

Bank Loan Case Study – Final Project Report

1

Project Description

This project involves analyzing bank loan data to identify patterns that predict loan defaults.

As a data analyst at a financial services company, your goal is to minimize the risks of approving loans for customers.

You've performed a complete Exploratory Data Analysis (EDA) using Excel.

2

Approach

The project was divided into five core tasks:

- Task A: Missing Value Analysis
- Task B: Outlier Detection
- Task C: Data Imbalance Assessment
- Task D: Univariate, Segmented Univariate, and Bivariate Analysis
- Task E: Correlation Analysis by Scenario

Each task was conducted using Excel functions, pivot tables, statistical formulas, and visualizations.

3

Tech Stack Used

- Microsoft Excel 2021
- Functions: COUNTBLANK, IF, CORREL, QUARTILE, COUNTIF, etc.
- Tools: Pivot Tables, Charts, Conditional Formatting, Filters

4

Key Insights

- Missing values were mostly found in columns like OCCUPATION_TYPE and EXT_SOURCE_1, which were imputed.
- Outliers were identified in income, annuity, and credit columns using the IQR method.
- TARGET was highly imbalanced (~92% non-defaulters), requiring careful treatment in predictive modeling.
- Applicants with low external scores and short employment history showed higher default risk.
- Approved vs refused applications in past loans correlated with certain loan purposes and channels.

5

Final Result