LENDING CLUB CASE STUDY

Group Members:
1. **Susrita Das**
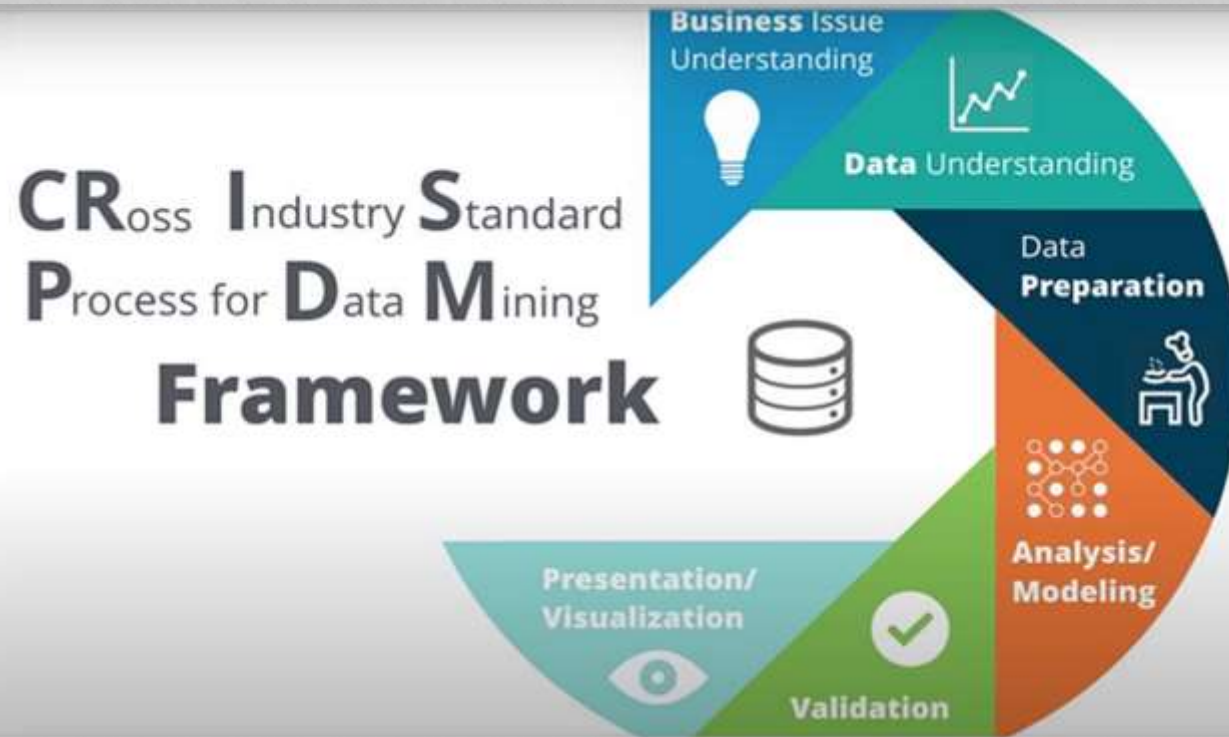2. **Digvijay Kadam**

Case Study Objectives

Identification of Patterns among loan applicants' data who tend to 'default' paying back

Understand the Factors and concerning points behind Loan Default phenomena

Suggesting better parameters that got overlooked to reduce number of defaulters to maximum extent

Problem solving methodology using CRISP-DM

Data Cleaning Steps

1. **Column Deletion:** Eliminate columns that are not required for the analysis.

2. **Outlier Removal:** Exclude extreme values, both high and low, that could unduly impact the analysis results.

3. **Handling Missing Data:** Address missing values using suitable methods or techniques.

4. **Duplicate Data Handling:** Get rid of identical rows and eliminate rows where certain columns have identical values.

5. **Row Filtering:** Apply filters based on specific segments or date ranges to retain only the rows pertinent to the analysis.
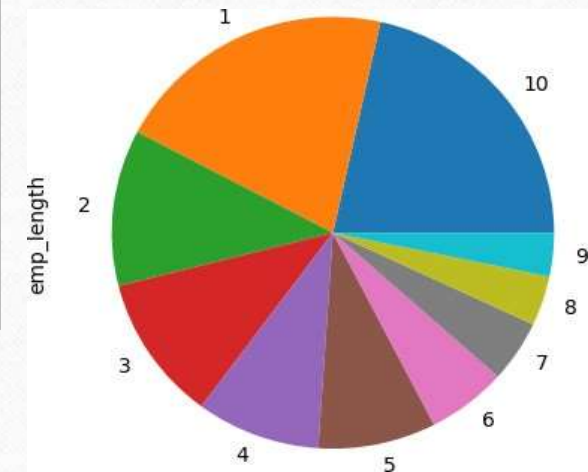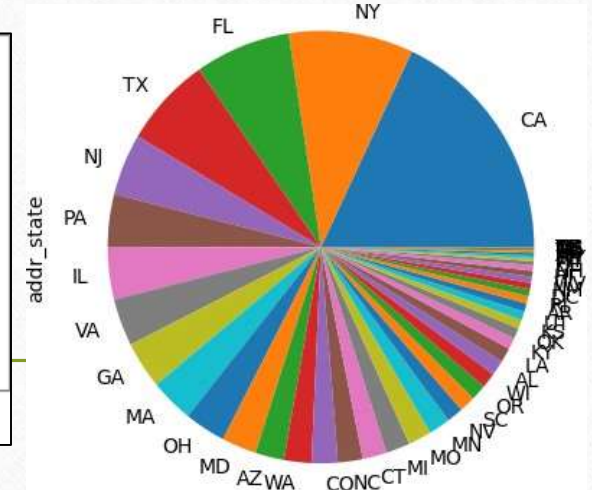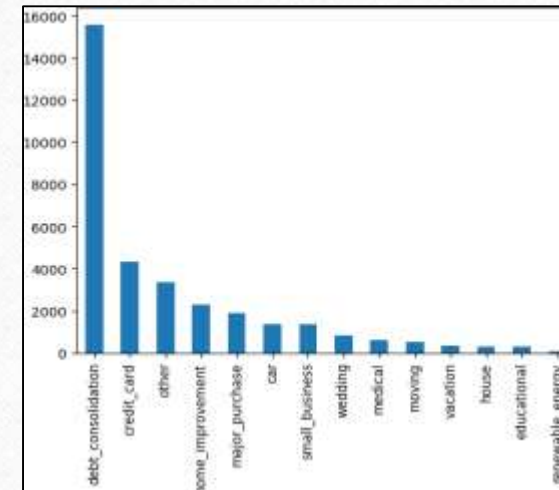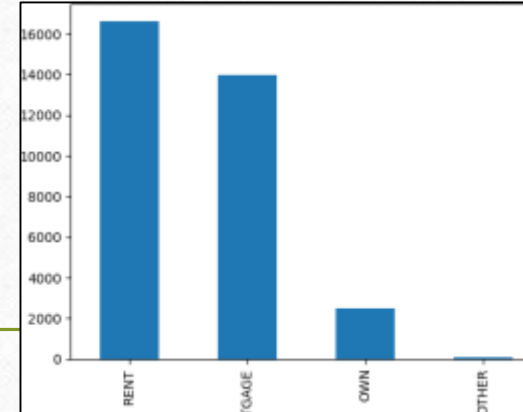
Analysis

- The project's core objective is to assess the impact of consumer and loan attributes on default tendencies.
- We conducted data cleaning, including NA value imputation, for the Loan dataset.
- We introduced two new columns: "Profit and Loss" and "Ratio of funded amount to annual income."
- Univariate analysis involved using line charts and bar charts to examine variable distributions and box plots to identify outliers.
- Multivariate analysis was carried out to explore the relationships and interactions between various variables.

Univariant Analysis
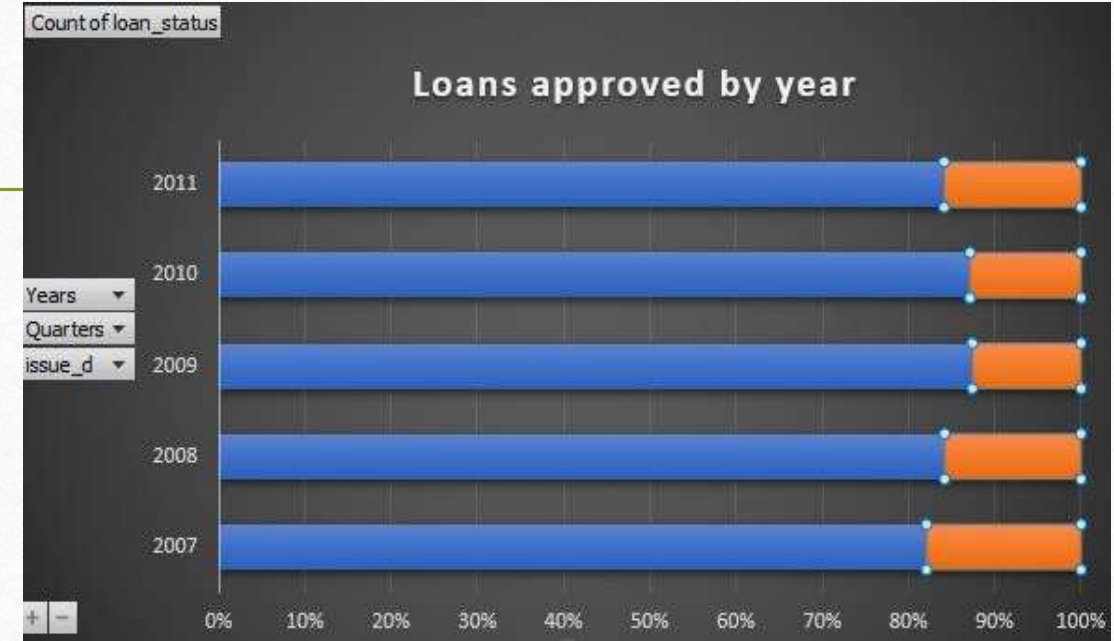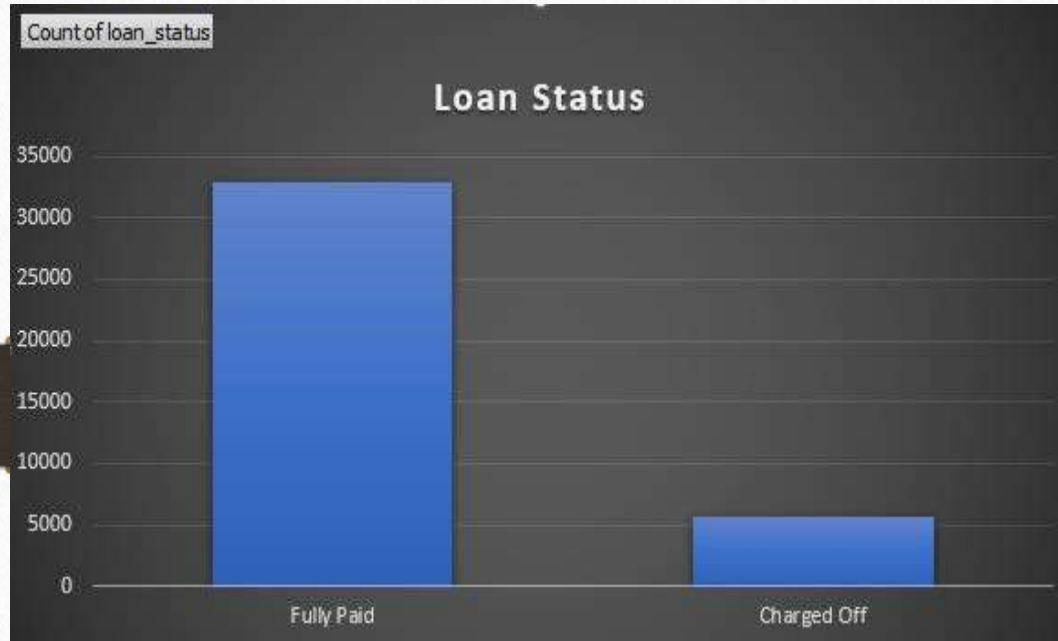
Unordered & Ordered Categorical Variable Analysis

Observations:

- Majority of loan applicants are either living on Rent or on Mortgage.
- Most of the loan applicants are for debt_consolidations.
- Most of the Loan applicants are from CA(State).
- Most of the applications are having 10+ yrs. of Exp.

Loan Status Analysis

Observations:
1. Most of the loans are Fully Paid.
2. About 14% of loan are having status as defaulters.
3. The loan has been increasing exponentially over the years with sudden increase in defaulters in year 2011

Bivariate Analysis

Annual income vs Charged Off

Observations:

- Income range 80000+ has less chances of charged off.
- Income range 0-20000 has high chances of charged off.
- Notice that with increase in annual income charged off proportion got decreased.

Annual Income vs Chargedoff Proportion
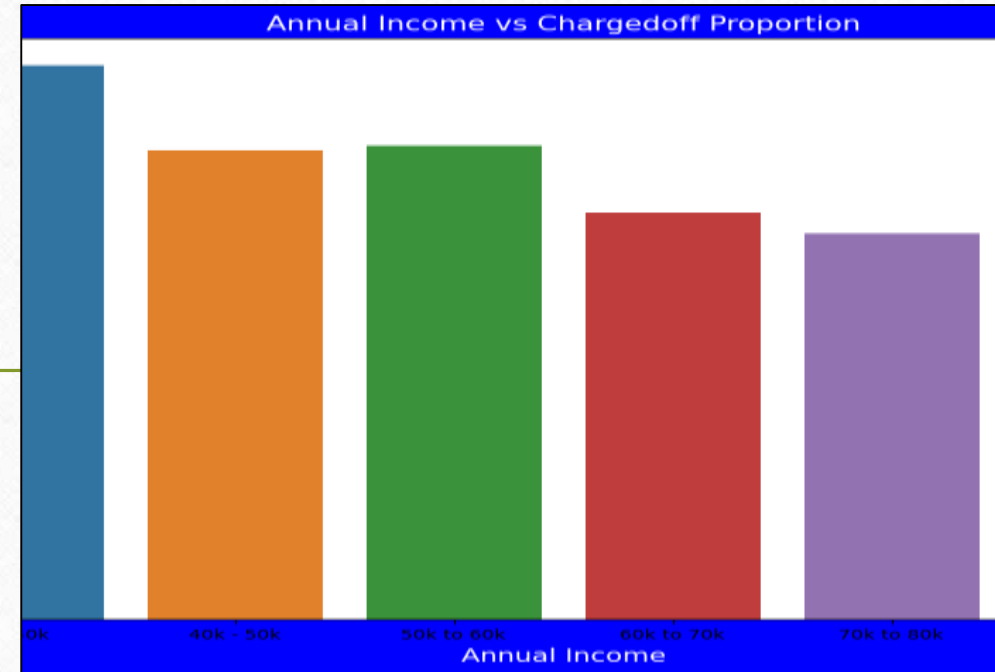
| loan_status	int_rate_b	Charged Off	Fully Paid	Total	Chargedoff_Proportion
3	Very High	1670	4751	6421	0.260084
0	High	985	4851	5836	0.168780
2	Moderate	961	5638	6599	0.145628
1	Low	579	4983	5562	0.104099
4	Very Low	519	8254	8773	0.059159

Loan Grade Analysis
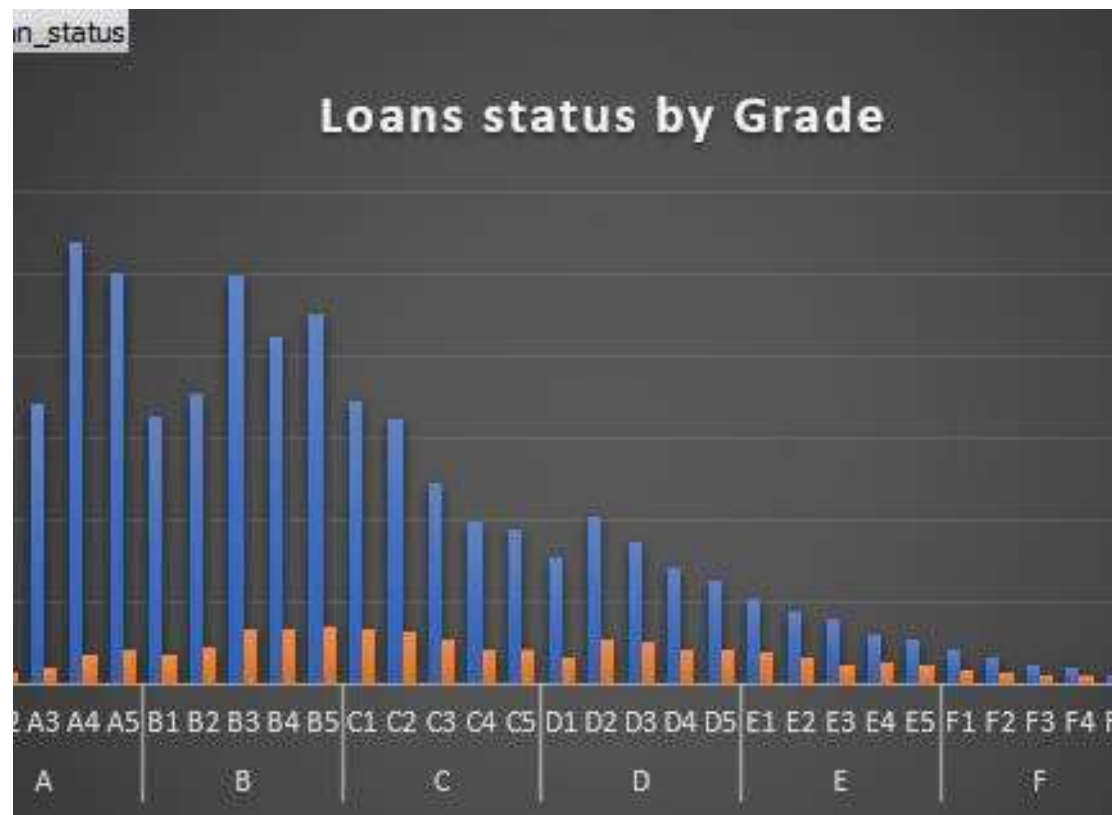
Observation:

- Most loans fall into the A and B grade categories, suggesting that a significant portion of loans are of high quality.

- Loans with higher quality grades generally come with lower interest rates.

- Consequently, Faulty loans shows a peak in B & C grade.

Loan Applicants work ExperienceAnalysis

Loans status by experience
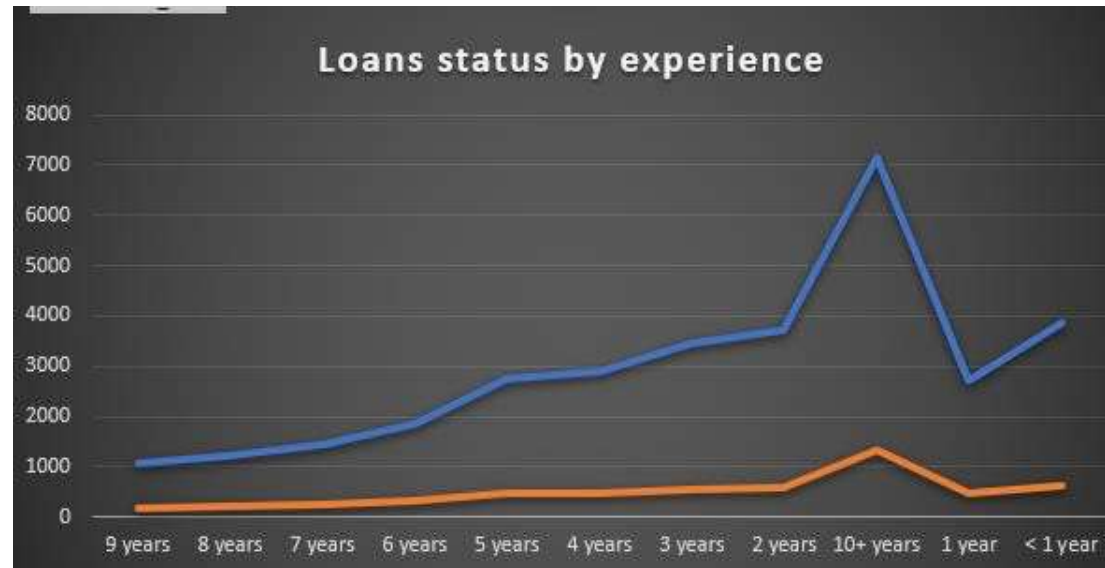
Observation:

- Majority of employees applying for the loan have more than 10 years of experience

- Tendency of person to default the loan with 10 years of experience is also high. Also, in case of percentage less than 1 year has a sudden spike in both categories.

Loan Amount Analysis

Loans status by experience

Observation:

- Loan Amount peaks towards center amount with 10K being most common.

- Among top amounts, not verified has the maximum counts of loans.

Top 3 amount Loan status with verification s
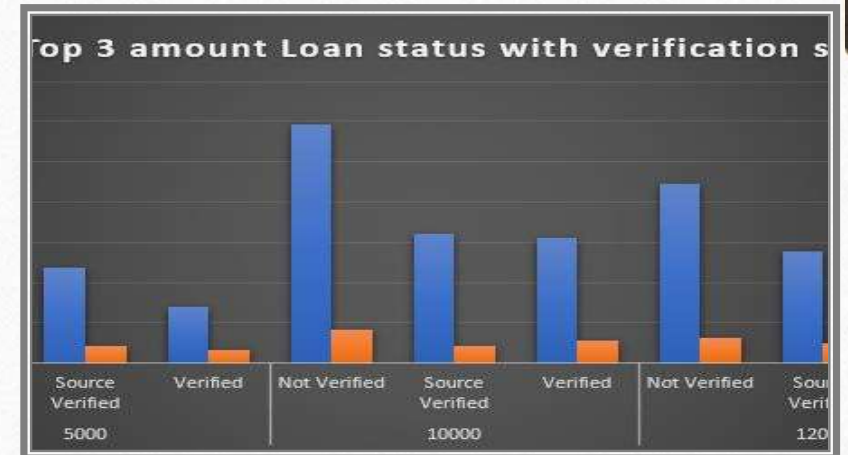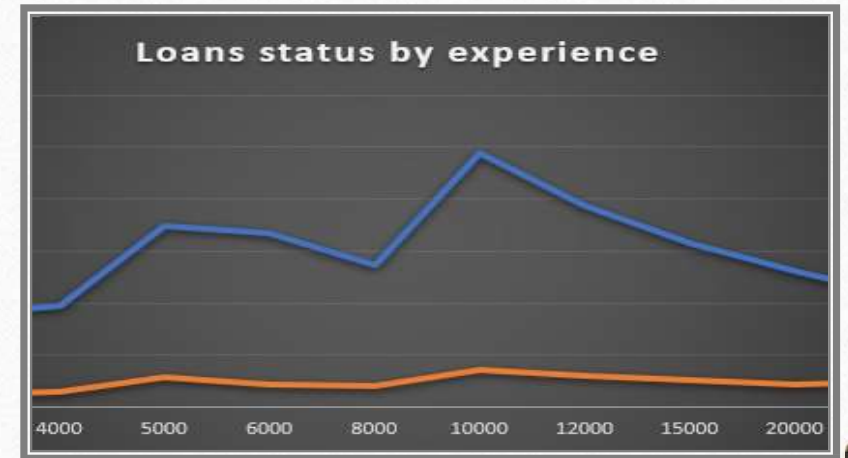
Income Analysis for the loan applicants

Observation:

- Most of the applicants have income between 4 to 7 lakhs.

- There are two applicants having income above 30 lakhs

- Maximum defaulters can be seen in 5-6 lakh category with sudden peak on 10k loan amount.

Top 10 income with loan status

3 income by Loan amount with loan status
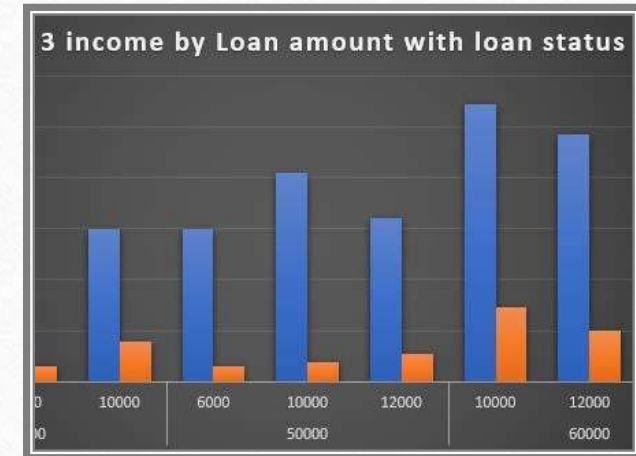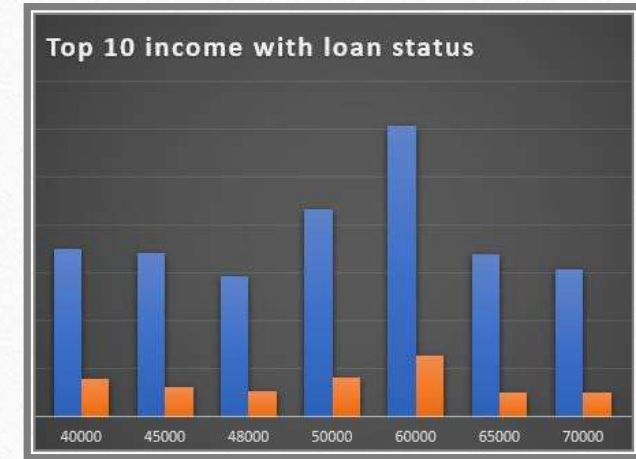
Loan Defaulters by State Analysis
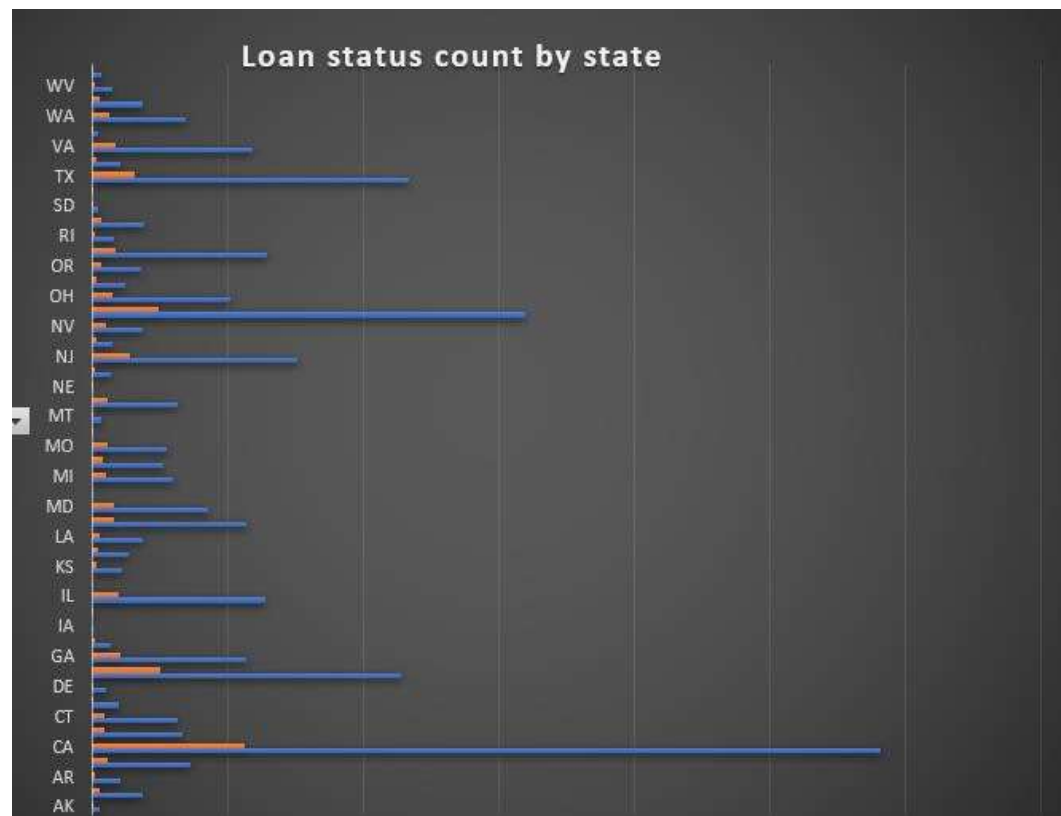
Observation:

- Highest number of loans are from CA
- Highest defaulters being in NV by 23% & lacking numbers but with 60% defaulters in NE.

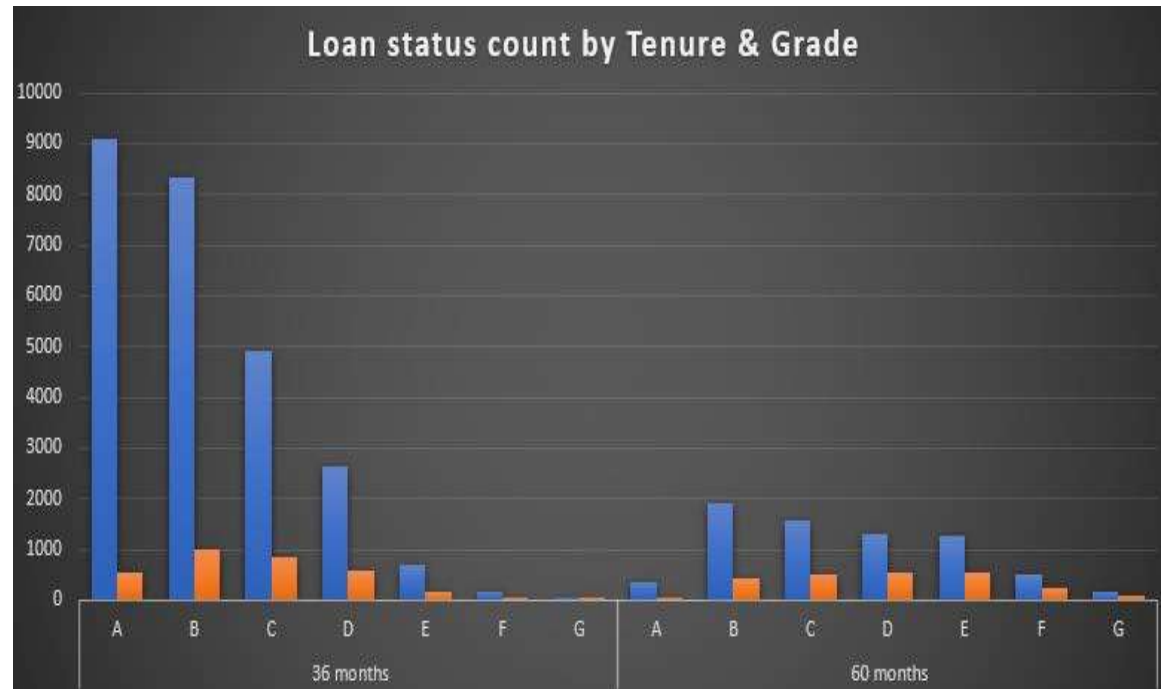Loan status count by state

Purpose for applying Loan Analysis

Observation:

- Most of the loans are granted were for debt consolidation.

- Applicants having higher income tends to default the loan.

- Applicants with lower grade tends to show higher percentage of defaulters in case of top 3 purposes.
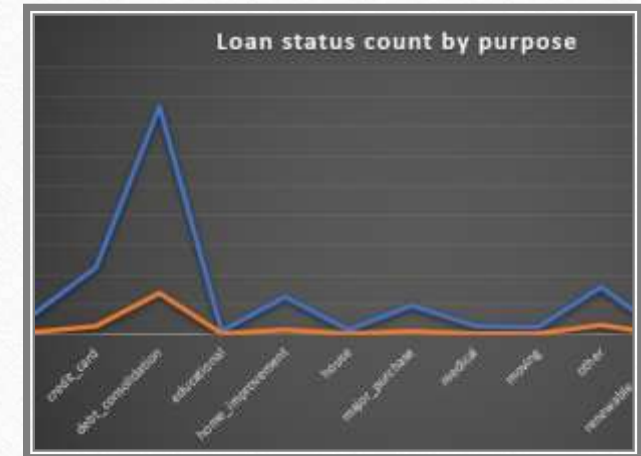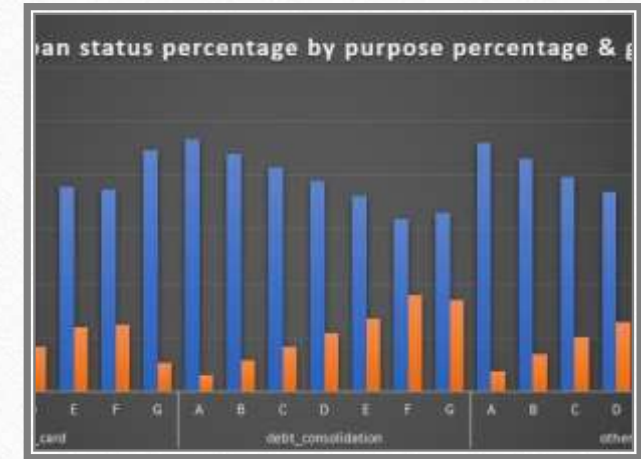
Loan Status vs Loan Tenure


Loan status count by Tenure & Grade

Observation:

- Most of the loans granted were of 36 months.

- Loans granted for 36 months have slightly high tendency to default in mid grades while max count comes out of Top grades.

- Debt Consolidation is the most common purpose and loans are granted for 36 months and 60 months.

Loan Status by Home Ownership

Observation:

• Applicants having there home rented or on mortgage have almost equal tendency to default the loan.

• Although when sub divided into verification status, Percentage of defaulters is shown higher in verified. Indicating towards flaw in verification system.
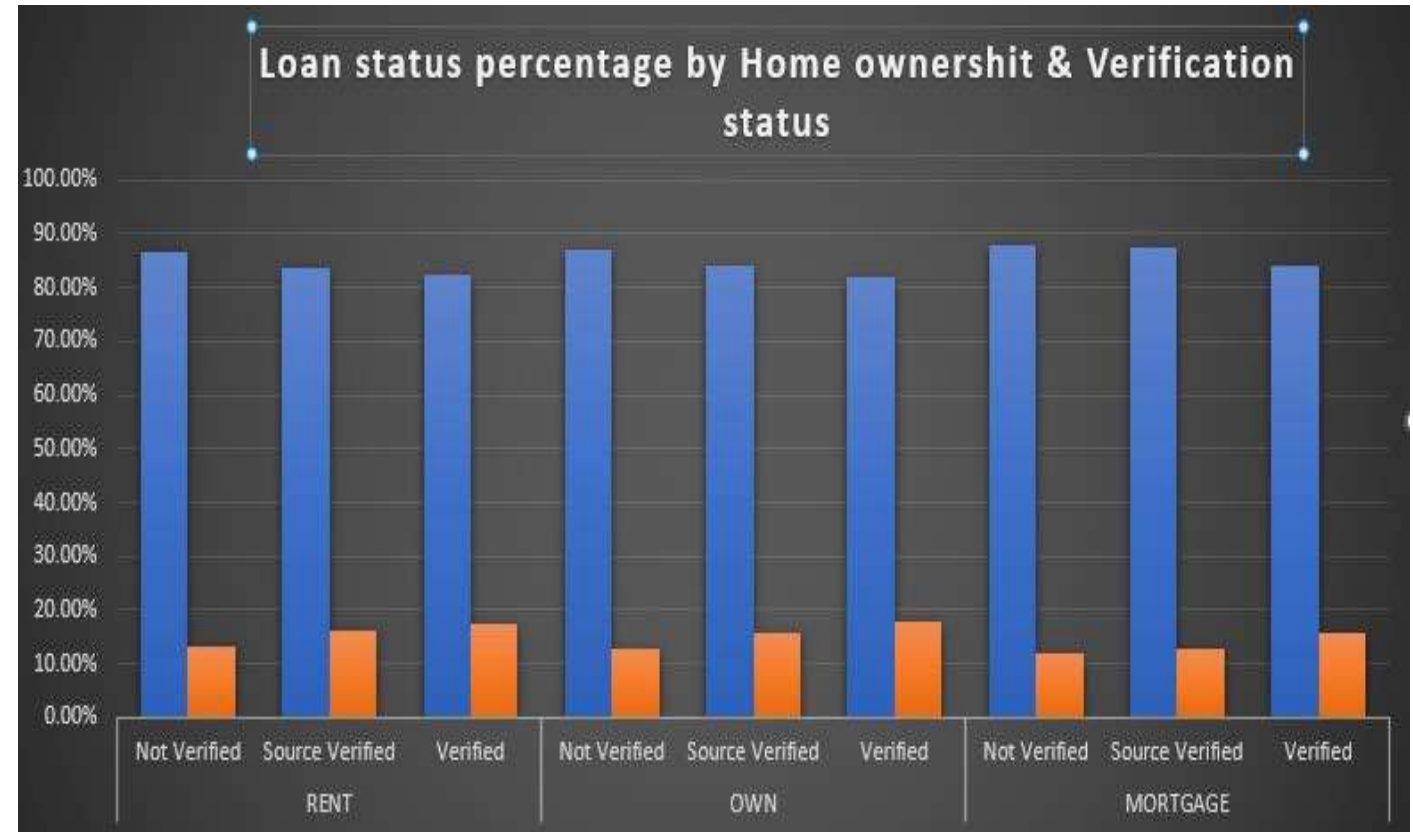
Conclusion

Results

1. Low-grade loans exhibit a greater propensity for default, indicating the effectiveness of the grading system.

2. Loans with elevated interest rates are associated with an increased likelihood of default. It is crucial to conduct a comprehensive background check on applicants with high-interest loans.

3. Applicants hailing from California (CA) should undergo additional scrutiny due to their elevated default risk.

4. For loans intended for debt consolidation, it is imperative to exercise thorough applicant assessment, as this purpose is linked to a heightened default tendency.

5. Small business loans are defaulted more. So, issuing loans to them should be stop/reduce.