

Hasil analisis data

Oleh :Febrio

Data hasil analisis ini merupakan data yang bersumber dari [:https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales?resource=download](https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales?resource=download)

Dikarenakan data yang cukup besar maka kita harus mengecek dulu apakah data tersebut ada yang kosong atau terduplikasi maka kita bisa cek terlebih dahulu dengan preprocessing data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   Invoice ID              1000 non-null  object  
1   Branch                  1000 non-null  object  
2   City                    1000 non-null  object  
3   Customer type           1000 non-null  object  
4   Gender                  1000 non-null  object  
5   Product line            1000 non-null  object  
6   Unit price              1000 non-null  float64  
7   Quantity                1000 non-null  int64  
8   Tax 5%                  1000 non-null  float64  
9   Total                   1000 non-null  float64  
10  Date                    1000 non-null  object  
11  Time                    1000 non-null  object  
12  Payment                 1000 non-null  object  
13  cogs                    1000 non-null  float64  
14  gross margin percentage  1000 non-null  float64  
15  gross income            1000 non-null  float64  
16  Rating                  1000 non-null  float64  
dtypes: float64(7), int64(1), object(9)
memory usage: 132.9+ KB
```

Gambar 1.0

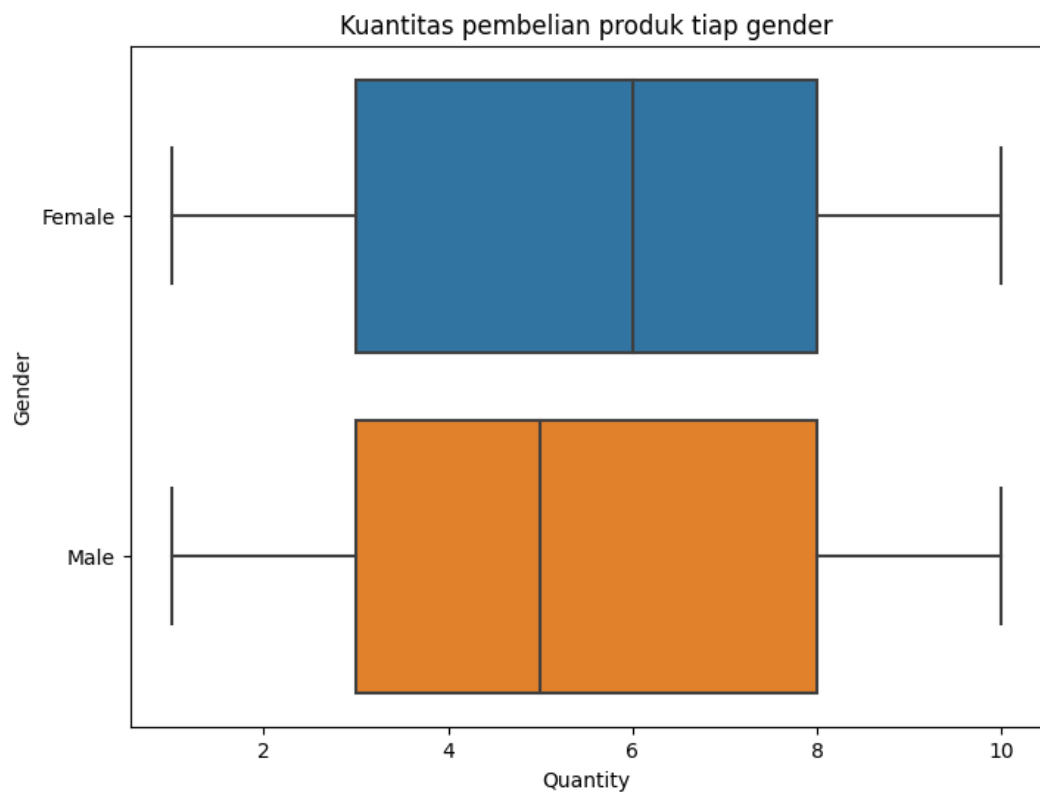
```
data.duplicated()
[7] ✓ 0.1s
...
0   False
1   False
2   False
3   False
4   False
...
995  False
996  False
997  False
998  False
999  False
Length: 1000, dtype: bool
```

Gambar 1.1

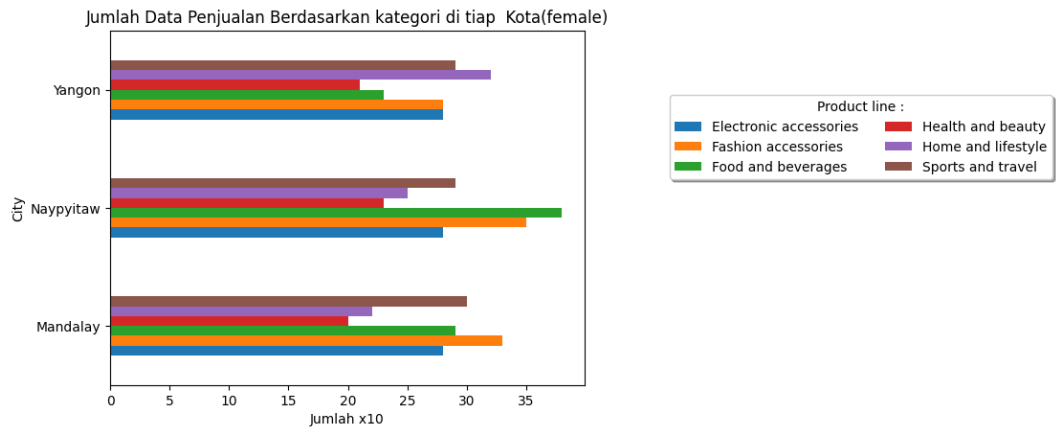
```
data.isnull().sum()
[5] ✓ 0.3s Python
... Invoice ID      0
      Branch      0
      City        0
      Customer type 0
      Gender       0
      Product line 0
      Unit price   0
      Quantity     0
      Tax 5%       0
      Total        0
      Date         0
      Time         0
      Payment      0
      cogs         0
      gross margin percentage 0
      gross income 0
      Rating       0
      dtype: int64
```

Gambar 1.3

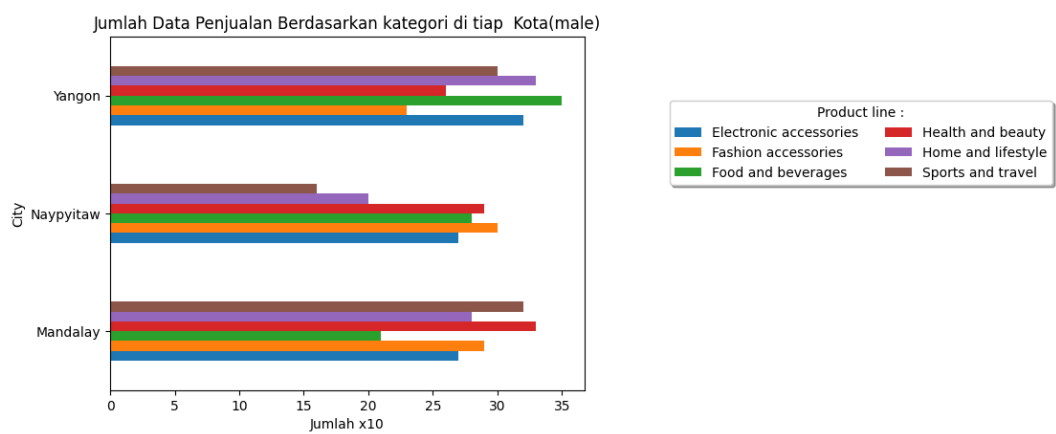
Setelah di preprocessing ternyata data tidak ada yang kosong,terduplikasi dan error sehingga bisa dilanjutkan.



Gambar 1.4



Gambar 1.5



Gambar 1.6

Male	
	Quantity
Jumlah	499
mean	5.292585
std	2.947084
min	1
25%	3
50%	5
75%	8
max	10

Tabel 1.0

Female	
	Quantity
Jumlah	501
mean	5.726547

std	2.886360
min	1
Quartil 1	3
Quartil 2	6
Quartil 3	8
max	10

Tabel 1.1

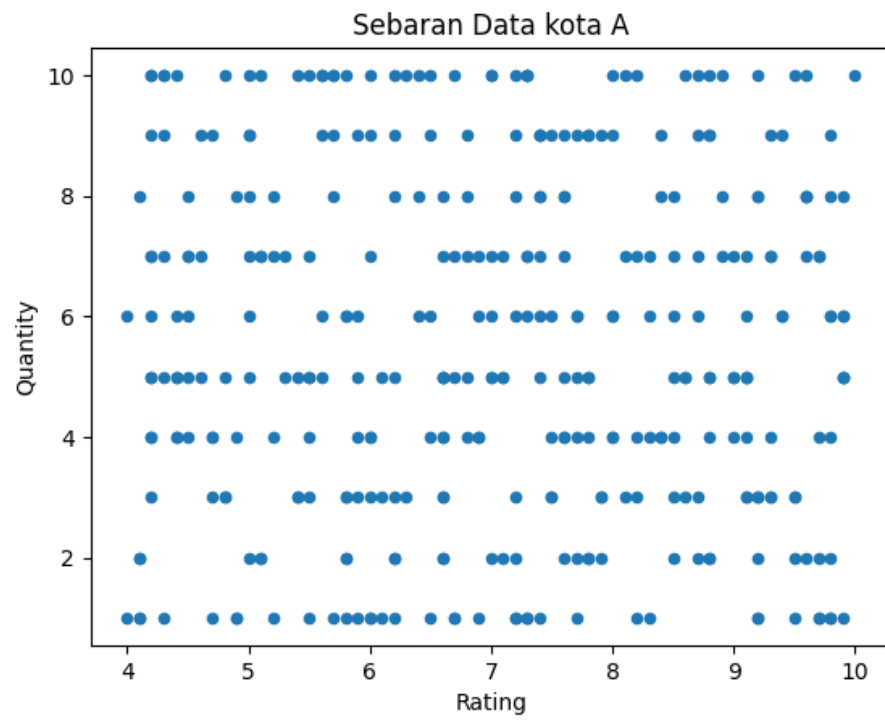
Dari data tersebut diketahui bahwa rata rata pembelian produk oleh wanita lebih tinggi dibandingkan pria dimana wanita $5.726547 \approx 6$ dan pria $5.292585 \approx 5$ dan untuk standar deviasinya relative kecil dan artinya untuk pembelian tidak jauh dari kuantitas rata rata untuk wanita 6 dan untuk pria 5 barang total jumlah barang yang dibeli oleh pria adalah 2641 dan wanita 2869.

Di plotbar diketahui disetiap kota memiliki penjualan kuantitas yang menonjol masing masingnya di kota Yangon ada home and lifestyle untuk wanita dan food and beverages untuk pria sehingga disetiap kota bisa dilihat prioritas produk yang laku disana.

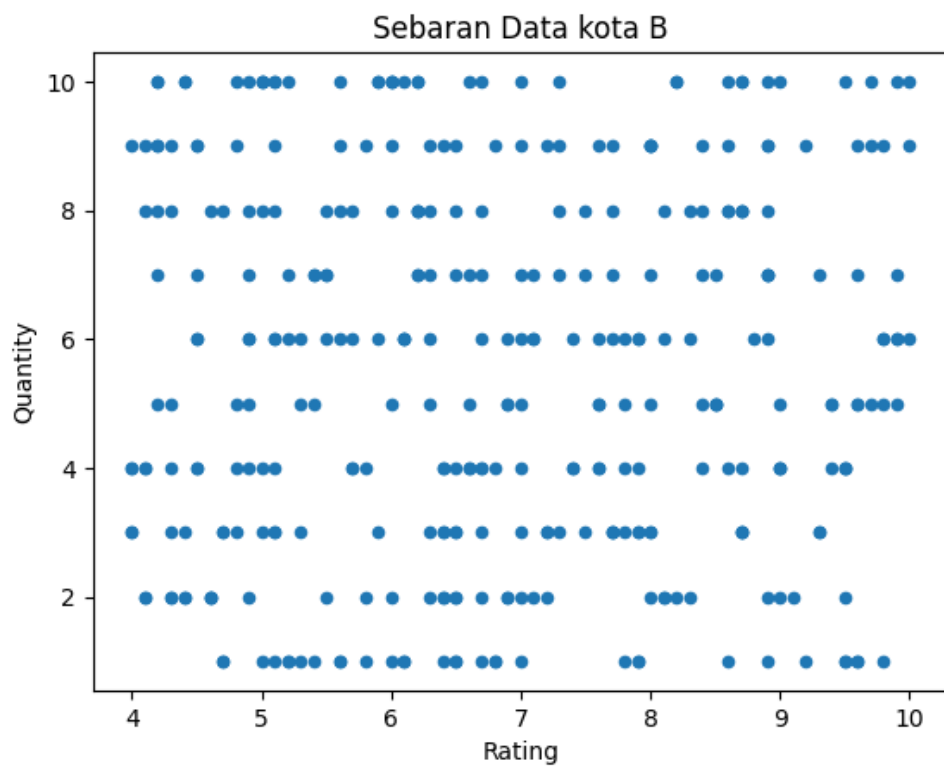
	Rating		
Kota	A	B	C
Jumlah	340	332	328
mean	7.027059	6.818072	7.072866
std	1.731345	1.713719	1.704526
min	4	4	4
25%	5	5.3	5.6
50%	7.1	6.7	7.1
75%	8.5	8.2	8.5
max	10	10	10
Kuantitas	1859	1820	1831

Tabel 1.2

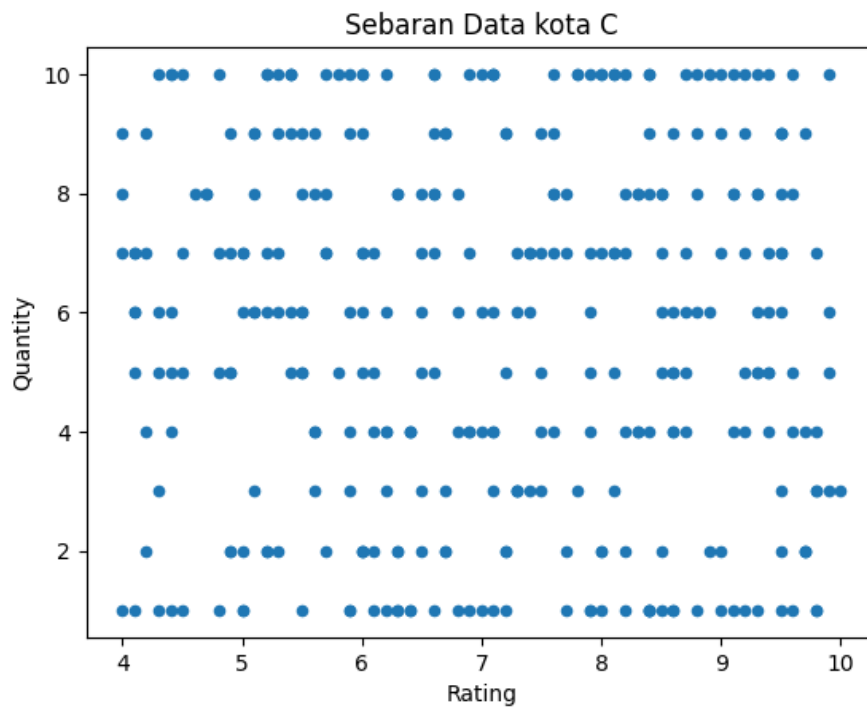
Data tersebut adalah rating yang didapatkan dari tiap kota dan kuantitas penjualan dari tabel tersebut agak bias apakah rating dan jumlah kuantitas penjualan saling memengaruhi tidaknya maka bisa dicek dengan korelasi.



Gambar 1.6



Gambar 1.7



Gambar 1.8

Korelasi Kota A		
	Quantity	Rating
Quantity	1	-0.015909
Rating	-0.015909	1

Tabel 1.2

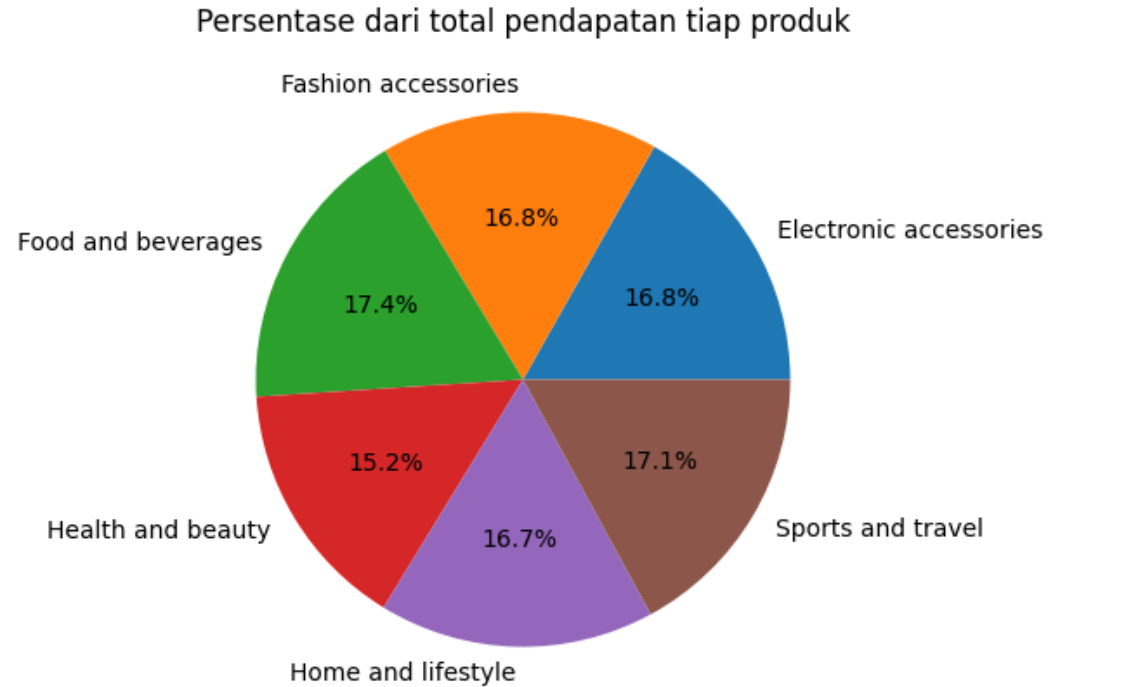
Korelasi Kota B		
	Quantity	Rating
Quantity	1	0.004792
Rating	0.004792	1

Tabel 1.3

Korelasi Kota C		
	Quantity	Rating
Quantity	1	-0.023955
Rating	-0.023955	1

Tabel 1.4

Dari ke 3 kota tersebut setelah di tes korelasi cenderung lebih mendekati 0 jadi tidak terlalu saling memengaruhi rating yang didapatkan dengan kuantitas yang dijual dan persebaran data cukup luas

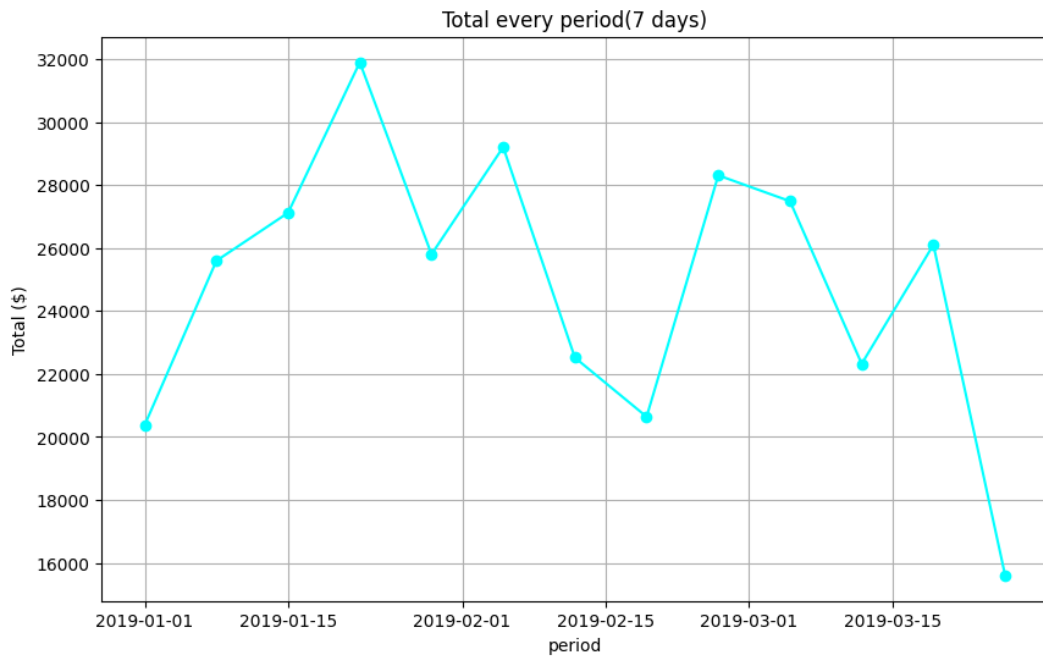


Gambar 1.9

	Pendapatan(\$)
Electronic accessories	54.337,532
Fashion accessories	54.305,895
Food and beverages	56.144,844
Health and beauty	49.193,739
Home and lifestyle	53.861,913
Sports and travel	55.122,827
Total	322.966,749

Tabel 1.5

Dari grafik tersebut diketahui Food and beverages menyumbang total pendapatan paling besar dan health and beauty yang paling rendah sehingga bisa dimaksimalkan untuk bagian food and beverage agar bisa lebih meningkat atau stagnan.



Gambar 2.0

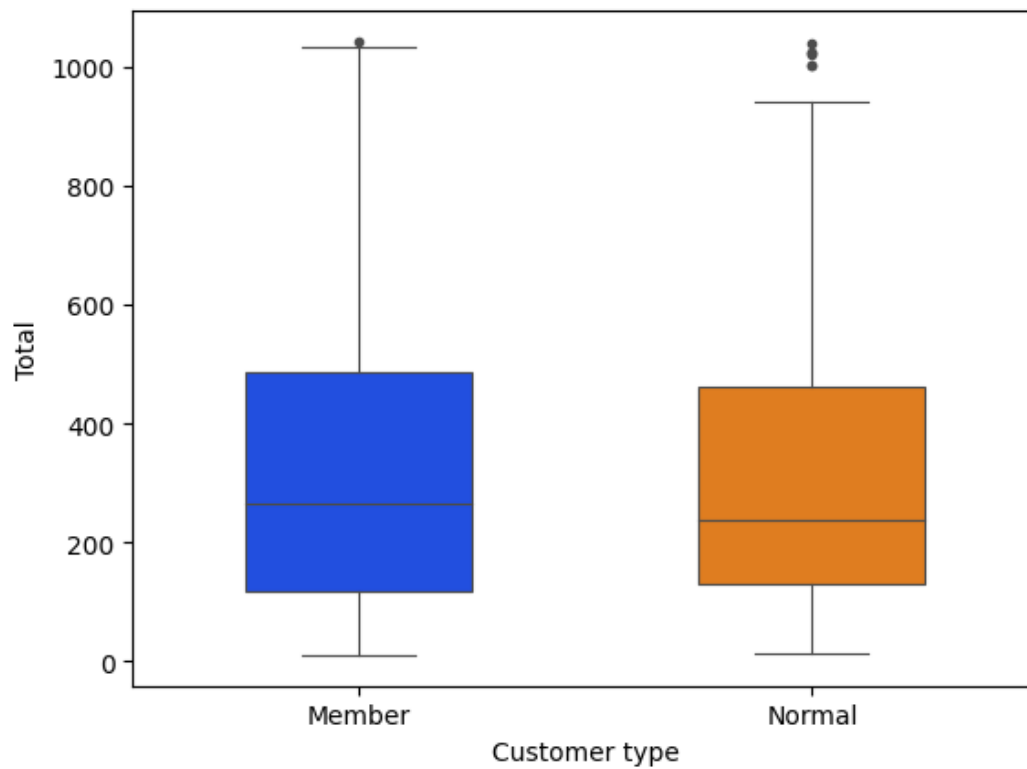
Tabel pendapatan per 7 hari	
	Total(s)
2019-01-01	20377.6335
2019-01-08	25593.5715
2019-01-15	27118.8435
2019-01-22	31894.4955
2019-01-29	25800.2325
2019-02-05	29204.4900
2019-02-12	22517.4705
2019-02-19	20639.8710
2019-02-26	28306.9185
2019-03-05	27485.6505
2019-03-12	22321.8030
2019-03-19	26100.7320
2019-03-26	15605.0370

Tabel 1.6

Informasi pendapatan per 7 hari	
	Total
Jumlah	13 pekan
mean	\$24843.596077
std	\$4384.440204
min	\$15605.037000
25%	\$ 22321.803000
50%	\$25800.232500
75%	\$27485.650500
max	\$31894.495500

Tabel 1.7

Dari tabel dan grafik tersebut penurunan paling tinggi ada di periode akhir pada 2019-03-19 sampai 2019-03-26 turun sebesar 40% dan kenaikan paling tinggi saat periode 1 dan 2 sebesar 25% dan untuk pendapatan per periodenya cukup beragam naik dan turun



Gambar 2.1

	Member	Normal
Jumlah	499	501
mean	\$327.791305	\$318.122856
std	\$247.993782	\$243.902800
min	\$10.678500	\$12.694500
25%	\$118.251000	\$129.234000
50%	\$266.028000	\$237.426000
75%	\$485.572500	\$462.609000
max	\$1042.650000	\$1039.290000

Tabel 1.8

Diketahui member lebih banyak berkontribusi disbanding normal walau tidak beda jauh jumlah uang dari member ialah \$ 164223.444000000002 dan normal \$ 158743.305