

Data Catalog - BCSB: Accubbranch Analysis

Project Overview

3 files have been produced for Accubbranch to perform their analysis

- 1. **Account Data** (`account_data.parquet`):
- 2. **Transaction Data** (`transaction_cleaned.parquet`):
- 3. **Historical FYE Deposits per Branch** (`five_yr_history.parquet`):

Notes

1. `account_data.parquet`

This file contains a detailed snapshot of all active customer accounts as of 6/30/2025. It is enriched with customer, organization, and primary address details to match requested format.

- **Source:** Generated by the Account Data Pipeline (`accubbranch.core.process_account_data`).
- **Description:** Provides a comprehensive "who and what" view of the bank's customer base, linking customers to their accounts and associated branches.
- **Table Schema:**

Column	Description	Source Field(s)
Primary Key	A unique identifier for the tax-reporting customer.	<code>taxrptforpersnbr</code> ,
	Prefixed with 'P' for individuals, 'O' for businesses.	<code>taxrptfororgnbr</code>
Address	The primary street address of the customer.	<code>text1</code> , <code>text2</code> , <code>text3</code>
City	The city of the customer's primary address.	<code>cityname</code>
State	The state of the customer's primary address.	<code>statedcd</code>
Zip	The zip code of the customer's primary address.	<code>zipcd</code>
Branch Associated	The branch associated with the customer's account.	<code>branchname</code>
Account Type	A cleaned product type. See Account Type Mapping below.	<code>mjaccttypcd</code> , <code>loanofficer</code>

Column	Description	Source Field(s)
Date Account Opened	The date the account contract was signed.	contractdate
Current Balance	The net balance of the account as of the snapshot date.	Net Balance
Original Balance (Loans)	The original principal amount for loan accounts; null for deposit accounts.	orig_ttl_loan_amt
Date of Birth	The date of birth for individual customers; null for businesses.	datebirth

2. transaction_cleaned.parquet

This file contains a detailed log of all completed customer transactions over a one-year period.

- **Source:** Generated by the Transaction Data Pipeline (transactions.core.process_transaction_data).
- **Table Schema:**

Column	Description	Source Field(s)
Customer Unique ID	A unique identifier for an entity.	P/O + persnbr/orgnbr
Date of Transaction	The date the transaction occurred.	actdatetime
Time of Transaction	The time the transaction occurred.	actdatetime
Branch of Transaction	The branch where the transaction was processed. All digital transactions are assigned to Main Office.	orgname
Type of Teller	The channel through which the transaction was initiated. See Transaction Channel Mapping below.	rtxnsourcecd
Type of Transaction	A description of the transaction type	rtxntypdesc
Account Type	The cleaned product type of the account involved in the transaction.	mjaccttypcd, loanofficer

3. five_yr_history.parquet

This file provides a summarized, time-series view of total deposit balances by branch over five years.

- **Source:** Generated by the Historical Analysis Pipeline
(`accubbranch.core.process_historical_data`).
- **Table Schema:**

Column	Type	Description
(Index)	string	branchname (The name of the branch)
2020-12-31	float	Total deposit balance for the branch on this date.
2021-12-31	float	Total deposit balance for the branch on this date.
2022-12-30	float	Total deposit balance for the branch on this date.
2023-12-29	float	Total deposit balance for the branch on this date.
2024-12-31	float	Total deposit balance for the branch on this date.

Key Business Logic & Definitions

Account Type Mapping

The `mjaccttypcd` from the source system is mapped to a standard `Account Type` using the following logic.

```
# System Code -> Standard Name
ACCOUNT_TYPE_MAPPING = {
    'CML': 'Commercial Loan',
    'MLN': 'Commercial Loan',
    'CNS': 'Consumer Loan',
    'MTG': 'Residential Loan',
    'CK': 'Checking',
    'SAV': 'Savings',
    'TD': 'CD'
}
```

- Any product major codes not in this dictionary are mapped to 'Other' and are excluded from the account & transaction detail.

Small Business Loan Identification

A loan is classified as a `Small Business Loan` if it meets two conditions:

1. Its initial `Account Type` is `Commercial Loan`.

2. The assigned loanofficer is one of the following:

```
SMALL_BUSINESS_OFFICERS = ['EBL PROGRAM ADMIN', 'SBLC LOAN OFFICER']
```

Transaction Channel (Type of Teller) Mapping

The `rtxnsourcecd` field is mapped to the Type of Teller to identify the transaction channel.

- **In-Person / Branch:** ONLI
- **ATM:** ATM
- **Digital Banking:** WWW (Note: All WWW transactions are assigned to the 'BCSB - Main Office' branch)
- **Phone Banking:** VRU
- **Third-Party / API:** API (e.g., card systems, vendor-generated)
- **Real Time Payment:** RTP
- **Batch Processing:** BAT (System-generated, often internal processing. Should be used with caution as 'COCC' is the branch for over 99% of these records).

ITM Operations

To classify ITM Operations, there is a Branch assigned to all ITM transactions:

- BCSB - ITM OPS

Data Exclusions

The following records are explicitly filtered out during the pipeline:

- **Accounts:** Non-active accounts (not in 'ACT', 'NPFM', 'DORM' status).
- **Organization Types:** Municipal (MUNI) and Trust (TRST) customer types are excluded from the account analysis.
- **Product Types:** ACH Manager products (CI07) are excluded.
- **Transactions:** Only completed transactions (`CURRTXNSTATCD = 'C'`) are included.