

Loan Default Risk Analysis

SQL- Based Lending Risk Insights

Abhishek Verma

Data Analytics Portfolio Project

Project Overview & Objective

This project leverages a comprehensive LendingClub dataset to unearth critical risk factors in personal loans using SQL. By transforming raw data into actionable insights, we aim to demonstrate proficiency in data cleaning, SQL querying, and analytical interpretation for robust risk assessment.

Uncover Key Risk Factors

Identify which loan attributes correlate most strongly with default rates.

Generate Actionable Insights

Provide data-driven recommendations for improving loan underwriting processes.



Tools & Dataset

Dataset: LendingClub Loan Data

A publicly available dataset from Kaggle, containing millions of past loan applications with various attributes and their default status.

This rich dataset provides a realistic scenario for analyzing financial risk in consumer lending.



Tools Used for Analysis



SQL (SQLite)

For efficient data querying, cleaning, and transformation.



Excel

Initial data cleaning and preliminary exploration.



DB Browser

Visual interface for SQLite database management.



GitHub

Version control and project portfolio showcase.



Data Cleaning & Transformation

The raw LendingClub dataset required significant cleaning and transformation to ensure accuracy and usability for analysis. This critical step involved:

Handling Missing Values

Identified and addressed nulls, especially in critical financial fields.

Parsing & Formatting

Extracted numerical values from text fields (e.g., loan term from "36 months").

Categorization

Grouped continuous data (like DTI, income) into meaningful bands for easier analysis.

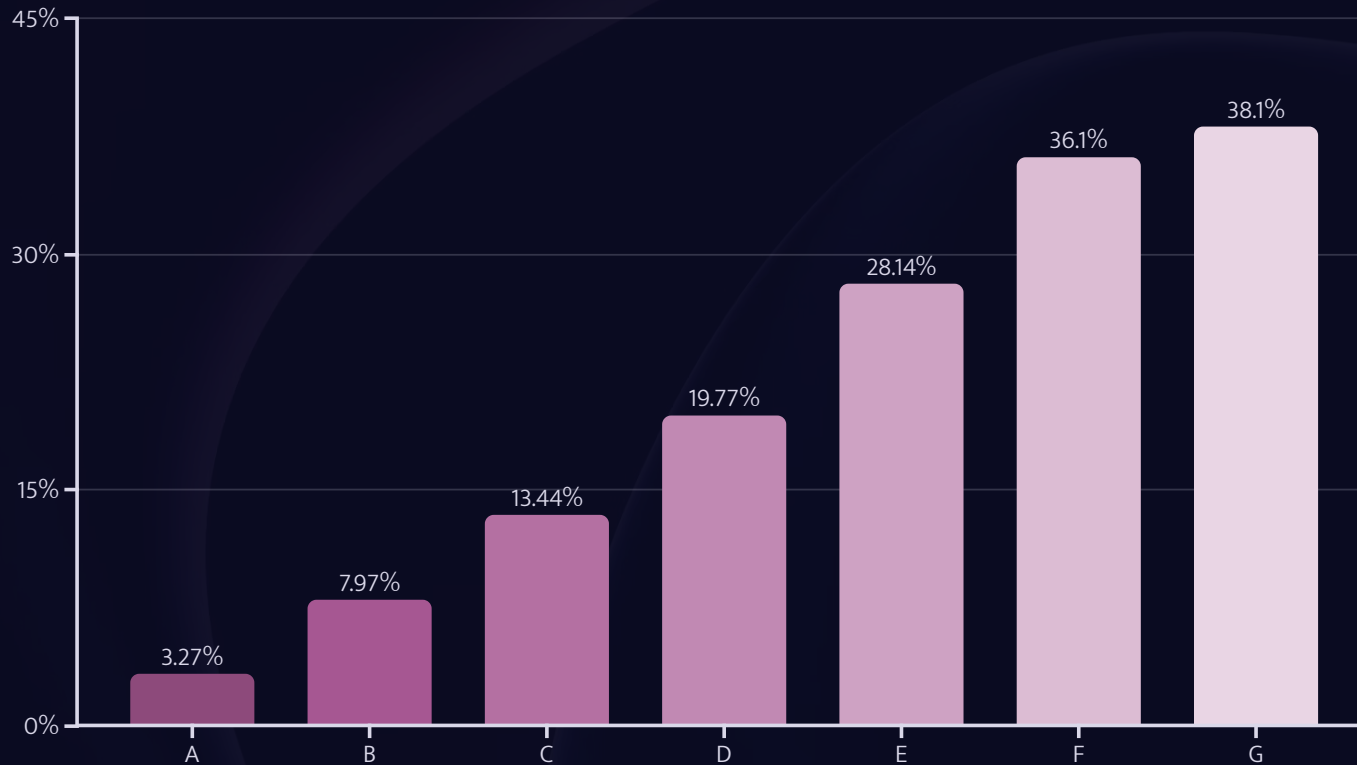
Schema Refinement

Simplified and optimized table structure for efficient querying.

Key Business Insights

Default

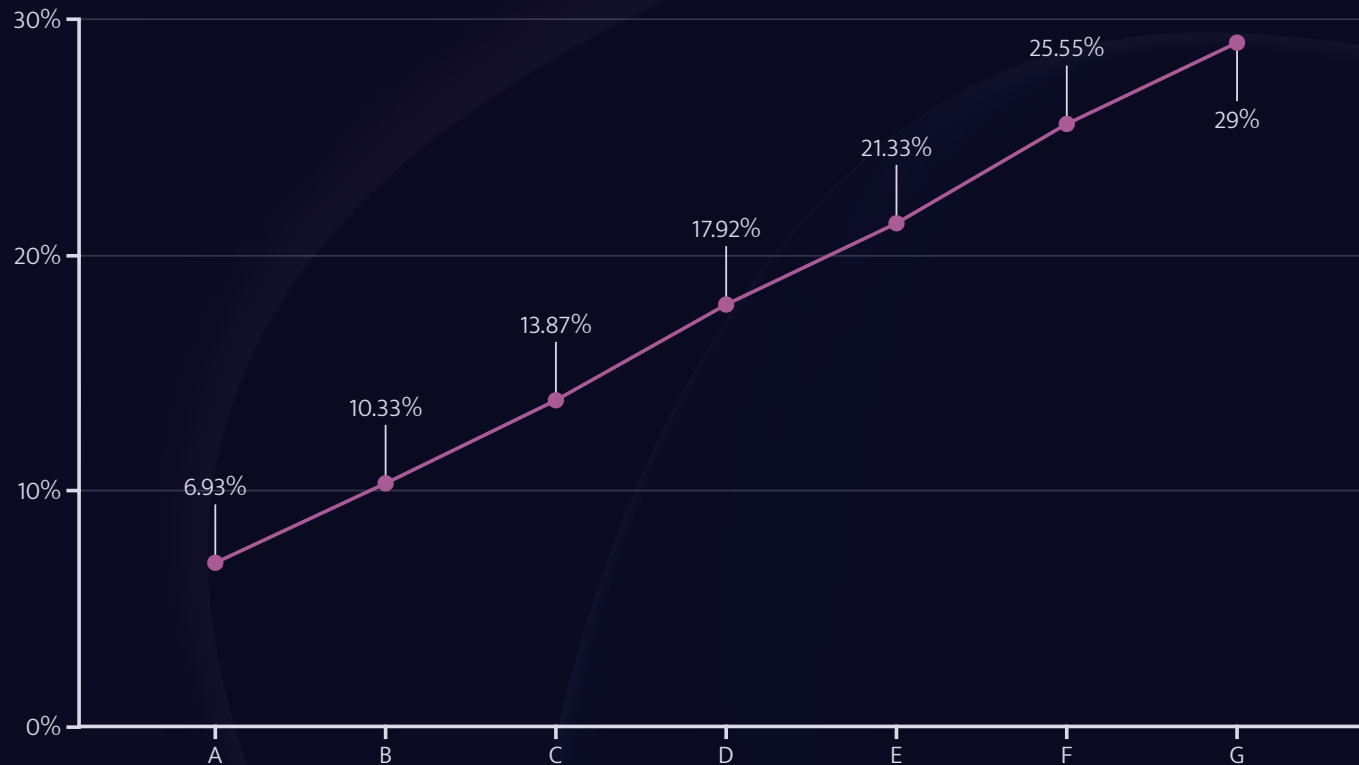
Loan grade is a primary indicator of risk, with a stark contrast in default rates across the spectrum.



⊗ **Observation:** Default rates escalate significantly with lower loan grades, indicating higher inherent risk.

Average Interest Rate by Loan Grade

Interest rates are directly correlated with loan grades, reflecting the risk premium for less creditworthy borrowers.



Observation: Higher risk (lower grade) corresponds to significantly higher average interest rates, as expected.

Loan Purpose vs. Default Risk

The purpose for which a loan is taken significantly impacts its likelihood of default.



Small Business

Highest Default Rate ~17.9%

Often unpredictable cash flows, higher inherent business risk.



Debt Consolidation

High Default Rate: ~13.2%

Borrowers potentially over-leveraged, seeking to simplify payments.



Home Improvement

Moderate Default Rate: ~10.4%

Often secured by property, but still subject to borrower's financial stability.



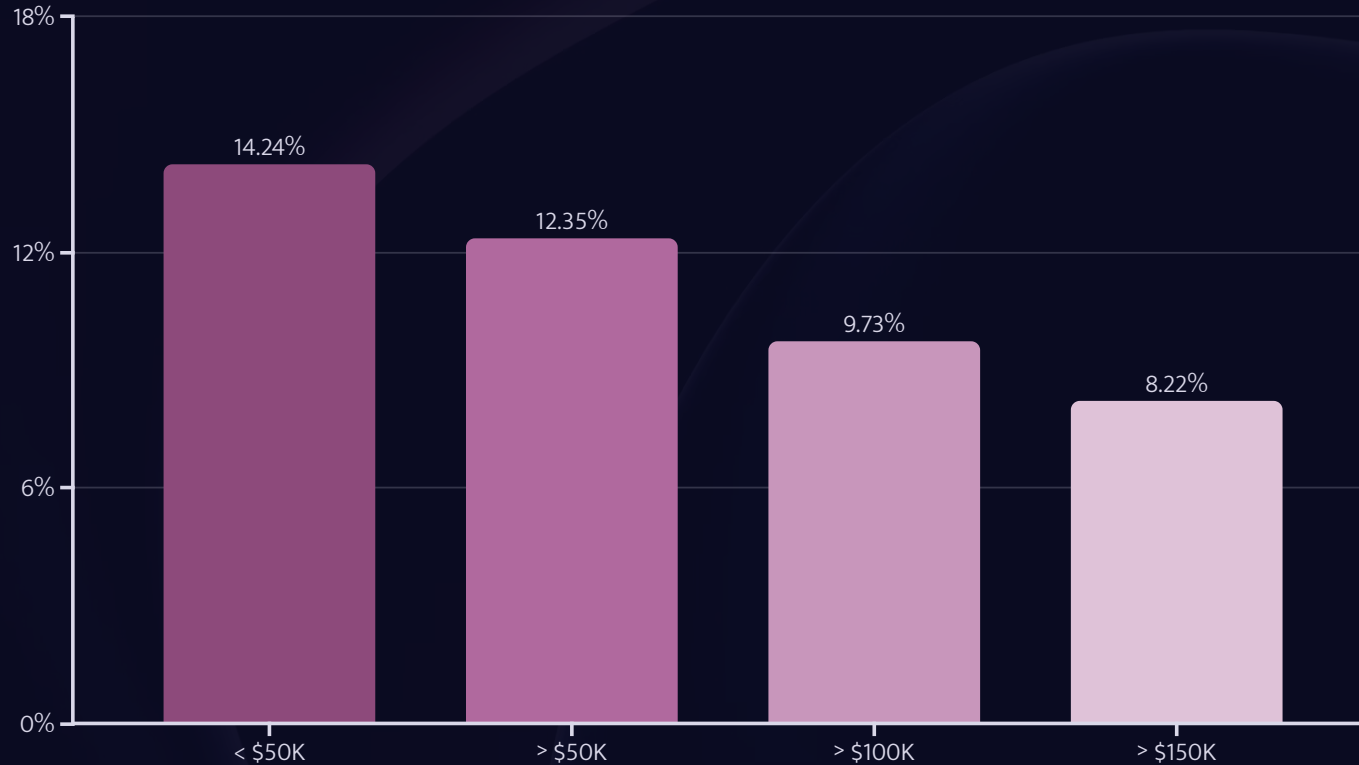
Car Financing

Lower Default Rate: ~9%

Typically secured, and vehicles are essential for daily life.

Income Group vs. Default Rate

Income level plays a role in a borrower's ability to repay, though not always as straightforward as expected due to other factors.



Observation: Generally, higher income bands correlate with lower default rates, as expected, reinforcing financial stability as a key factor.

US State-wise Default Risk Map

Geographic location can reveal regional economic and demographic factors influencing default risk.

Top Risky States:

Arkansas (AR)

Alabama (AL)

South Dakota (SD)

Lowest Risk States:

Maine (ME)

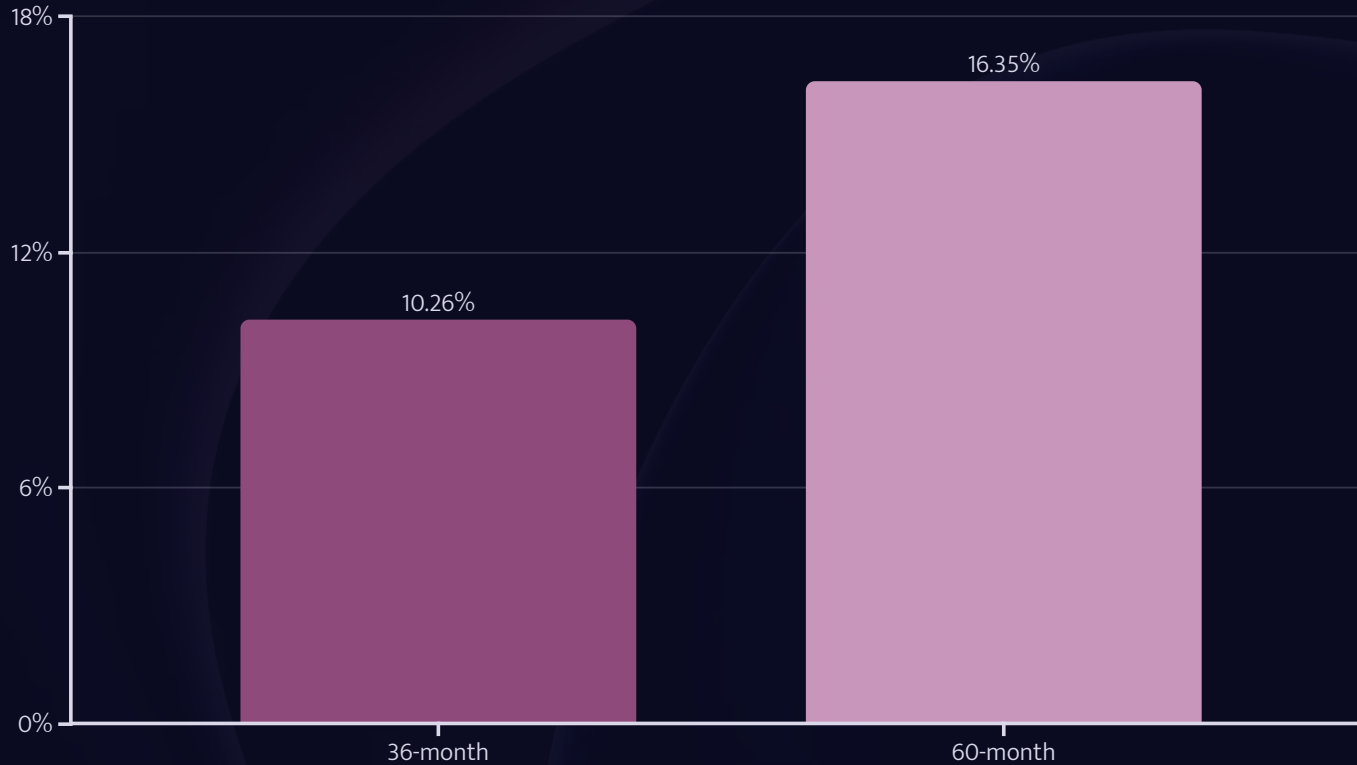
Vermont (VT)

Idaho (ID)

- ⊗ **Insight:** States with lower average incomes or higher unemployment rates tend to show elevated default risks.

Loan Term vs. Default Rate

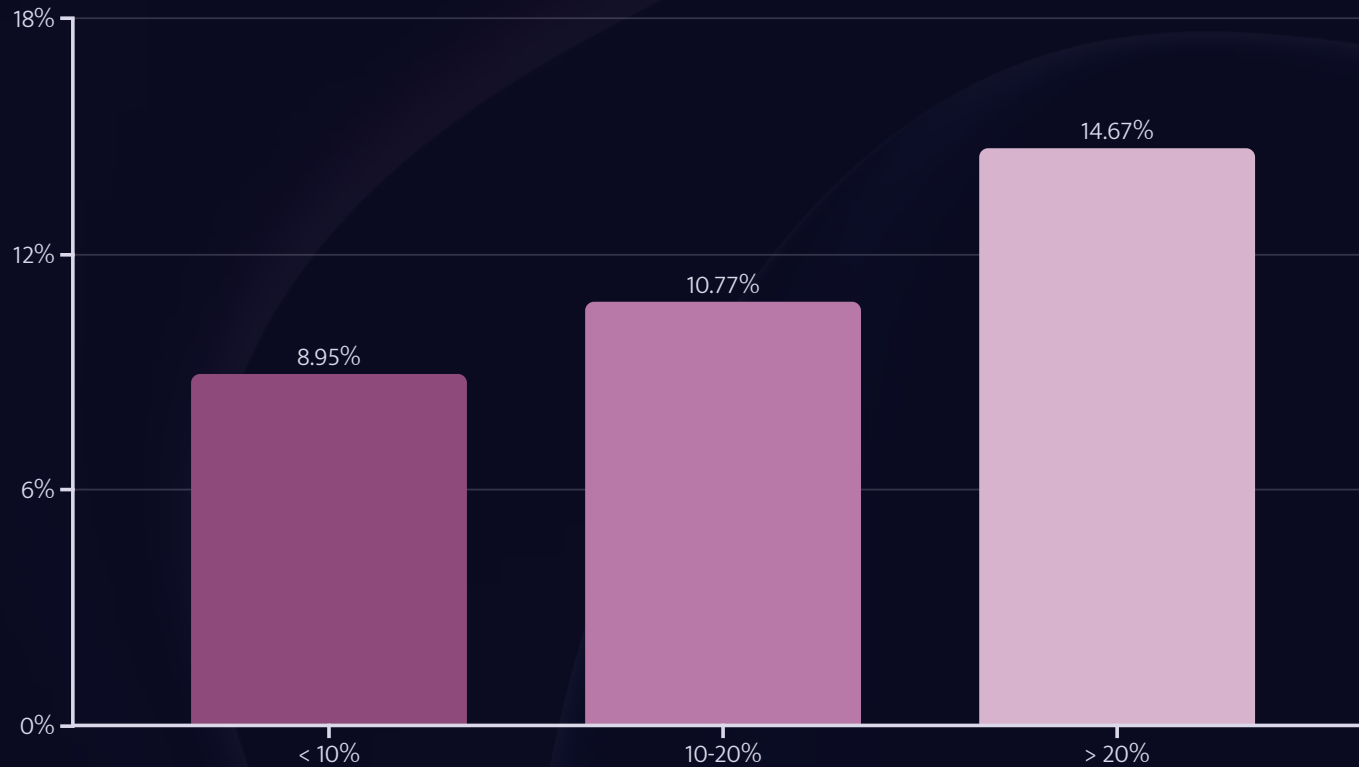
The duration of a loan directly impacts its associated risk. Longer terms generally increase exposure to unforeseen financial events.



⊗ **Observation:** 60-month term loans exhibit significantly higher default rates than 36-month terms, nearly doubling the risk.

DTI Band vs. Default Rate

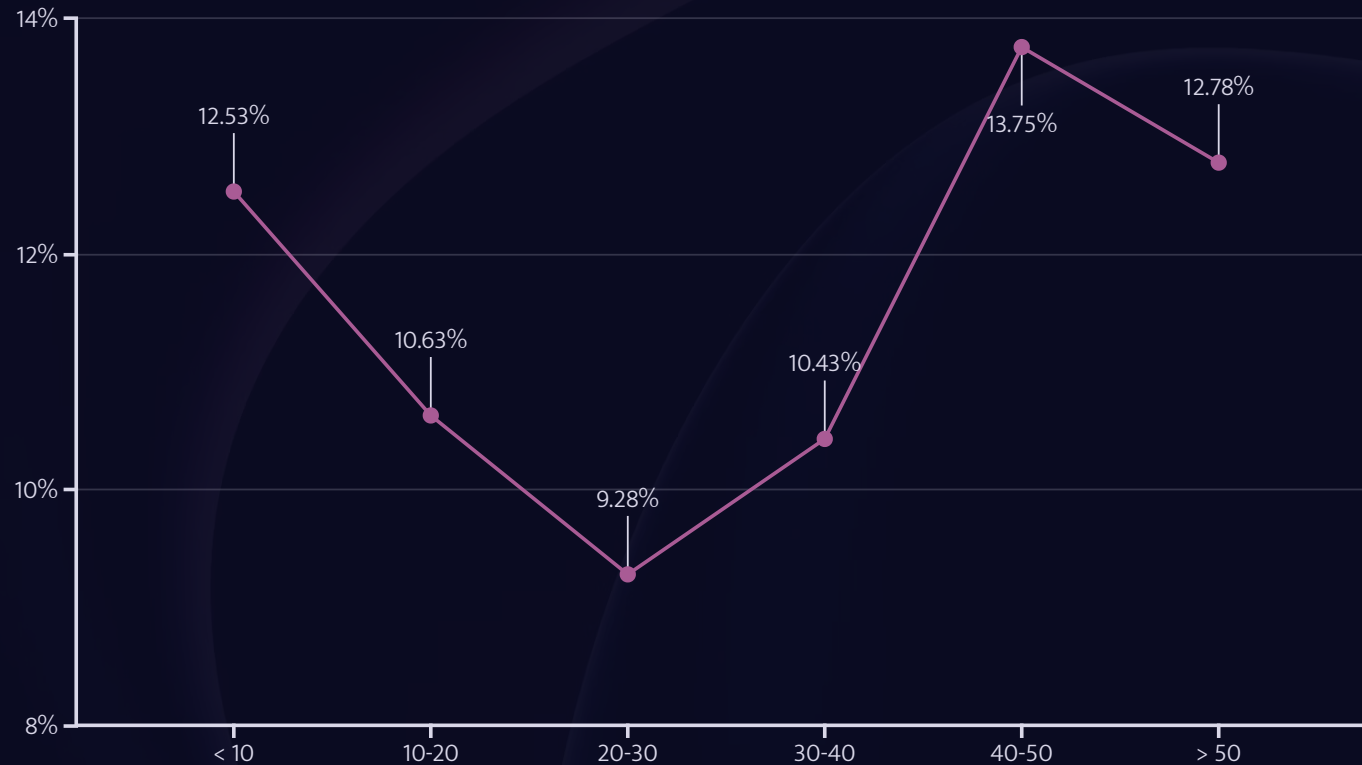
Debt-to-Income (DTI) ratio is a crucial metric, indicating a borrower's capacity to manage debt payments.



⊗ **Observation:** Higher DTI bands correlate with increased default rates, with DTI over 30% showing a clear elevation in risk.

Credit History Age vs. Default Rate

The length of a borrower's credit history often reflects financial maturity and stability, influencing default risk.



✔ **Insight:** Borrowers with 20–40 years of credit history show the lowest default rates, emphasizing the reduced credit risk associated with long-term financial stability.

Key Recommendations

Actionable Insights for Underwriting

Based on the analysis, here are key recommendations for refining loan underwriting processes to mitigate default risk.

Risk-Adjusted Lending

Implement dynamic interest rates and loan terms strictly based on loan grade and DTI.

Purpose-Specific Vetting

Apply stricter criteria for "Small Business" and "Debt Consolidation" loans, considering their higher default rates.

Geographic Risk Scoring

Integrate state-level economic indicators into risk models, with special attention to high-risk states identified.

Credit History Emphasis

Prioritize applicants with longer, established credit histories, indicating proven financial responsibility.

Skills & Contact

My SQL & Data Analytics Skills



Advanced SQL Querying

Complex joins, subqueries, window functions.



Statistical Analysis

Identifying correlations, trends, and anomalies.



Business Acumen

Connecting data insights to practical business recommendations.



Data Visualization

Translating complex data into clear, impactful visuals.



ETL Processes

Extracting, transforming, and loading data for analysis.



Version Control (Git)

Managing code and collaborating on projects effectively.

Let's Connect!



Abhishek Verma

Data Analyst | SQL Enthusiast

✉ abhiv0040@gmail.com

Explore the Project

🔗 GitHub: github.com/abhiv0040/loan-risk-analysis

👤 LinkedIn: linkedin.com/in/averma2025

"Data is not just about numbers; it's about telling a compelling story and solving real-world problems."