SQL Query Documentation

Project Information

Project Name:	Credit_Risk_Scoring
Database Name:	Credit_Risk_Scoring_DB

Data Preparation Summary

The dataset consisted of over 2.2 million rows and initially failed to import into PostgreSQL due to structural inconsistencies, special characters, and data type mismatches. Excel was unable to load the file due to its size, so I used Notepad and Power BI Power Query for large-scale text-level cleanup, including column selection, encoding issues, and delimiter fixes. I also adjusted PostgreSQL memory settings and used manual debugging at the row level to fix malformed entries. Successfully imported the cleaned dataset into PostgreSQL for analysis.

Query Documentation

Query 1: Total Loan Amount Disbursed

Objective:

What is the Total Loan amount disbursed across all records?

SQL Query:

```
SELECT SUM(loan_amnt) AS total_loan_amount

FROM loan_data;
```

Result:

Row no	total_loan_amount
1	34016115925

Business Insights:

The total loan amount disbursed stands at **$34.02 billion**, indicating the scale of lending operations and overall credit exposure. This KPI is crucial for assessing the company's financial outreach and portfolio size.

Query 2: Total Number of Loan Isuued

Objective:

How many loans have been disbursed overall?

SQL Query:

```
SELECT COUNT(*) AS total_loan_issued

FROM loan_data;
```

Result:

Row no	total_loan_issued
1	2260701

Business Insights:

A total of **2,260,701 loans** have been issued, reflecting the scale of lending operations and customer outreach. This figure helps evaluate business growth, borrower volume, and operational activity.

Query 3: Total Outstanding Revolving Balance

Objective:

What is the total unpaid revolving balance across all active loans?

SQL Query:

```
SELECT SUM(revol_bal) AS total_outstanding_balance

FROM loan_data;Result:
```

Row no	"total_outstanding_balance"
1	37659243106

Business Insights:

The total outstanding revolving balance stands at **$37.66 billion**, indicating the cumulative unpaid credit amount. This metric is vital for assessing credit exposure and portfolio risk.

Query 4: Average Interest Rate

Objective:

What is the average interest rate applied across all loans?

SQL Query:

```
SELECT ROUND(AVG(interest_rate_decimal) * 100, 1) || '%' AS avg_interest_rate

FROM loan_data;
```

Result:

Row no	avg_interest_rate
1	13.1%

Business Insights:

The average interest rate across all loans is **13.1%**, reflecting the overall cost of borrowing. This KPI is key for assessing the profitability of loan products and competitive positioning in the market..

Query 5: Loan Default Rate

Objective:

What percentage of loans are in default or at risk of default?

SQL Query:

```
SELECT

  ROUND(

    100.0 * COUNT(*) FILTER (

      WHERE loan_status IN (

        'Charged Off',

        'Default',

        'Late (31-120 days)',

        'Late (16-30 days)',

        'In Grace Period'
```

```
        )
    )::NUMERIC / COUNT(*),
    1
  ) || '%' AS default_rate
FROM loan_data;
```

Result:

Row no	default_rate
1	13.4%

Business Insights:

The loan default rate is 13.4%, indicating the proportion of loans that are either charged off, in default, or at risk due to being late or in a grace period. This metric is crucial for assessing credit risk and the overall quality of the loan portfolio.

Query 6: Loan Issued Over Time in 2018

Objective:

How many loans were issued each month in 2018?

SQL Query:

```
SELECT D.month,
        D.year,
        COUNT(L.loan_amnt) AS loan_count
        FROM loan_data AS L
        JOIN date_table AS D
        ON L.issue_d=D.date
        WHERE D.year= 2018
        GROUP BY D.year, D.month
        ORDER BY D.year, D.month;
```

Result:

Row no	month	year	loan_count

1	April	2018	42928
2	August	2018	46079
3	December	2018	40134
4	February	2018	32746
5	January	2018	36347
6	July	2018	43089
7	June	2018	41533
8	March	2018	38771
9	May	2018	46311
10	November	2018	41973
11	October	2018	46305
12	September	2018	39026

Business Insights:

Loan issuance fluctuated throughout 2018, with **May** and **October** leading the charge, each issuing **46,311** and **46,305** loans, respectively. This insight highlights peak months for loan activity and can be used to identify trends and inform strategic planning..

Visualized via Line chart for clear Trend analysis..

Query 7: Loan Amount by Grade

Objective:

What are the top-performing loan grades in terms of total loan disbursement, and what does this indicate about the company's lending strategy?

SQL Query:

```
SELECT grade, SUM(loan_amnt) AS loan_amnt_by_grade

FROM loan_data

WHERE loan_amnt IS NOT NULL

GROUP BY grade

ORDER BY SUM(loan_amnt) DESC LIMIT 5;
```

Result:

Row no	grade	loan_amnt_by_grade
1	C	9775551175
2	B	9404817775
3	A	6323641900
4	D	5097344375
5	E	2367318100

Business Insights:

Grades C, B, and A received the highest loan disbursements, led by Grade C with over $9.7B.
This suggests a focus on moderate to low-risk borrowers for portfolio growth.
Grades B and C balance risk and return, attracting substantial lending volume.
The lending strategy reflects strong risk management and income optimization.

Query 8: Default Rate over Time in 2018

Objective:

How did the loan default rate trend month-over-month in 2018?

SQL Query:

```
SELECT
  D.month,
  D.monthnumber,
  ROUND(
    100.0 * COUNT(CASE WHEN L.loan_status = 'Charged Off' THEN 1 END)
    / NULLIF(COUNT(*), 0), 2
  ) AS default_rate_percent
FROM
  loan_data L
JOIN
  date_table D
  ON L.issue_d = D.date
WHERE
  D.year = 2018
GROUP BY
  D.month,
  D.monthnumber
ORDER BY
```

D.monthnumber;

Result:

Row no	Month	monthnumber	default_rate_percent
1	January	1	4.60
2	February	2	4.02
3	March	3	3.60
4	April	4	3.17
5	May	5	2.60
6	June	6	1.88
7	July	7	1.23
8	August	8	0.68
9	September	9	0.44
10	October	10	0.18
11	November	11	0.06
12	December	12	0.03

Business Insights:

Default rates steadily declined throughout 2018, from 4.6% in January to just 0.03% in December.
This sharp drop suggests improved borrower quality, risk controls, or policy changes over time.
The trend reflects growing credit discipline and effective underwriting strategies.
Such patterns are key indicators of portfolio health for financial institutions.

Query 9: Interest Rate by Term

Objective:

What percentage of total interest rates is contributed by each loan term?

SQL Query:

SELECT

```
  term,
  ROUND(
    (SUM(int_rate) / (SELECT SUM(int_rate) FROM loan_data))* 100,2
  ) AS int_rate_by_term_percentage
FROM
  loan_data
WHERE int_rate IS NOT NULL
GROUP BY
  term
ORDER BY
  int_rate_by_term_percentage DESC;
```

Result:

Row no	term	"int_rate_by_term_percentage"
1	36 months	63.71
2	60 months	36.29

Business Insights:

Loans with a 36-month term contribute 63.71% of the total interest rate volume.
This suggests shorter-term loans are more prevalent or carry higher individual interest rates.
Understanding term-based interest contributions helps optimize lending strategies and portfolio risk.
Focus on term preferences can guide targeted loan offerings and marketing..

Query 10: Loan Purpose Distribution

Objective:

Which loan purposes are most common among borrowers?

SQL Query:

```
SELECT purpose, COUNT(loan_amnt) AS loan_amnt_distributed
FROM loan_data
GROUP BY purpose
```

ORDER BY COUNT(loan_amnt) DESC LIMIT 5;

Result:

Row no	purpose	loan_amnt_distributed
1	debt_consolidation	1277877
2	credit_card	516971
3	home_improvement	150457
4	other	139440
5	major_purchase	50445

Business Insights:

Debt consolidation leads with 1.27M loans, followed by credit card and home improvement loans.
This indicates a high borrower demand for refinancing and managing existing debts.
Lenders can focus on these categories to design tailored financial products and risk models.

.

Query 11: Annual Income Distribution by Credit Grades

Objective:

How does the total annual income of borrowers vary across credit grades?

SQL Query:

```
SELECT grade, SUM(annual_inc) AS annual_inc_by_grade

FROM loan_data

WHERE annual_inc IS NOT NULL

GROUP BY grade

ORDER BY SUM(annual_inc) DESC LIMIT 5;
```

Result:

Row no	grade	annual_inc_by_grade
1	B	52297294653.31
2	C	48344630505.41
3	A	38929808311.69
4	D	23043895899.80

5	E	9746410196.62

Business Insights:

Grade B borrowers contribute the highest total income (~$52.3B), followed by Grades C and A.
This suggests that mid-grade borrowers (B & C) form a substantial portion of the earning customer base.
Lenders can optimize offerings for these segments to balance income potential and credit risk.

Query 12: Loan purpose vs Default Rate

Objective:

Which loan purposes have the highest default rates?

SQL Query:

```
SELECT
  purpose,
  ROUND(
    COUNT(CASE WHEN loan_status = 'Charged Off' THEN 1 END) * 100.0 / COUNT(*),
    0
  ) AS default_rate_percentage
FROM
  loan_data
GROUP BY
  purpose
ORDER BY
  default_rate_percentage DESC LIMIT 5;
```

Result:

Row no	purpose	default_rate_percentage
1	small_business	19
2	renewable_energy	15
3	moving	14

4	educational	13
5	debt_consolidation	13

Business Insights:

Loans for **small businesses** show the highest default rate (19%), followed by **renewable energy** and **moving**.
These high-risk categories may need stricter underwriting or adjusted interest rates.
Focusing on risk-prone purposes can improve overall loan portfolio health.

Project Summary

This project involved analyzing over 2.2 million loan records to uncover key insights into loan distribution, interest rates, customer behavior, and default patterns using PostgreSQL and Power BI.
KPIs, interactive charts, and business-driven questions were answered with actionable insights, ensuring the project aligns with real-world financial analytics scenarios.
The final dashboard is optimized for decision-making and demonstrates strong SQL skills, data visualization expertise, and business acumen—suitable for high-level data analyst roles.