Credits EDA Case Study

Minimise risk of losing money while lending to customers

The problem

Company

A consumer finance company specializing in lending various types of loans to urban customers.

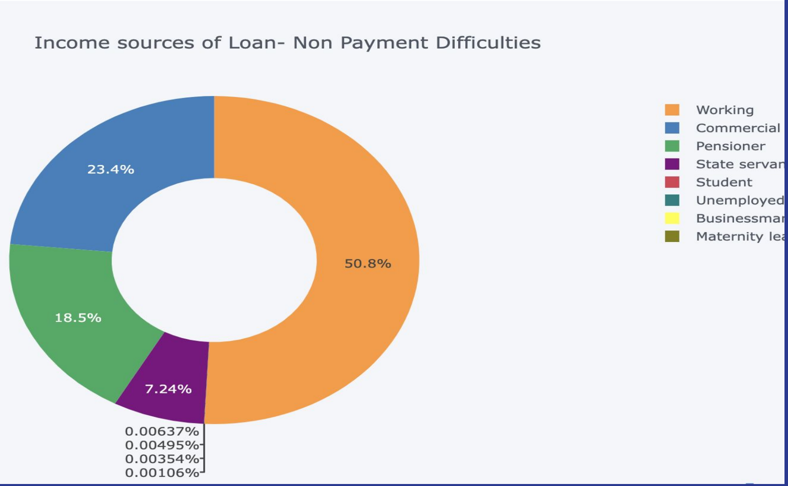
Problem statement

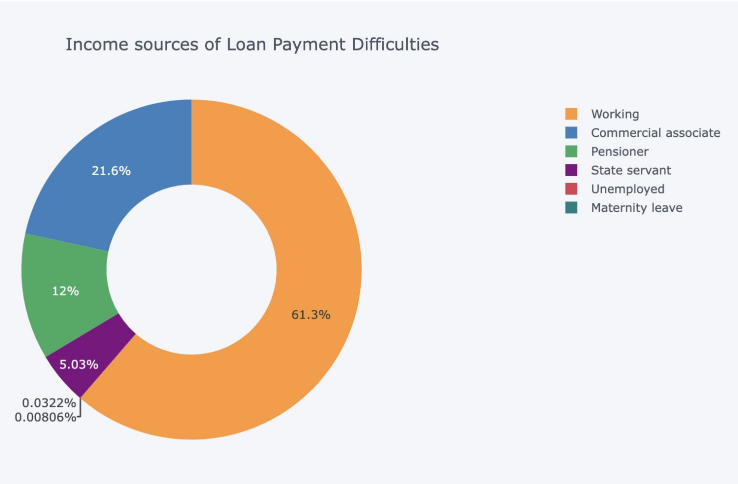
The company wants to understand the driving factors behind loan default. The company can use this knowledge for portfolio and risk assessment.

Implementation

- Read CSV and analyse data
- Clean data - Negative values for days variables, no values as XNA
- Bin the data - Bin age groups and income range
- Check for data imbalance
- Univariate and Bivariate Analysis
- Read, analyse and clean previous application data
- Merge application data
- Univariate and Bivariate Analysis of merged data
- Recommendation and Risks

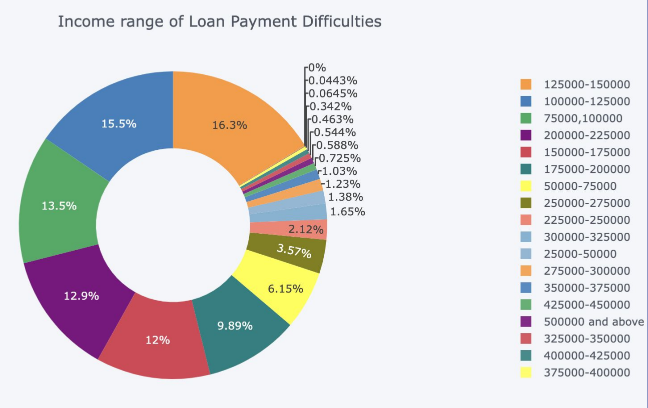
Univariate Analysis - Income Source

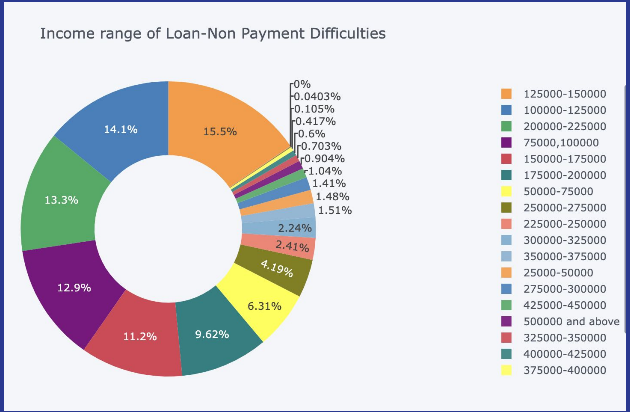




Inference-
State servant & pensioner likely to repay the loan of all income sources

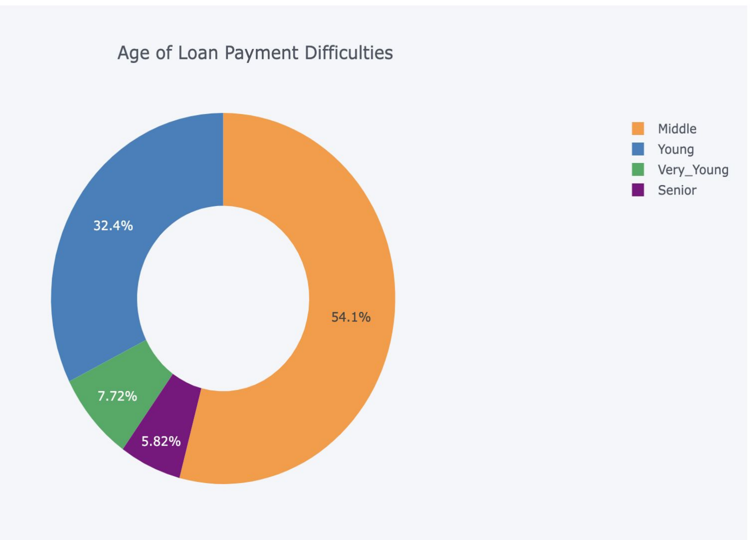
Univariate Analysis - Income range

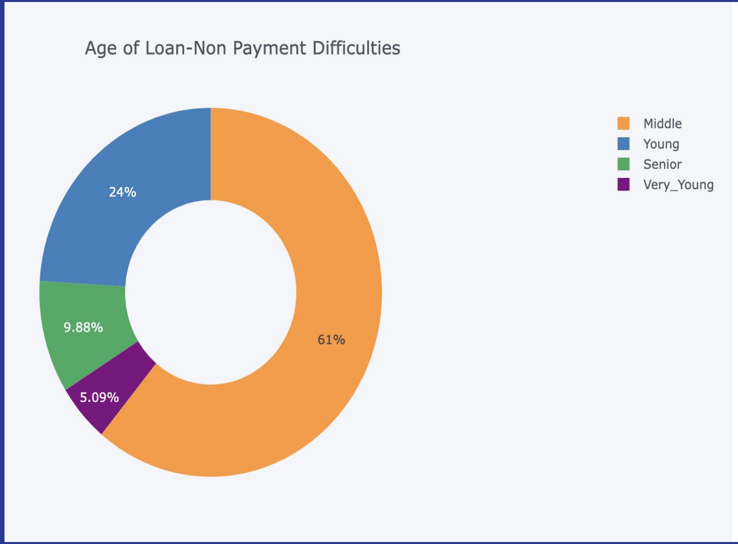




Inference-
Clients with low income (<200000) likely to default on loan

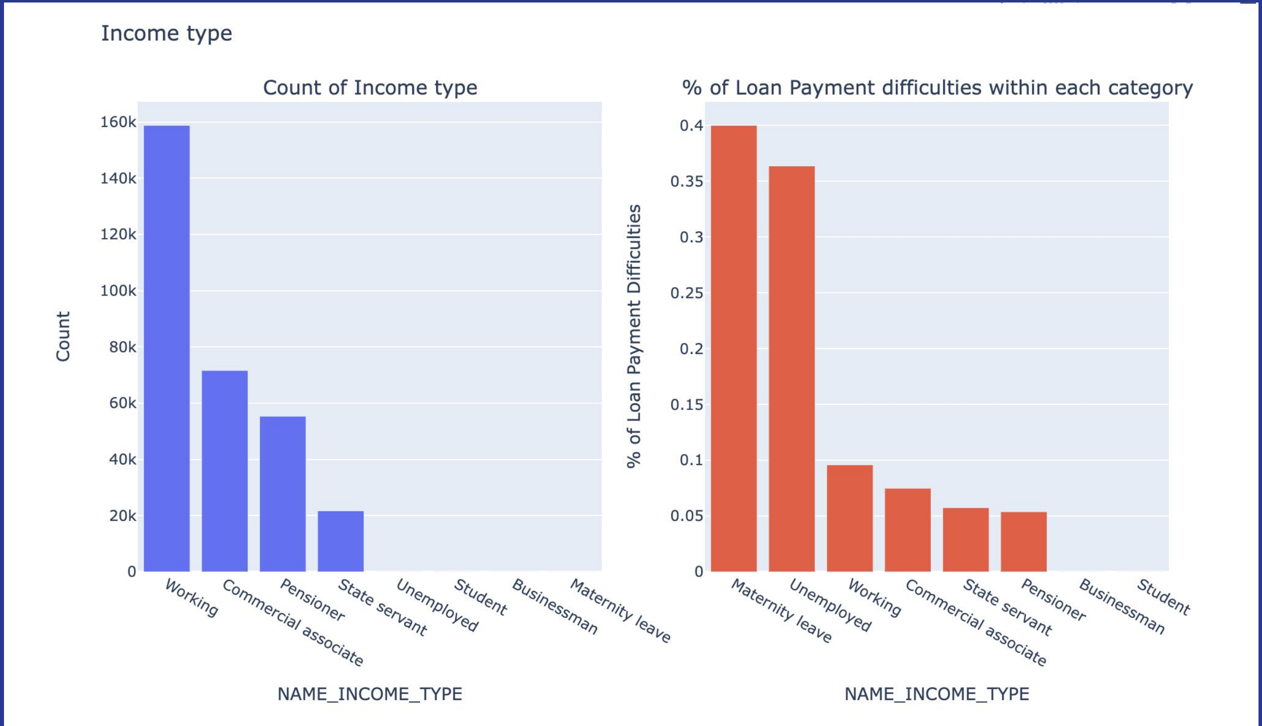
Univariate Analysis - Age





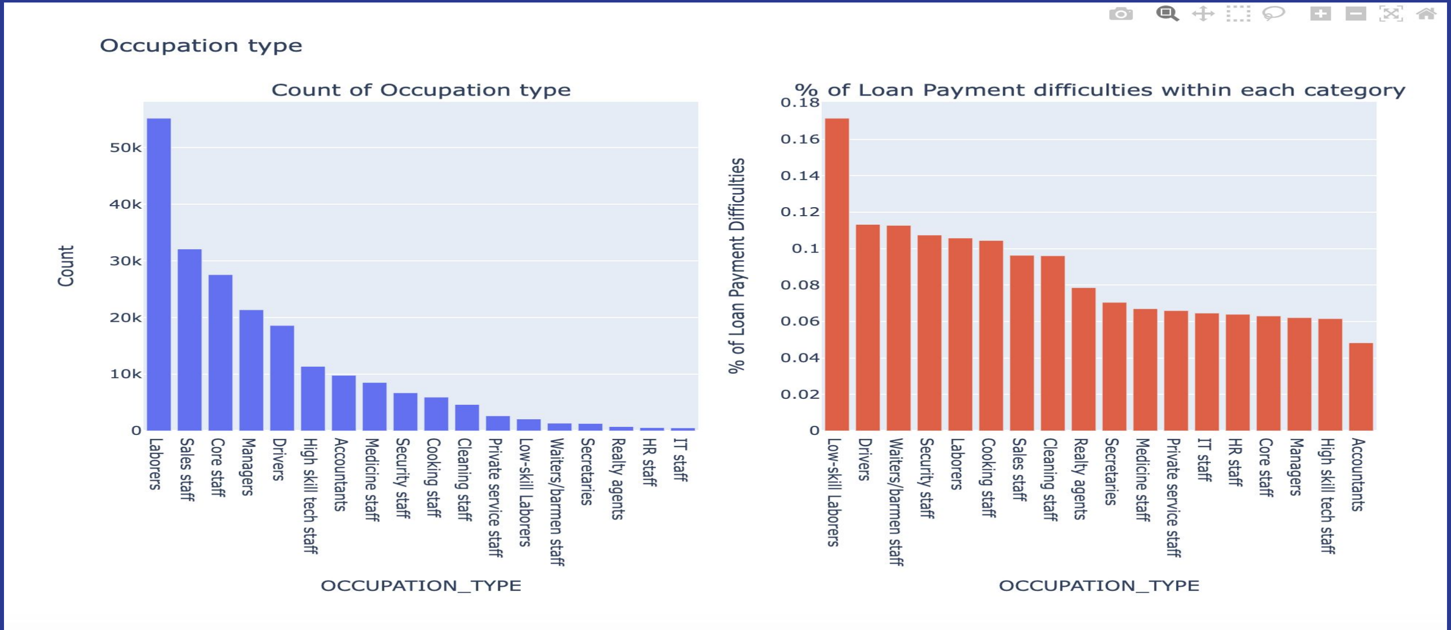
Inference-
Seniors less likely to default on loan

Bivariate Analysis



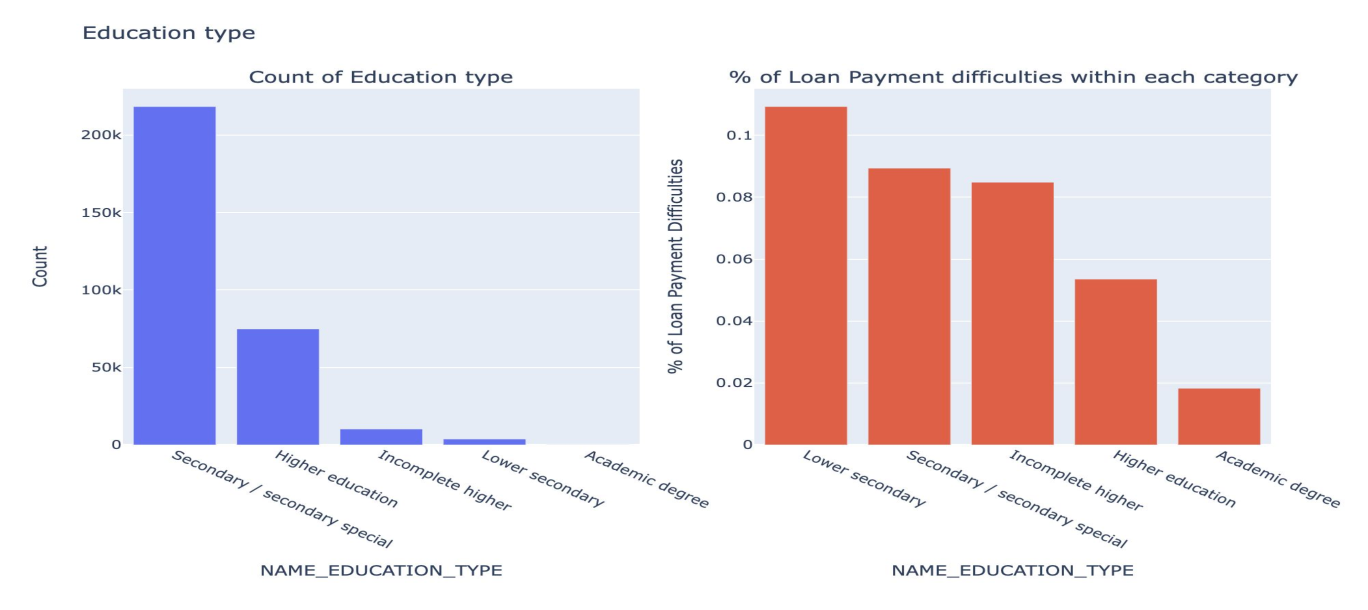
About 40% of clients with income type maternity leave are defaulter. This should be driving factor for loan defaulters

Bivariate Analysis



Low Skilled Labourers have maximum percent of being defaulters. This should be driving factor for loan defaulters

Bivariate Analysis

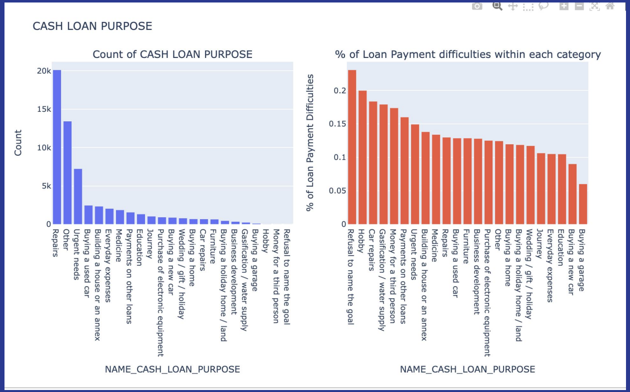


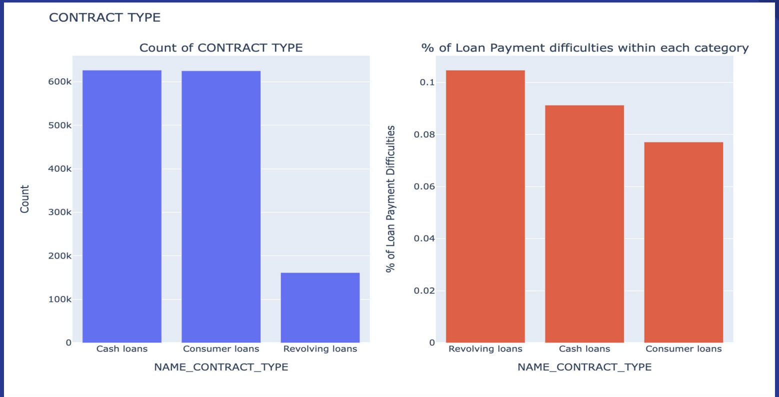
Lower Secondary have maximum percent of being defaulters. This should be driving factor for loan defaulters

Top 10 Correlation for client with difficulties

	VAR1	VAR2	CORRELATION	CORR_ABS
56	AMT_CREDIT	AMT_GOODS_PRICE	0.982783	0.982783
16	AMT_ANNUITY	AMT_GOODS_PRICE	0.752295	0.752295
58	AMT_CREDIT	AMT_ANNUITY	0.752195	0.752195
35	DAYS_BIRTH	DAYS_EMPLOYED	0.582441	0.582441
44	DAYS_REGISTRATION	DAYS_BIRTH	0.289116	0.289116
52	DAYS_ID_PUBLISH	DAYS_BIRTH	0.252256	0.252256
51	DAYS_ID_PUBLISH	DAYS_EMPLOYED	0.229090	0.229090
43	DAYS_REGISTRATION	DAYS_EMPLOYED	0.192455	0.192455
32	DAYS_BIRTH	AMT_GOODS_PRICE	0.135532	0.135532
60	AMT_CREDIT	DAYS_BIRTH	0.135070	0.135070

Conclusion after merging current and previous applications





The count of ‘Refusal to name the goal’ in ‘NAME_CASH_LOAN_PURPOSE’ is comparatively very less and it also has maximum % of payment difficulties- around 23%. Clients who have ‘Refused to name the goal’ for cash loan in previous application are the driving factors for Loan Defaulters.

The count of ‘Revolving Loans’ in ‘NAME_CONTRACT_TYPE’ is comparatively very less and it also has maximum % of payment difficulties- around 10%. Hence, client with contract type as ‘Revolving loans’ in previous application are the driving factors for Loan Defaulters.

The count of ‘Refused’ in ‘NAME_CONTRACT_STATUS’ is comparatively less and it also has maximum % of payment difficulties- around 12%. Hence, client with contract status as ‘Refused’ in previous application are the driving factors for Loan Defaulters.

Clients with 'Revolving loans' and with 'Refused' previous application tend to have more % of payment difficulties in current application. Since the count of both 'Revolving loans' and ‘Refused’ is comparatively less(from the graphs in previous slide), clients with ‘Revolving Loans’ and ‘Refused’ previous application are driving factors for Loan Defaulters