Loan Default Analysis

1. Project Overview

This project focuses on analyzing loan default behavior using customer demographic, financial, and loan-related data. The objective is to identify high-risk customer segments, understand key drivers of loan defaults, and support data-driven credit risk decisions for financial institutions.

The analysis leverages historical loan data to uncover relationships between default rates and factors such as age group, income level, credit score, employment stability, loan amount, interest rate, and debt-to-income (DTI) ratio. The insights are presented through an interactive Power BI dashboard to enable stakeholders to monitor risk patterns and take preventive actions.

---

2. Dataset Summary

- **Total Customers:** ~30,000
- **Total Defaults:** ~30,000
- **Default Rate:** ~11.6%
- **Key Features:**
    - Customer demographics (Age Group, Income Group)
    - Financial indicators (Credit Score, DTI Level)
    - Employment attributes (Employment Stability)
    - Loan characteristics (Loan Amount, Interest Rate, Loan Purpose)
    - Target variable (Default Flag)
- **Data Type:** Structured tabular dataset suitable for risk segmentation and predictive analysis

---

3. Exploratory Data Analysis using Python

Data preparation and exploratory analysis were performed in Python to ensure data quality and readiness for downstream analysis.

Key Steps

- **Data Loading:** Imported the dataset using Pandas.
- **Data Validation:** Checked data types, value ranges, and consistency.
- **Missing Value Treatment:** Identified and handled null or inconsistent values.
- **Univariate Analysis:** Analyzed distributions of income groups, age groups, credit scores, and loan amounts.
- **Bivariate Analysis:** Examined default rate variations across key dimensions such as income, age, credit score, and employment stability.
- **Feature Transformation:** Prepared clean and analysis-ready data for SQL and Power BI.

---

4. Feature Engineering

To support meaningful segmentation and business analysis, additional features were created:

- **Age Groups:** 18–25, 26–35, 36–45, 46–55, 56–65, 65+
- **Income Groups:** Low, Lower-Mid, Upper-Mid, High
- **Credit Score Categories:** Poor, Fair, Good, Excellent
- **Employment Stability Groups:** New, Early Career, Stable, Very Stable
- **DTI Buckets:** Low, Moderate, High, Very High
- **Loan Amount Buckets:** Small, Medium, Large, Very Large

---

5. Data Analysis using SQL (Business Transactions)

Structured SQL queries were executed in PostgreSQL to answer key business questions:

1. **Default Rate by Income Group** – Identified significantly higher default rates among low-income customers.
2. **Default Rate by Age Group** – Younger age groups (18–25, 26–35) showed elevated risk.

3. **Credit Score vs Default** – Clear inverse relationship between credit score and default rate.
4. **Employment Stability vs Default** – New and early-career customers exhibited higher default risk.
5. **Loan Amount vs Default** – Larger loan amounts showed comparatively higher default rates.
6. **Interest Rate vs Default** – Higher interest rates correlated with increased defaults.
7. **DTI Level vs Default** – Customers with high DTI ratios were more prone to default.

---

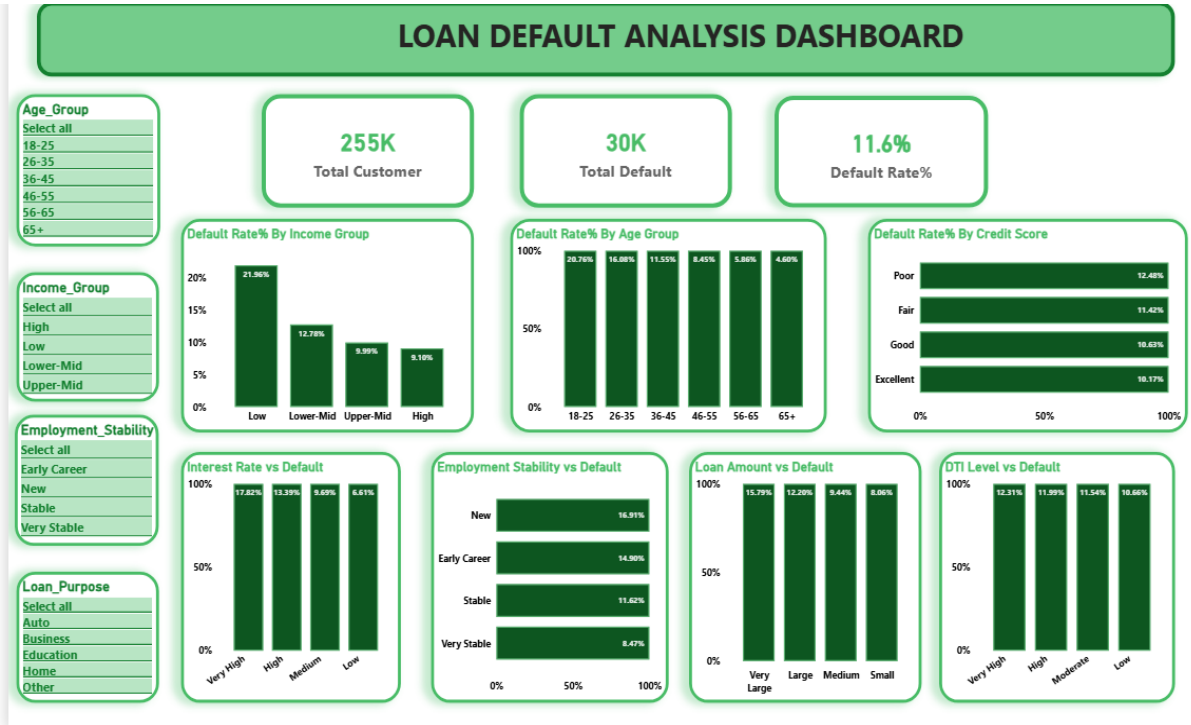
6. Power BI Dashboard

An interactive **Loan Default Analysis Dashboard** was developed to visualize insights and enable dynamic exploration.

Dashboard Highlights

- KPI cards showing **Total Customers**, **Total Defaults**, and **Default Rate**
- Default rate breakdowns by:
    - Age Group
    - Income Group
    - Credit Score
    - Employment Stability
    - Loan Amount
    - Interest Rate
    - DTI Level
- Interactive slicers for:
    - Age Group
    - Income Group
    - Employment Stability
    - Loan Purpose

This dashboard allows stakeholders to quickly identify high-risk segments and drill down into specific customer profiles.



---

7. Key Insights & Business Recommendations

Key Insights

- Low-income and early-career customers have the highest default rates.
- Credit score is the strongest predictor of default behavior.
- High DTI levels significantly increase default probability.
- Larger loan amounts and higher interest rates carry higher risk.

Business Recommendations

- **Strengthen Credit Screening:** Apply stricter checks for low credit score and high DTI applicants.
- **Risk-Based Pricing:** Adjust interest rates based on customer risk profiles.
- **Targeted Monitoring:** Closely monitor early-career and low-income segments.
- **Loan Amount Controls:** Introduce tighter limits for high-risk borrowers.
- **Portfolio Optimization:** Use dashboard insights to continuously track and manage credit risk.
