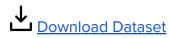# Banking Data Analysis for Strategic Insights

## Introduction

A Banking Data Analysis project involves leveraging data analytics techniques to extract meaningful insights from a bank's data. This type of project can provide valuable information for decision-making
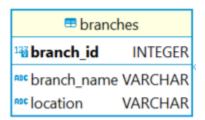
## Dataset Information

### Customers:

⬇ Download Dataset

**Description :** Represents information about bank customers, such as their name and date of birth.

**Source Type** : Database

**Schema Information**



### Accounts:

⬇ Download Dataset

**Description :** Stores details about customer accounts, including the account type (e.g., savings or checking) and the account balance.

**Source Type** : JSON

### Transactions:

⬇ Download Dataset

**Description :** Records transactions associated with customer accounts, specifying the transaction type (e.g., deposit, withdrawal, or transfer), amount, and date.

**Source Type** : CSV

## Employees:

⤓ Download Dataset

**Description :** Contains details about bank employees, including their name and position.

**Source Type** : XML

**Schema Information**

| employee_id | branch_id | first_name | last_name | position |
|---|---|---|---|---|
| INTEGER | INTEGER | STRING | STRING | STRING |

## Loans:

⤓ Download Dataset

**Description :** Introduces the concept of loans, including the loan amount, interest rate, start date, end date, and status.

**Source Type** : Parquet

## Payment History:

⤓ Download Dataset

**Description :** Records the payment history for loans, including the payment date and the amount paid.

**Source Type** : Parquet

## Branches:

⤓ Download Dataset

**Description :** Represents different bank branches, storing information such as branch name and location.

**Source Type** : JSON

# Summary of the Reports Required

## Basic Reports

### Problem Statement

Write a spark dataframe to show the balance amount for an account_id = 1:

**Expected Output**

| 123 balance |
|---|
| 5,000 |

### Problem Statement

List Transactions for an account_id = 1:

**Expected Output**

```
|transaction_id|account_id|transaction_type|amount|transaction_date   |
+--------------+----------+----------------+------+-------------------+
             1|         1|Deposit         |  1000|2023-01-15 08:30:00|
             2|         1|Withdrawal      |   500|2023-02-02 12:45:00|
            10|         1|Deposit         |  1200|2023-10-01 10:00:00|
```

### Problem Statement

List Accounts with a zero balance:

**Expected Output**

### Problem Statement

Find the Oldest Customer:

**Expected Output**

```
customer_id|first_name|last_name|date_of_birth|
-----------+----------+---------+-------------+
          3|Alice     |Johnson  |1975-02-10   |
```

## 🧑‍💻 Problem Statement

Calculate the Total Interest Earned Across All Accounts:

**Expected Output**

```
total_interest_earned|
---------------------+
                     |
```

# Accounts Reports

1.  List All Accounts with Customer Information:

    **Expected Output**

```
customer_id|first_name|last_name|account_id|account_type|balance|
-----------+----------+---------+----------+------------+-------+
          1|John      |Doe      |         1|Savings     |   5000|
          1|John      |Doe      |         2|Checking    |   1000|
          2|Jane      |Smith    |         3|Savings     |   8000|
          3|Alice     |Johnson  |         4|Checking    |   3000|
          2|Jane      |Smith    |         5|Checking    |   2500|
          3|Alice     |Johnson  |         6|Savings     |   6000|
          4|Sarah     |Jones    |         7|Checking    |  12000|
          5|David     |Brown    |         8|Savings     |   3000|
          1|John      |Doe      |         9|Savings     |   5000|
          1|John      |Doe      |        10|Checking    |   1000|
          2|Jane      |Smith    |        11|Savings     |   8000|
          3|Alice     |Johnson  |        12|Checking    |   3000|
          2|Jane      |Smith    |        13|Checking    |   2500|
          3|Alice     |Johnson  |        14|Savings     |   6000|
          4|Sarah     |Jones    |        15|Checking    |  12000|
          5|David     |Brown    |        16|Savings     |   3000|
```
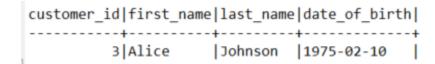
2.  Calculate Total Balance for Each Customer:

    **Expected Output**

```
customer_id|first_name|last_name|total_balance|
-----------+----------+---------+-------------+
          1|John      |Doe      |        12000|
          2|Jane      |Smith    |        21000|
          3|Alice     |Johnson  |        18000|
          4|Sarah     |Jones    |        24000|
          5|David     |Brown    |         6000|
```

3. Find Customers with Multiple Accounts:

   **Expected Output**

```
customer_id|num_accounts|
-----------+------------+
          1|           4|
          2|           4|
          3|           4|
          4|           2|
          5|           2|
```

# Customer Transactions Reports

1. List Transactions with Account and Customer Information:

   **Expected Output**

```
transaction_id|transaction_type|amount|transaction_date   |customer_id|first_name|last_name|account_id|account_type|
--------------+----------------+------+-------------------+-----------+----------+---------+----------+------------+
             1|Deposit         |  1000|2023-01-15 08:30:00|          1|John      |Doe      |         1|Savings     |
             2|Withdrawal      |   500|2023-02-02 12:45:00|          1|John      |Doe      |         1|Savings     |
             3|Deposit         |  2000|2023-03-10 15:20:00|          1|John      |Doe      |         2|Checking    |
             4|Withdrawal      |  1000|2023-04-05 10:10:00|          2|Jane      |Smith    |         3|Savings     |
             5|Deposit         |  1500|2023-05-20 09:00:00|          3|Alice     |Johnson  |         4|Checking    |
             6|Deposit         |  2000|2023-06-12 11:30:00|          2|Jane      |Smith    |         5|Checking    |
             7|Withdrawal      |   800|2023-07-08 14:15:00|          2|Jane      |Smith    |         3|Savings     |
             8|Deposit         |  3000|2023-08-22 16:45:00|          1|John      |Doe      |         2|Checking    |
             9|Withdrawal      |  1500|2023-09-14 09:30:00|          3|Alice     |Johnson  |         4|Checking    |
            10|Deposit         |  1200|2023-10-01 10:00:00|          1|John      |Doe      |         1|Savings     |
```

2. Calculate Average Transaction Amount:

   **Expected Output**

```
avg_transaction_amount|
----------------------+
                1450.0|
```

3. Identify High-Value Customers with Total Balance:

**Expected Output**

```
customer_id|first_name|last_name|total_balance|
-----------+----------+---------+-------------+
          1|John      |Doe      |        12000|
          2|Jane      |Smith    |        21000|
          3|Alice     |Johnson  |        18000|
          4|Sarah     |Jones    |        24000|
```

4. List Employees and Their Assigned Customers:

**Expected Output**

```
employee_id|first_name|last_name|position|customer_id|customer_first_name|customer_last_name|
-----------+----------+---------+--------+-----------+-------------------+------------------+
          1|Mike      |Johnson  |Manager |          3|Alice              |Johnson           |
          1|Mike      |Johnson  |Manager |          3|Alice              |Johnson           |
          2|Emily     |Williams |Teller  |          1|John               |Doe               |
          2|Emily     |Williams |Teller  |          1|John               |Doe               |
          2|Emily     |Williams |Teller  |          2|Jane               |Smith             |
          2|Emily     |Williams |Teller  |          2|Jane               |Smith             |
          2|Emily     |Williams |Teller  |          5|David              |Brown             |
          2|Emily     |Williams |Teller  |          5|David              |Brown             |
          3|Robert    |Davis    |Teller  |          1|John               |Doe               |
          3|Robert    |Davis    |Teller  |          1|John               |Doe               |
          3|Robert    |Davis    |Teller  |          3|Alice              |Johnson           |
          3|Robert    |Davis    |Teller  |          3|Alice              |Johnson           |
          4|Olivia    |Wilson   |Teller  |          2|Jane               |Smith             |
          4|Olivia    |Wilson   |Teller  |          2|Jane               |Smith             |
          4|Olivia    |Wilson   |Teller  |          4|Sarah              |Jones             |
          4|Olivia    |Wilson   |Teller  |          4|Sarah              |Jones             |
          5|Daniel    |Johnson  |Analyst |           |                   |                  |
          6|Sophia    |Clark    |Manager |           |                   |                  |
          7|Mike      |Johnson  |Manager |           |                   |                  |
          8|Emily     |Williams |Teller  |           |                   |                  |
          9|Robert    |Davis    |Teller  |           |                   |                  |
         10|Olivia    |Wilson   |Teller  |           |                   |                  |
         11|Daniel    |Johnson  |Analyst |           |                   |                  |
         12|Sophia    |Clark    |Manager |           |                   |                  |
```

5. Calculate the Total Number of Transactions for Each Account Type:

**Expected Output**

```
account_type|num_transactions|
------------+----------------+
Checking    |               5|
Savings     |               5|
```

6. Find Customers with No Accounts:

   **Expected Output**

```
account_id|latest_transaction_date|
----------+-----------------------+
         1|2023-10-01 10:00:00    |
         2|2023-08-22 16:45:00    |
         3|2023-07-08 14:15:00    |
         4|2023-09-14 09:30:00    |
         5|2023-06-12 11:30:00    |
         6|                       |
         7|                       |
         8|                       |
         9|                       |
        10|                       |
        11|                       |
        12|                       |
        13|                       |
        14|                       |
        15|                       |
        16|                       |
```

7. Find Customers with No Accounts:

   **Expected Output**

```
customer_id|first_name|last_name|
-----------+----------+---------+
          6|Emma      |Miller   |
```

8. List the Latest Transaction for Each Account:

   **Expected Output**

```
customer_id|first_name|last_name|total_withdrawals|
-----------+----------+---------+-----------------+
          1|John      |Doe      |              500|
          2|Jane      |Smith    |             1800|
          3|Alice     |Johnson  |             1500|
```

9.  Calculate the Total Withdrawals for Each Customer:

    **Expected Output**

```
customer_id|first_name|last_name|total_withdrawals|
-----------+----------+---------+-----------------+
          1|John      |Doe      |              500|
          2|Jane      |Smith    |             1800|
          3|Alice     |Johnson  |             1500|
```

10. Find Duplicate Transactions:

    **Expected Output**

```
transaction_id|duplicate_count|
--------------+---------------+
```