**Bank\_Entry Dataset Overview**

* **Rows (records):** 21,012
* **Columns (features):** 20
* Represents **bank customers** with their demographics, financial details, account activity, and loan/credit information.

**🔹 1. Customer Demographics**

* **id** → Unique record ID (1–21,012).
* **customer\_id** → Customer code (some repeated = same customer with multiple entries).
* **age** → 18–75 years, average ≈ 45 years.
* **gender** → Male, Female, Other (Female most common ~7,040).
* **region** → 5 categories (e.g., South = most frequent ~4,650).
* **marital\_status** → 4 categories (Single, Married, Divorced, Widowed; “Divorced” most common ~5,436).
* **dependents** → Number of dependents (0–5, avg ≈ 2.4).
* **education** → 7 categories (PhD most common ~4,959).
* **employment\_status** → 5 categories (Employed, Unemployed, Self-Employed, Retired, Student; “Self-Employed” most common ~4,954).

**🔹 2. Financial Information**

* **income** → Annual income:
  + Range: 20k – 200k
  + Avg ≈ 99k
  + Std dev ≈ 48k (varied income levels).
* **account\_type** → 6 categories (Savings most common ~6,460).
* **balance** → Account balance:
  + Range: ~15 – 999k
  + Avg ≈ 255k
  + Median ≈ 134k
  + Very wide spread (many low balances, few very high).
* **num\_transactions** → 0–500 per period (avg ≈ 131).
* **last\_transaction\_date** → Dates from 2023–2025 (customers are active recently).

**🔹 3. Loan & Credit Profile**

* **loan\_status** → 4 categories (Approved, Pending, Rejected, Defaulted).
  + “Rejected” is most frequent (~6,480).
* **loan\_amount** → Loan values:
  + Range: 0 – 499k
  + Avg ≈ 80k
  + Median ≈ 31k (many small loans, few very large).
* **credit\_score** → Classic FICO-like scores:
  + Range: 300–850
  + Avg ≈ 576 (slightly below “good credit”).
  + Distribution suggests mix of risky & safe customers.

**🔹 4. Banking Behavior**

* **card\_type** → 10 card types (Visa most common ~3,815, followed by MasterCard & Amex).
* **online\_banking** → Boolean (~50% customers use online banking).
* **email\_alerts** → Boolean (~50% customers use email alerts).

**📌 Key Insights**

1. **Customers are middle-aged** (avg 45) with 2–3 dependents.
2. **Income is varied**: wide gap between low (20k) and high earners (200k).
3. **Balances & loans are skewed**: most have moderate amounts, but a few extremely high.
4. **Credit scores average ~576**, so many are borderline risky borrowers.
5. **Loan rejection is high**, possibly linked to lower credit scores.
6. **Digital banking adoption is ~50%** (scope for growth).