Bank Loan Analysis Report

Executive Summary

The Bank Loan Analysis Report aims to provide actionable insights into loan applications, approvals, and related metrics.

Data Source

- The loan data is sourced from a link shared by a you tube account, for the purpose of showing analysis skills utilizing Postgres.

Data Fields and Usage

The data consists of 38,576 rows and following columns:

1. **Loan ID**: Unique identifier for loans.
2. **Address State**: Borrower location for regional analysis.
3. **Employment Length**: Indicates employment stability.
4. **Employee Title**: Job title for income source verification.
5. **Grade/Sub Grade**: Creditworthiness and risk classification.
6. **Home Ownership**: Housing status for financial stability assessment.
7. **Issue Date**: Loan origination date.
8. **Loan Status**: Current state of the loan for performance tracking.
9. **Purpose**: Loan reason for segmentation and customization.
10. **Term**: Loan duration.
11. **Verification Status**: Status of financial information verification.
12. **Annual Income**: Yearly earnings for creditworthiness.
13. **DTI**: Debt burden relative to income.
14. **Instalment**: Monthly repayment amount.
15. **Interest Rate**: Cost of borrowing.
16. **Loan Amount**: Principal amount borrowed.

Methodology

I used Postgres SQL to analyze the data. The data cleaning posed little challenge as data was found be mostly normal and clean. Following changes were made before the analysis:

- I changed the data style of key date columns to 'ISO,DMY'.
- I changed the datatype of these date columns to varchar to import the data and after cleaning, they were changed into date type.
- The annual income column was changed to float data type.

Approach

There were a lot of possibilities for the data, but I decided to focus on following aspects of analysis:

1. **Descriptive Statistics and Data Summary:**
    - Calculate the average of KPIs such as annual income, debt to income ratio, installment, interest rate, total amount, and payments etc.
2. **Loan Status Analysis:**
    - Investigate the distribution of loan statuses (loan_status).
    - Compare default rates for different loan statuses.
    - Explore reasons for loan delinquency or default.
3. **Credit Score Analysis:**
    - Group loans by credit grades (grade and sub_grade) and analyze their performance.
    - Calculate average interest rates (int_rate) for each credit grade.
4. **Temporal Analysis:**
    - Analyze the growth, defaults on monthly basis.
5. **Geographical Analysis:**
    - Group loans by address_state and analyze loan characteristics by state.

Key Findings:

Descriptive Analysis

- Average KPIs:

KPI	Average Value
Annual Income	$69,644
Debt to Income Ratio	0.13
Installment	$326
Interest Rate	12.4%
Loan Amount	$11,296
Total Payments	$12,263

- While there were only 2 categories of loan terms, 36 months and 60 months, average is 42 months and median is 36 months.
- CA has the highest number of applicants, 6894 about 17.9% and ME has the lowest number of applicants, only 3.
- Grade B has the higest number of applicants, 11674 which is about 30% of total applicants and Grade G has the lowest, 313 about 0.8%
- Amongst the applicants, those who rent are the highest in number, 18,439 about 48% and those who have mortgage are the second highest, 17,198, about 45%. Home owners are lowest about 0.3%.

- 42.7% (16,464) of the applicants are Not Verified, 32% (12335) are Verified and about 25.3% (9777) are Source Verified.
- Employee with 10+ years of experience has most applications under him, about 23% while employee with less than 1 year of experience has the second highest number of applications, 12 %

Credit Risk Analysis

Grade	Total Defaults	% of total defaults	% within grade
"A"	552	10.400	5.700
"B"	1343	25.200	11.500
"C"	1266	23.700	16.000
"D"	1072	20.100	20.700
"E"	691	13.000	24.800
"F"	311	5.800	30.300
"G"	98	1.800	31.300

- 13.8% loans were charged off.
- In terms of absolute numbers, B had the highest number (about 25%) of defaults but in terms of percentage within each grade, it was grade G with most defaults in category.
- While 49% of overall total bad loans come from purpose ‘debt consolidation’, the highest default ratio belongs to small business, (about 25% of loans defaulting)
- Bad loans are more common in 60 months term loans than 36 months term loans.

Temporal Analysis

Overall growth in application over the months:

Month	New applications	Change from prev. month
"January"	2332	
"February"	2279	-53
"March"	2627	348
"April"	2755	128
"May "	2911	156
"June"	3184	273
"July"	3366	182
"August "	3441	75
"September"	3536	95
"October "	3796	260

"November"	4035	239
"December"	4314	279

DEFAULT RATES over the Months.

Month	Default Rate
"January "	13.300
"February"	11.600
"March"	12.700
"April"	12.800
"May"	15.100
"June"	14.200
"July"	13.500
"August"	13.100
"September"	14.700
"October"	14.400
"November"	13.900
"December"	15.000

- Months of May and December had the highest default rate.

Geographical Analysis

- CA has the highest percentage of applications, about 18% and ME has the lowest.
- In terms of default ratio, NE, NV and AK have the highest default ratios while WY has the lowest.

Conclusion:

As mentioned in the report, there are many factors that may affect the outcome of a loan, from its timing to its location or the purpose of it. The next step is the visual analysis of data to go deeper into regions and timing.