**Graphical user interface, text

Description automatically generated**

**EDA-PART\_1:**

**[1]\*\*/All object data-type attributes have been checked for accuracy, relevance (referential integrity check - performed against customer IDs between Transactions and CustomerDemographics table), duplication and completeness\*\*/**

Findings:

* Let us consider instances where values in the order\_status column == §approved§ and other details like class and size are NaN –Probable Invalid Transactions (PIT).

Number of transactions with order status approved and product details are not not available – [PIT] COUNT: 15

* Instances where values in the Order\_Status column == §Cancelled§ - in this scenario, product details are irrelevant as they probably did not make it out of the warehouse. -- cancelled transactions (CT).

Number of transactions with order status cancelled and product details not available

[CT] Count : 1

**[2]Exploring further PITs and CTs and determining Mismanagement rate**

Validity\_check: These transactions were validated against the id's in customer\_demographic dataset.

(Out of the 15 transactions,

**Case-1:**

If cust\_id is not found against name in customer\_demographic list - possible fraud (in the scenario where a customer could have possibly been a customer in the new\_customer\_list table (where ID's have not been assigned), the nature of the table suggests a customer cannot have been assigned an id without placing an order----unless validated otherwise by company policy.

**Probability(fraud) :** x/15 , where x is inv\_trans without references in cust\_demographic table

**IN THIS CASE: 0** -- there are no customer IDs that could not be referenced back to cust\_demographic table.

**Case-2:**

Customer ID could be referenced back to customer\_demographic table.

**IN THIS CASE: 15 --** all these customer ID's deemed part of invalid transactions can be referenced back to the cust\_demographic table (with their cust\_name, DOB and deceased indicator retrieved). This is clearly the case of data mismanagement.

**Mismanagement rate:**

number of orders mismanaged (invalid transactions with order iD and no details of the order)/total number of orders.

 mismanagement rate -- 0.00075

[3] Customer DOBs

**Checked DOB ranges:**

**When checked for range, an outlier with YOB 1843 was found; identified cust\_id and Name -- customer made 59 purchases in the last 3 years and deceased indicator N – considering this is a high value customer from the stats above, record can be looked into.**

Chart, scatter chart

Description automatically generated

[0-25] 501

[25-50] 2337

[50-75] 1069

[75-100] 5

From the graph above, most customers lie in the age range 25-50.

**[4] Wealth segments.**

Graphical user interface

Description automatically generated with medium confidence

High networth and affluent customers are about the same number(s) and all up closely to the total count of the mass customer (2000 ct.).

Mass Customer 2000

High Net Worth 1021

Affluent Customer 979