Hyoju Kang
04/10/2024

Lending Club Loan Analysis using querying (MySQL)

Objective

The objective of this project is to analyze data from the Lending Club dataset using SQL queries. The project aims to gain insights into various aspects of loans such as

- Loan Purposes
- Risk Assessment
- Behavioral Insights
- Loan Term Analysis
- Utilization Analysis

Data Dictionary

Field	Description
credit_policy	1 if the customer meets the credit underwriting criteria of LendingClub.com, and 0 otherwise.
purpose	The purpose of the loan (takes values "credit_card", "debt_consolidation", "educational", "major_purchase", "small_business", and "all_other").
int_rate	The interest rate of the loan, as a proportion (a rate of 11% would be stored as 0.11).
installment	The monthly installments owed by the borrower if the loan is funded.
log_annual_inc	The natural log of the self-reported annual income of the borrower.
dti	The debt-to-income ratio of the borrower (amount of debt divided by annual income).
fico	The FICO credit score of the borrower.
days_with_cr_line	The number of days the borrower has had a credit line.
revol_bal	The borrower's revolving balance (amount unpaid at the end of the credit card billing cycle).
revol_util	The borrower's revolving line utilization rate (the amount of the credit line used relative to total credit available).
inq_last_6mths	The borrower's number of inquiries by creditors in the last 6 months.
delinq_2yrs	The number of times the borrower had been 30+ days past due on a payment in the past 2 years.

Hyoju Kang
04/10/2024

pub_rec	The borrower's number of derogatory public records (bankruptcy filings, tax liens, or judgments).

Questions

1. Loan Purpose Analysis:

- What is the distribution of loans by purpose?
- Which loan purpose has the highest average interest rate, and how does it correlate with FICO score?

```
select purpose, count(*) as num
from loan_data
group by purpose
order by num desc;
```

purpose	num
debt_consolidation	3957
all_other	2331
credit_card	1262
home_improvement	629
small_business	619
major_purchase	437
educational	343

: The most common purpose of a loan is to consolidate debt into a single loan, as it allows the borrower to secure a lower interest rate compared to when it is spread across multiple debts. Additionally, the smallest number of loans is made for educational purposes.

```
select purpose, round(avg(int_rate),2) as avg_int_rate,
round(avg(fico), 2) as avg_fico
from loan_data
group by purpose
order by avg_int_rate desc;
```

purpose	avg_int_rate	avg_fico
small_business	0.14	719.99
debt_consolidation	0.13	703.87
credit_card	0.12	709.63
all_other	0.12	715.34
home_improvement	0.12	724.81
educational	0.12	708.28
major_purchase	0.11	722.5

Hyoju Kang
04/10/2024

: Clearly, the average interest rate for small business loans is the highest due to a higher default risk compared to other purposes. Additionally, there's no clear relationship between interest rates and FICO scores, as loans for major purchases have high FICO scores and the lowest interest rates, whereas small business loans have the highest interest rates despite similarly high FICO scores.

2. Risk Assessment:

- What is the relationship between debt-to-income ratio (dti) and the likelihood of not fully paying the loan (not_fully_paid)?

```
with cte as (
select *, case when dti between 0 and 5 then '0-5'
when dti between 5 and 10 then '5-10'
when dti between 10 and 15 then '10-15'
when dti between 15 and 20 then '15-20'
when dti between 20 and 25 then '20-25'
when dti > 25 then '>25'
else 'Unknown' end as dti_range
from loan_data)

select dti_range,
avg(case when not_fully_paid = 0 then 1 else 0 end) as avg_fully_paid,
avg(case when not_fully_paid = 1 then 1 else 0 end) as avg_not_fully_paid,
count(*) as total_loans
from cte
group by dti_range
order by avg_not_fully_paid desc
```

dti_range	avg_fully_paid	avg_not_fully_p...	total_loans
>25	0.7500	0.2500	152
5-10	0.8474	0.1526	2012
20-25	0.8237	0.1763	1475
15-20	0.8321	0.1679	2067
10-15	0.8450	0.1550	2271
0-5	0.8570	0.1430	1601

: By grouping rows into different DTI ranges and calculating the average fully paid and unpaid rates, I investigated the correlation between DTI and payment behavior. It's been proven that as the DTI range increases, the average full payment rate decreases. For example, when the DTI (debt-to-income ratio) exceeded 25, the average percentage of full payment was the lowest.

3. Behavioral Insights:

Hyoju Kang
04/10/2024

- Analyze the impact of the number of inquiries in the last 6 months (inq_last_6mths) on loan approval (credit_policy). Is there a threshold that significantly affects approval rates?
- What is the effect of public records (pub.rec) and past delinquencies (delinq.2yrs) on the interest rate and the likelihood of full payment?

```
select credit_policy as Approved,
round(avg(inq_last_6mths),5) as avg_inq_6
from loan_data
group by credit_policy
```

Approved	avg_inq_6
1	0.99754
0	3.97109

: As expected, the average number of credit inquiries within 6 months was significantly higher (four times) in disapproved loan cases compared to approved ones. The average number of credit inquiries was only 1 for approved loan cases.

```
select pub_rec, delinq_2yrs,
round(avg(int_rate),2) as avg_int_rate,
avg(case when not_fully_paid = 1 then 1 else 0 end) as rate_not_fully_paid
from loan_data
group by pub_rec, delinq_2yrs
order by pub_rec, delinq_2yrs;
```

pub_rec	delinq_2yrs	avg_int_rate	rate_not_fully_paid
0	0	0.12	0.1519
0	1	0.13	0.1732
0	2	0.14	0.1899
0	3	0.14	0.1935
0	4	0.14	0.1667
0	5	0.15	0.0000
0	6	0.13	0.0000
0	7	0.13	0.0000
0	8	0.15	0.0000
0	13	0.14	0.0000
1	0	0.13	0.2648
1	1	0.14	0.2093
1	2	0.15	0.0000
1	3	0.14	0.3333
1	4	0.14	1.0000
1	5	0.13	0.0000
1	11	0.19	0.0000
2	0	0.14	0.1250
2	1	0.15	0.0000
2	2	0.13	0.0000
3	0	0.15	0.0000
3	1	0.13	0.0000
4	0	0.16	0.0000
5	0	0.14	0.0000

: I was curious if public records and delinquency numbers within 2 years affect the interest rate and full payment rate of loans. It turns out that there is no specific relationship between them.

4. Loan Term Insights:

- How does the length of the credit line (days_with_cr_line) correlate with loan approval and the interest rate? Are longer credit lines associated with better loan terms?

Hyoju Kang
04/10/2024

- Calculate the average installment amount by purpose and by FICO score range. How does the reason for the loan and creditworthiness affect the installment plans?

```
with cte as(
select case when days_with_cr_line between 0 and 5000 then '0-5000'
            when days_with_cr_line between 5000 and 10000 then '5000-10000'
            when days_with_cr_line between 10000 and 15000 then '10000-15000'
            when days_with_cr_line between 15000 and 20000 then '15000-20000'
            when days_with_cr_line between 20000 and 25000 then '20000-25000'
            else '>25000' end as day_range,
count(*) as cnt, avg(int_rate) as avg_int_rate, avg(credit_policy) as
loan_approved
from loan_data
group by day_range)

select day_range as how_long,
round(avg_int_rate,3) as avg_int_rate, round(loan_approved,3) as loan_approved
from cte
```

how_long	avg_int_rate	loan_approved
5000-10000	0.12	0.833
0-5000	0.125	0.791
10000-15000	0.116	0.816
15000-20000	0.116	0.8

: It proves that the interest rate decreases as the number of years from opening a credit line increase. Also, the percentage of approved loans is highest when the range of years is between 5000 and 15000. This might be because individuals with fewer than 5000 days or over 15000 days are perceived as riskier, possibly due to being too young or too old.

5. Utilization Analysis:

- Examine the relationship between revolving balance (revol_bal), revolving line utilization rate (revol_util), and loan approval. Is there a pattern that suggests higher utilization rates affect credit policy decisions?
- What are the characteristics of borrowers who have high revolving balances but low utilization rates?

```
select credit_policy as approved,
avg(revol_bal) as avg_rev_balance, avg(revol_util) as avg_util
from loan_data
group by credit_policy;
```

Hyoju Kang
04/10/2024

approved	avg_rev_balance	avg_util
1	13798.4	45.31
0	29773.15	52.93

: This proves that approved loan customers tend to exhibit lower average revolving balance amounts and utilization rates. This correlation makes sense, as high revolving balances signify potential high debt and reflect negatively on credit evaluation. Additionally, lower utilization rates are preferable, indicating that customers are using less of their available credit limit, which reflects positively on their financial management.

```
select
  avg(fico) as avg_fico,
  avg(dti) as avg_dti,
  avg(pub_rec) as avg_pub_rec
from loan_data
where revol_bal > (select avg(revol_bal) from loan_data)
and revol_util < (select avg(revol_util) from loan_data);
```

avg_fico	avg_dti	avg_pub_rec
733.841726618705	13.508359712230225	0.008633093525179856

: The aggregated FICO, DTI, and public record values are provided above. I aimed to discern the characteristics of a group exhibiting high revolving balances and low utilization rates.

My assumption before calculating the values were people with high revolving balances but low utilization rate can be categorized into two different groups.

1. people with very high credit limits may spend substantial amounts during specific periods, often due to seasonal business activities.
2. People utilize diverse accounts to manage their credit scores effectively, striving to maintain low utilization rates across all accounts to potentially increase their credit limits based on their credit scores.

Overall, the combination of a high revolving balance and low utilization rates suggests that borrowers are effectively managing their finances within appropriate credit limits, indicating financial stability. In essence, they maintain financial flexibility by avoiding exceeding their credit limits and refraining from excessive credit use, thus preserving financial reserves. This is supported by their high FICO score (around 733), low DTI rate, and almost zero public records.