Nama : Mochamma Qussay Alhindi Achmadi
NIM : 1103213087

**Task Week 3**

**A. Superstore**

**Menampilkan 5 data**



```
First 5 rows of dataset:
        Order ID           Order Date              Ship Date  \
0  CA-2016-152156 2016-11-08 00:00:00+00:00 2016-11-11 00:00:00+00:00
1  CA-2016-152156 2016-11-08 00:00:00+00:00 2016-11-11 00:00:00+00:00
2  CA-2016-138688 2016-06-12 00:00:00+00:00 2016-06-16 00:00:00+00:00
3  US-2015-108966 2015-10-11 00:00:00+00:00 2015-10-18 00:00:00+00:00
4  US-2015-108966 2015-10-11 00:00:00+00:00 2015-10-18 00:00:00+00:00

       Ship Mode Customer ID    Customer Name    Segment        Country  \
0   Second Class    CG-12520      Claire Gute   Consumer  United States
1   Second Class    CG-12520      Claire Gute   Consumer  United States
2   Second Class    DV-13045  Darrin Van Huff  Corporate  United States
3  Standard Class    SO-20335    Sean O'Donnell  Consumer  United States
4  Standard Class    SO-20335    Sean O'Donnell  Consumer  United States

            City       State Postal Code Region    Product ID  \
0      Henderson    Kentucky       42420  South  FUR-BO-10001798
1      Henderson    Kentucky       42420  South  FUR-CH-10000454
2    Los Angeles  California       90036   West  OFF-LA-10000240
3  Fort Lauderdale    Florida       33311  South  FUR-TA-10000577
4  Fort Lauderdale    Florida       33311  South  OFF-ST-10000760

          Category Sub-Category  \
0        Furniture    Bookcases
1        Furniture       Chairs
2  Office Supplies       Labels
3        Furniture       Tables
4  Office Supplies      Storage

                                        Product Name     Sales  Quantity  \
0                   Bush Somerset Collection Bookcase  261.9600      2.00
1                  "Hon Deluxe Fabric Upholstered Stacking Chairs       NaN    731.94
2  Self-Adhesive Address Labels for Typewriters b...   14.6200      2.00
3         Bretford CR4500 Series Slim Rectangular Table  957.5775      5.00
4                    Eldon Fold 'N Roll Cart System   22.3680      2.00

   Discount    Profit
0      0.00   41.9136
1      3.00    0.0000
2      0.00    6.8714
3      0.45 -383.0310
4      0.20    2.5164
```

**Informasi dari data set**

```
 1    <class 'pandas.core.frame.DataFrame'>
 2    RangeIndex: 9994 entries, 0 to 9993
 3    Data columns (total 20 columns):
 4     #   Column          Non-Null Count  Dtype
 5    ---  ------          --------------  -----
 6     0   Order ID        9994 non-null   object
 7     1   Order Date      9994 non-null   datetime64[ns, UTC]
 8     2   Ship Date       9994 non-null   datetime64[ns, UTC]
 9     3   Ship Mode       9994 non-null   object
10     4   Customer ID     9994 non-null   object
11     5   Customer Name   9994 non-null   object
12     6   Segment         9994 non-null   object
13     7   Country         9994 non-null   object
14     8   City            9994 non-null   object
15     9   State           9994 non-null   object
16    10   Postal Code     9994 non-null   object
17    11   Region          9994 non-null   object
18    12   Product ID      9994 non-null   object
19    13   Category        9994 non-null   object
20    14   Sub-Category    9994 non-null   object
21    15   Product Name    9994 non-null   object
22    16   Sales           7767 non-null   float64
23    17   Quantity        9280 non-null   float64
24    18   Discount        9827 non-null   float64
25    19   Profit          9964 non-null   float64
26    dtypes: datetime64[ns, UTC](2), float64(4), object(14)
27    memory usage: 1.5+ MB
```

**Analisis statistik dasar**

```
              Sales      Quantity      Discount        Profit
count    7767.000000   9280.000000   9827.000000   9964.000000
mean      246.118528     34.850892      8.176369     26.646715
std       676.046656    154.844305    123.079546    230.926293
min         0.444000      1.000000      0.000000  -6599.978000
25%        16.768000      2.000000      0.000000      0.200000
50%        55.176000      4.000000      0.200000      5.150850
75%       220.856000      7.000000      0.700000     22.508000
max     22638.480000   4404.900000   9099.930000   8399.976000
```

Berikut beberapa poin menarik dari hasil deskripsi statistik:

- Sales:
  - Rata-rata: 246.12
  - Maksimum: 22,638.48 (kemungkinan outlier)
- Quantity:
  - Rata-rata: 34.85
  - Maksimum: 4,404.9 (sangat tinggi, kemungkinan outlier)
- Discount:
  - Maksimum: 9,099.93 (nilai tidak masuk akal, perlu diperiksa)
- Profit:

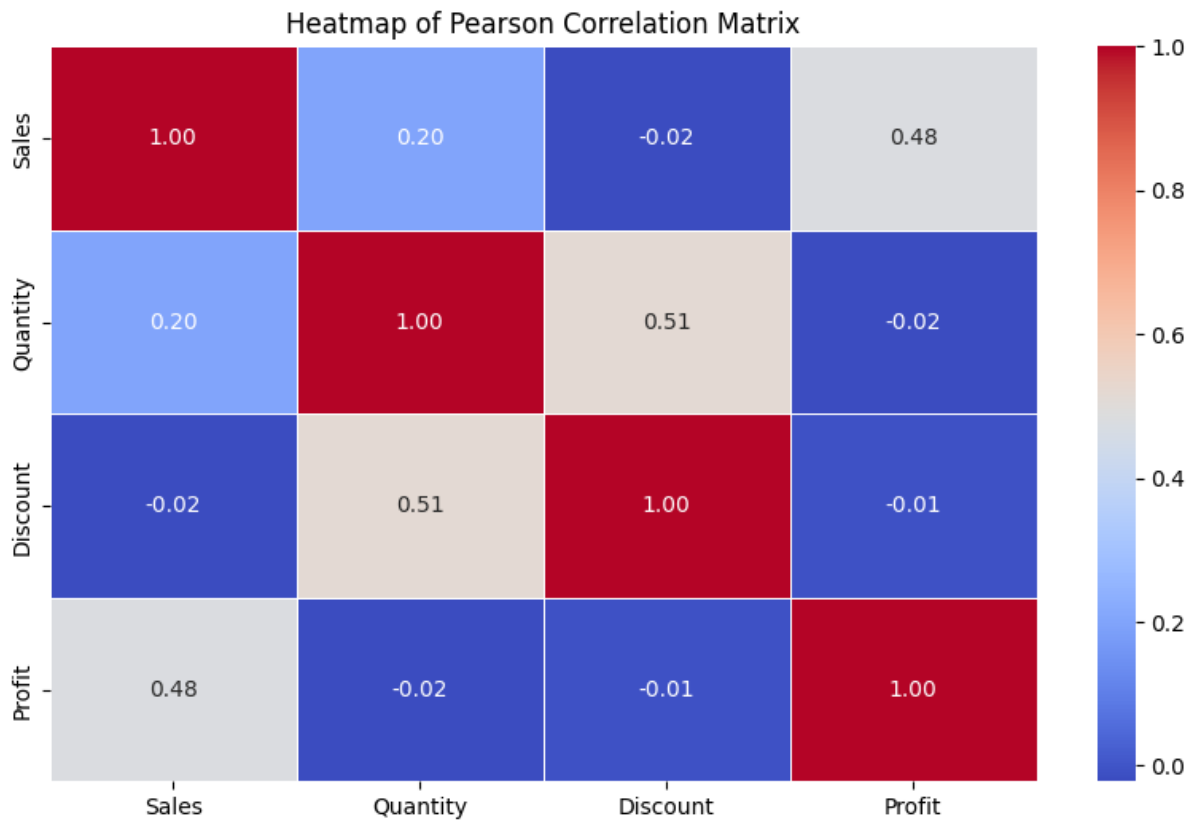- Minimum: -6,599.98 (rugi besar, kemungkinan outlier)
- Maksimum: 8,399.98

**Dengan Tukey's Fence**

```
1    Outliers in Sales: 928
2    Outliers in Quantity: 1278
3    Outliers in Discount: 1924
4    Outliers in Profit: 1841
5    |
```

- Sales → 928 data outlier
- Quantity → 1,278 data outlier
- Discount → 1,924 data outlier (kemungkinan ada kesalahan input)
- Profit → 1,841 data outlier

**Hasil Korelasi Pearson**

```
1    |
2    Pearson Correlation Matrix:
3               Sales  Quantity  Discount    Profit
4    Sales     1.000000  0.200140 -0.021627  0.478165
5    Quantity  0.200140  1.000000  0.514447 -0.021725
6    Discount -0.021627  0.514447  1.000000 -0.006146
7    Profit    0.478165 -0.021725 -0.006146  1.000000
8
```

Heatmap of Pearson Correlation Matrix

- Sales & Profit → Korelasi positif kuat (0.48) → Makin besar penjualan, makin besar profit.

- Quantity & Discount → Korelasi positif sedang (0.51) → Makin banyak jumlah barang, makin besar diskon yang diberikan.

- Sales & Quantity → Korelasi lemah (0.20) → Jumlah barang tidak selalu berbanding lurus dengan total penjualan.

- Discount & Profit → Korelasi negatif sangat lemah (-0.006) → Diskon tidak terlalu mempengaruhi profit secara langsung.

**B. Customer – Churn**