✅ Overall Insights

1. Data Loading & Structure

- Three datasets are loaded:
    - **accounts_df** – account-level details (account_id, customer_id, account_type, status, dates).
    - **customer_df** – customer demographics (region, occupation, etc.).
    - **transactions_df** – transaction details (txn_date, amount, channel, etc.).
- These datasets are then merged to form a **final master dataframe**:
  txn_finaldf = accounts + customers + transactions

✅ 2. Basic Data Exploration

Preview & Structure

- head(), info(), describe() used to understand:
    - Number of rows & columns
    - Data types
    - Summary statistics
    - Presence of categorical fields like region, occupation, status, account_type

Value Counts

- Most common:
    - **customer_id** occurrence in accounts
    - **account_id** frequency in transactions

Insight:

- Some customers have multiple accounts.
- Some accounts have high transaction density.

✅ 3. Regional Customer Insights

Region-wise distribution

- txn_finaldf['region'].value_counts()

Insight:

- Shows which regions hold the largest share of customers.

Region × Occupation grouping

- groupby(['region','occupation'])['occupation'].count()

Insight:

- Helps identify **occupation mix in each region** (e.g., salaried vs self-employed concentration).
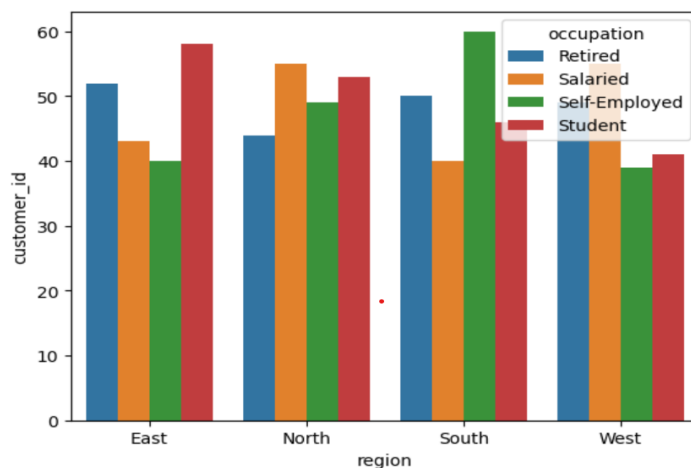
Pivot Table (Region × Occupation)

- Displays **customer count** across region and occupation combinations.

✅ 4. Visual Insights Using Seaborn

Barplot: Region vs Customer Count (with Occupation hue)

`<Axes: xlabel='region', ylabel='customer_id'>`

Insight:

- Reveals which **occupation dominates each region**.
- Allows comparison of customer volume across regions.

Heatmap: Region × Occupation

`<Axes: xlabel='occupation', ylabel='region'>`

Insight:

- Visually highlights regions with significantly higher customers of certain occupations.

✅ 5. Account Type Insights

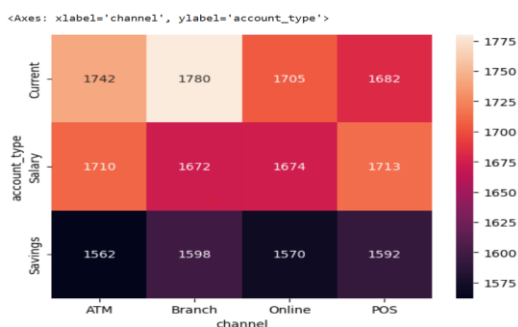**Grouping by Account Type & Channel**

Metrics computed:

- **txn_count** – number of transactions

- **total_amt** – total amount

- **avg_amt** – average amount

Insights:

- Identifies which **account types** (Savings, Current, etc.) generate:

    o   The most transactions

    o   The highest total amount

    o   Highest average ticket size

Pivot Table: Account Type × Channel (Transaction Count)

`<Axes: xlabel='channel', ylabel='account_type'>`

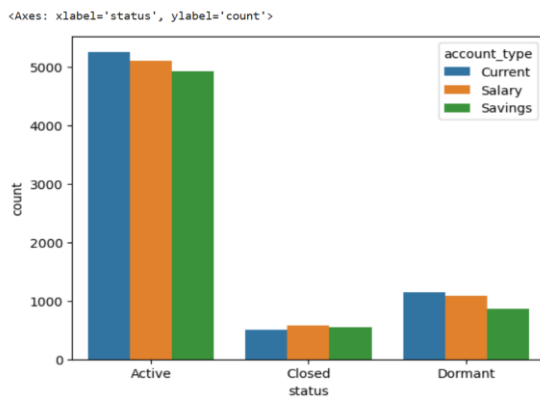- Heatmap visualizes which **transaction channels** (ATM, Online, Branch) are used most by each account type.

Insight:

- Current accounts use Branch channels more.

- POS dominate for salary accounts.

✅ 6. Account Status Insights (Open vs Closed)

Group by Status & Account Type

- Value counts shown using a barplot.

Insights:

- Shows Salary account type has **higher closure rates**.

✅ 7. Account Closure Tenure

For closed accounts:

- Tenure calculated:
  Tenure = close_date – open_date

Insights:

- Helps understand:

  - Typical account lifespan

  - Whether certain accounts close sooner

  - Whether some regions/customers churn quickly

✅ 8. Transaction Time Analysis

Transaction Month Extraction

- txn_date converted to datetime

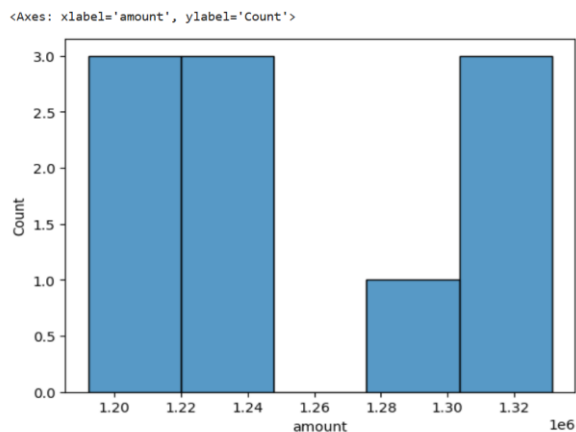- Month column created (Period format)

Monthly Amount Grouping

- Top 10 months with highest transaction volume identified.

Insights:

- Identifies **peak business months**

- Useful for seasonality and forecasting.

Histogram Plot

- Shows distribution of monthly transaction amounts.

Insight:

- Reveals monthly amounts are skewed

Final Summary of Insights

- The notebook performs a full **end-to-end exploratory analysis** on banking data.

- Covers:

  o Customer segmentation

  o Regional & occupational comparisons

  o Account type performance

  o Transaction behaviour

- Churn analysis through closed accounts
- Monthly transaction trends
- Visualization (barplots & heatmaps) highlight:
  - Customer concentration
  - Channel usage
  - Behavioural patterns across categories