FINANCIAL RISK ANALYSIS WITH PYTHON – BARCLAYS

PROJECT OVERVIEW

This project focuses on financial transaction analysis and risk identification using real-world banking transaction data from Barclays. The objective is to understand customer transaction behavior, identify financially risky accounts, and derive data-driven insights that can support risk monitoring and customer management decisions.

The dataset contains detailed information on customers, accounts, transaction types, dates, balances, inflows, outflows, and risk indicators. Using Python-based data analysis and statistical techniques, the project performs exploratory data analysis (EDA), behavioral segmentation, financial risk detection, and hypothesis testing.

OBJECTIVES

The main objectives of this project are to analyze transactional banking data and extract meaningful insights related to customer behavior and financial risk. Specifically, the project aims to:

- Analyze customer transaction patterns across accounts, products, and time periods
- Identify dormant and low-activity accounts based on transaction gaps and frequency
- Detect accounts with financial risk, including overdrafts, negative balances, and high balance volatility
- Track credit and debit trends over time to understand cash flow behavior
- Detect anomalous transactions using statistical techniques such as IQR and Z-score methods
- Segment customers based on average balance and transaction volume
- Test hypotheses to validate whether transaction volume impacts account balance levels
- Provide data-backed recommendations for customer engagement and risk monitoring

DATA CLEANING & VALIDATION

Before performing analysis, the dataset was thoroughly cleaned and validated to ensure accuracy and reliability of insights.

1. Special Character & Numeric Validation

- Special characters in numeric columns: **0 found**
- Currency symbols detected: **0**
- Decimal separator verified: **(point) – correct**
- All **800 rows** successfully converted to **float64**

2. Negative Values Check

- Negative transactions identified: **21**
  - Interpreted as valid payments/refunds
- Negative balances identified: **14**
  - Treated as valid overdraft cases

3. Null Value Handling

- Null values across all columns: **0**
- No rows dropped or imputed

4. Date Validation & Transformation

- Original format converted: DD-MM-YYYY → datetime64[ns]
- Conversion success rate: **800 / 800 (100%)**
- Invalid or corrupted dates: **0**
- Extracted components:
    - Year
    - Month
    - Day
    - Day of Week
    - Week of Year

5. Date Range Confirmation

- Start Date: 2023-01-02
- End Date: 2024-06-22
- Total Duration: 537 days

6. Standardization of Categorical Columns

- **Account Type:** 4 categories
    - Savings, Credit, Loan, Current
- **Transaction Type:** 4 categories
    - Deposit, Payment, Withdrawal, Transfer
- **Product:** 5 categories
    - Credit Card, Mutual Fund, Personal Loan, Home Loan, Savings Account
- Extra spaces removed: **0**
- Inconsistent labels found: **0**

The dataset is **fully clean, standardized, and analysis ready**.

EXPLORATORY DATA ANALYSIS (EDA)

The EDA reveals clear patterns:

- Customer activity is decreasing over time.
- Majority of accounts are low-activity but moderate balance holders.
- Significant risk indicators exist (overdrafts, high CV accounts, anomaly transactions).
- No statistical link exists between frequency of transactions and account balances.
- Dormancy is widespread → immediate operational attention required.

INSIGHTS & FINDINGS

This section summarizes the key analytical findings derived from the cleaned and processed Barclays financial dataset (800 transactions across 15 variables). Insights are organized by analytical tasks for clarity.

1. Data Quality Insights

- **No missing values** in the dataset, ensuring reliability of analysis (0 nulls across all columns).
- **No special characters** or corrupted entries found in numeric fields such as TransactionAmount and AccountBalance.
- All dates successfully converted, enabling accurate time-series analysis.
- **14 accounts** were found with **negative balances**, representing overdrafts—important early indicators of financial stress.

- Numeric ranges validated:
    - TransactionAmount ranges from **–59,669** to **166,083**
    - AccountBalance ranges from **–30,766** to **184,008**

Overall Finding:
The data is clean, structured, and trustworthy for financial risk analytics.

2. Transaction Behaviour Insights

Monthly Trends

- Highest monthly transaction activity occurred in **June 2023 (41 transactions, ₹18.9 lakh)**.
- Year-wise totals:
    - **2023:** 209 transactions (₹1.14 crore)
    - **2024:** 113 transactions (₹55.0 lakh)

Insight:
Customer engagement decreased by ~46% from 2023 to 2024, indicating a possible decline in activity or user base.

3. Net Inflow Performance

Top Net Inflow Accounts

- Highest positive net inflow: **ACC76549 – ₹1,78,514**
- Other strong performers include: **ACC39544 (₹1,42,078)**, **ACC64022 (₹1,20,503)**, **ACC95164 (₹1,16,682)**

Bottom/Zero Net Inflow Accounts

- Nearly **70% of accounts** show **no positive inflow** (Total_Inflow = Total_Outflow = 0).
- Many accounts appear **inactive or dormant**.

Insight:
A major portion of customers are not actively adding funds, suggesting low engagement or dormant status.

4. Dormancy & Activity Analysis

- Accounts with **> 60-day inactivity** flagged as dormant:
    - Examples: **ACC10117 (298 days), ACC11062 (235 days), ACC15228 (329 days)**.
- Activity segmentation:
    - **100% of the sample shown** falls under **Low Activity** (<10 transactions).

Insight:
Dormancy risk is very high; multiple customers show long gaps between transactions.

5. Segmentation Findings (Balance & Volume)

- Majority of customers fall under:

    - **Medium Balance (₹50k–₹1 lakh)**
    - **Low Volume (<10 transactions)**
- Example segments:
    - **High Balance, Low Volume:** CUST1555 (₹1.03 lakh), CUST1749 (₹1.25 lakh)
    - **Low Balance, Low Volume:** CUST1467 (₹44,806), CUST1547 (₹42,838)

Insight:
Even high-balance customers remain low-engagement, indicating missed cross-sell or up-sell opportunities.

6. Financial Risk Insights

Overdraft & Negative Balance Accounts (14 identified)

Examples:

- ACC42467: –₹30,766
- ACC19156: –₹9,649
- ACC72197: –₹14,934

These accounts require monitoring due to potential credit risk.

Volatility Findings (Coefficient of Variation – CV)

- High volatility accounts (CV > 0.7):
    - ACC11285 (CV = 1.12)
    - ACC12334 (CV = 0.74)
    - ACC19156 (CV = 0.95)

Insight:
These accounts show unstable balances → risk of sudden negative turns.

Anomaly Detection

- **7 IQR anomalies** & **4 Z-score anomalies** detected.
- Very high deviation examples:
    - Debit of **–₹59,669**
    - Transfer of **₹1,66,083**
    - Credit of **₹1,42,081**

Insight:
Irregularly large movements indicate possible unusual or risky transactions.

7. Hypothesis Testing Insights

Test 1: Do high-volume accounts have higher balances?

- p-value = **0.869**
- **Result:** Fail to reject H0

→ No evidence that transaction volume increases balance.

Test 2: Segment Comparison (Low vs Medium vs High balance groups)

- p-values: **0.56–0.57**
- **Result:** No significant difference found.

Insight:
Balance levels and transaction volume are **not statistically related** in this dataset.

FINAL SUMMARY OF FINDINGS

- Customer transaction activity is **decreasing significantly** over time.
- Most accounts display **very low transaction volume**, indicating disengagement.
- **Dormancy risk is high** (multiple accounts inactive >200 days).
- **Overdrafts and high-volatility accounts** highlight financial risk pockets.
- **Net inflow performance** is heavily skewed — only a few accounts contribute positively.
- High-value customers are often **low-engagement**, showing cross-sell opportunity.
- There is **no statistical relationship** between transaction volume and balance.

FINAL RECOMMENDATIONS

- Proactively engage low-activity accounts with reminders, offers, and targeted nudges.
- Introduce periodic alerts for accounts showing long inactivity (>200 days) to reduce dormancy risk.
- Monitor overdraft and high-volatility accounts with stricter controls and automated risk flags.
- Strengthen credit controls for customers frequently touching negative balances.
- Prioritize retention campaigns for high-balance but low-engagement customers (cross-sell opportunity).
- Improve transaction monitoring dashboards to track monthly inflow, outflow, and anomalies.
- Encourage digital payments and auto-debit setups to boost transaction frequency.
- Review the customer journey for inactive users to identify friction points and service gaps.

Video Link:

https://drive.google.com/file/d/1ItXhXRkqWyfsF2B0iFwEsTPG_s-HqPMF/view?usp=sharing