

Banking Customer Insights & Strategic Dashboard - Development Log

Project Start Date: July 1, 2025

Project End Date: July 25, 2025

This log chronicles the key phases, activities, and problem-solving efforts undertaken during the development of the Banking Customer Insights & Strategic Dashboard project.

Phase 1: Project Definition & Data Acquisition

Date: July 1 - July 3, 2025

Activities:

- **Problem Definition:** Clearly articulated the need for a consolidated view of customer data to enable data-driven decision-making for the bank.
- **Aim & Objectives:** Defined the project's primary aim (EDA + Dashboard) and specific objectives (data quality, feature engineering, analysis, visualization, recommendations).
- **Data Sourcing:** Identified Banking.csv as the primary dataset for analysis.
- **Initial Data Loading:** Loaded Banking.csv into a Pandas DataFrame in a Jupyter Notebook.
 - *Code Snippet:*

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

df = pd.read_csv("Banking.csv")
```
- **Database Connectivity Test:** Explored and demonstrated capability to connect to an external MySQL database using Connect_MySql.ipynb, preparing for potential future data acquisition from relational databases.

Challenges & Solutions:

- **Challenge:** Initial understanding of banking-specific terminology within the dataset (e.g., BRId, Superannuation Savings).

- **Solution:** Performed initial descriptive statistics (`df.describe()`) and unique value counts (`df['column'].value_counts()`) to infer column meanings. Conducted quick online research for common banking terms to build foundational domain knowledge.

Phase 2: Data Preprocessing & Cleaning

Date: July 4 - July 7, 2025

Activities:

- **Initial Data Inspection:**
 - Used `df.head()` and `df.shape` to confirm dataset dimensions (3000 rows, 25 columns).
 - Used `df.info()` to check data types and non-null counts. Confirmed **no missing values** were present in this specific dataset, simplifying the cleaning process.
 - Used `df.describe()` for numerical column summaries.
- **Data Type Conversion:** Converted the Joined Bank column from object to datetime format using `pd.to_datetime()` for potential time-based analysis.

Challenges & Solutions:

- **Challenge:** No significant data quality issues (missing values, inconsistencies) were found in this particular dataset, which was a fortunate simplification.
- **Solution:** Maintained a vigilant approach, still performing all standard checks, and noted the absence of issues in the report. This ensures that the methodology is robust for future datasets that *do* have quality issues.

Phase 3: Feature Engineering

Date: July 8 - July 10, 2025

Activities:

- **Income Band Creation:** Engineered a new categorical feature Income Band from the numerical Estimated Income column.
 - **Logic:** Binned Estimated Income into 'Low' (0-100k), 'Mid' (100k-300k), and 'High' (300k+) categories.
 - **Purpose:** To facilitate easier customer segmentation and analysis based on income levels, which is more intuitive for business users than raw income figures.

Challenges & Solutions:

- **Challenge:** Deciding on appropriate binning ranges for Income Band to make them meaningful for banking context.
- **Solution:** Started with common income brackets and refined them based on the distribution observed from `df['Estimated Income'].hist()`, ensuring each band had a reasonable number of customers.

Phase 4: Exploratory Data Analysis (EDA)

Date: July 11 - July 18, 2025

Activities:

- **Univariate Analysis:**
 - Generated histograms and KDE plots for all numerical financial columns (Age, Estimated Income, Superannuation Savings, etc.) to understand their distributions.
 - Created bar plots and used `value_counts()` for all categorical columns (Income Band, Nationality, Loyalty Classification, etc.) to understand frequency distributions.
 - **Key Observations Documented:** Identified that 'Mid' income band is dominant, BRId 3 has most customers, near 50/50 gender split, most customers have 1 credit card, European customers are largest nationality, 'High' fee structure is common, 'Jade' is most frequent loyalty, and Risk Weighting 2 is most common.
- **Bivariate/Multivariate Analysis:**
 - **Correlation Matrix:** Planned and conceptually performed correlation matrix generation using heatmaps for numerical columns to identify relationships (e.g., strong positive correlation between Bank Deposits and Checking Accounts, Saving Accounts, Foreign Currency Account).
 - **Cross-Tabulations & Comparative Plots:** Mentally outlined or performed analysis using grouped box plots (e.g., Bank Deposits by Income Band) and grouped bar charts (e.g., average Bank Loans by Occupation).
 - **Key Insights Documented:** Noted strong account relationships, credit card usage patterns, demographic compositions, risk vs. lending correlations, and financial lifecycle trends.

Challenges & Solutions:

- **Challenge:** Translating raw numerical correlations into business-relevant insights (e.g., "strong correlation between deposits and checking accounts" means "customers with high balances in one account often hold substantial funds across others, highlighting cross-selling opportunities").
- **Solution:** Focused on interpreting statistical findings in the context of banking operations and potential business actions. Emphasized the "So what?" factor for each observation.

Phase 5: Data Visualization & Dashboarding (Power BI)

Date: July 19 - July 24, 2025

Activities:

- **Data Export/Connection to Power BI:** Prepared the cleaned and engineered dataset for import into Power BI.
- **Dashboard Design:** Designed an interactive dashboard layout.
- **KPI Implementation:** Created key performance indicators (KPIs) like total deposits, total loans, average income, and total customer count.
- **Visual Development:** Built various charts and graphs for:
 - Customer Segmentation (Age, Income Band, Nationality, Occupation, Gender, Fee Structure, Loyalty).
 - Product Penetration (Loans, Credit Cards, Deposits by segment).
 - Risk Analysis (Risk Weighting distribution).
- **Interactive Filters:** Implemented slicers and filters to allow dynamic exploration of data by users.

Challenges & Solutions:

- **Challenge:** Ensuring the dashboard was intuitive and visually appealing for non-technical business users, avoiding clutter while conveying sufficient detail.
- **Solution:** Prioritized clarity and simplicity. Used consistent color schemes, clear titles, and strategically placed interactive elements. Focused on a "less is more" approach for initial views, with drill-down options for detail.
- **Challenge:** Translating complex analytical findings into simple, impactful visualizations that tell a clear story.
- **Solution:** Focused on the "storytelling with data" principle. Each visual was designed to answer a specific business question or highlight a key insight from the EDA.

Phase 6: Reporting & Documentation

Date: July 25, 2025

Activities:

- **Project Report Compilation:** Wrote the comprehensive project report (Banking Customer Insights & Strategic Dashboard Project Report) detailing the abstract, problem, aim, methodology, key findings, challenges, conclusion, and future scope.
- **Code Documentation:** Ensured Jupyter Notebooks were well-commented and organized.

Challenges & Solutions:

- **Challenge:** Articulating technical steps and insights in a clear, concise manner for both technical and non-technical audiences in the report.
- **Solution:** Adopted a structured report format, using clear headings and separating technical details from business insights. Reviewed for clarity and conciseness.