

# Banking Exploratory Data Analysis

## 1. Project Overview

This project focuses on analysing real-world banking data to uncover meaningful insights related to customer behaviour, financial trends, and operational performance. The objective is to transform raw banking data into actionable insights using **Exploratory Data Analysis (EDA)** and **interactive visual dashboards**. The project combines **Python-based data analysis** with **Power BI visualization**, enabling stakeholders to make informed, data-driven decisions through intuitive dashboards and KPIs.

## 2. Dataset Summary

- **Rows:** 3,900
- **Columns:** 18

### ◊ Key Features:

- **Customer Demographics:** Age, Gender, Location, Subscription Status
- **Purchase Details:** Item Purchased, Product Category, Purchase Amount, Season, Size, Colour
- **Shopping Behaviour:**  
Discount Applied, Promo Code Used, Previous Purchases, Purchase Frequency, Review Rating, Shipping Type

### ◊ Missing Data:

- **Review Rating:** 37 missing values

## 3. Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

- **Data Loading:** Imported the dataset using pandas.

	Client ID	Name	Age	Location ID	Joined Bank	Banking Contact	Nationality	Occupation	Fee Structure	Loyalty Classification	...	Bank Deposits	Checking Accounts	Saving Accounts	Foreign Currency Account	Business Lending
0	IND81288	Raymond Mills	24	34324	06-05-2019	Anthony Torres	American	Safety Technician IV	High	Jade	...	1485828.64	603617.88	607332.46	12249.96	1134475.3
1	IND65833	Julia Spencer	23	42205	10-12-2001	Jonathan Hawkins	African	Software Consultant	High	Jade	...	641482.79	229521.37	344635.16	61162.31	2000526.1
2	IND47499	Stephen Murray	27	7314	25-01-2010	Anthony Berry	European	Help Desk Operator	High	Gold	...	1033401.59	652674.69	203054.35	79071.78	548137.5
3	IND72498	Virginia Garza	40	34594	28-03-2019	Steve Diaz	American	Geologist II	Mid	Silver	...	1048157.49	1048157.49	234685.02	57513.65	1148402.2
4	IND60181	Melissa Sanders	46	41269	20-07-2012	Shawn Long	American	Assistant Professor	Mid	Platinum	...	487782.53	446644.25	128351.45	30012.14	1674412.1

- **Initial Exploration:** Used df.info() to check structure and .describe() for summary statistics.

```

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3000 entries, 0 to 2999
Data columns (total 25 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Client ID        3000 non-null   object  
 1   Name              3000 non-null   object  
 2   Age               3000 non-null   int64   
 3   Location ID      3000 non-null   int64   
 4   Joined Bank       3000 non-null   object  
 5   Banking Contact  3000 non-null   object  
 6   Nationality       3000 non-null   object  
 7   Occupation        3000 non-null   object  
 8   Fee Structure     3000 non-null   object  
 9   Loyalty Classification 3000 non-null   object  
 10  Estimated Income 3000 non-null   float64 
 11  Superannuation Savings 3000 non-null   float64 
 12  Amount of Credit Cards 3000 non-null   int64   
 13  Credit Card Balance 3000 non-null   float64 
 14  Bank Loans        3000 non-null   float64 
 15  Bank Deposits    3000 non-null   float64 
 16  Checking Accounts 3000 non-null   float64 
 17  Saving Accounts   3000 non-null   float64 
 18  Foreign Currency Account 3000 non-null   float64 
 19  Business Lending  3000 non-null   float64 
 20  Properties Owned  3000 non-null   int64   
 21  Risk Weighting    3000 non-null   int64   
 22  BRId              3000 non-null   int64   
 23  GenderId          3000 non-null   int64   
 24  IAId              3000 non-null   int64   

dtypes: float64(9), int64(8), object(8)
memory usage: 586.1+ KB

```

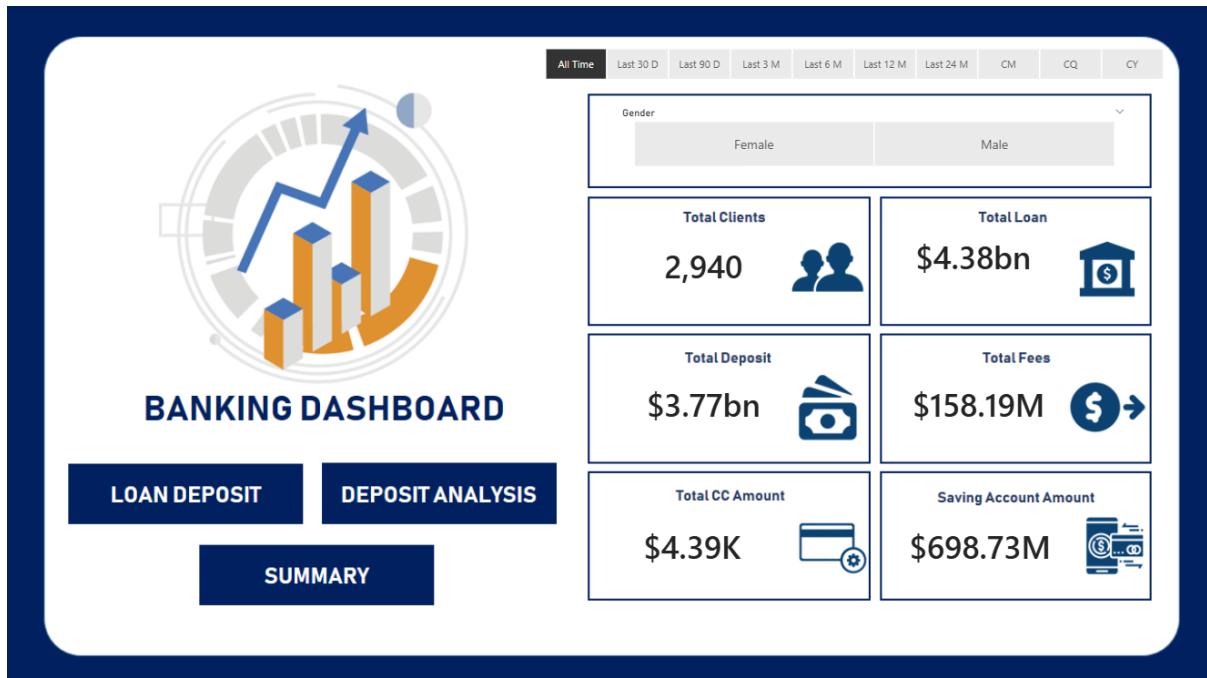
**4. Data Quality and Missing Value Analysis:** Missing value analysis was conducted using the df.info () function. All columns contained complete non-null values, and no missing data was detected. As a result, no imputation or data removal was required before analysis.

- **Database Integration:** Connected Python script to MySQL and loaded the cleaned Data Frame into the database for SQL analysis
- **Column Standardization:** Renamed all dataset columns to `snake_case` format to improve readability, maintain naming consistency, and support better documentation and code maintainability.

## 5. Dashboard in Power BI:

Finally, we built an interactive dashboard in **Power BI** to present insights visually

- **BANKING DATA ANALYTICS DASHBOARD:** Loan & Deposit Analysis using Power BI.

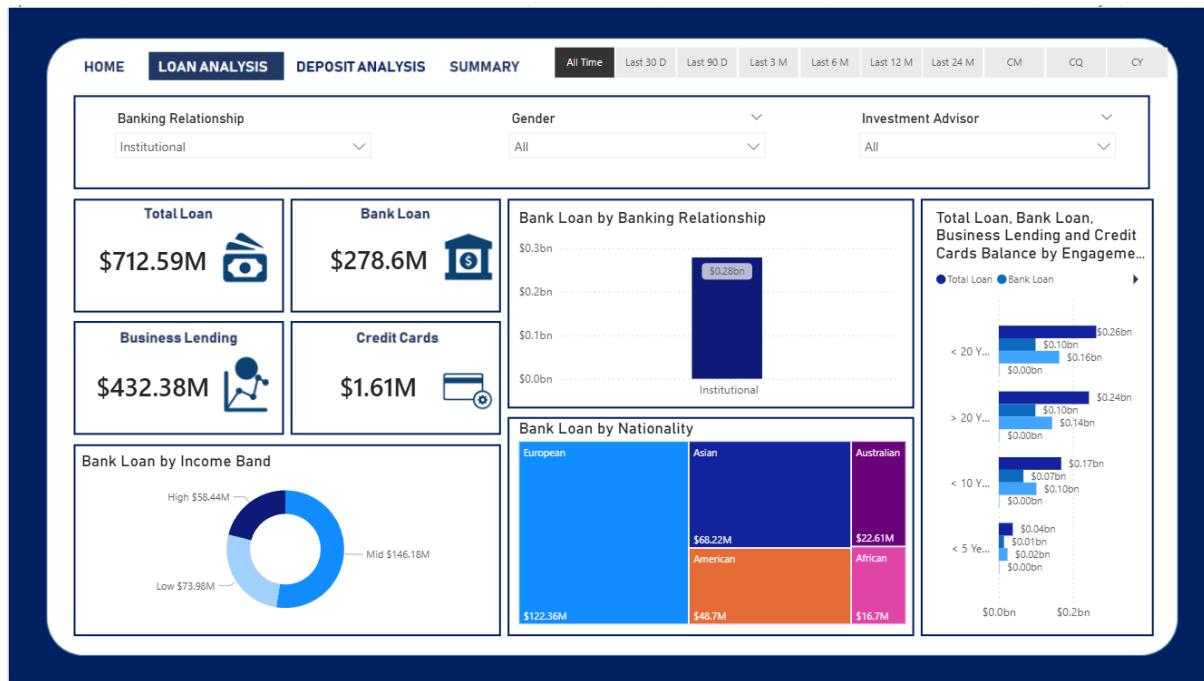


This summary dashboard provides a high-level overview of key banking metrics, including total clients, loans, deposits, fees, and account balances for quick performance monitoring.

### Key Metrics

- **Total Clients:** 2,940
- **Total Loan:** \$4.38 Bn
- **Total Deposit:** \$3.77 Bn
- **Total Fees:** \$158.19 M
- **Credit Card Amount:** \$4.39 K
- **Savings Account Balance:** \$698.73 M

- **LOAN ANALYSIS DASHBOARD:** Customer Loan Distribution & Risk Insights.

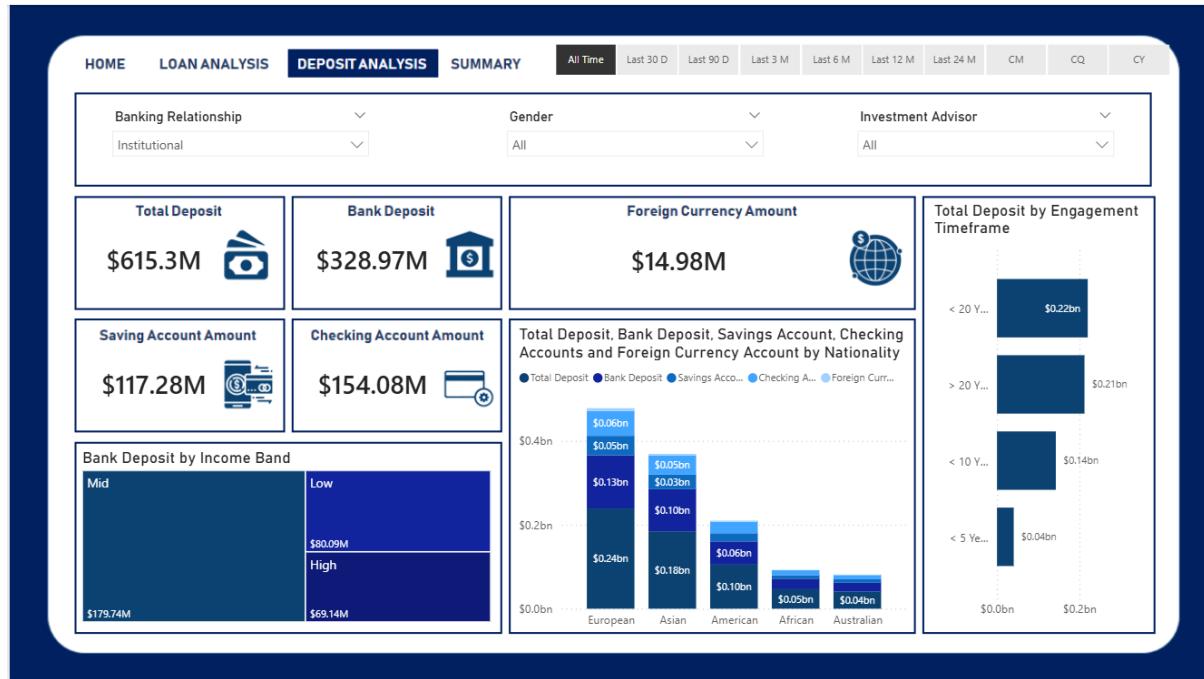


The Loan Analysis dashboard provides a detailed breakdown of total loans across banking relationships, income bands, nationalities, and customer age groups. It helps identify high-value loan segments, lending patterns, and potential risk areas to support data-driven credit decisions.

#### Key Loan Metrics

- **Total Loan Amount:** \$712.59M
- **Bank Loans:** \$278.6M
- **Business Lending:** \$432.38M
- **Credit Card Loans:** \$1.61M

- **DEPOSIT ANALYSIS DASHBOARD:** Customer Deposit Distribution & Savings Insights.

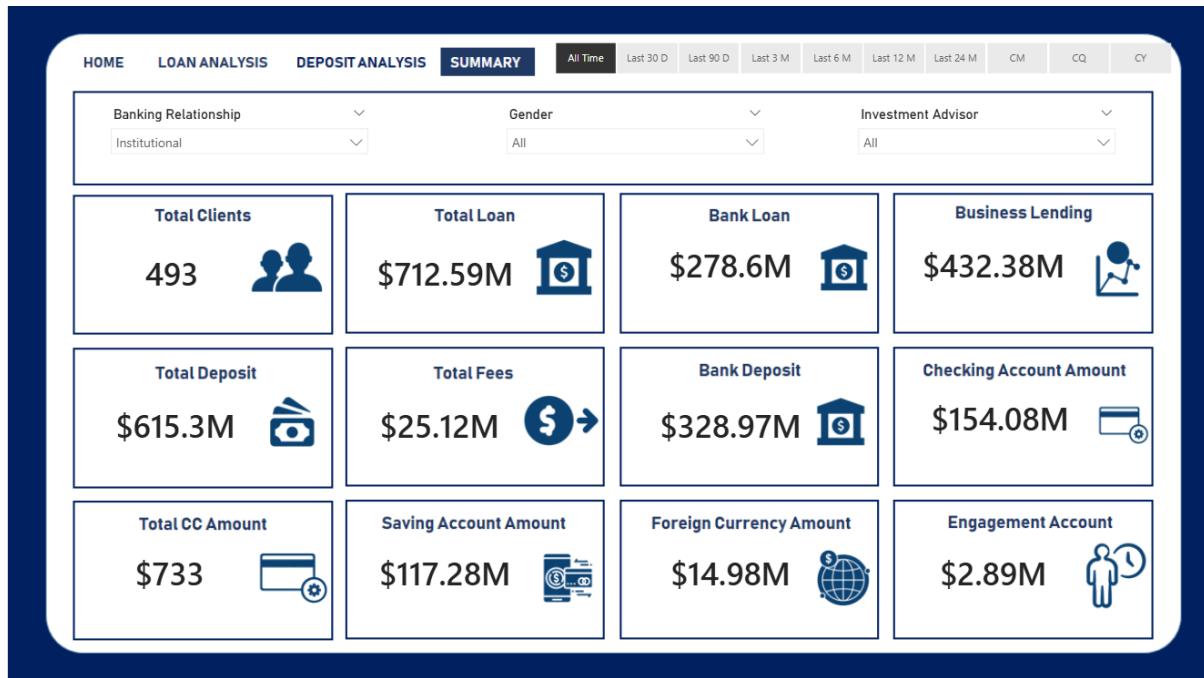


The Deposit Analysis dashboard provides a comprehensive view of customer deposits across account types, income bands, nationalities, and engagement timeframes. It helps identify savings behaviour, deposit concentration, and high-value customer segments.

### Key Deposit Metrics

- **Total Deposit:** \$615.3M
- **Bank Deposit:** \$328.97M
- **Savings Account Amount:** \$117.28M
- **Checking Account Amount:** \$154.08M
- **Foreign Currency Amount:** \$14.98M

- **SUMMARY DASHBOARD:** A consolidated overview of customers, loans, deposits, and revenue performance.



A consolidated snapshot of customer activity, lending exposure, and deposit distribution across key banking products. A consolidated snapshot of customer activity across the bank. It highlights overall lending exposure and deposit distribution. The view supports quick assessment of key banking products.

### Key Deposit Metrics

- **Total Deposits:** \$615.3M
- **Bank Deposits:** \$328.97M
- **Savings Account Balance:** \$117.28M
- **Checking Account Balance:** \$154.08M
- **Foreign Currency Deposits:** \$14.9

- **DRILL-THROUGH ANALYSIS:** Detailed customer-level view of fees, engagement, and credit card activity.

The screenshot shows a detailed analysis page with the following sections:

- Header:** HOME, LOAN ANALYSIS, DEPOSIT ANALYSIS, SUMMARY. Time filters: All Time, Last 30 D, Last 90 D, Last 3 M, Last 6 M, Last 12 M, Last 24 M, CM, CQ, CY.
- Filters:** Banking Relationship (Institutional), Gender (All), Investment Advisor (All).
- Table:** A grid showing customer details like Name, Investment Advisor, Engagement Length, Total Fees, Total CC Amount, and Credit Cards Balance. A summary row at the bottom shows Totals: \$979,348.00, \$6,746,952.00, \$242.00, and \$401,863.22.
- Visualizations:**
  - Total Fees by Loyalty Classification:** Bar chart showing fees for Jade (\$3.16M), Silver (\$1.51M), Gold (\$1.36M), and Platinum (\$0.72M).
  - Total Fees and Engagement Length by Nationality:** Bar chart showing fees for European (\$2.64M), Asian (\$1.92M), American (\$1.14M), Australian (\$0.73M), and African (\$0.31M).

This drill-through page provides a detailed, customer-level breakdown of fees, engagement duration, and credit card balances, enabling deeper investigation behind high-level dashboard insights.

### Key Deposit Metrics

- **Total Deposits:** \$615.3M
- **Bank Deposits:** \$328.97M
- **Savings Account Balance:** \$117.28M
- **Checking Account Balance:** \$154.08M