TITLE: Credit Risk Analytics: Identifying Drivers of Loan Default

Summer Internship 2025
Role: Data Analyst Intern
Organization: Quant Lake Lending

Project Title:

Credit Risk Analytics: Identifying Drivers of Loan Default

Objective:

To analyze demographic and financial attributes of loan applicants and identify the key factors that influence loan default risk. The project aims to assist QuantLake Lending in understanding which customer segments are most at risk and provide actionable recommendations for reducing loan default.

Methodology:

- Cleaned and pre-processed the raw Kaggle data for accuracy, handled missing values, and engineered features for analysis.
- Used Excel/Python for exploratory analysis and Tableau to build interactive dashboards to visualize key risk factors

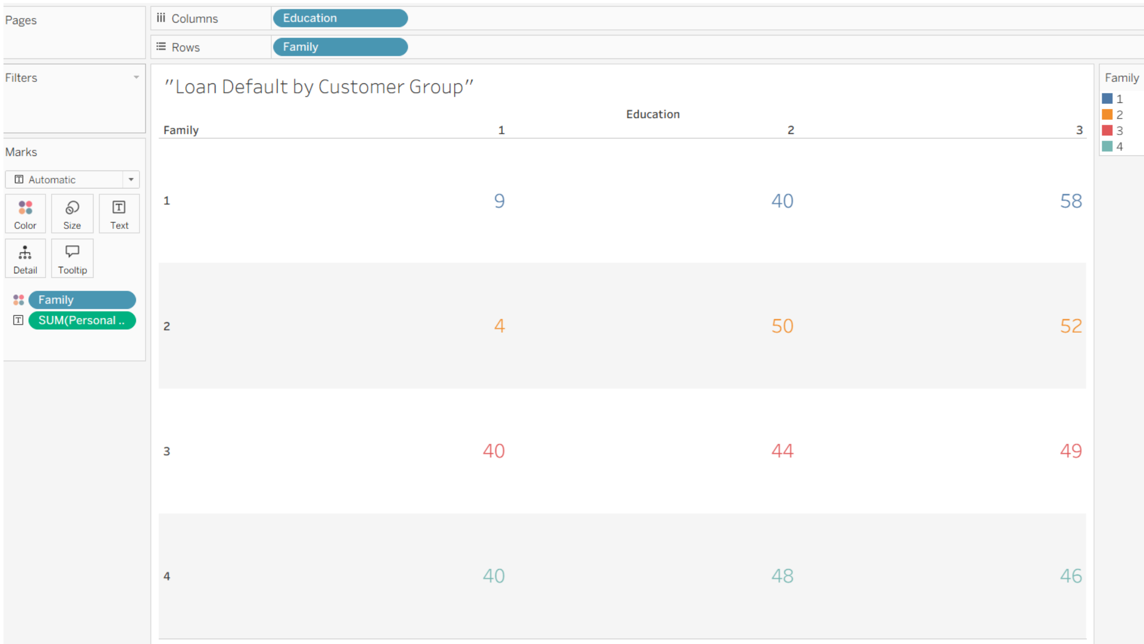
Dataset Description:

Source: Kaggle Dataset by itsmesunil - Bank Loan Modelling
Attributes include:
- Age
- Income
- Family size
- Education level
- Online banking usage
- Credit card ownership
- Securities and CD account

- Mortgage amount
- Personal loan (target variable)

Dashboard Visualizations and Interpretations:

1. Loan Default by Customer Group:
 Shows how default rates vary across family size and education level. Higher default rates were observed among customers with lower education and larger families.



Family	Education 1	Education 2	Education 3
1	9	40	58
2	4	50	52
3	40	44	49
4	40	48	46

This chart highlights that loan default rates are higher among individuals with lower education levels and larger family sizes. "This suggests that both educational background and family size significantly influence repayment ability".

2. Loan Applications: Online vs Offline:

The bar chart compares online and offline loan applicants. Online users applied more frequently and showed slightly lower default rates. This indicates that online applicants might be more financially aware or digitally literate, and that promoting online loan platforms could help attract lower-risk borrowers.

3. Income Distribution by Age:

This visualization shows how income varies across age groups. Average income peaks between ages 40 to 50, suggesting that mid-aged individuals have stronger earning power. Younger and older customers tend to have lower income, which may correspond to higher loan default risk and should be evaluated accordingly.

Key Insights:

1. Education level significantly impacts loan default – Lower-educated individuals are more likely to default.
2. Online applicants tend to be more reliable – Slightly lower default rates.
3. Middle-aged customers (40-50) have the highest average income and lowest default probability.
4. Larger families correlate with higher default risk, possibly due to higher financial burden.
5. Income is a strong predictor – Lower income brackets show higher default likelihood.

Recommendations:

- Focus lending offers toward customers aged 35–50 with higher education and income.
- Use online platforms as the preferred loan channel.
- Set stricter screening for applicants with large families or low income.
- Incorporate education and income scorecards into risk models.

Tools Used:

Tableau for dashboard and visual analysis
Excel/Python for preprocessing and feature exploration.

Prepared By: Shaik Esmail
Gmail: shaikesmail998@gmail.com
Intern ID: IN-25-1005