Lending Club Case Study

LOAN RISK ANALYSIS

Problem Statement

- For a consumer finance company specializing in providing various types of loans to urban customers. When a loan application is received, the company must evaluate the applicant's profile to determine loan approval.
- Given the data dataset includes details of past loan applicants and whether they defaulted.
- The objective is to identify patterns that can predict the likelihood of default, helping in decisions such as rejecting loans, lowering loan amounts, or offering loans to high-risk applicants at higher interest rates.
- 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.

Analysis and Approach

Dataset Overview:

- The dataset contains over 100+ attributes with various data types.
- Initial review revealed unnecessary and completely null columns, which were removed using Excel during early exploration.

Data Loading and Preparation:

- Loaded the dataset into Python for further analysis.
- Renamed columns for consistency and readability.
- Identified and handled null values in essential columns to ensure data quality.
- Converted columns into appropriate data types for accurate analysis.

- **Data Cleaning**:
- Identified and removed duplicate records, if any.
- Standardized inconsistent formatting across categorical and numerical fields.
- Imputed missing values using statistical methods or domain-specific logic.

Outlier Detection and Handling:

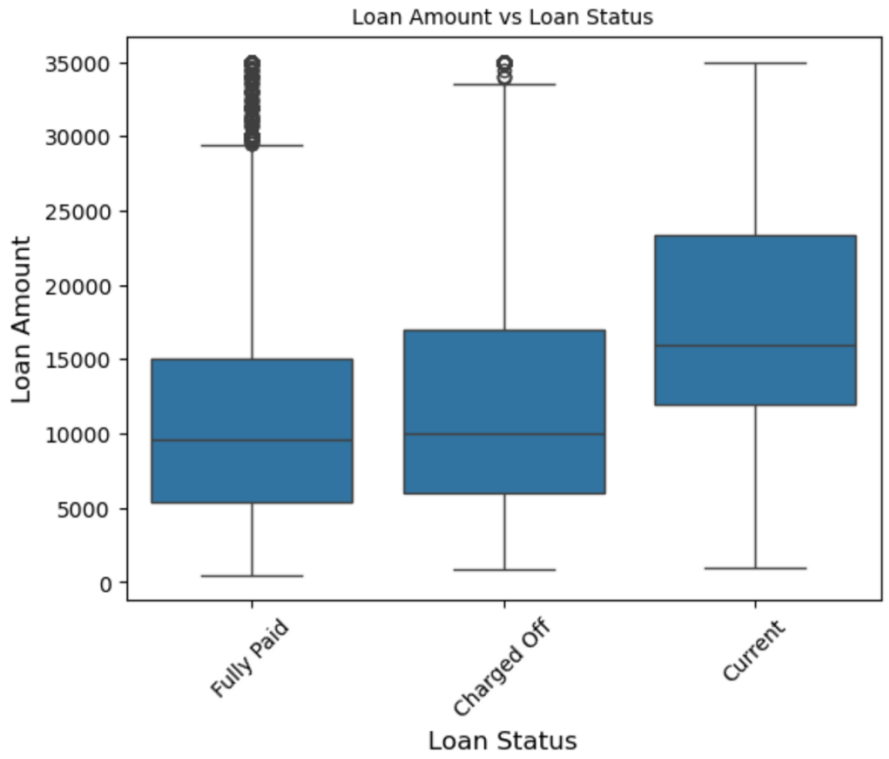
- Detected outliers using statistical techniques (e.g., IQR, Z-scores).
- Visualized outliers to understand their impact on analysis.
- Decided on retaining or removing outliers

Visualization and Insights:

- Generated visualizations to understand data distributions and relationships between different attribute
- Identified key trends and anomalies in the data.
- Correlation heat map has been made

Visualizations and Results

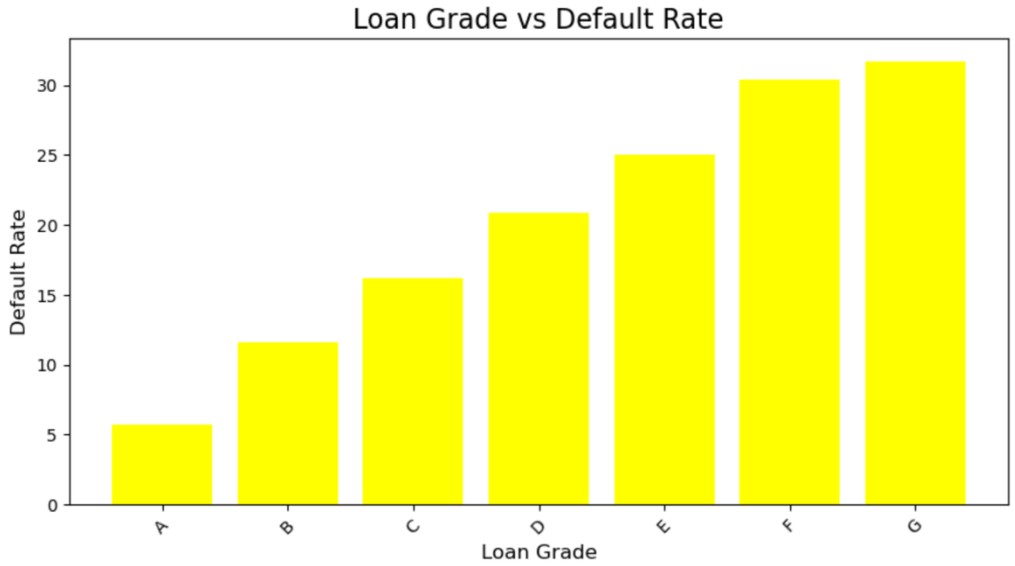
- The graph above shows the overall distribution of loan statuses based on loan amounts. Fully paid loans are evenly distributed, with most being on the lower side. The data indicates that moderate loan amounts are more likely to be fully repaid. However, defaulted loans also fall within a similar range but have a slightly lower median. The data distribution suggests that even smaller loan amounts do not ensure repayment, as some borrowers still default.



Visualizations and Results

Loan Grade vs Default Rate

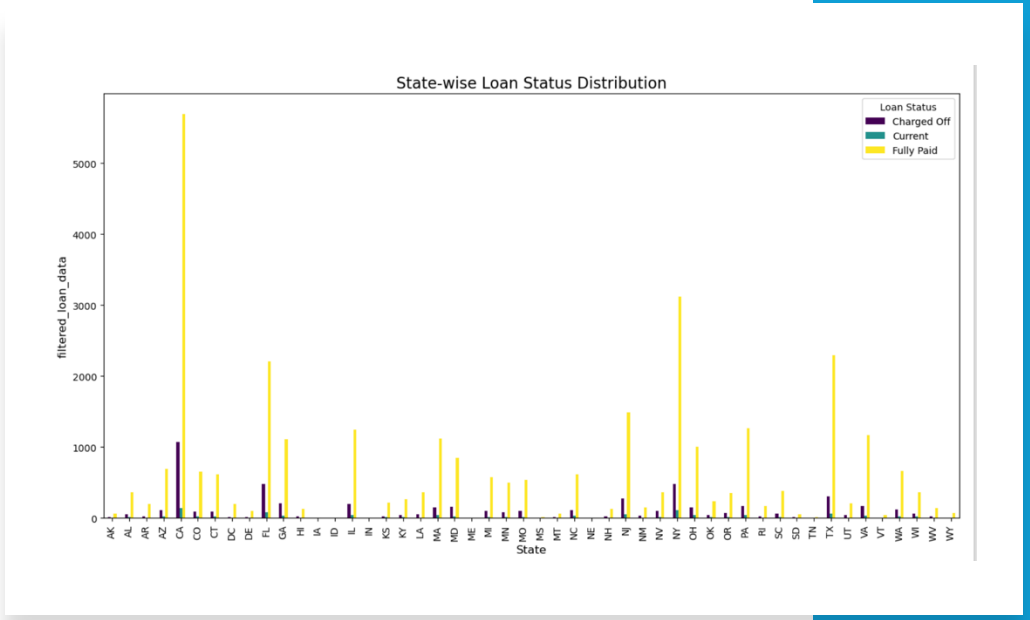
- The graph shows the relationship between loan grades and loan statuses. It reveals that lower-grade loans are associated with a significantly higher risk of default, highlighting the impact of loan grade on repayment behaviour. Where G indicating Higher risk and A low.

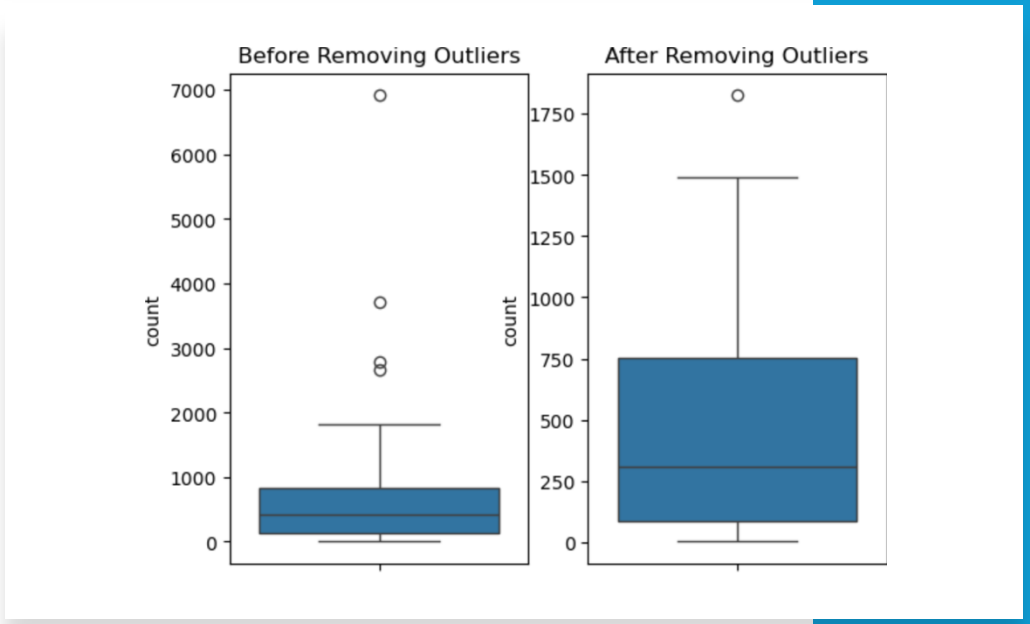


Visualizations and Results

Sate wide Analysis of how the borrowers loan status has been

- The plots highlight the state-wise loan distribution before and after removing outliers. Initially, states such as California (CA), Texas (TX), New York (NY), and Florida (FL) show significantly higher loan distributions compared to others, with extreme outliers skewing the data. These outliers represent unusually high loan amounts concentrated in a few states.
- After handling the outliers, the distribution becomes more uniform, as depicted in the adjusted box plot.

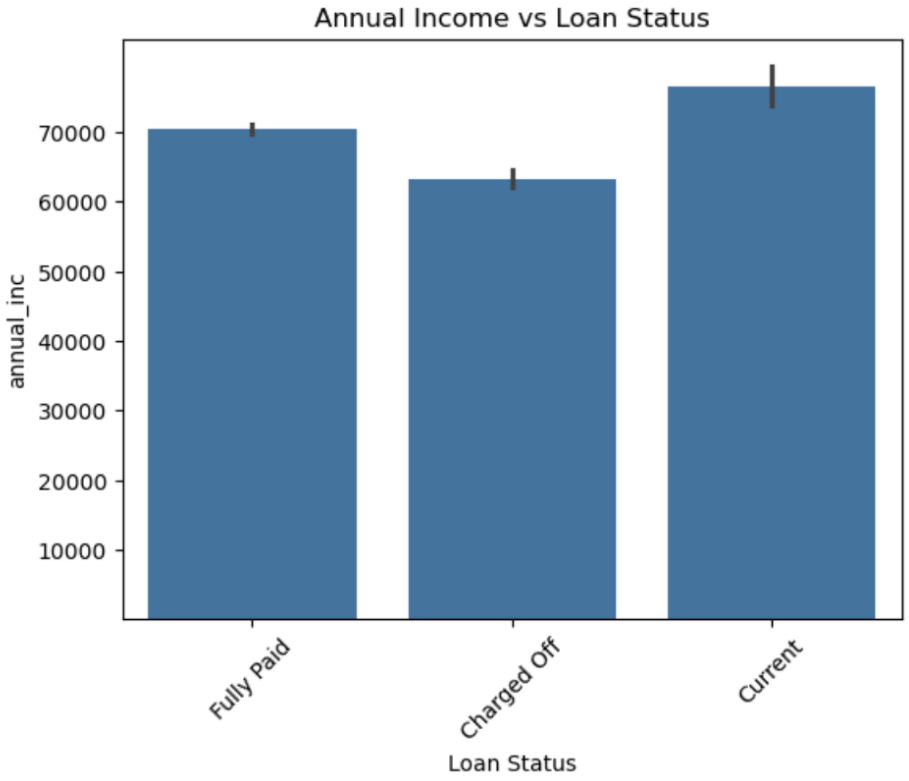




Visualizations and Results

Annual Income vs Loan Status

- The graph above shows clear difference between borrowers who fully repaid their loans and those who defaulted. Borrowers with higher annual incomes are significantly more likely to repay their loans compared to those with lower annual incomes. This suggests that annual income is a good indicator of a borrower's ability to meet repayment obligations.



Visualizations and Results

- The mean interest rate for Charged Off loans is 13.88%, which is higher than the Fully Paid loans 11.64% coming to the quartile percentages The 75th percentile 16.45% for Charged Off loans is higher than Fully Paid loans 13.99%, which suggests that riskier loans carry higher interest rates. Higher interest rates are associated with a higher likelihood of default. The minimum interest rate for all loan statuses is around 5.42%, but the maximum for Charged Off loans reaches 24.40%, indicating riskier loans get higher rates.

Interest Rate vs Loan Status

loan_status	count	mean	std	min	25%	50%	75%	max
Charged Off	5399.0	13.881943	3.639045	5.42	11.46	13.61	16.45	24.40
Current	1098.0	15.099326	3.577011	6.00	11.99	14.65	17.99	24.59
Fully Paid	32145.0	11.641071	3.591229	5.42	8.59	11.49	13.99	24.11

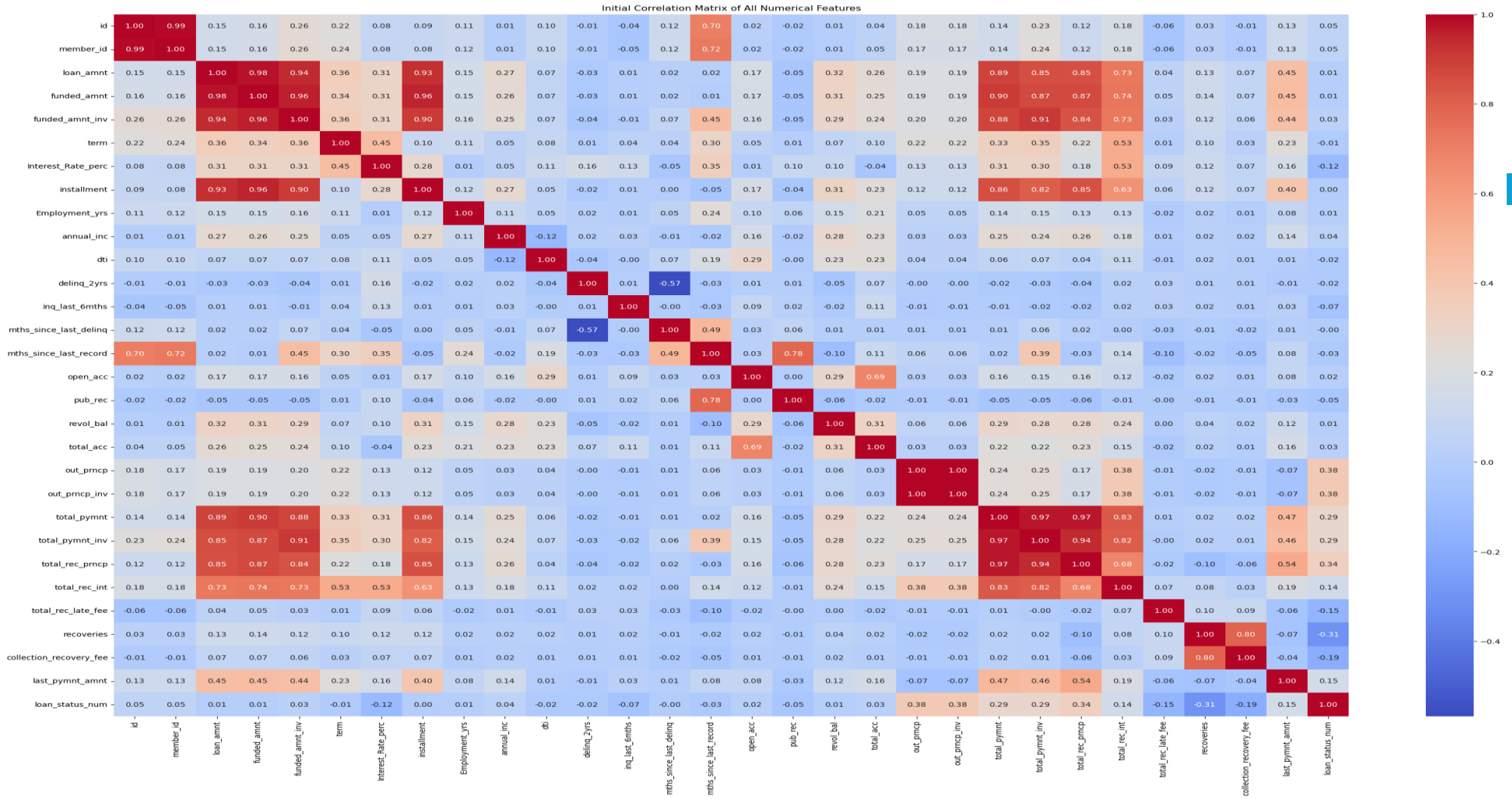
Visualizations and Results

Annual Income vs Loan Status

loan_status	count	mean	std	min	25%	50%	75%	max
Charged Off	5399.0	63331.098537	48103.561033	4080.0	38244.5	54000.0	75000.0	1250000.0
Current	1098.0	76487.814208	47048.648026	10020.0	48000.0	65000.0	90000.0	650000.0
Fully Paid	32145.0	70427.580899	67006.653953	4000.0	42000.0	60000.0	84996.0	6000000.0

- The mean annual income for Fully Paid borrowers is 70,427, but Charged Off borrowers still have a significant mean income of 63,331. This suggests that higher income does not always lead to successful loan repayment The minimum annual income across all categories is relatively low 4,000 - 10020. However, maximum income is significantly high, especially for Fully Paid loans 6,000,000, suggesting that a few high-income individuals fully repay their loans. coming to the Income distribution differences of quartile ranges: Charged Off Loans: 25% of defaulters earn less than 38,244. 75% of them earn up to 75,000, meaning most defaulters are in the low to middle-income range. Fully Paid Loans: The median income is 60,000, which is higher than Charged Off loans. 75% of fully paid borrowers earn up to 84,996, which is higher than defaulters.
- Higher annual income borrowers tend to have fully paid or current loans. Charged Off loans are more common among lower-income borrowers, but income alone does not determine loan repayment behavior. income along with certain other attributes together having effected to default.

Visualizations and Results



Heat map overall loan data attribute relationship

Visualizations and Results

- The heat map is the overview of the numerical variable relationships in the dataset. Showing the correlation values for each if we see the summary from the map the loan amount is positively correlated with the instalments and total payment. Expectation with larger loans have higher monthly payments and correlated with interest rates moderately. and the interest rate is positively correlated with DTI ratio showing the borrowers with hight debt to get higher interest rates. and negatively correlated with loan grade where the interest rates higher for the low grades (E,F,G). and DTI negatively correlated with income where higher income borrowers are having low DTI ratios.

Summary and Outcomes

Critical Attributes Driving Loan Defaults:

Interest Rate: The Higher interest rates are linked with the higher defaulters rate

Loan Amount: Loan amount is also taken a place impacting the risk of defaults

State of Residence: Some states show higher default rates, so here it concludes loans some specific region is also taking the part to impact the loan payment on time / correctly

Annual Income: A moderate predictor of repayment capability.

Debt-to-Income Ratio: A the borrowers tend to be defaulter with higher DTI ratio

Loan Grade - lower-grade loans are associated with a significantly higher risk of default, highlighting the impact of loan grade on repayment behavior.

Measures for good portfolio’s

- Limit loan amounts for borrowers with lower credit scores or high debt-to-income ratios.
- Ensure borrowers' debt obligations do not exceed a safe percentage of their income. Monitor ratios during the loan term and offer assistance if financial strain increases.
- Offer low interest rates to low-risk borrowers to encourage repayment. Avoid setting excessively high rates for high-risk borrowers, as this increases their chances of default.
- Tailor lending policies based on geographic insights, particularly in states with higher default rates. Diversify lending across multiple regions to avoid over-concentration of high-risk loans.
- Approve loans based on a borrower’s annual income to ensure they can manage the repayments comfortably.
- Consider offering smaller loans or extended repayment periods for lower-income applicants.