Loan Default Risk & Portfolio Analytics

Project Author: Thulaganyo Outule

Objective

The objective of this project was to analyze loan default risks within a lending portfolio using historical data, and to develop an interactive dashboard that provides insights into portfolio distribution, risk exposure, borrower characteristics, and potential areas of concern for financial institutions.

Aim

The aim of the project was to:

•Identify default risk patterns across different demographic and financial groups.
•Build a fraud/risk-focused loan analysis framework with reproducible Python workflows.
•Present findings through an interactive Looker Studio dashboard for clear decision-making.

Project Components

1. Jupyter Notebook – End-to-end data pipeline for cleaning, wrangling, analysis, and visualization (Python).
2. Raw Dataset – Original loan data in `.csv` format.
3. Clean Dataset – Preprocessed and analysis-ready dataset in `.csv` format.
4. Looker Studio Dashboard – Final interactive dashboard summarizing portfolio risk insights.

Methodology

1. Data Preprocessing

•Imported raw data (`.csv` format).
•Cleaned null values, outliers, and inconsistent data formats.
•Performed feature engineering (e.g., age grouping, risk banding).
•Standardized numeric fields (loan amount, interest rate, probability of default).

2. Data Wrangling

•Aggregated metrics (average interest rates, expected losses).
•Created categorical variables for risk bands and age groups.
•Encoded gender and product types for analysis.

3. Data Visualization & Dashboard

•Built multiple visualizations in Looker Studio.
•Integrated filters/slicers for employment and gender.
•Added scorecards for KPIs (Total Loans, Average Interest Rate, Default Probability, Expected Loss).

Dashboard Components & Insights

Scorecards (KPIs)

•Total Loan Amount – Portfolio exposure in monetary terms.
•Average Interest Rate – Overall lending cost.
•Portfolio Default Probability – Predicted average risk.
•Expected Loss – Key risk-adjusted performance measure.

Visualizations

1. Portfolio Risk Distribution (Pie Chart)

- Shows the distribution of loans across risk bands (low, medium, high).
- Helps stakeholders see portfolio concentration and exposure to high-risk loans.

2. Expected Loss by Age Group (Line Chart)

- Plots average expected loss against borrower age groups.
- Identifies which age ranges contribute most to portfolio risk.

3. Default Probability Distribution (Bubble Chart)

- Each bubble represents a cluster of loans with a probability of default.
- Bubble size reflects loan amount.
- Useful for spotting high-risk, high-value borrowers.

4. Borrower Profile Table (Table)

- Displays Age, Loan Amount, Interest Rate, Expected Loss.
- Allows for granular borrower-level risk assessment.

5. Gender vs Interest Rate (Pie Chart)

- Compares average interest rate distribution across genders.
- Useful for checking fairness, bias, or portfolio skewness.

6. Product Penetration vs Predicted Default Probability (Bar Chart)

● Shows which loan products have higher penetration and how they align with predicted defaults.
● Helps financial institutions adjust product strategies.

Outcomes & Responsibilities (What I Did)

•Conducted data cleaning, wrangling, and preprocessing of raw loan data.
•Performed exploratory data analysis (EDA) using Python.
•Designed new features (risk band, age groupings) for better segmentation.
•Built multiple visualizations in Looker Studio to communicate insights.
•Delivered a dashboard with KPIs, filters, and interactive insights for decision-making.
•Packaged final outputs in multiple formats:

* Raw dataset: `.csv`
* Clean dataset: `.csv`
* Notebook: `.ipynb`
* Dashboard: Looker Studio share link

Relevance to Business Intelligence & Data Analysis

This project highlights key Business Intelligence (BI) and Data Analysis processes:

1. Data Processing & Structuring

•Raw loan and customer datasets were cleaned, standardized, and structured for analysis.
•Missing values, outliers, and inconsistencies were handled to ensure data quality, a core BI practice.

2. Feature Engineering & Modelling

•New features such as risk bands, age groups, debt-to-income ratios, and product penetration were created to enhance insights.
•Predictive models (Logistic Regression, Random Forest, Gradient Boosting) were applied to estimate default probabilities, demonstrating the analytics and modeling component of BI.

3. Visualization & Dashboarding

•Visualizations (pie charts, line charts, bubble charts, tables, and bar charts) were designed to communicate insights effectively to stakeholders.
•Interactive filters and scorecards allow dynamic exploration, aligning with BI objectives of actionable, data-driven decision-making.

4. Insight Generation & Business Impact

- The dashboard provides high-level KPIs (expected loss, default probability) and detailed borrower-level insights.
- Organizations can leverage these insights to mitigate risks, adjust product strategies, and optimize lending portfolios, reflecting the strategic value of BI in real-world business contexts.

5. End-to-End Workflow

- The combination of data cleaning, wrangling, modeling, and dashboarding demonstrates a full BI lifecycle from raw data ingestion to actionable business insights.
- This structured approach mirrors real-world BI and Data Analytics projects, where data-driven decision-making is paramount.

Conclusion

This project demonstrates how data analytics can uncover risk drivers in loan portfolios. By combining Python data wrangling with Looker Studio visualization, the workflow provides a scalable framework for risk assessment, expected loss estimation, and portfolio monitoring.