MORTGAGE-BACKED SECURITIES PREPAYMENT RISK ANALYSIS

OBJECTIVES

- Analyze the relationship between key loan attributes such as loan-to-value, debt-to-income, and interest rate and their impact on loan prepayment.
- Analyze customers/borrowers and loan characteristics, such as repayment range, loan purpose, and mortgage insurance percentage.
- Determine factors contributing to prepayment risk.
- Identify the high-risk factors for delinquency status.

DATA DESCRIPTION

The dataset consists of 291451 entries and 23 columns. Here is the description:

1. **Customer_ID:** Unique customer ID.
2. **Credit_Score_Range:** Credit score group ‘Excellent’, ‘Fair’, ‘Good’, ‘Poor’.
3. **Credit_Score:** Borrowers credit score during loan process.
4. **First_Payment_Date:** First payment date of loan.
5. **First_Time_Home_Buyer:** Whether the borrower is a first-time loan buyer or not.
6. **Maturity_Data:** Loan maturity date.
7. **Mortgage_Insurance_Percentage:** Percentage of mortgage insurance coverage obtained.
8. **Occupancy:** Property occupancy status at the time of the loan.
9. **Debt_to_Income:** Total monthly debt expense by the total monthly income of the borrower.

DATA DESCRIPTION

10. **Debt_to_Income_Range:** Debt-to-Income group by ‘High’, ‘Medium’, ‘Low’.
11. **Orig_UnPaid_Balance:** The unpaid amount of loan.
12. **Loan_To_Value_Range:** Loan-to-value group by ‘high’, ‘Medium’, ‘Low’.
13. **Loan_To_Value:** The amount of the loan at origination divided by the value of the property.
14. **Interest_Rate:** The interest rate of the loan.
15. **Channel:** Channel used by the party that delivered the loan to the issuer.
16. **Property_State:** The location of the property securing the loan.
17. **Property_Type:** The type of property that secures the loan.
18. **Loan_Purpose:** The purpose of the loan. 0 = Purchase, 1 = Refinance, 2 = Cash-Out

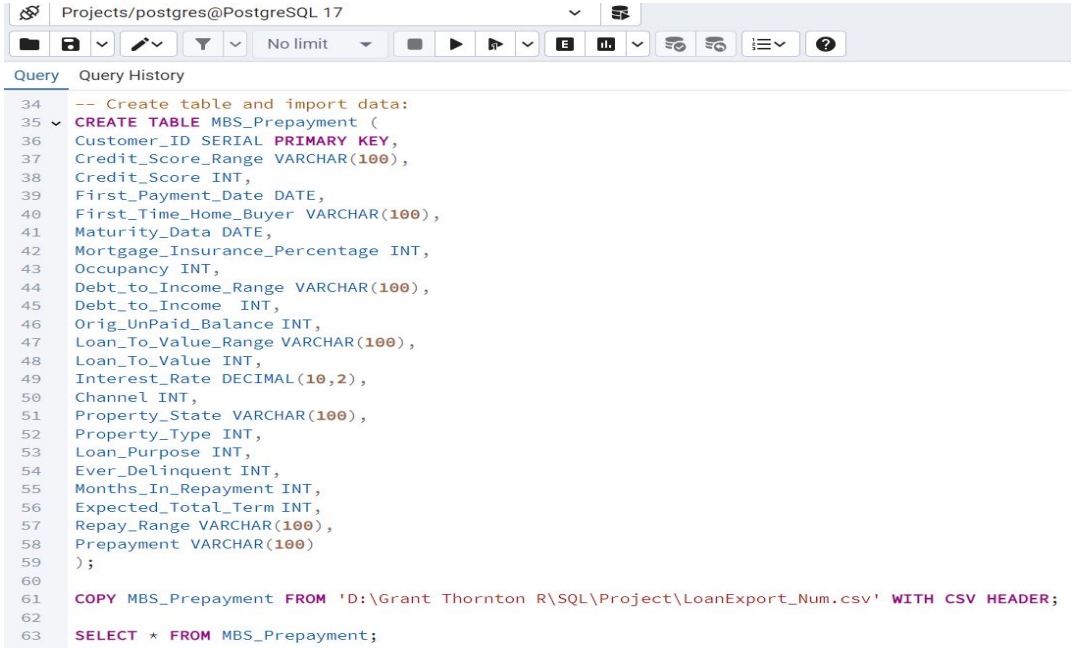
DATA DESCRIPTION

19. **Ever_Delinquent:** A mortgage loan has been reported as delinquent. 1 = Delinquent, 0 = Not-delinquent.
20. **Months_In_Repayment:** Total months for repayment of loan.
21. **Expected_Total_Term:** Expected months to repay.
22. **Repay_Range:** Years to repay loan.
23. **Prepayment:** Loan paid before maturity. 'Prepaid' and 'Not-Prepaid'

(Prepayment and Expected total term were calculated in Excel.)

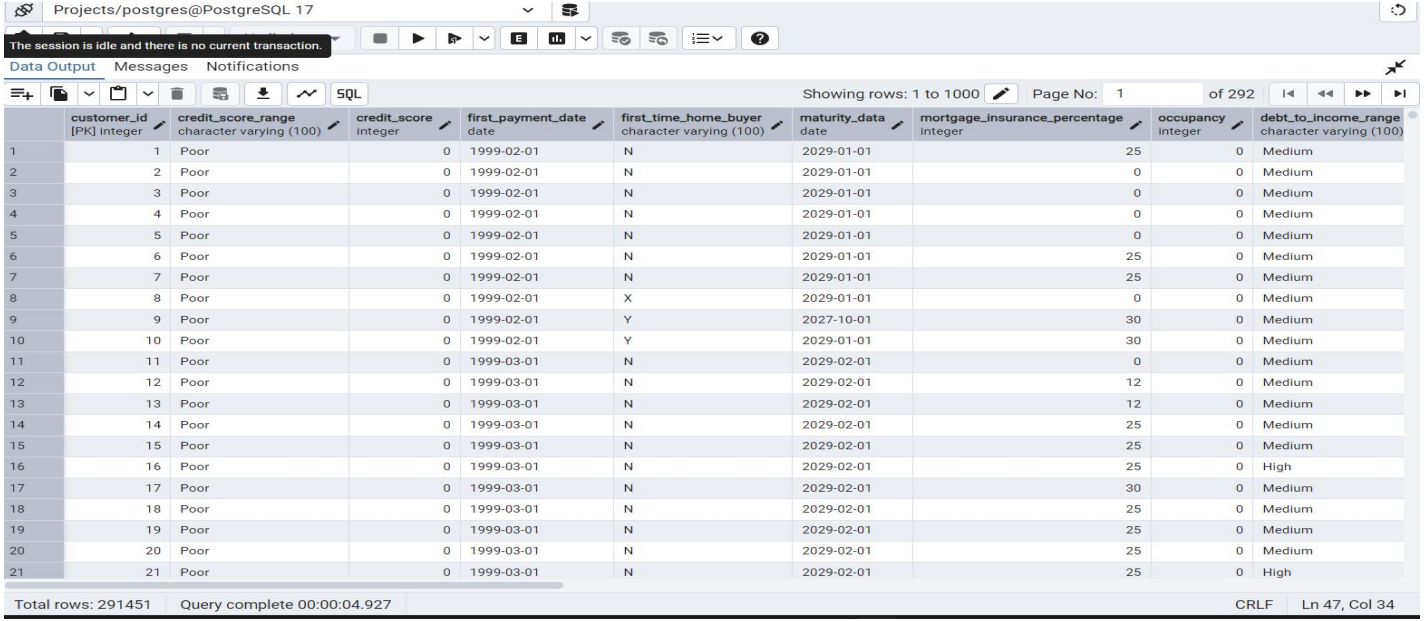
ANALYSIS

- Create a table and import data in the query editor:



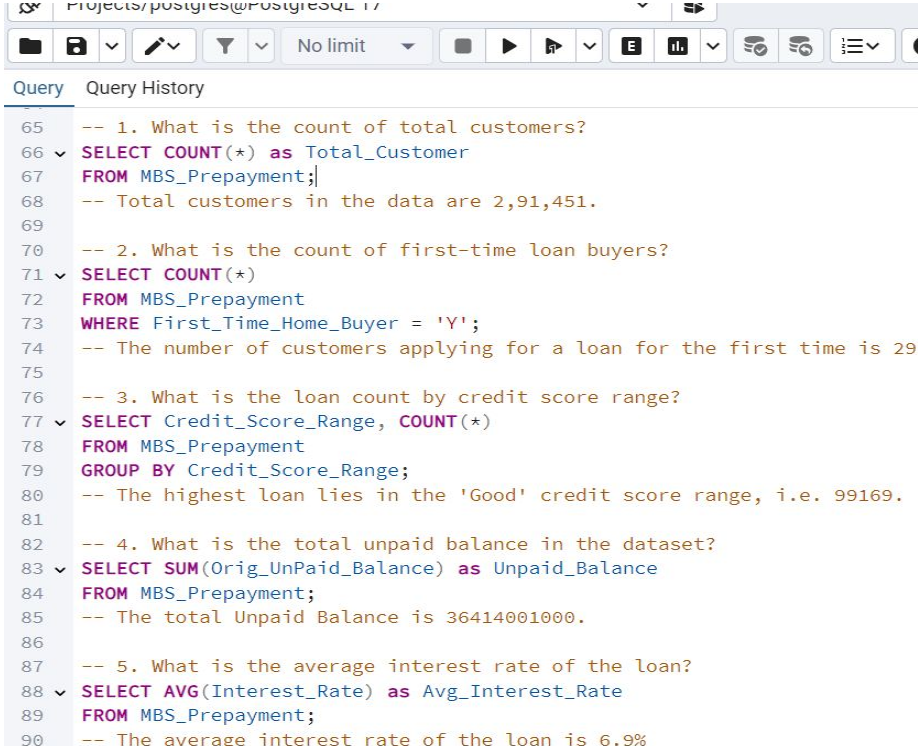
- View records of the data imported:

  SELECT * FROM MBS_Prepayment;



	customer_id [PK] integer	credit_score_range character varying (100)	credit_score integer	first_payment_date date	first_time_home_buyer character varying (100)	maturity_data date	mortgage_insurance_percentage integer	occupancy integer	debt_to_income_range character varying (100)
1	1	Poor	0	1999-02-01	N	2029-01-01	25	0	Medium
2	2	Poor	0	1999-02-01	N	2029-01-01	0	0	Medium
3	3	Poor	0	1999-02-01	N	2029-01-01	0	0	Medium
4	4	Poor	0	1999-02-01	N	2029-01-01	0	0	Medium
5	5	Poor	0	1999-02-01	N	2029-01-01	0	0	Medium
6	6	Poor	0	1999-02-01	N	2029-01-01	25	0	Medium
7	7	Poor	0	1999-02-01	N	2029-01-01	25	0	Medium
8	8	Poor	0	1999-02-01	X	2029-01-01	0	0	Medium
9	9	Poor	0	1999-02-01	Y	2027-10-01	30	0	Medium
10	10	Poor	0	1999-02-01	Y	2029-01-01	30	0	Medium
11	11	Poor	0	1999-03-01	N	2029-02-01	0	0	Medium
12	12	Poor	0	1999-03-01	N	2029-02-01	12	0	Medium
13	13	Poor	0	1999-03-01	N	2029-02-01	12	0	Medium
14	14	Poor	0	1999-03-01	N	2029-02-01	25	0	Medium
15	15	Poor	0	1999-03-01	N	2029-02-01	25	0	Medium
16	16	Poor	0	1999-03-01	N	2029-02-01	25	0	High
17	17	Poor	0	1999-03-01	N	2029-02-01	30	0	Medium
18	18	Poor	0	1999-03-01	N	2029-02-01	25	0	Medium
19	19	Poor	0	1999-03-01	N	2029-02-01	25	0	Medium
20	20	Poor	0	1999-03-01	N	2029-02-01	25	0	Medium
21	21	Poor	0	1999-03-01	N	2029-02-01	25	0	High

- Basic Counts and Aggregations:



	total_customer bigint
1	291451

	unpaid_balance bigint
1	36414001000

	avg_interest_rate numeric
1	6.9289700498540063

	credit_score_range character varying (100)	count bigint
1	Excellent	77311
2	Fair	74558
3	Good	99169
4	Poor	40413

	first_time_home_buyer character varying (100)	average_interest_rate numeric
1	N N = No	6.932770833107073 4
2	X X = Not Defined	6.9050571044029994
3	Y Y = Yes	6.9687774059148965

	ever_delinquent integer	count bigint
1	Not Delinquent 0	233788
2	Delinquent 1	57663

	first_time_home_buyer bigint
1	29282

Query History

```sql
-- 8. What is the loan count, Interest Rate and Unpaid Balance by PropertyState?
SELECT Property_State,
     COUNT(Property_State) AS loan_Count,
     AVG(Interest_Rate) AS Average_Interest_Rate ,
     SUM(Orig_UnPaid_balance) as Unpaid_Balance
FROM MBS_Prepayment
GROUP BY Property_State
ORDER BY Loan_Count DESC;
--LIMIT 5;
-- The top 5 states where the number of loans is high are California, Florida, Michigan, Illinois, and Texas.
-- There is no significant difference between interest rates across the state.
-- The highest loan count state also seems to have the highest unpaid balance.

-- 9. What is the average loan-to-value for each Credit Score Range?
SELECT Credit_Score_Range, round(AVG(Loan_to_Value),2) as Avg_loan_to_Value
FROM MBS_Prepayment
GROUP BY Credit_Score_Range;
-- Average loan to value for credit score range is:
-- Excellent = 71.03, Fair = 80, Good = 77, Poor = 81

-- 10. What is the total unpaid balance for each Debt to Income Range?
SELECT Debt_to_Income_Range, SUM(Orig_UnPaid_Balance) as Unpaid_balance
FROM MBS_Prepayment
GROUP BY Debt_to_Income_Range;
-- The medium debt-to-income range has the highest unpaid loan amount.
```

	property_state character varying (100)	loan_count bigint	average_interest_rate numeric	unpaid_balance bigint
1	CA	43327	6.95	6834875000
2	FL	19133	7.00	2015456000
3	MI	16285	6.98	1843107000
4	IL	13621	6.99	1667285000
5	TX	13092	6.93	1539799000

	credit_score_range character varying (100)	avg_loan_to_value numeric
1	Excellent	71.06
2	Fair	80.24
3	Good	77.27
4	Poor	81.27

	debt_to_income_range character varying (100)	unpaid_balance bigint
1	High	7711954000
2	Low	2843268000
3	Medium	25858779000

Query History

```sql
-- 11. Is there a relationship between Loan To Value Range and Interest Rate?
SELECT
    Loan_to_Value_Range,
    AVG(Interest_Rate) AS Average_Interest_Rate,
    MIN(Interest_Rate) AS Minimum_Interest_Rate,
    MAX(Interest_Rate) AS Maximum_Interest_Rate
FROM MBS_Prepayment
GROUP BY Loan_to_Value_Range;
-- For the High loan-to-value-range min interest is = 4 and max is 12.
-- For the Low loan-to-value-range min interest is = 5.63 and max is 10.85.
-- For the Medium loan-to-value-range min interest is = 5.13 and max is 9.13.
-- The average interest rate is 6.9 but the interest is varying as we can see the min & max for each range.
```

	loan_to_value_range character varying (100)	average_interest_rate numeric	minimum_interest_rate numeric	maximum_interest_rate numeric
1	High	6.93	4.00	12.35
2	Low	6.92	5.63	10.85
3	Medium	6.88	5.13	9.13

```sql
-- 12. Is there a relationship between Debt To Income Range and Interest Rate?
SELECT
    Debt_to_Income_Range,
    MIN(Interest_Rate) AS Min_Interest_Rate,
    MAX(Interest_Rate) AS Max_Interest_Rate,
    AVG(Interest_Rate) AS Avg_Interest_Rate
FROM MBS_Prepayment
GROUP BY Debt_to_Income_Range;
-- For the Low debt-to-income-range min interest is = 4.75 and max is 12.35.
-- For the High debt-to-income-range min interest is = 5.25 and max is 10.38.
-- For the Medium debt-to-income-range min interest is = 4 and max is 11.50.
```

	debt_to_income_range character varying (100)	min_interest_rate numeric	max_interest_rate numeric	avg_interest_rate numeric
1	High	5.25	10.38	6.97
2	Low	4.75	12.35	6.82
3	Medium	4.00	11.50	6.93

Query History

```sql
-- 13. What is the delinquency rate for each Debt to Income Range?
SELECT
    Debt_to_Income_Range,
    COUNT(*) AS Total_Loans,
    (COUNT(CASE WHEN Ever_Delinquent = 1 THEN 1 END) * 100.0 / COUNT(*)) AS Delinquency_Rate
FROM MBS_Prepayment
GROUP BY Debt_to_Income_Range
ORDER BY Delinquency_Rate DESC;
-- Customers having a 'High' debt-to-income have the highest delinquency rate i.e. 23.21%.
```

	debt_to_income_range character varying (100)	total_loans bigint	delinquency_rate numeric
1	High	59242	23.2115728705985618
2	Medium	209267	19.2820654952763694
3	Low	22942	15.5217505012640572

```sql
-- 14. Which combination of Credit Score Range and Debt to Income Range has the highest delinquency rate?
SELECT
    Credit_Score_Range,
    Debt_to_Income_Range,
    COUNT(CASE WHEN Ever_Delinquent = 1 THEN 1 END) AS Total_Delinquent,
    COUNT(*) AS Total_Loans,
    (COUNT(CASE WHEN Ever_Delinquent = 1 THEN 1 END) * 100.0 / COUNT(*)) AS Delinquency_Rate
FROM MBS_Prepayment
GROUP BY Credit_Score_Range, Debt_to_Income_Range
ORDER BY Delinquency_Rate ASC;
-- LIMIT 1
-- The 'Poor' credit score range and 'High' debt-to-income range have the highest delinquency rate at 42%.
-- The 'Excellent' credit score range and 'Low' debt-to-income range have the lowest delinquency rate at 7.44%.
-- Customers with a poor credit score and high debt to income are likely to be delinquent.
```

	credit_score_range character varying (100)	debt_to_income_range character varying (100)	total_delinquent bigint	total_loans bigint	delinquency_rate numeric
1	Poor	High	3846	9059	42.4550171100562976
2	Poor	Medium	11607	28999	40.0255181213145281
3	Poor	Low	936	2355	39.7452229299363057
4	Fair	High	5236	18519	28.2736648847129975

Query History

```sql
-- 15. For loans with a repayment range, what is LoanPurpose?
SELECT
    Repay_Range,
    COUNT(CASE WHEN Loan_Purpose = 0 THEN 0 END) AS Cash_IN,
    COUNT(CASE WHEN Loan_Purpose = 2 THEN 2 END) AS Cash_OUT
FROM MBS_Prepayment
GROUP BY Repay_Range
ORDER BY Cash_IN DESC;
-- The Loan purpose for the overall repayment range is cash-in(purchase).
```

	repay_range character varying (100)	cash_in bigint	cash_out bigint
1	0 - 4 Years	65614	27924
2	4 - 8 Years	41400	19546
3	8 - 12 Years	7373	4705
4	12 - 16 Years	4719	2957
5	16 - 20 Years	4298	2879

```sql
-- 16. Which is the most common repayment range for loans with a Loan To Value greater than 90?
SELECT Repay_Range, COUNT(Loan_to_Value)
FROM MBS_Prepayment
WHERE Loan_to_Value > 90
GROUP BY Repay_Range
--LIMIT 1;
-- The repayment range for the loan-to-value above 90 is 0-4 Years.
```

	repay_range character varying (100)	loan_to_value bigint
1	0 - 4 Years	25969
2	12 - 16 Years	1250
3	16 - 20 Years	1432
4	4 - 8 Years	16311
5	8 - 12 Years	1985

```sql
-- 17. What is the total number and unpaid balance of loans for customers with a Mortgage Insurance Percentage greater than 25%?
SELECT COUNT(*) AS Total_Loans,
SUM(Orig_UnPaid_Balance) AS Total_Unpaid_Balance
FROM MBS_Prepayment
WHERE Mortgage_Insurance_Percentage > 25;
-- The total count whose mortgage insurance percentage is more than 25 is 44,684.
-- The total unpaid balance is 5,55,49,89,000.
```

	total_loans bigint	total_unpaid_balance bigint
1	44684	5554989000

Query History

```sql
-- 18. What is the count and percentage of prepaid loans?
SELECT Prepayment, COUNT(*)
FROM MBS_Prepayment
GROUP BY Prepayment;
-- The total Prepaid Loan is 577.
```

	prepayment character varying (100)	count bigint
1	Not Prepaid	290874
2	Prepaid	577

```sql
-- 19. Is there a correlation between loan-to-value and prepayment risk?
SELECT Loan_to_Value_Range,
COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) AS Prepaid_Count,
COUNT(*) AS Total_Count,
(COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) * 100.0) / COUNT(*) AS Prepayment_Percentage
FROM MBS_Prepayment
GROUP BY Loan_to_Value_Range;
-- Customers with 'High' loan-to-value are likely to make a prepayment. The payment percentage is 0.199%.
```

	loan_to_value_range character varying (100)	prepaid_count bigint	total_count bigint	prepayment_percentage numeric
1	High	540	270470	0.20
2	Low	4	1953	0.20
3	Medium	33	19028	0.17

```sql
-- 20. How do loans with a higher debt-to-income compare in terms of prepayment risk?
SELECT Debt_to_Income_Range,
    COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) AS Prepaid_Count,
    COUNT(*) AS Total_Count,
    (COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) * 100.0) / COUNT(*) AS Prepayment_Percentage
FROM MBS_Prepayment
GROUP BY Debt_to_Income_Range
ORDER BY Prepaid_count DESC;
-- Customers with 'Medium' debt-to-income are likely to make prepayment.
```

	debt_to_income_range character varying (100)	prepaid_count bigint	total_count bigint	prepayment_percentage numeric
1	Medium	398	209267	0.19
2	High	90	59242	0.15
3	Low	89	22942	0.39

Query History

```
-- 21. Are first-time home buyers more or less likely to prepay their loans?
SELECT First_Time_Home_Buyer,
    COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) AS Prepaid_Count,
    COUNT(*) AS Total_Count,
    round(((COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) * 100.0) / COUNT(*)),2) AS Prepayment_
FROM MBS_Prepayment
GROUP BY First_Time_Home_Buyer
ORDER BY First_Time_Home_Buyer;
-- first-time home buyers are less likely to prepay the loan.

-- 22. Are loans with more extended repayment periods more prone to prepayment?
SELECT Repay_Range,
    COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) AS Prepaid_Count,
    COUNT(*) AS Total_Count,
    round(((COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) * 100.0) / COUNT(*)),2) AS Prepayment
FROM MBS_Prepayment
GROUP BY Repay_Range
ORDER BY Repay_Range;
-- The lower repayment range, i.e. 0 - 4 Years, will likely prepay the loan early.

-- 23. Is there any significant relationship between delinquency and prepayment behavior?
SELECT Ever_Delinquent,
    COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) AS Prepaid_Count,
    COUNT(*) AS Total_Count,
    (COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) * 100.0) / COUNT(*) AS Prepayment_Percenta
FROM MBS_Prepayment
GROUP BY Ever_Delinquent
ORDER BY Ever_Delinquent;
-- There is no significant relationship between the delinquent and prepayment.
```

	first_time_home_buyer character varying (100)	prepaid_count bigint	total_count bigint	prepayment_percentage numeric
1	N	357	184154	0.19
2	X	177	78015	0.23
3	Y	43	29282	0.15

N = No

X = Not Defined

Y = Yes

	repay_range character varying (100)	prepaid_count bigint	total_count bigint	prepayment_percentage numeric
1	0 - 4 Years	479	148859	0.32
2	12 - 16 Years	1	12575	0.01
3	16 - 20 Years	0	11740	0.00
4	4 - 8 Years	96	98320	0.10
5	8 - 12 Years	1	19957	0.01

	ever_delinquent integer	prepaid_count bigint	total_count bigint	prepayment_percentage numeric
1	Not Delinquent 0	517	233788	0.22
2	Delinquent 1	60	57663	0.10

Query History

```sql
-- 24. Which loan purpose has the highest prepayment risk across the state?
SELECT Loan_Purpose, Property_State,
(SELECT COUNT(*) FROM MBS_Prepayment AS subquery
  WHERE subquery.Loan_Purpose = MBS_Prepayment.Loan_Purpose
  AND subquery.Property_State = MBS_Prepayment.Property_State
  AND subquery.Prepayment = 'Prepaid') AS Prepaid_Count,
  COUNT(*) AS Total_Count,
  round((SELECT COUNT(*) FROM MBS_Prepayment AS subquery
  WHERE subquery.Loan_Purpose = MBS_Prepayment.Loan_Purpose
  AND subquery.Property_State = MBS_Prepayment.Property_State
  AND subquery.Prepayment = 'Prepaid') * 100.0 / COUNT(*),2) AS Prepayment_Percentage
FROM MBS_Prepayment
GROUP BY Loan_Purpose, Property_State
ORDER BY Prepayment_Percentage DESC;
--LIMIT 5;
-- State Wisconsin, ALALABAMA, VERMONT, irrespective of the loan purpose, has a high prepayment risk.
-- Used subquery to get accurate results.

-- 25. How does prepayment risk differ between loans with varying loan purposes?
SELECT Loan_Purpose,
    COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) AS Prepaid_Count
FROM MBS_Prepayment
GROUP BY Loan_Purpose
ORDER BY Prepaid_count DESC;
-- Customers with a loan purpose of 'Purchase' are more likely to prepay the loan than others.
```

	loan_purpose integer	property_state character varying (100)	prepaid_count bigint	total_count bigint	prepayment_percentage numeric
1	0	WI	25	1764	1.42
2	0	VT	4	319	1.25
3	1	AL	16	1307	1.22
4	1	VT	5	427	1.17
5	1	SC	9	1249	0.72
6	0	FL	79	11317	0.70
7	0	SC	17	2443	0.70
8	1	MI	41	5937	0.69

	loan_purpose integer		prepaid_count bigint
1	Purchase	0	280
2	Refinance	1	270
3	Cash-Out	2	27

```sql
-- 26. How does a high interest rate affect prepayment risk?
SELECT
CASE WHEN Interest_Rate > 7 THEN 'High Interest Rate'
ELSE 'Low Interest Rate'
END AS Interest_Rate_Category,
COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) AS Prepaid_Count
FROM MBS_Prepayment
GROUP BY Interest_Rate_Category;
-- The loans with high interest rates are likely to be prepaid.

-- 27. How do interest rate, debt-to-income, and credit score affect prepayment risk?
SELECT Debt_to_Income_Range, Credit_Score_Range,
CASE
WHEN Interest_Rate > 7 THEN 'High Interest Rate'
ELSE 'Low Interest Rate' END AS Interest_Rate_Category,
COUNT(CASE WHEN Prepayment = 'Prepaid' THEN 1 END) AS Prepaid_Count
FROM MBS_Prepayment
GROUP BY Interest_Rate_Category, Debt_to_Income_Range, Credit_Score_Range
ORDER BY Prepaid_count DESC;
-- Loans are prepaid, having a 'Medium' Debt-to-income range and 'Good' and 'Excellent'
-- credit score range, whether interest rate is high or low.
```

	interest_rate_category text	prepaid_count bigint
1	High Interest Rate	325
2	Low Interest Rate	252

	debt_to_income_range character varying (100)	credit_score_range character varying (100)	interest_rate_category text	prepaid_count bigint
1	Medium	Good	Low Interest Rate	82
2	Medium	Excellent	High Interest Rate	66
3	Medium	Good	High Interest Rate	62
4	Medium	Excellent	Low Interest Rate	55
5	Medium	Fair	High Interest Rate	53
6	Medium	Poor	High Interest Rate	39
7	Medium	Fair	Low Interest Rate	27
8	Low	Excellent	High Interest Rate	17
9	High	Excellent	High Interest Rate	16

```sql
-- 28. What is the delinquent count by repayment range?
SELECT
    Repay_Range,
    COUNT(CASE WHEN Ever_Delinquent = 0 THEN 0 END) AS Not_Delinquent,
    COUNT(CASE WHEN Ever_Delinquent = 1 THEN 1 END) AS Delinquent
FROM MBS_Prepayment
GROUP BY Repay_Range
ORDER BY Delinquent DESC;
-- the repayment range of 4 - 8 Years shows more delinquency.
```

	repay_range character varying (100)	not_delinquent bigint	delinquent bigint
1	4 - 8 Years	76525	21795
2	0 - 4 Years	129228	19631
3	16 - 20 Years	5807	5933
4	8 - 12 Years	14228	5729
5	12 - 16 Years	8000	4575

The repayment range varying from 0 to 8 years shows more delinquency compared to the long-term repayment range.

KEY INSIGHTS

- Loans with higher loan-to-value ranges have variation in interest rates, with minimum rates starting as low as 4% and maximum rates reaching up to 12%.
- Lower loan-to-value ranges generally lower interest rates, indicating reduced risk.
- For higher debt-to-income ranges interest rates are elevated, indicating increased borrower risk.
- Higher prepayment risk is observed in loans with ‘high’ interest rates and ‘less’ repayment periods, likely due to refinancing opportunities.

KEY INSIGHTS

- Delinquency rates are higher in loans with ‘high’ debt-to-income ratios with poor credit scores, indicating these factors significantly contribute to borrower financial stress.
- Loans with repayment periods more than 12 years are most frequently associated with certain purposes, such as refinancing or purchasing a property. Loans with longer repayment terms often have a risk of delinquency.
- Certain states may have a higher concentration of long-term loans.
- Loans with ‘less’ repayment terms show a higher risk of delinquency.

RECOMMENDATIONS

- For loans with high loan-to-value and debt-to-income ratios, implement strict criteria to reduce delinquency risk.
- Loans with high interest rates and lower repayment ranges have prepayment risk.
- Encourage borrowers for lower loan-to-value loans by offering incentives like reduced interest rates or flexible terms.

THANK YOU!