1. **TransactionID:** Unique identifier for each transaction.
2. **CustomerID:** Unique identifier for each customer.
3. **CustomerDOB:** Date of birth of the customer.
4. **CustGender:** Gender of the customer.
5. **CustLocation:** Location or address of the customer.
6. **CustAccountBalance:** Account balance of the customer.
7. **TransactionDate:** Date of the transaction.
8. **TransactionTime:** Time of the transaction.
9. **STATUS\_:** Status of the transaction.
10. **bplayer0:** ?
11. **product\_id:** ID of the product involved in the transaction.
12. **category\_id:** ID of the category to which the product belongs.
13. **category\_code:** Code representing the category of the product.
14. **brand:** Brand of the product.
15. **price:** Price of the product.
16. **PaymentMode:** Payment mode used for the transaction.
17. The dataset contains 100,104 rows and 16 columns.
18. Most columns have non-null values, but some columns have missing values:
    * **CustomerDOB**: 99,727 non-null values
    * **CustLocation**: 100,088 non-null values
    * **CustAccountBalance**: 99,867 non-null values
    * **brand**: 100,103 non-null values
    * **price**: 100,103 non-null values
    * **PaymentMode**: 100,103 non-null values
19. The data types of the columns are as follows:
    * Object: 10 columns
    * Float64: 3 columns
    * Int64: 3 columns
20. The **TransactionDate** column contains dates as objects. It might be beneficial to convert this column to datetime format for time-based analysis.
21. The **TransactionTime** column contains integers. Depending on the format, it might need further processing for time-based analysis.
22. The **STATUS\_** column seems to represent the status of transactions and is encoded as float64.
23. The **price** column represents the price of products and is encoded as float64

* **CustAccountBalance**: The average customer account balance is approximately 115,155, with a standard deviation of approximately 753,757. The minimum account balance is 0, and the maximum is approximately 115,035,500.
* **TransactionTime**: The average transaction time is around 156,644, with a standard deviation of approximately 51,080. The minimum transaction time is 0, and the maximum is 235,959.
* **STATUS\_**: All values are 1, indicating a single status throughout the dataset.
* **product\_id**: The average product ID is approximately 5,376,388, with a standard deviation of approximately 8,127,078. The minimum product ID is 1,001,588, and the maximum is approximately 53,900,010.
* **category\_id**: The average category ID is approximately 2.055506e+18, with a standard deviation of approximately 1.520079e+16. The minimum category ID is 2.053014e+15, and the maximum is approximately 2.172371e+18.
* **price**: The average price is approximately 32,633, with a standard deviation of approximately 35,987. The minimum price is 79, and the maximum is 231,666.
* **DOB\_Day**: The average day of birth is approximately 14, with a standard deviation of approximately 9.24. The minimum day is 1, and the maximum is 31.
* **DOB\_Month**: The average month of birth is approximately 6, with a standard deviation of approximately 3.54. The minimum month is 1, and the maximum is 12.
* **DOB\_Year**: The average year of birth is approximately 85, with a standard deviation of approximately 9.06. The minimum year is 0, and the maximum is 99.

CustGender value counts:

Male 399191

Female 277484

Name: CustGender, dtype: int64

CustLocation value counts:

BANGALORE 81032

MUMBAI 67442

NEW DELHI 55424

GURGAON 47263

KOLKATA 45562

...

VIDYAGIRI 1

DESAI RD MUMBAI 1

MUMBAI ANDHERI EAST 1

SHERI TALUKA HAVELI VIMAN NGR PUNE 1

KUNNAMKULAM 1

Name: CustLocation, Length: 8217, dtype: int64

CustAccountBalance value counts:

0 4525

17069 1750

1 1149

2 1072

3 1033

...

24776 1

15012 1

601851 1

77388 1

103220 1

Name: CustAccountBalance, Length: 69262, dtype: int64

TransactionDate value counts:

2023-07-08 27256

2023-08-13 26916

2023-06-08 26578

2023-08-14 25594

2023-08-15 24165

2023-08-21 22982

2023-12-08 22433

2023-11-08 21830

2023-09-08 21817

2023-10-08 21646

2023-08-27 21465

2023-08-20 21333

2023-08-17 21113

2023-05-08 21104

2023-08-08 21038

2023-02-08 20945

2023-04-08 20675

2023-03-08 20612

2023-01-08 20430

2023-08-16 20412

2023-08-18 19451

2023-08-19 18712

2023-08-22 18554

2023-08-25 18541

2023-08-24 18022

2023-08-26 17738

2023-08-23 17367

2023-08-28 17347

2023-08-29 16388

2023-09-26 12460

2023-08-31 11125

2023-03-09 8936

2023-09-25 8164

2023-09-27 7447

2023-09-22 6971

2023-10-21 3656

2023-09-23 3485

2023-09-30 1951

2023-08-30 13

2023-10-16 3

Name: TransactionDate, dtype: int64

STATUS\_ value counts:

1.0 676675

Name: STATUS\_, dtype: int64

category\_code value counts:

electronics.smartphone 294001

electronics.clocks 31559

computers.notebook 30903

electronics.audio.headphone 28160

electronics.video.tv 23865

...

apparel.skirt 15

country\_yard.furniture.bench 12

apparel.shorts 8

apparel.costume 8

apparel.belt 8

Name: category\_code, Length: 122, dtype: int64

brand value counts:

samsung 126017

apple 109761

xiaomi 64033

huawei 29844

lg 10876

...

chenri 1

blg 1

vasin 1

bonvini 1

worwo 1

Name: brand, Length: 1222, dtype: int64

price value counts:

87800 8261

87801 7443

11762 6833

127393 6539

14573 4943

...

7752 1

72117 1

48195 1

375 1

42948 1

Name: price, Length: 10454, dtype: int64

PaymentMode value counts:

Online Payment 307604

Credit Card 115980

Cash on Delivery 107868

Debit Card 62220

Cash 44858

Bank Transfer 38145

Name: PaymentMode, dtype: int64

1. **CustGender**:
   * Male: 399,191
   * Female: 277,484
2. **CustLocation**:
   * BANGALORE: 81,032
   * MUMBAI: 67,442
   * NEW DELHI: 55,424
   * GURGAON: 47,263
   * KOLKATA: 45,562
   * ... (and more)
3. **CustAccountBalance**: There are many unique values. The most frequent ones are 0, 17,069, 1, 2, and 3.
4. **TransactionDate**: The dates are spread across different days, with some dates having significantly more transactions than others.
5. **STATUS\_**: All transactions have a status of 1.
6. **category\_code**: There are many unique categories, with "electronics.smartphone" being the most common.
7. **brand**: There are many unique brands, with "samsung" and "apple" being the most common.
8. **price**: The prices vary, with some prices being more common than others.
9. **PaymentMode**: The most common payment mode is "Online Payment", followed by "Credit Card" and "Cash on Delivery".
10. **CustGender Counts**:
    * Male: 399,191
    * Female: 277,484
11. **CustLocation Counts**:
    * BANGALORE: 81,032
    * MUMBAI: 67,442
    * NEW DELHI: 55,424
    * GURGAON: 47,263
    * KOLKATA: 45,562
12. **CustAccountBalance Counts**: (Assuming we're considering the top 5 most frequent balances)
    * 0: 4,525
    * 17,069: 1,750
    * 1: 1,149
    * 2: 1,072
    * 3: 1,033
13. **TransactionDate Counts**: (Assuming we're considering the top 5 most frequent dates)
    * 2023-07-08: 27,256
    * 2023-08-13: 26,916
    * 2023-06-08: 26,578
    * 2023-08-14: 25,594
    * 2023-08-15: 24,165
14. **CategoryCode Counts**: (Assuming we're considering the top 5 most frequent category codes)
    * electronics.smartphone: 294,001
    * electronics.clocks: 31,559
    * computers.notebook: 30,903
    * electronics.audio.headphone: 28,160
    * electronics.video.tv: 23,865
15. **Brand Counts**: (Assuming we're considering the top 5 most frequent brands)
    * samsung: 126,017
    * apple: 109,761
    * xiaomi: 64,033
    * huawei: 29,844
    * lg: 10,876
16. **Price Counts**: (Assuming we're considering the top 5 most frequent prices)
    * 87,800: 8,261
    * 87,801: 7,443
    * 11,762: 6,833
    * 127,393: 6,539
    * 14,573: 4,943
17. **PaymentMode Counts**: (Assuming we're considering the top 5 most frequent payment modes)
    * Online Payment: 307,604
    * Credit Card: 115,980
    * Cash on Delivery: 107,868
    * Debit Card: 62,220
    * Cash: 44,858

**1. Dataset Overview:**

* The dataset comprises 100,104 rows and 16 columns.
* It includes transaction-related information such as TransactionID, CustomerID, TransactionDate, TransactionTime, and PaymentMode.
* Customer details like CustomerDOB, CustGender, and CustLocation are also present.
* Product-related information includes product\_id, category\_id, category\_code, brand, and price.

**2. Data Quality:**

* Missing Values: Some columns have missing values, including CustomerDOB, CustLocation, CustAccountBalance, brand, price, and PaymentMode.
* Data Types: TransactionDate needs to be converted to datetime format for time-based analysis.
* Redundancy: The STATUS\_ column contains only one unique value (1.0), indicating potential redundancy.

**3. Insights from Data:**

* **Customer Information:**
  + Gender Distribution: Male customers (399,191) outnumber female customers (277,484).
  + Location Distribution: Bangalore (81,032), Mumbai (67,442), and New Delhi (55,424) are among the top locations.
  + Account Balance: The most frequent account balances are 0, 17,069, 1, 2, and 3.
* **Transaction Information:**
  + Date Analysis: The dataset contains transactions across various dates, with some dates having significantly more transactions than others.
  + Status: All transactions have a status of 1.
* **Product Information:**
  + Category: The dataset includes a wide range of product categories, with "electronics.smartphone" being the most common.
  + Brand: Samsung (126,017) and Apple (109,761) are the top brands.
  + Price: Prices vary widely, with some being more common than others.
* **Payment Information:**
  + Payment Mode: Online Payment (307,604) is the most common payment mode, followed by Credit Card and Cash on Delivery.

**4. Exploratory Analysis:**

* Further analysis can be performed to explore patterns in transaction dates, account balances, product categories, and payment modes.
* Visualization techniques such as histograms, bar plots, and pie charts can help visualize distributions and relationships in the data.

**5. Recommendations:**

* Preprocessing: Address missing values and convert data types for accurate analysis.
* Redundancy Check: Investigate the need for the STATUS\_ column if it contains only one unique value.
* Visual Exploration: Utilize visualizations to gain insights into customer behavior, product preferences, and transaction trends.

**6. Conclusion:**

* The dataset provides valuable insights into customer transactions, product preferences, and payment behavior.
* Further analysis and visualization can uncover hidden patterns and trends, aiding decision-making processes