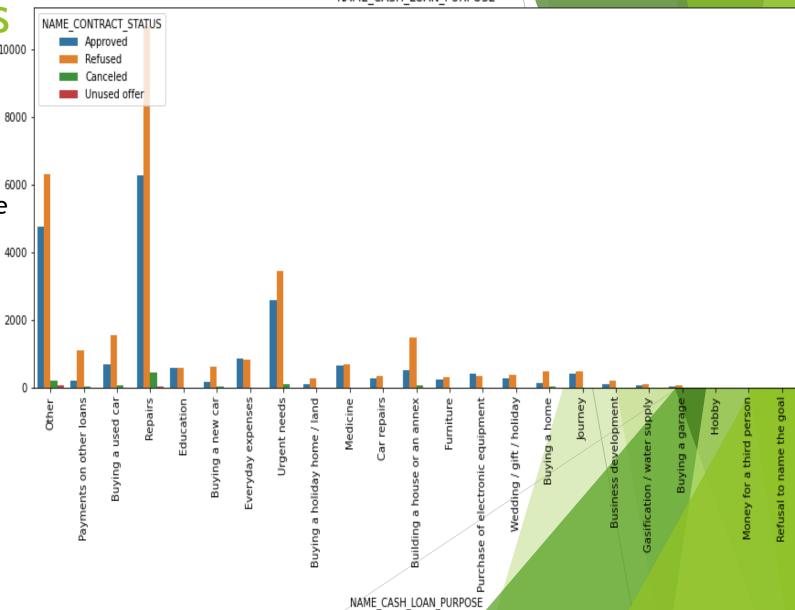


Low-skill workers like laborers, 40000-drivers, waiters, and barmen are likely to default, while accountants have low defaults due to awarensess.

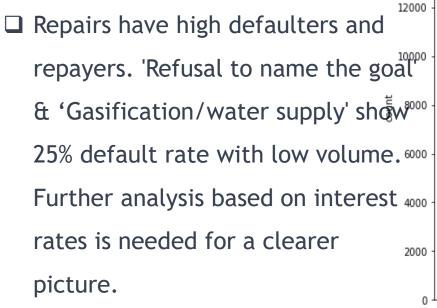
Core/HR/IT/Manager/Medicine staffshow low default rates.

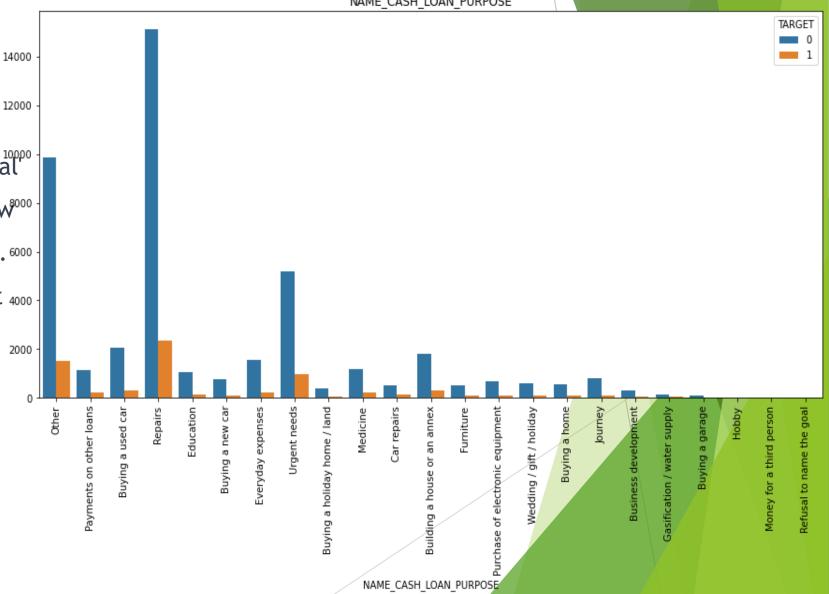


Most of the customers applied for the purpose of Repairs and the same type has most of the refusals and cancellations

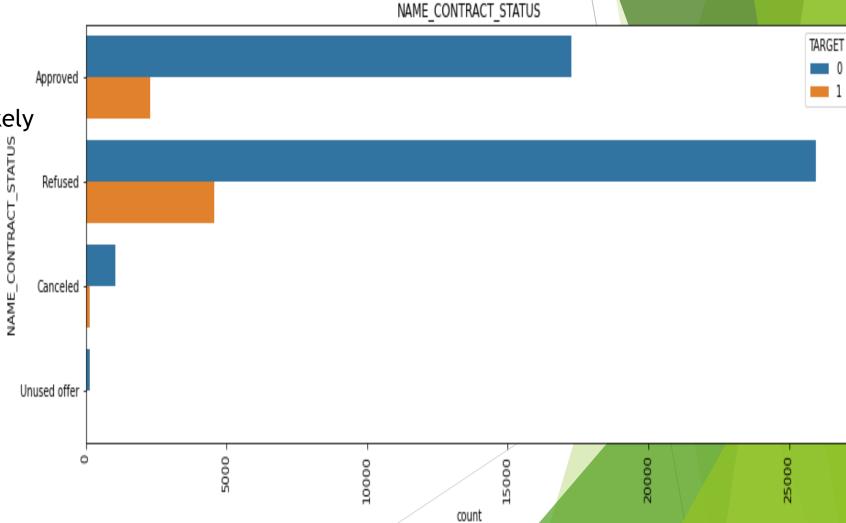


NAME_CASH_LOAN_PURPOSE

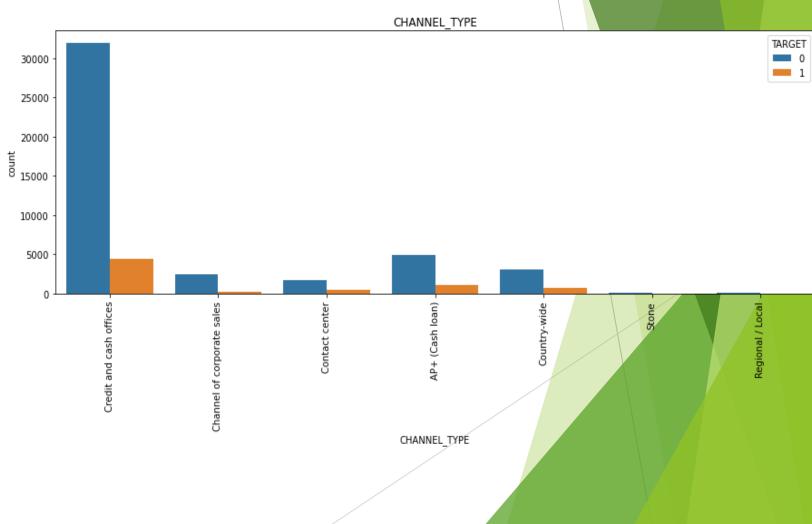




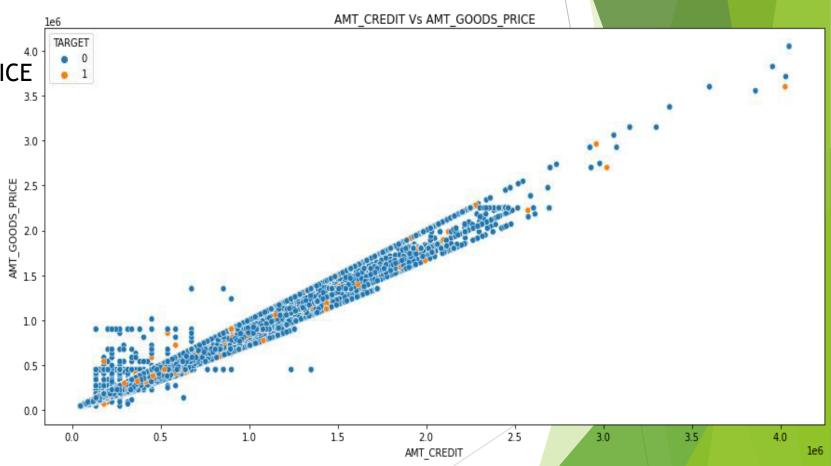
Applicants whose loans were previously refused are more likely to be defaulters



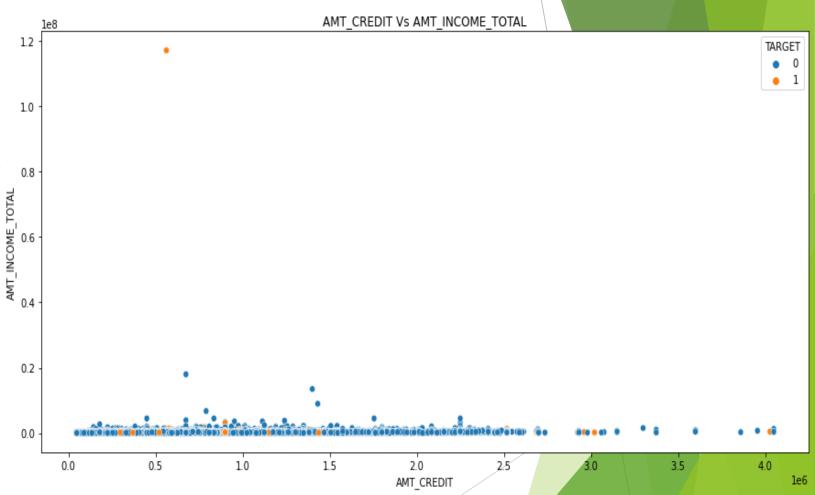
☐ The most effective channels for acquiring new customers with a lower rate of defaulters are corporate sales and regional/local channels.



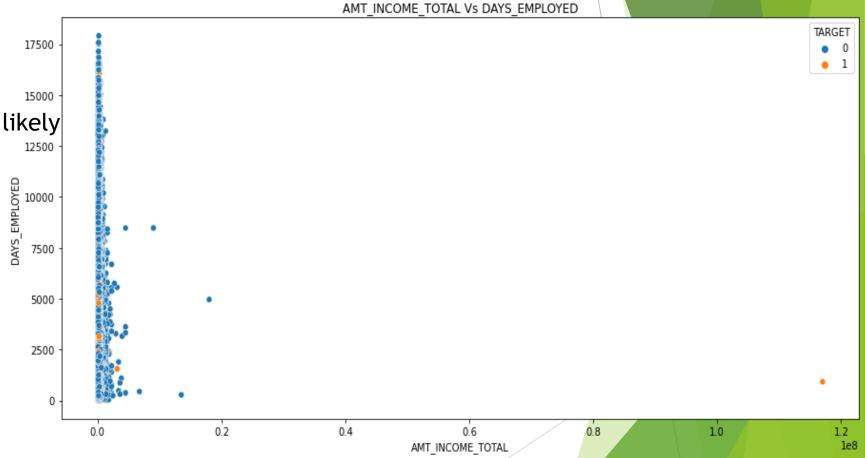
AMT_CREDIT and AMT_GOODS_PRICE are linearly corelated. The defaulters got decreased as the credit amount is increasing.



■ Lower-income individuals take more loans; defaulters decrease as income and credit amounts rise.



□ Customers with extended
employment duration are less likely
to be defaulters.



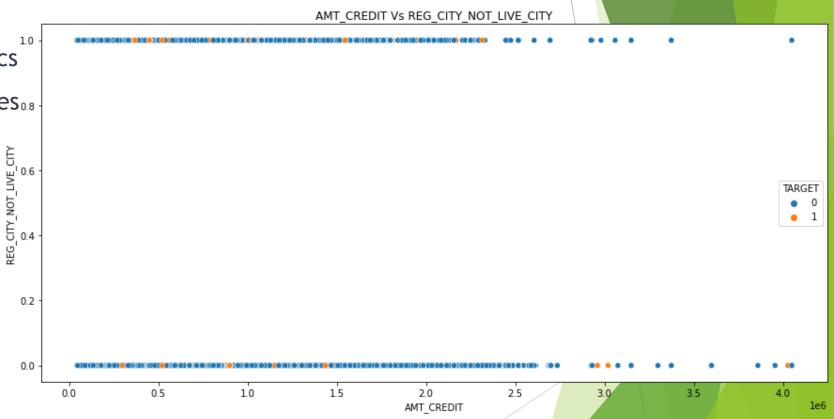
☐ Customers with credit under 5 lacs

and mismatched contact addresses

have higher default likelihood

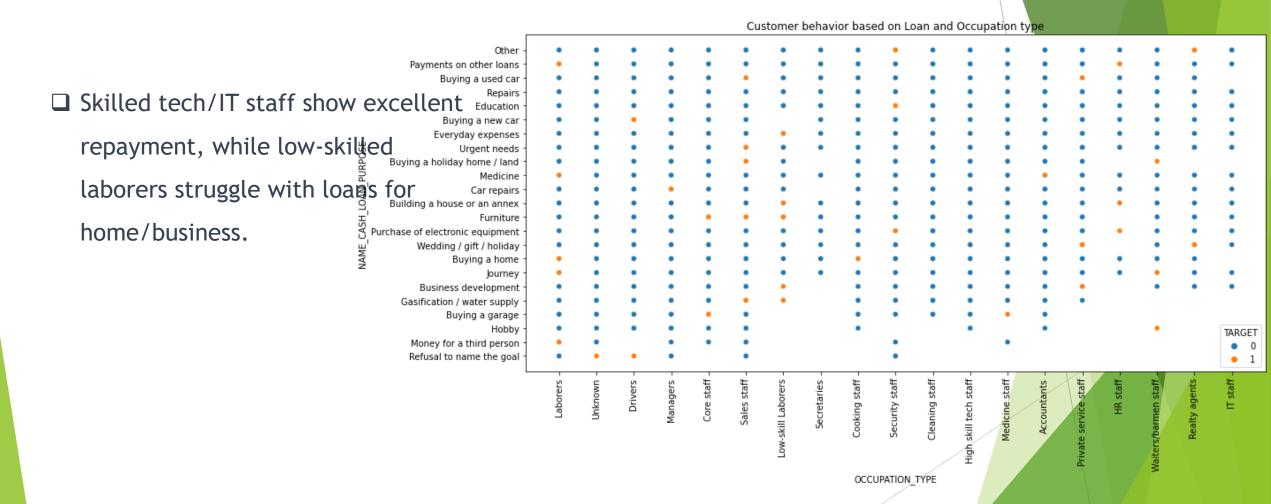
¡

]

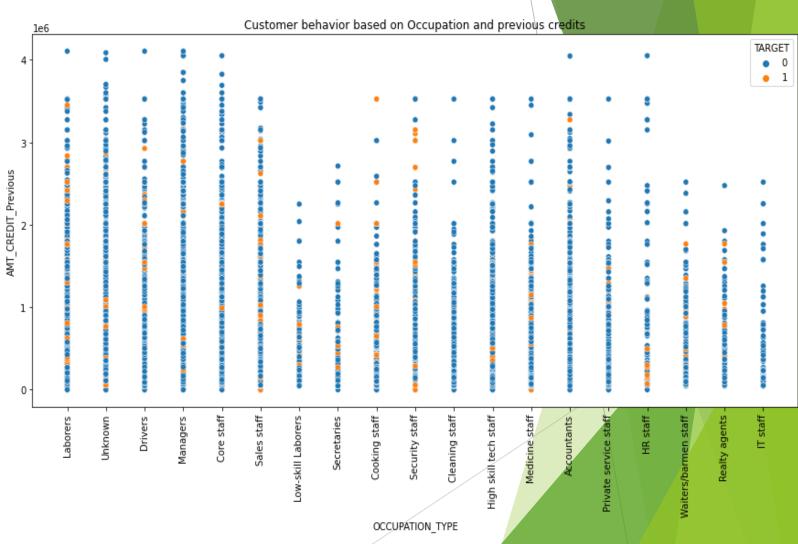


☐ Customers with medical and new car/home/garage loans have low default rates, suggesting safe lending decisions for these segments.





☐ Higher credit amounts lead to increased default rates among laborers, drivers, and security staff.



Multivariate Analysis

with higher income are linked to
lower defaults. Certain jobs like

drivers, laborers, and security seaffow-skill Laborers - 0.28 0.15

Managers - 0.089 0.16

Medicine staff - 0.24 0.14

Core staff - 0.24 0.14

Core staff - 0.087 0.14

Drivers - 0.08 0.27

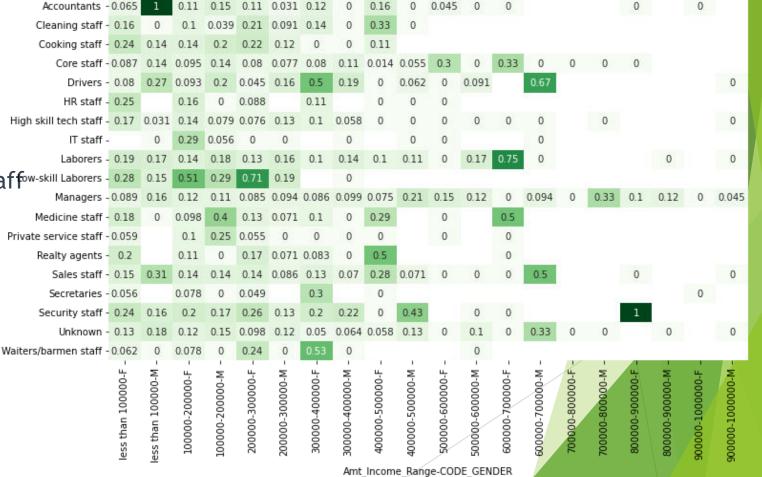
High skill tech staff - 0.17 0.031

Laborers - 0.19 0.17

Managers - 0.089 0.16

Medicine staff - 0.18 0

Customer behavior based on Occupation,Income and Gender
0.15 0.11 0.031 0.12 0 0.16 0 0.045 0 0



Multivariate Analysis

☐ Payment issues noted among specific groups: co-op apartment residents without cars (e.g., drivers, medical staff), low-skilled laborers, \(\frac{1}{2} \) and barmen in municipal apartments. Caution is advised for loans to these demographics.

Customer behavior based on Occupation, Housing type and own car

0	0.11	0.27	0.083	0.4	0.18	0	0.095	0.18		0	0.027
	0.13	0.17	0	0.17	0.22		0.13	0	0		0
	0.18	0.19	0.22	0.32	0.21		0.16	0.14		0.25	0.25
0	0.1	0.089	0.18	0.039	0.084	0	0.09	0.075	0	0	0.046
0.5	0.19	0.17	0	0.14	0.23	0.3	0.17	0.16	0.15	0.2	0.29
	0.13	0	0		0.4		0.11	0			
0	0.11	0.14	0	0.13	0.18	0	0.1	0.12	0	0	0.1
	0.2	0			0.25		0				0
0.067	0.18	0.18	0.022	0.24	0.26	0.43	0.12	0.17	0.067	0.17	0.12
	0.3	0.8		0.062	0.43		0.16	0		1	0.27
1	0.11	0.23	0.03	0.12	0.065	0	0.097	0.14	0	0.054	0.2
1	0.14	0.093	0.18	0.44	0.19	0	0.074	0		0	0.04
0	0.082	0.067		0	0.22		0.017	0		0	0
	0.13				0		0.18	0		0	0
0	0.14	0.14	0.095	0.16	0.21	0	0.13	0.31	0	0.029	0.19
	0.071	0	0	0	0.29	0	0.086	0			0
0	0.21	0.16	0.12	0.12	0.29		0.16	0	0	0	0.095
0.3	0.12	0.13	0.077	0.15	0.14	0	0.12	0.031	0	0.087	0.091
	0.096	0.67		0	0		0.21	0			0
N-Co-op apartment -	N-House / apartment -	N-Municipal apartment –	N-Office apartment -	N-Rented apartment -	N-With parents -	Y-Co-op apartment -	YHouse / apartment -	Y.Municipal apartment -	YOffice apartment	YRented apartment -	YWith parents -
	0 0.5 0 0.067 1 1 0 0	0.13 0.18 0 0.1 0.5 0.19 0.13 0 0.11 0.2 0.067 0.18 0.3 1 0.11 1 0.14 0 0.082 0.13 0 0.14 0.071 0 0.21 0.3 0.12	0.13	0.13	0.13	0.13	0.13	0.13 0.17 0 0.17 0.22 0.13 0.18 0.19 0.22 0.32 0.21 0.16 0 0.1 0.089 0.18 0.039 0.084 0 0.09 0.5 0.19 0.17 0 0.14 0.23 0.3 0.17 0.13 0 0 0.4 0.11 0.11 0.11 0.04 0.11 0 0.11 0.14 0 0.13 0.18 0 0.1 0.2 0 0.25 0 0 0.25 0 0 0.067 0.18 0.18 0.022 0.24 0.26 0.43 0.12 1 0.11 0.23 0.03 0.12 0.065 0 0.097 1 0.14 0.093 0.18 0.44 0.19 0 0.074 0 0.082 0.067 0 0.22 0.017 0.13 0.071 <	0.13	0.13	013 0.17 0 0.17 0.22 0.13 0 0 0 0 0.18 0.19 0.22 0.32 0.21 0.16 0.14 0.25 0 0 0.1 0.089 0.18 0.039 0.084 0 0.09 0.075 0 0 0 0.5 0.19 0.17 0 0.14 0.23 0.3 0.17 0.16 0.15 0.2 0.13 0 0 0 0.4 0.11 0 0 0.11 0.14 0 0.13 0.18 0 0.1 0.12 0 0 0 0.067 0.18 0.18 0.022 0.24 0.26 0.43 0.12 0.17 0.067 0.17 0.067 0.17 0.3 0.8 0.062 0.43 0.16 0 1 0.14 0.093 0.18 0.44 0.19 0 0.074 0 0 0.054 1 0.14 0.093 0.18 0.44 0.19 0 0.074 0 0 0.13 0.13 0.13 0 0 0 0.14 0.14 0.095 0.16 0.21 0 0.18 0 0 0.007 0.14 0 0.009 0.071 0 0 0 0.22 0.017 0 0 0.009 0.071 0 0 0 0.29 0 0.086 0 0 0.009 0.14 0.14 0.095 0.16 0.21 0 0.13 0.31 0 0.029 0.0086 0 0 0.0096

FLAG OWN CAR-NAME HOUSING

Multivariate Analysis

☐ High credit customers with mismatched permanent and contact addresses, especially among security staff and realty agents, exhibit higher default tendencies.

Customer behavior based on Occupation, Credit range and region mismatch Accountants - 0 0.15 Cleaning staff - 0.11 0.13 0.13 0.091 0.18 0.055 0.21 Cooking staff - 0.02 0.12 Core staff - 0.039 0.076 Drivers - 0.11 HR staff - 0 0.28 0.17 0.21 0.17 0.073 0 0.15 0.23 0.33 0.094 0.084 0.2 High skill tech staff - 0 IT staff -0.33 0 0.42 0.18 0.36 0.13 0.33 0.15 Laborers - 0.14 0.21 0.18 0.17 Low-skill Laborers - 0.22 0.16 0.27 0 0.17 Managers - 0.036 0.15 0.15 0.14 0.022 0.17 0.12 0.026 0.034 0.012 Private service staff - 0 0.031 0.15 0 0.031 0.25 0.053 Realty agents - 0 0.12 0.38 0.05 0.048 0.12 0.13 0.05 0.11 0.33 0.12 Security staff - 0 0.15 0.16 0.21 0.21 Unknown - 0.092 0.11 0.11 Waiters/barmen staff - 0 0.14

Amt Credit Range-REG REGION NOT LIVE REGION

- 0.6

Findings

- Defaulters are fewer compared to non-defaulters and the data imbalance ratio is 10.55
- Few percentage of customers are having different contact and permanent addresses and they are turning out to be defaulters in cases of high credit
- Customers with cash loans tend to have higher default rates compared to revolving loans
- Customers with own cars are less likely to default may be because of financial stability
- Married individuals default less than single or civil-married counterparts and parental statuses are also impacting default rates

Findings

- Customers having decent job tenurity having good behavior those on maternity leave present higher default rates
- Lower-educated individuals tend to default more, while higher-educated clients show better repayment behavior
- Property owners have lower default rates
- Certain occupations, like low-skilled laborers and drivers, are prone to higher default rates
- Higher incomes correlate with lower defaults, but extremely high incomes again have higher default risk

Conclusion

Considering the findings, banks should focus on customers with moderate/highly skilled occupations, stable income, and tenure. Married individuals owning properties and cars are preferred, while caution is advised with low-skilled and less educated customers. Attention is needed for accuracy in documents and matching contact regions.

Thank You