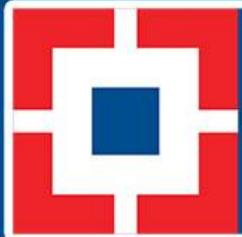# HDFC DATASET ASSESSMENT DOCUMENT (DAD)

1.Dataset Overview

Item	Details
Dataset Title	HDFC Bank Customer, Transaction & Branch Data
Domain / Theme	Banking & Financial Analytics
Source	Sample project dataset (user-provided Excel file)
File Type	Excel (.xlsx)
No. of Rows	~1,0000+ (approximate)
No. of Columns	15

2.Data Structure Summary

Column Name	Data Type	Dimension / Measure	Description
Customer_ID	String / Number	Dimension	Unique ID for each customer

Column Name	Data Type	Dimension / Measure	Description
Customer_Type	String	Dimension	Type of customer (Individual / Business / Employee)
Age	Number	Dimension	Customer's age
City	String	Dimension	City or branch location
Branch_ID	Number	Dimension	Unique identifier for each branch
Region	String	Dimension	Region name (North, South, East, West)
Transaction_ID	String	Dimension	Unique ID for each transaction
Transaction_Date	Date	Dimension	Date of transaction
Transaction_Amount	Number	Measure	Amount of transaction
Account_Balance	Number	Measure	Current account balance
Revenue	Number	Measure	Branch-level revenue
Expense	Number	Measure	Branch-level expense
Profit	Number	Measure	Calculated profit (Revenue - Expense)
Investment_Amount	Number	Measure	Investment made by customer
Account_Type	String	Dimension	Type of account (Savings / Current / Business)

3. Data Quality Assessment

Aspect Checked	Observations	Action Needed
Missing Data	Some blank values in Age and Investment columns	Replace with average or "Not Available"
Duplicates	Few duplicate Customer_IDs or Branch_IDs	Use Tableau "Remove Duplicates" / Excel Power Query
Outliers	High transaction or balance values	Apply filters or box plot analysis
Incorrect Data Types	Transaction_Date in text format in some rows	Convert to proper date format
Consistency Issues	Inconsistent city or region spellings	Standardize values (Mumbai/MUMBAI → Mumbai)

4.Key Metrics & KPIs Identified (Initial)

KPI	Why It Matters
Total Revenue	Measures total income generated by all branches
Total Profit	Shows bank profitability
Total Expense	Helps evaluate cost efficiency
Total Transactions	Reflects banking activity level
Total Investments	Indicates customer trust and growth potential
Average Account Balance	Represents customer financial strength
No. of Customers	Measures customer base size
No. of Branches	Shows bank presence and coverage
Business Customer %	Indicates corporate relationship strength
Average Transaction Value	Shows spending and transaction size pattern

5.Assumptions / Risks

- • Some customer details (like age or investment) are incomplete.
- • Dataset may represent a sample, not full bank data.
- • Data cleaning and standardization are required before analysis.
- • Profit values depend on accurate expense and revenue entries.