



SQL Insight Questionnaire

Dataset Overview

 How many total records and unique loan applications are there? Are there any missing values in key columns like Gender, Married, Self_Employed, Dependents, and Credit_History?


Query: SELECT COUNT(*) AS total_records, COUNT(DISTINCT Loan_ID) AS unique_loan_ids FROM loan_data;

Overall Business Health

 What is the overall loan approval rate in the dataset?


Query: SELECT Loan_Status, COUNT(*) AS total, ROUND(COUNT(*) * 100.0 / (SELECT COUNT(*) FROM loan_data), 2) AS percentage FROM loan_data GROUP BY Loan_Status;

Demographic Analysis By Gender:

 Loan approval rate for males vs females.


Query: SELECT Gender, Loan_Status, COUNT(*) AS total FROM loan_data GROUP BY Gender, Loan_Status;

By Marital Status:

 Do married applicants get loans more frequently?

Query: SELECT Married, Loan_Status, COUNT(*) AS total FROM loan_data GROUP BY Married, Loan_Status;

By Education:

 Do graduates have higher approval rates?


Query: SELECT Education, Loan_Status, COUNT(*) AS total FROM loan_data GROUP BY Education, Loan_Status;

By Self Employment:

 Are self-employed applicants less likely to get loans?

Query: SELECT Self_Employed, COUNT(*) AS total_applicants, SUM(CASE WHEN Loan_Status = 'Y' THEN 1 ELSE 0 END) AS approved_loans, ROUND(SUM(CASE WHEN Loan_Status = 'Y' THEN 1 ELSE 0 END) * 100.0 / COUNT(*), 2) AS approval_percentage FROM loan_data WHERE Self_Employed IN ('Yes', 'No') GROUP BY Self_Employed;

Property Analysis

 Which areas (Urban/Rural/Semiurban) see the most loan approvals?

Query: SELECT Property_Area, Loan_Status, COUNT(*) AS total FROM loan_data GROUP BY Property_Area, Loan_Status;

Loan & Income Analysis By Education and LoanAmount:

 Does education affect average loan amount?


Query: SELECT Education, ROUND(AVG(LoanAmount), 2) AS avg_loan_amount FROM loan_data GROUP BY Education;

By Income Group:

 Do higher-income applicants get loans more easily?

Query: SELECT CASE WHEN ApplicantIncome < 2500 THEN 'Low Income' WHEN ApplicantIncome BETWEEN 2500 AND 5000 THEN 'Middle Income' WHEN ApplicantIncome BETWEEN 5001 AND 10000 THEN 'Upper-Middle Income' ELSE 'High Income' END AS income_group, Loan_Status, COUNT(*) AS total FROM loan_data GROUP BY income_group, Loan_Status;

Coapplicant Income:

 Does average coapplicant income vary between approvals and rejections?

Query: SELECT Loan_Status, ROUND(AVG(CoapplicantIncome), 2) AS avg_coapplicant_income FROM loan_data GROUP BY Loan_Status;