Insights from Loan Data

Contents

Approach & Analysis

I'm taking a hands-on, step-by-step approach to delve into the nuances of loan data, with an emphasis on clarity and actionable insights. Here's my game plan:

Data Preparation and Cleaning:

Data Importation: I started by importing the dataset into a SQL environment ready for in-depth analysis.

Sanitization: The initial phase involved cleaning the data, which meant addressing missing values, rectifying any discrepancies, and ensuring uniform data types.

Normalization: An interesting challenge was the need to change column names containing periods to underscores to prevent issues in SQL syntax, which streamlined subsequent data handling and analysis.

Exploratory Data Analysis (EDA):

Descriptive Statistics: I dove into SQL queries to extract descriptive statistics, getting a preliminary understanding of the dataset's characteristics, from average FICO scores and DTI ratios to the range of interest rates.

In-Depth Statistical Analysis:

Correlation Analysis: To address my hypotheses, I investigated the relationships between borrower traits and their loan repayment performance, as well as how these characteristics correlate with the interest rates assigned.

Insights:

Synthesis: I'm synthesizing my findings to transform complex datasets into straightforward, digestible insights.

Questions & Hypothesis

1. What factors most strongly predict whether a loan will be fully repaid?

<u>Hypothesis</u>: Borrowers with higher FICO scores, lower debt-to-income ratios, and loans for certain purposes such as credit card repayment are more likely to fully repay their loans.

2. How do borrower characteristics (e.g., FICO score, income level) influence the interest rate offered on loans?

<u>Hypothesis</u>: Borrowers with higher FICO scores and higher income levels are offered loans at lower interest rates due to their lower perceived risk.

3. Are there discernible patterns in loan approval rates or conditions based on the purpose of the loan (e.g., debt consolidation vs. credit card)?

<u>Hypothesis</u>: Loans for debt consolidation are more likely to be approved with favorable conditions compared to loans for small business due to the latter's higher risk of non-repayment.

4. What is the relationship between debt-to-income ratio and loan repayment behavior?

<u>Hypothesis</u>: A higher debt-to-income ratio is associated with a higher likelihood of loan non-repayment, as a higher percentage of the borrower's income is already committed to debt servicing.

Q1. Repayment Predictors

	purpose	percent_fully_paid	percent_not_fully_paid
1	small_business	72.213247172859	27.786752827140
2	educational	79.883381924198	20.116618075801
3	home_improvement	82.988871224165	17.011128775834
4	all_other	83.397683397683	16.602316602316
5	debt_consolidation	84.761182714177	15.238817285822
6	credit_card	88.431061806656	11.568938193343
7	major_purchase	88.787185354691	11.212814645308

	not_fully_paid	average_fico_score
1	0	713.326911124922
2	1	697.828440965427

	not_fully_paid	average_dti
1	0	12.4944126786824
2	1	13.1958382257012

These results suggest that loans with higher FICO scores have a higher likelihood of being fully repaid, which supports part of hypothesis #1. Although the difference is not large, it still indicates that loans with a lower DTI are slightly more likely to be fully repaid, also supporting part of hypothesis #1. This part of the analysis reveals that loan purpose does indeed affect repayment rates, with certain purposes such as small business loans being associated with a higher risk of non-repayment.

Q2. Borrower Impact

The data supports the hypothesis that FICO scores influence the interest rate offered on loans, as seen by the general trend of decreasing average interest rates with higher FICO score ranges. However, an anomaly is observed with the 800+ score range, which warrants further investigation. There isn't sufficient data provided regarding income levels to confirm its impact on interest rates. The data reveals no clear trend between income categories and interest rates, suggesting other factors may also significantly influence the rates borrowers receive.

	ficorange	average_interest_rate
1	650-699	0.14041077694235
2	700-749	0.11459853539463
3	750-799	0.0938304918032783
4	800+	0.129760638297872

	income_category	average_interest_rate	number_of_loans
1	High	0.122242424242424	33
2	Low	0.121850261780105	573
3	Medium	0.123613855860229	4121
4	Medium-High	0.128582530120482	332
5	Medium-Low	0.121418499668068	4519

Q3. Approval Patterns

	purpose	total_loans	fully_paid	not_fully_paid	percent_fully_paid	percent_not_fully_paid
1	small_business	619	447	172	72.213247172859	27.78675282714
2	educational	343	274	69	79.883381924198	20.116618075801
3	home_improvement	629	522	107	82.988871224165	17.011128775834
4	all_other	2331	1944	387	83.397683397683	16.602316602316
5	debt_consolidation	3957	3354	603	84.761182714177	15.238817285822
6	credit_card	1262	1116	146	88.431061806656	11.568938193343
7	major_purchase	437	388	49	88.787185354691	11.212814645308

From the results, it's evident that loans taken out for small business purposes have the highest proportion of not being fully paid, followed by educational and home improvement loans. These insights can inform risk assessments for different loan purposes. The analysis suggests that small business loans might be riskier than those for other purposes, which could be due to various factors, such as the inherent risks in starting or maintaining a small business.

Q4. Debt Behavior

	not_fully_paid	average_dti	min_dti	max_dti
1	0	12.4944126786824	0	29.95
2	1	13.1958382257012	0	29.96

The data shows that loans not fully paid have a slightly higher average DTI ratio compared to those that are fully paid, which could suggest that higher DTI is associated with a higher likelihood of not fully repaying a loan, supporting hypothesis #4. However, the difference in average DTI between the two groups is relatively small, indicating that while there may be a relationship, DTI alone may not be a strong predictor of repayment without considering other factors.

Results

1. What factors most strongly predict whether a loan will be fully repaid?

The findings indicate that loans with higher FICO scores and lower DTIs are more likely to be repaid, partially confirming hypothesis #1. Additionally, the analysis shows that loan purpose influences repayment rates, with some purposes like small business loans posing a higher non-repayment risk.

2. How do borrower characteristics (e.g., FICO score, income level) influence the interest rate offered on loans?

The data indicates that higher FICO scores generally lead to lower interest rates on loans, supporting the hypothesis about FICO scores affecting interest rates, though an anomaly in the 800+ range calls for further exploration. Without enough data on income levels, their effect on interest rates remains unconfirmed, as no definite trend links income to rates, hinting at other influential factors.

3. Are there discernible patterns in loan approval rates or conditions based on the purpose of the loan (e.g., debt consolidation vs. credit card)?

The analysis shows small business loans have the highest default rate, followed by educational and home improvement loans, indicating they're riskier. This insight could help refine risk assessments for loan purposes, considering the inherent risks associated with small businesses.

4. What is the relationship between debt-to-income ratio and loan repayment behavior?

The data shows that loans not fully paid have a slightly higher average DTI ratio compared to those that are fully paid, which could suggest that higher DTI is associated with a higher likelihood of not fully repaying a loan, supporting hypothesis #4. However, the difference in average DTI between the two groups is relatively small, indicating that while there may be a relationship, DTI alone may not be a strong predictor of repayment without considering other factors.