CREDIT RISK ANALYTICS Using Python

This presentation outlines the findings of a comprehensive credit risk analytics project conducted for Bondora Bank. The project aimed to address the challenges faced by the bank in assessing credit risk and to develop strategies for improving loan approval processes and minimizing financial losses.

Evoastra Ventutres Major Project

Internship Presentation

By Team A

Team Leads

Sachin Sharma
Team Lead

Nidhi Gummaraju
Team Lead

Muhammad Umar
Co - Team Lead

Santosh Shinde
Co - Team Lead

Team Members

Mohit	Syed Musa	Solanki
Akansha	Srinivas	Dixit
Akshaya	Sahil	Hannah
Pranay	Rikesh	Mahesh
Vignesh	Sandeep	Raval

Content

INTRODUCTION AND OBJECTIVE

DATA COLLECTION

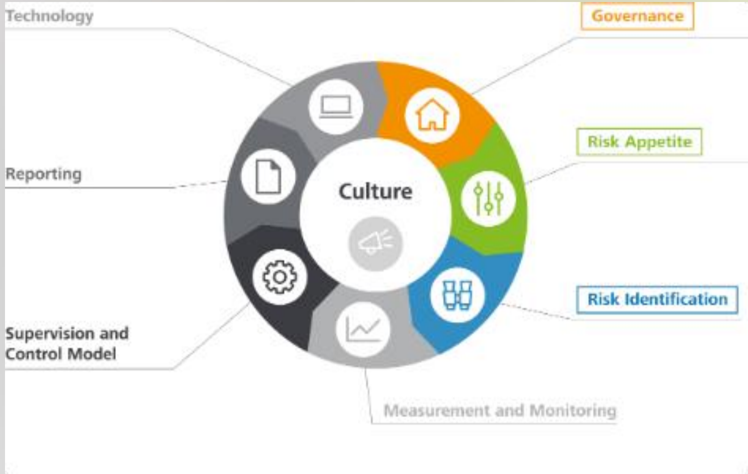
DATA PROCESSING AND CLEANING

DATA ANALYSIS AND INSIGHTS

CHALLENGES AND CONCLUSION

Objectives

Our objectives are to collect and analyze credit risk data from Bondora Bank, ensuring data accuracy through cleaning and processing, and providing valuable insights to support informed lending decisions and minimize financial risks.



Problem Statement

Loan Default Rates

Bondora Bank faces significant challenges in managing loan default rates, which impact profitability and overall financial stability.

Accurate Risk Assessment

The bank needs to improve its ability to accurately assess the creditworthiness of loan applicants to make informed lending decisions.

Fraud Detection

Preventing fraudulent loan applications is crucial to protect the bank's assets and maintain customer trust.

Research and Planning

1. Analyzing Data
2. Data Cleaning Strategies
3. Anticipate Challenges

Data Understanding and Exploration

Data Sources

The project utilized three primary datasets: application_data.csv, previous_application.csv, and columns_description.csv. These datasets contain information about loan applications, previous loan history, and detailed descriptions of each data field.

Data Cleaning and Preparation

The data underwent thorough cleaning and preparation, addressing missing values, outliers, and inconsistencies. This ensured data quality and reliability for subsequent analysis.

Exploratory Data Analysis (EDA)

EDA involved univariate, bivariate, and segmented analyses to uncover patterns and relationships within the data. Visualizations such as histograms, scatter plots, and box plots were used to gain insights into key variables.

Data Cleaning and Organization

Eliminate Duplicates

Implement deduplication algorithms to ensure the dataset is free from redundant information.

Handle Missing Data

Fill in gaps or replace missing values using appropriate techniques, such as imputation or interpolation.

Standardize Data Format

Normalize the data structure and ensure consistent formatting across different data sources.

Organize for Analysis

Store the cleaned and structured data in a format suitable for subsequent data analysis and visualization.

Key Findings and Insights

1. **Applicant Demographics**

   The analysis revealed a correlation between certain demographic factors, such as age, education level, and employment status, and loan default risk.

2. **Financial Information**

   Income, credit score, debt-to-income ratio, and loan amount were found to have a significant impact on the likelihood of loan default.

3. **Loan Characteristics**

   Loan type, term, interest rate, and purpose were also identified as influential factors in repayment behavior.

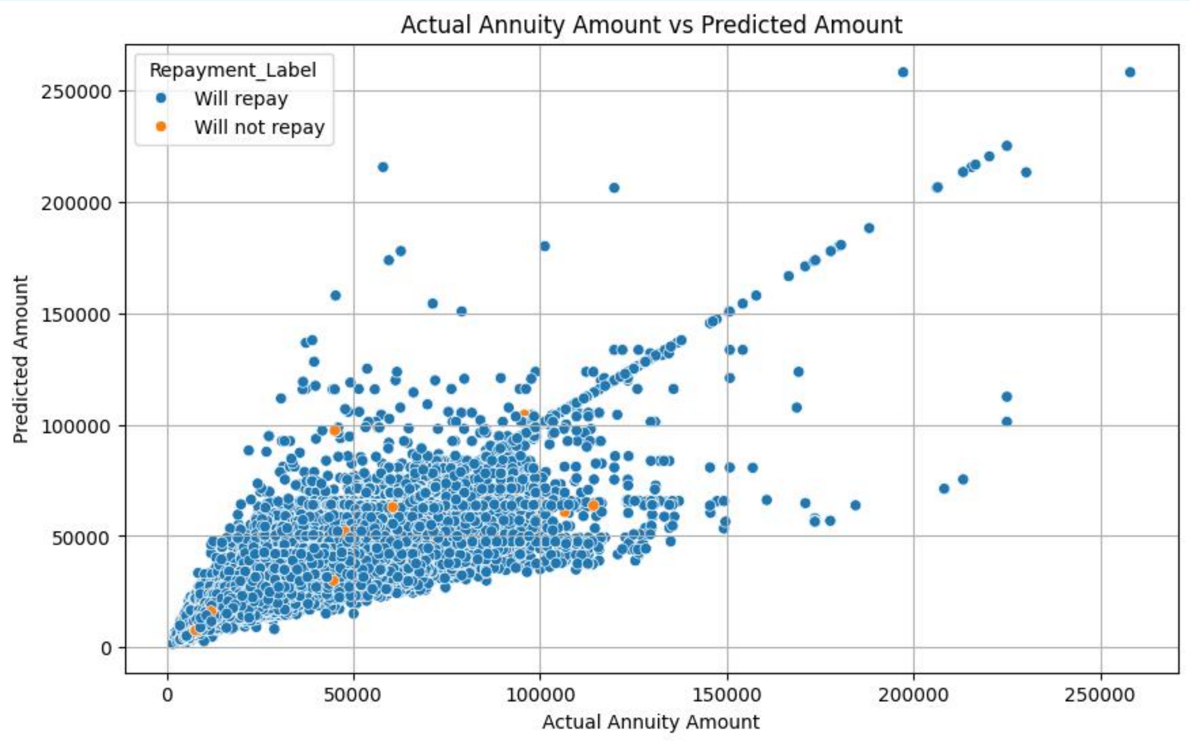
4. **Previous Loan Performance**

   A strong correlation was observed between past loan performance and future default risk, highlighting the importance of considering historical data.

Data Analysis 01

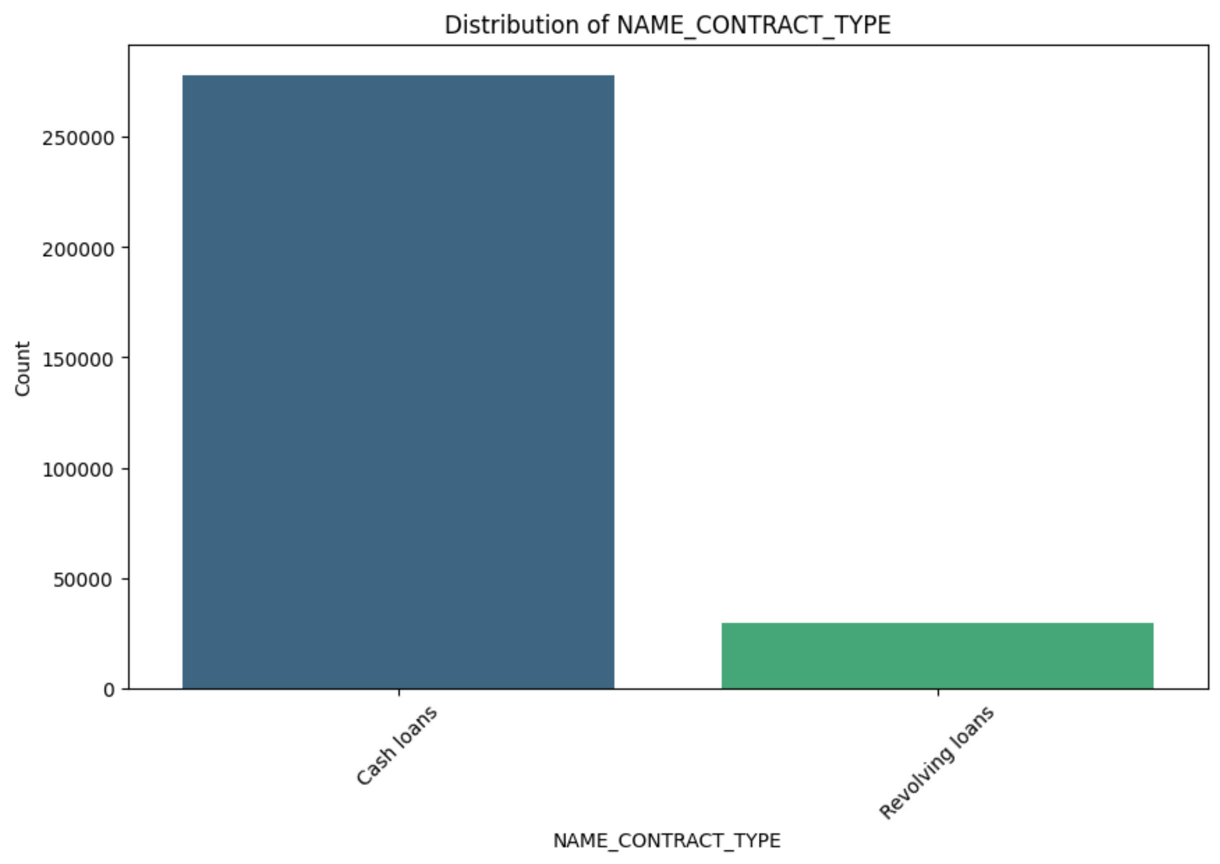
Bar Chart of Actual Annuity Amount:

- The scatter plot provides valuable insights into the relationship between actual and predicted annuity amounts. While the model shows a general trend of predicting higher amounts for individuals with higher actual annuities, there is room for improvement in terms of accuracy and addressing outliers.
- **Description:** This scatter plot visualizes the relationship between actual and predicted annuity amounts. The blue dots represent individuals who repaid their loans, while orange dots indicate those who did not.



Data Analysis 02

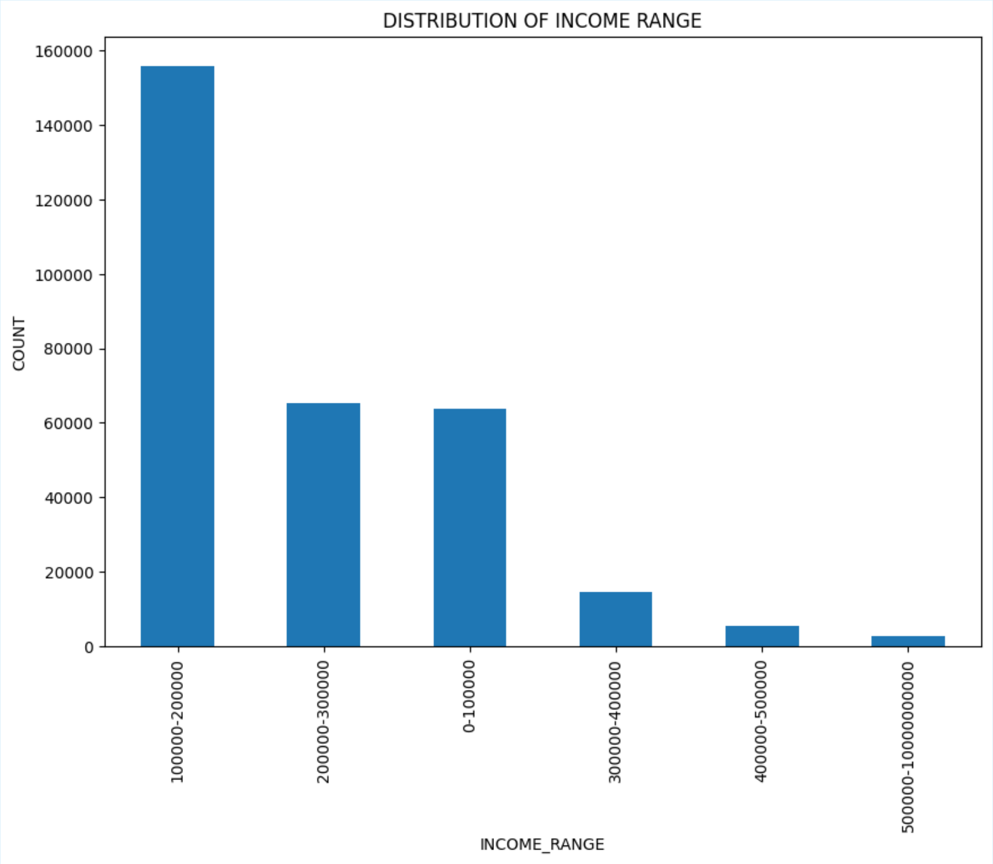
Loan type



Description: This bar chart illustrates the distribution of contract types within the dataset. The x-axis represents the different contract types, while the y-axis indicates the count of occurrences for each type.

Data Analysis 03

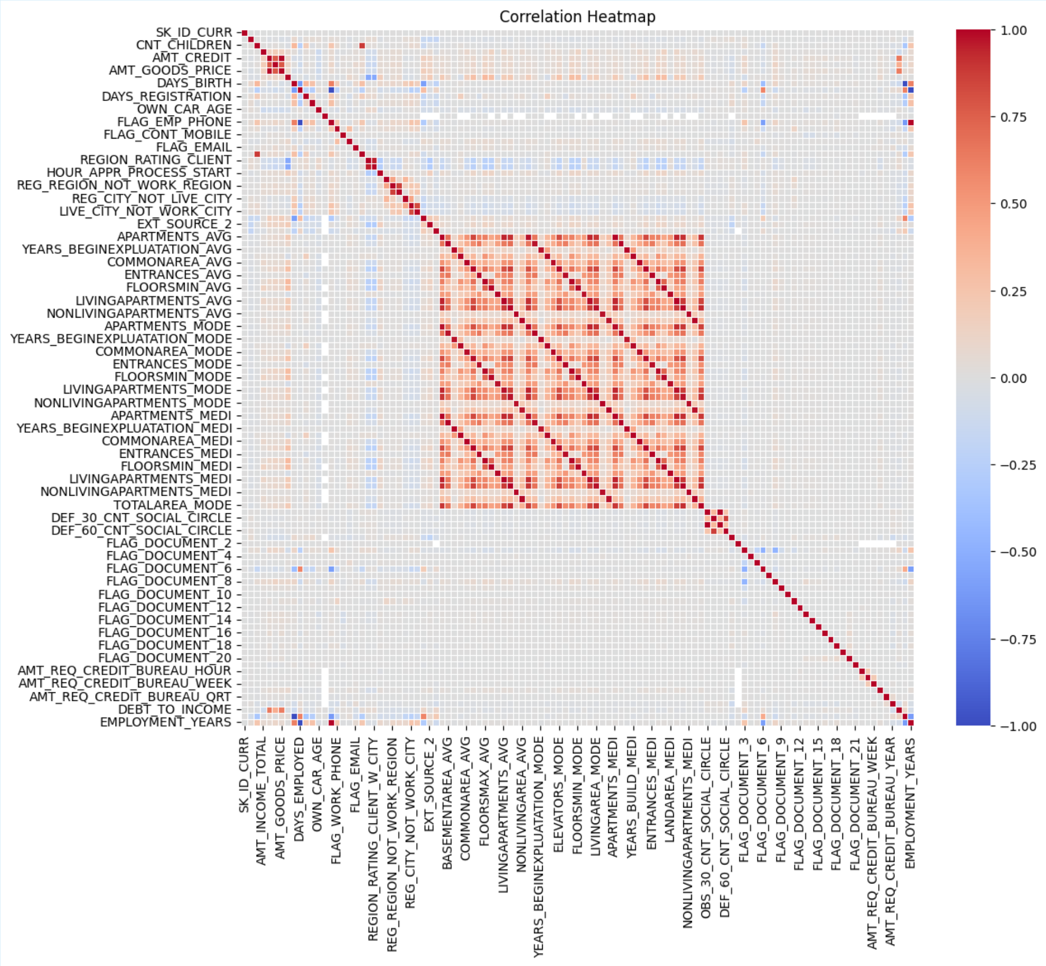
Income range



The bar chart illustrates the distribution of income across various ranges, with the majority of individuals falling within the **100,000-200,000** income range. This range has the highest count, approximately **150,000**. Other significant income groups include 200,000-300,000 and 0-100,000, both with counts around 60,000. As income increases, the frequency decreases, with much smaller counts observed in the 300,000-400,000 range and very few individuals in the higher income ranges of 400,000-500,000 and 500,000-1,000,000,000**.**

Data Analysis 04

Correlation heatmap



This heatmap visually highlights the relationships between different variables in the dataset, with stronger correlations (positive or negative) shown in red and blue. The information can be useful in identifying which variables are closely related, which can assist in feature selection or further analysis.

Modeling and Evaluation

1 — Model Selection

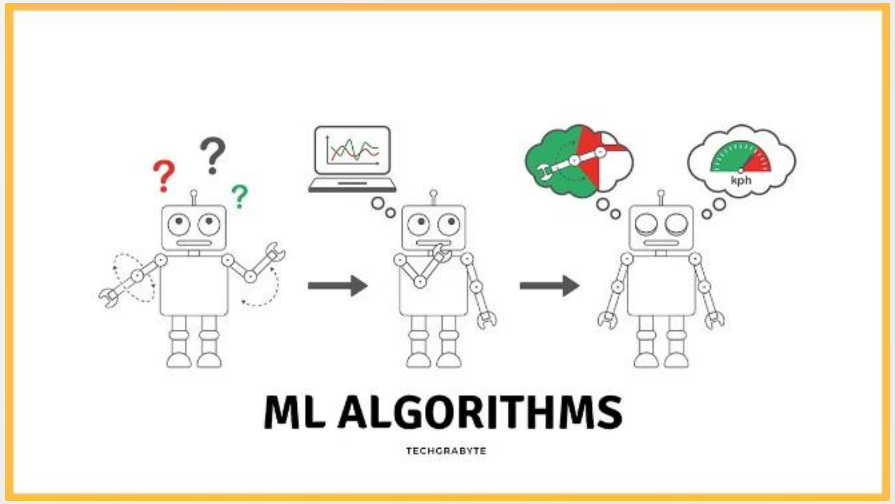
The project explored various statistical models and machine learning algorithms, including logistic regression, decision trees, and random forests, to predict loan default probability.

2 — Feature Engineering

New features were engineered from existing data to enhance model performance. These features captured complex relationships and provided additional insights into credit risk.

3 — Model Training and Evaluation

The selected models were trained and evaluated using metrics such as accuracy, precision, recall, and F1-score. The models demonstrated strong predictive capabilities, achieving high levels of accuracy in identifying potential defaulters.



Recommendations and Action Plans

Risk Assessment	Implement a scoring system based on the developed models to quantify credit risk for each loan applicant.
Loan Pricing	Adopt dynamic pricing strategies that adjust interest rates based on individual risk assessments, ensuring fair pricing and profitability.
Underwriting Policies	Review and adjust underwriting criteria to reflect the insights gained from the analysis, focusing on key risk factors.
Customer Segmentation	Develop targeted marketing and loan offerings based on customer risk profiles, optimizing customer acquisition and retention strategies.

Conclusion

1

Summary of Key Findings

The analysis revealed significant insights into the factors influencing loan default risk, including applicant demographics, financial information, loan characteristics, and previous loan performance.

2

Impact and Benefits

Implementing the recommendations is expected to improve loan approval processes, reduce default rates, enhance profitability, and strengthen the bank's financial stability.

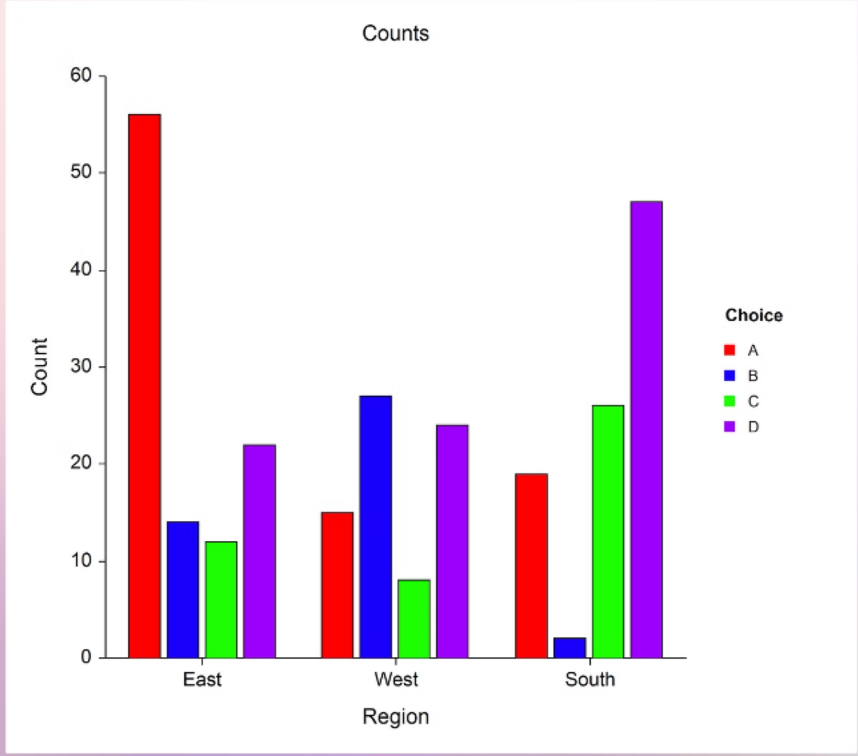
3

Future Directions

Further research could explore the use of advanced machine learning techniques, such as deep learning, to enhance risk prediction capabilities and develop more sophisticated risk management strategies.

Appendix: Model Evaluation Metrics

The selected models were evaluated using metrics such as accuracy, precision, recall, and F1-score.
These metrics provide a comprehensive assessment of the models' performance in predicting loan default probability.



THANK YOU

Evoastra Ventutres Mini Project

Internship Presentation

By Team A