**Bank Loan Data Analysis Report**

**Introduction**  
 The analysis covers loan applications, funded amounts, received payments, interest rates, and borrower demographics, categorized by loan status, grade, purpose, company type, gender, race, and credit score. It provides insights to support management in refining lending policies, improving risk evaluation, and enhancing customer segmentation for better financial performance.

**Data Preparation/Cleaning**

The following was done to prepare and clean the data before it was ready for analysis.

1. Standardizing the date format to MDY in ‘last\_credit\_pull\_date’, ‘last\_payment\_date’, ‘next\_payment\_date’ columns
2. Adding a Column for ‘grade’. This was extracted from the ‘sub\_grade ‘field
3. Adding a ‘address\_state\_abb’, which shows abbreviations for all the states
4. Adding a ‘Good vs Bad Loan’ column

**Key Insights**

The key insights gathered from the analysis of the bank loan data are listed below.

Loan Application Trends

* The highest loan applications occurred in November (9,903) and October (5,477).
* January to May had very few applications, suggesting seasonal lending trends.

Loan Performance (Good vs. Bad Loans)

* Good Loan % = 50.14%, and Bad Loan % = 49.86%, meaning almost half of loans are at risk.
* Total Funded Amount is nearly equal for Good ($264.85M) and Bad ($264.78M) loans, indicating a 50-50 lending risk ratio.

Interest Rate & Loan Status

* Defaulted loans have the highest average interest rate (12.13%), followed by Charged Off (12.04%).
* Fully Paid loans have a slightly lower interest rate (11.98%), indicating riskier borrowers are charged more interest.

Funded vs. Received Amount

* Loan Funding: Defaulted & Charged Off loans have nearly the same funded amount as Good Loans, signaling risk in lending policies.
* Total Payment Received: Defaults and Charged-Off loans contribute lower returns, affecting profitability.

Loan Purpose Breakdown

* Top 3 Loan Purposes: Car Financing (2.80K), Vacation (2.78K), Home Improvement (2.78K)

Higher-risk loans (Defaulted & Charged-Off off) might be linked to discretionary spending.

Loan Applications by Employment Length

* Employees with 6-10+ years of experience had more loan applications (~1.7K each).
* <1 year experience had the highest loan applications (~1.77K), which is a Possible risk factor for default.

Correlation Analysis

* Credit Score vs. Good/Bad Loan: Weak correlation (-0.0051), meaning credit score alone doesn’t determine loan performance.
* Annual Income vs. Good/Bad Loan: Very weak correlation (-0.0020), indicating income doesn’t strongly impact default risk.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Good Vs Bad Loan - Numeric* | *credit\_score* | *annual\_income* |
| Good Vs Bad Loan - Numeric | 1 |  |  |
| credit\_score | -0.005157374 | 1 |  |
| annual\_income | -0.001999932 | -0.005795998 | 1 |

State-wise Loan Application Distribution

* The top 5 states with the highest loan applications are:
  + Michigan (763 loans)
  + Pennsylvania (420 loans)
  + Virginia (394 loans)
  + Georgia (417 loans)
  + Washington (398 loans)
* These states could be experiencing higher demand for credit, possibly due to economic conditions or local lending policies.

**Recommendations**

The bank should refine its Loan Approval process by strengthening the approval criteria for browsers with low employment experience(<1) and tightening lending rules for loans with high default rates (Discretionary Spending like Car, Vacation, etc.)

The Bank should Adjust interest rates based on Loan Risk. Higher interest rates should be imposed on categories most likely to default and provide lower interest rate incentives for borrowers with longer employment histories.

The bank should use machine learning models for better prediction of loan performance and defaults. It should also factor in multiple risk indicators like employment length, credit scores, and loan purpose.