Banking Customer Insights & Strategic Dashboard Project Report

Project Date: July 2025

1. Abstract / Executive Summary

This project focuses on a comprehensive Exploratory Data Analysis (EDA) of a banking customer dataset, culminating in the development of an interactive Power BI dashboard. The primary objective was to transform raw customer and financial data into actionable insights, enabling the bank to better understand its customer base, analyze financial product usage, assess risk profiles, and support data-driven strategic decision-making. Through meticulous data cleaning, preprocessing, and a multi-faceted analytical approach, key demographic and financial patterns were uncovered and presented visually, providing a clear and accessible overview of the bank's client portfolio.

2. Introduction

2.1. Problem Statement

In today's competitive financial landscape, banks collect vast amounts of customer data. However, this raw data often remains siloed and unanalyzed, preventing stakeholders from gaining a holistic and actionable understanding of their customer base. The core problem addressed by this project is the **lack of a clear, consolidated, and easily digestible view of customer demographics, financial behaviors, and risk profiles.** This deficiency can lead to suboptimal decision-making in areas such as targeted marketing, new product development, and risk management, as strategies may be based on intuition rather than concrete data.

2.2. Project Aim

The primary aim of this project was to **extract meaningful patterns and actionable insights from a raw banking customer dataset through comprehensive Exploratory Data Analysis (EDA), and to effectively communicate these insights via an intuitive and interactive Power BI dashboard.**

Specific objectives included:

- To ensure data quality through thorough data loading and initial inspection.
- To create new, relevant features that enhance analytical capabilities.
- To perform detailed univariate and bivariate analyses to uncover customer characteristics and relationships between financial variables.

- To visualize key findings in an interactive dashboard format for non-technical business users.
- To provide data-driven recommendations that support strategic planning for the bank.

3. Dataset Description

The project utilized the Banking.csv dataset, comprising **3000 records and 25 columns**. This rich dataset provides a wide array of attributes pertaining to bank clients and their financial interactions.

Key categories of data points include:

- **Client Identification:** Client ID, Name, Location ID.
- **Demographics:** Age, Nationality, Occupation, GenderId.
- **Bank Relationship Details:** Joined Bank, Banking Contact, Fee Structure, Loyalty Classification.
- **Financial Metrics:** Estimated Income, Superannuation Savings, Amount of Credit Cards, Credit Card Balance, Bank Loans, Bank Deposits, Checking Accounts, Saving Accounts, Foreign Currency Account, Business Lending.
- **Risk & Assets:** Properties Owned, Risk Weighting.
- **Internal Identifiers:** BRId, IAId.

4. Methodology

The project followed a structured data analysis methodology, encompassing data loading, preprocessing, exploratory analysis, feature engineering, and visualization.

4.1. Data Collection / Loading

The project began by loading the Banking.csv file into a Pandas DataFrame in a Jupyter Notebook environment.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("Banking.csv")
```

Additionally, the Connect_MySql.ipynb notebook demonstrates the capability to connect to an external MySQL database and query data, which is a common practice

in real-world data analysis projects for data acquisition.

4.2. Data Preprocessing & Cleaning

Initial data inspection was performed to understand the dataset's structure and quality:

- df.head() and df.shape were used to view the first few rows and confirm the dataset dimensions (3000 rows, 25 columns).
- df.info() provided a concise summary, revealing that all columns had 3000 non-null entries, indicating **no missing values** in this particular dataset. Data types were also verified.
- df.describe() generated descriptive statistics for numerical columns, offering initial insights into central tendencies, distributions, and potential outliers.
- The Joined Bank column, initially an object type, was converted to a datetime format using pd.to_datetime() for potential time-series analysis or age calculation if needed.

4.3. Feature Engineering

A crucial step in enhancing the dataset's analytical utility was feature engineering:

- A new categorical column, Income Band, was created from the numerical Estimated Income column. This involved binning income values into 'Low' (0-100k), 'Mid' (100k-300k), and 'High' (300k+) categories. This transformation facilitates easier segmentation and analysis of customer groups based on their income levels.

4.4. Exploratory Data Analysis (EDA)

Comprehensive EDA was performed using Python libraries (Matplotlib and Seaborn) to understand the distribution of individual variables (univariate analysis) and the relationships between them (bivariate/multivariate analysis).

- **Univariate Analysis:**
  - **Numerical Columns:** Histograms and KDE plots were generated for Age, Estimated Income, Superannuation Savings, Credit Card Balance, Bank Loans, Bank Deposits, Checking Accounts, Saving Accounts, Foreign Currency Account, and Business Lending to visualize their distributions and identify any skewness or unusual patterns.
  - **Categorical Columns:** Bar plots and value_counts() were used to analyze the frequency distribution of Income Band, BRId, GenderId, Amount of Credit Cards, Nationality, Occupation, Fee Structure, Loyalty Classification, Properties Owned, and Risk Weighting.

        - **Key Observations:**
            - Income Band: A significant portion of customers fall into the 'Mid' income band.
            - BRId: BRId 3 has the highest customer count.
            - GenderId: A near 50/50 split between GenderId 1 and GenderId 2.
            - Amount of Credit Cards: Most customers possess only 1 credit card.
            - Nationality: European customers form the largest group, followed by Asian and American.
            - Fee Structure: 'High' fee structure is the most common.
            - Loyalty Classification: 'Jade' customers are the most frequent loyalty type.
            - Risk Weighting: Risk Weighting 2 is the most common category.
- **Bivariate/Multivariate Analysis:**
    - Correlation matrices using heatmaps were planned and are typically generated for numerical columns (e.g., Estimated Income, Superannuation Savings, Credit Card Balance, Bank Loans, Bank Deposits, Checking Accounts, Saving Accounts, Foreign Currency Account, Business Lending). This helps in identifying strong positive or negative relationships between financial variables. For instance, strong positive correlations between Bank Deposits and Checking Accounts / Saving Accounts / Foreign Currency Account were noted, indicating that customers with high balances in one account often hold substantial funds across others.
    - Further analysis would involve cross-tabulations and comparative plots (e.g., box plots of Bank Deposits by Income Band or Loyalty Classification) to understand relationships between categorical and numerical variables.

4.5. Data Visualization & Dashboarding

The insights generated from the EDA were translated into an interactive dashboard using **Microsoft Power BI**. This dashboard serves as the final, user-friendly output, enabling business stakeholders to explore the data dynamically.

A typical Power BI dashboard for this project would include:

- **Overview KPIs:** Displaying total deposits, total loans, average income, and total customer count.
- **Customer Segmentation:** Visualizing customer distribution by Age, Income Band, Nationality, Occupation, GenderId, Fee Structure, and Loyalty Classification.
- **Product Penetration:** Charts showing the breakdown of various financial product holdings (e.g., Bank Loans, Credit Cards, Checking Accounts) across different

customer segments.
- **Risk Analysis:** Visualizing the distribution of Risk Weighting and its correlation with financial product usage.
- **Interactive Filters:** Allowing users to filter data by any demographic or financial attribute to drill down into specific segments of interest.

5. Key Findings & Insights

The EDA revealed several critical insights into the bank's customer base:

- **Income Distribution:** The customer base is predominantly in the 'Mid' income band, suggesting a focus on this segment for product offerings.
- **Account Relationships:** A strong positive correlation exists between various deposit accounts (Checking, Saving, Foreign Currency), indicating that customers with high balances in one account tend to hold substantial funds across others. This highlights opportunities for cross-selling related deposit products.
- **Credit Card Usage:** The majority of customers hold only one credit card, suggesting potential for promoting additional credit products to existing clients.
- **Demographic Composition:** European customers are the largest nationality group, followed by Asian and American, which can inform targeted marketing campaigns.
- **Risk vs. Lending:** Business Lending shows a moderate correlation with Bank Loans, implying some customers may have both personal and business debts. However, its weak correlation with other deposit metrics suggests a distinct customer segment for business lending.
- **Financial Lifecycle Trends:** Moderate correlations of Age and Estimated Income with various balances (Superannuation, Savings, Checking) reflect that higher income earners and older individuals often accumulate more savings and retirement funds.

6. Challenges Encountered & Solutions

- **Challenge 1: Data Understanding & Domain Nuances:** As a fresher, initially understanding the specific meaning and implications of certain banking-specific columns (e.g., BRId, IAId, Superannuation Savings without a detailed data dictionary) was a challenge.
    - **Solution:** Overcame this by performing extensive descriptive statistics and univariate analysis to infer meanings, and by researching common banking terminology online. In a real-world scenario, this would involve direct consultation with domain experts.

- **Challenge 2: Feature Engineering for Actionable Insights:** Deciding how to transform raw numerical data (like Estimated Income) into more meaningful categorical features (Income Band) for segmentation.
    - **Solution:** Used domain-relevant binning strategies (e.g., standard income brackets) and iteratively assessed their usefulness in revealing patterns.
- **Challenge 3: Effective Visualization & Storytelling:** Presenting complex data relationships in a clear, concise, and actionable manner for non-technical stakeholders.
    - **Solution:** Focused on creating intuitive and interactive dashboards in Power BI, prioritizing key insights and using appropriate chart types. Emphasized a "storytelling" approach by guiding the user through the data.

7. Conclusion

This project successfully delivered a comprehensive Exploratory Data Analysis of a banking customer dataset, providing valuable insights into customer demographics, financial behaviors, and product usage. The creation of an interactive Power BI dashboard effectively translates these complex data points into an accessible format for strategic decision-making. The findings can directly support the bank in optimizing marketing strategies, identifying cross-selling opportunities, and refining risk assessment models. This project demonstrates strong foundational skills in data analysis, including data preprocessing, EDA, feature engineering, and data visualization.

8. Future Scope

- **Predictive Modeling:** Develop machine learning models (e.g., classification models for churn prediction, regression models for loan default risk) based on the identified features.
- **Advanced Segmentation:** Utilize clustering algorithms (e.g., K-Means) to identify more granular and distinct customer segments beyond simple income bands.
- **Time-Series Analysis:** Leverage the Joined Bank date to analyze customer tenure, product adoption over time, and identify trends in financial behavior.
- **Integration of External Data:** Incorporate external macroeconomic data, market trends, or social media sentiment to enrich the analysis and provide broader business context.
- **A/B Testing Analysis:** If experimental data were available, analyze the impact of different marketing campaigns or product features.

9. Tools & Technologies Used

- **Programming Language:** Python
- **Data Manipulation & Analysis:** Pandas, NumPy
- **Data Visualization:** Matplotlib, Seaborn, Microsoft Power BI
- **Development Environment:** Jupyter Notebook
- **Database Connectivity (Demonstrated):** MySQL Connector (for MySQL)