Banking Customer Data Analysis using Exploratory Data Analysis (EDA)

4th Year B.Tech Minor Project

BY:- VISHU RATHI

VISHAL RAWAT

NITIN BIJALWAN

AYUSH RATURI

Project Overview & Objectives

1

Objective

Analyze banking customer data to understand financial behaviors and relationships.

2

Dataset

Banking.csv with 11 numerical and 6 categorical features, providing a rich foundation for analysis.

3

Methodology

Comprehensive EDA including univariate, bivariate analysis, and in-depth correlation studies.

4

Goal

Extract actionable insights to inform and optimize banking strategies.

Dataset Description & Preprocessing

Data Source & Key Features

The dataset originates from banking customer records, encompassing vital attributes:

- Age, Income, Bank Deposits, Savings
- Credit Card Balance, Business Lending
- Properties Owned

Data Cleaning & Feature Engineering

Rigorous preprocessing was essential for data integrity:

- **Date Conversion:** 'Joined Bank' feature transformed to datetime format.
- **Missing Values:** Comprehensive analysis and strategic handling of missing data points.
- **Income Band Creation:** Categorization into Low (0-100K), Mid (100K-300K), and High (300K+) for segmented analysis.

Univariate Analysis: Unveiling Customer Distributions

Our univariate analysis reveals foundational patterns in customer demographics and financial behaviors.

Age Distribution

The age distribution indicates a diverse customer base, from younger professionals to retirees.

Income Segmentation

Income analysis distinctly segments customers into three identifiable groups, suggesting varied financial capacities.

Account Balances

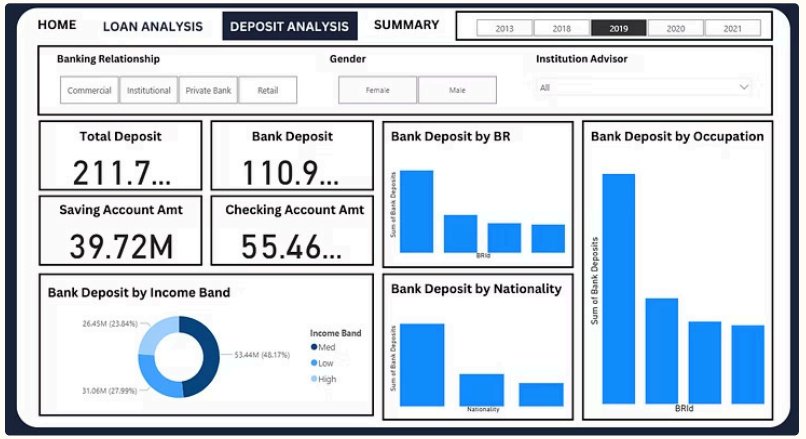
Varied account balances across deposits and savings highlight a spectrum of banking behaviors and needs.

These initial findings point towards multiple account types being common, reflecting a broad range of individual banking requirements and significant variation in savings and deposit patterns.

Correlation Analysis & Heatmap

A correlation heatmap vividly illustrates the interdependencies among the 11 numerical features, exposing critical relationships.

(Placeholder for Correlation Heatmap Visualization)



Key Correlations Identified

- **Bank Deposits ↔ Saving Accounts:** A high positive correlation suggests synchronized financial management.
- **Checking Accounts ↔ Saving Accounts:** A moderate correlation indicates linked transactional and saving habits.
- **Age ↔ Superannuation Savings:** Reflects a strong lifecycle correlation, with savings increasing with age.

Key Relationship Analysis: Unpacking Customer Dynamics

Deeper analysis reveals how various financial elements interlace, painting a clearer picture of customer segments.

Deposits and Savings Behavior

The strong correlation implies that customers often manage their deposits and savings in tandem, suggesting holistic financial planning rather than isolated actions.

Income and Age Impact

Higher income earners and older customers consistently demonstrate increased overall banking activity, indicating greater engagement with diverse banking products and services.

Business vs. Personal Banking

Business Lending shows a moderate link to Bank Loans, highlighting a distinction between entrepreneurial clients and those primarily engaged in personal finance. This segmentation is crucial for targeted product development.

Statistical Insights & Patterns

Our EDA uncovers significant statistical patterns driving banking customer behavior.

Financial Lifecycle Trends

Age exhibits a clear correlation with the accumulation of retirement (superannuation) savings, affirming a standard financial planning trajectory.

Property Investment

The low correlation with core banking variables suggests property investment decisions are often influenced by external market factors rather than immediate banking product engagement.

Customer Segmentation

Income bands reveal a balanced distribution across different risk categories, providing opportunities for tailored risk management and product offerings.

Account Usage

A strong relationship exists between different types of accounts, indicating customers typically leverage multiple banking products in a complementary manner.

Technical Implementation: Tools & Techniques

Our project leveraged industry-standard tools and robust analytical techniques to ensure accuracy and depth.

Python Libraries

- **Pandas:** For efficient data manipulation and structuring.
- **NumPy:** Essential for numerical operations and complex calculations.
- **Matplotlib/Seaborn:** Powerful tools for creating insightful data visualizations.

Analysis Techniques

- **Descriptive Statistics:** Summarizing main features of the data.
- **Correlation Matrix:** Quantifying relationships between variables.
- **Regression Analysis:** Modeling relationships between dependent and independent variables.

Data Visualization

- **Histograms:** To show data distribution.
- **Heatmaps:** For correlation patterns.
- **Scatter Plots:** With regression lines for bivariate relationships.

Feature Engineering

- **Income Categorization:** Creating meaningful income bands.
- **Categorical Analysis:** Deep dives into non-numerical attributes.

Business Applications & Future Scope

Key Business Applications

- **Customer Segmentation:** Enabling highly targeted marketing campaigns and personalized product offerings.
- **Risk Assessment:** Informing the development of more accurate risk models for lending and investment decisions.
- **Product Recommendation:** Enhancing customer experience through intelligent suggestions based on financial behavior.

Future Enhancements

Building upon this EDA, future work can expand into more advanced analytics:

- **Predictive Modeling:** Forecasting customer churn, product adoption, or financial trends.
- **Machine Learning Classification:** Automating customer categorization and behavior prediction.
- **Time Series Analysis:** Understanding dynamic changes in banking behaviors over time.

Project Value: This project lays a robust, data-driven foundation for advanced banking analytics.

Conclusion & Thank You

Our journey through banking customer data analysis has yielded significant insights and reinforced critical learning outcomes.

Key Achievements

- Comprehensive analysis of banking customer behavior.
- Statistical validation of financial product relationships.
- Generation of actionable insights for banking strategy.

Learning Outcomes

- Mastery of advanced Exploratory Data Analysis (EDA) techniques.
- Proficiency in data visualization for complex datasets.
- Application of academic rigor to real-world financial data.

Thank You!

Questions?