Loan Default Analysis Project

Data Analyst: Agustin A. Zavagnini

Purpose:

This project aims to analyze borrower demographics, financial status, credit history, and loan characteristics to understand the factors that influence loan repayment performance. Through comprehensive data cleaning, exploratory data analysis (EDA), and visual insights, the goal is to identify patterns associated with successful loan repayment versus default.

A secondary objective is to build a **baseline predictive model** to estimate the probability of loan default. The intention is *not* to perform hyperparameter tuning or pursue advanced optimization techniques, but rather to showcase the ability to implement, interpret, and evaluate a simple model within a data analysis workflow.

Scope / Major Project Activities:

Activity	Description
1. Data Understanding	Examine dataset structure, variable types, summary statistics, missing values, inconsistencies, and distributions across demographic, financial, loan, and credit history features. Prepare a clean data dictionary.
2. Data Cleaning	Standardize categories, handle nulls, convert data types, treat outliers (e.g., extreme incomes or credit limits), and ensure dataset consistency before analysis.
3. Exploratory Data Analysis (EDA)	Conduct univariate, bivariate, and multivariate analysis to explore relationships between borrower behavior, risk indicators, and repayment outcomes.
4. Risk & Borrower Behavior Insights	Identify characteristics strongly associated with default risk, such as high DTI, low credit score, unstable employment, high balances, or extensive delinquency history.
5. Baseline Predictive Modeling	Implement a simple predictive model (e.g., logistic regression, decision tree, or random forest baseline) to estimate loan default probability. Focus on model interpretability, metrics, and insights — *not* on hyperparameter optimization or model tuning.

6. Recommendations & Conclusions	Provide data-driven conclusions to improve risk assessment, borrower profiling, and loan approval strategies. Highlight high-impact variables based on the analysis and the model.
7. Visualization Assets	Create data visualizations including distributions, correlations, risk segmentation, and repayment patterns. Optionally develop a Tableau dashboard summarizing key insights.

This project does not include:

- Complex or highly optimized predictive models.
- Hyperparameter tuning or model performance optimization.
- Third-party data sources or external demographic/credit enrichment.
- Regulatory or compliance-focused analysis.

Deliverables:

Deliverable	Description / Details
2. Jupyter Notebook (Complete Workflow)	Notebook including data understanding, cleaning, EDA, baseline modeling, and interpretation of results.
4. Tableau Dashboard	Dashboard highlighting borrower segmentation, risk patterns, and repayment trends.