

EDA - Sest Consult

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com

Kelompok 5 :

Alfin Dwisatrio

Wildan Ryan

Dionisius Himando

Laurensia Vanida

Aldi Wachid Arifin

Kartika Novitasari

```
RangeIndex: 10999 entries, 0 to 10998
```

```
Data columns (total 12 columns):
```

#	Column	Non-Null Count	Dtype
0	ID	10999 non-null	int64
1	Warehouse_block	10999 non-null	object
2	Mode_of_Shipment	10999 non-null	object
3	Customer_care_calls	10999 non-null	int64
4	Customer_rating	10999 non-null	int64
5	Cost_of_the_Product	10999 non-null	int64
6	Prior_purchases	10999 non-null	int64
7	Product_importance	10999 non-null	object
8	Gender	10999 non-null	object
9	Discount_offered	10999 non-null	int64
10	Weight_in_gms	10999 non-null	int64
11	Reached.on.Time_Y.N	10999 non-null	int64

```
dtypes: int64(8), object(4)
```

```
memory usage: 1.0+ MB
```

Berdasarkan fungsi **.info()** bisa dilihat bahwa :

- Tipe data di setiap kolom terlihat sudah sesuai
- Tidak ada nilai kosong (null) di setiap kolomnya

Untuk melihat apakah ada nilai duplikat, digunakan fungsi **.duplicated().sum()**. Hasilnya 0. Jadi, tidak ada nilai duplikat.

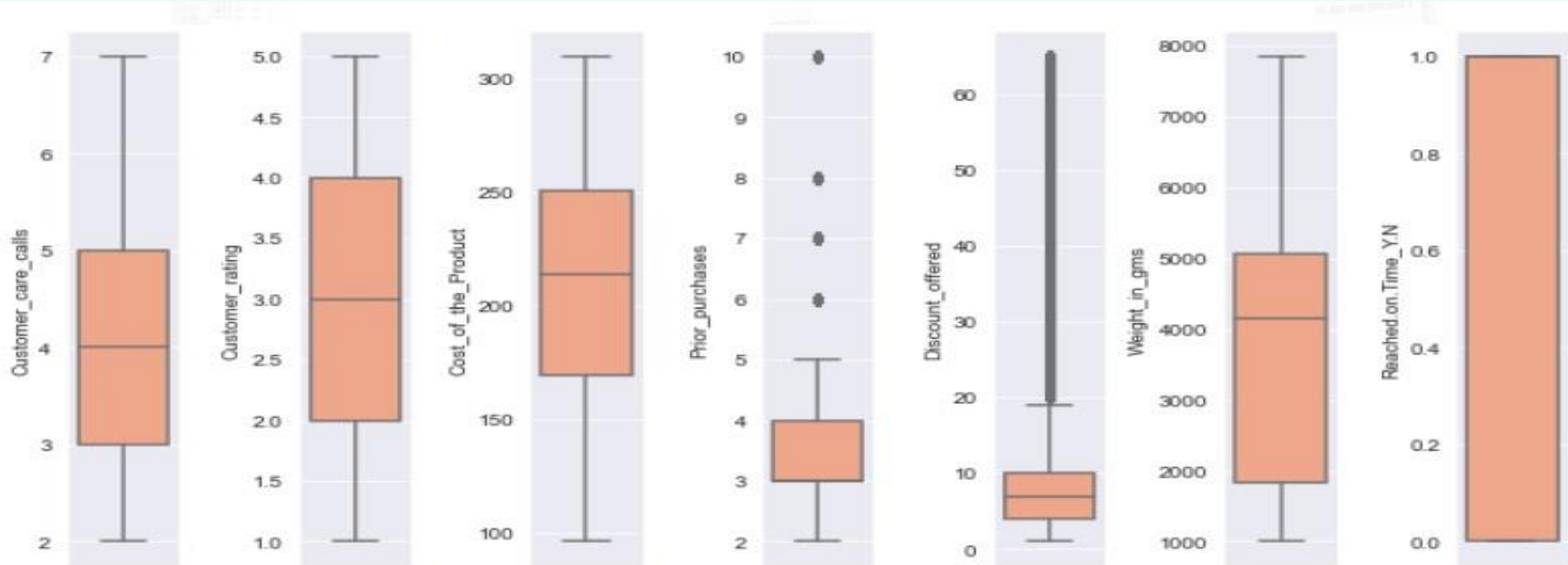
Descriptive Statistics

	ID	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases	Discount_offered	Weight_in_gms	Reached.on.Time_Y.N
count	10999.00000	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000
mean	5500.00000	4.054459	2.990545	210.196836	3.567597	13.373216	3634.016729	0.596691
std	3175.28214	1.141490	1.413603	48.063272	1.522860	16.205527	1635.377251	0.490584
min	1.00000	2.000000	1.000000	96.000000	2.000000	1.000000	1001.000000	0.000000
25%	2750.50000	3.000000	2.000000	169.000000	3.000000	4.000000	1839.500000	0.000000
50%	5500.00000	4.000000	3.000000	214.000000	3.000000	7.000000	4149.000000	1.000000
75%	8249.50000	5.000000	4.000000	251.000000	4.000000	10.000000	5050.000000	1.000000
max	10999.00000	7.000000	5.000000	310.000000	10.000000	65.000000	7846.000000	1.000000

Berdasarkan fungsi **.describe()** bisa dilihat bahwa :

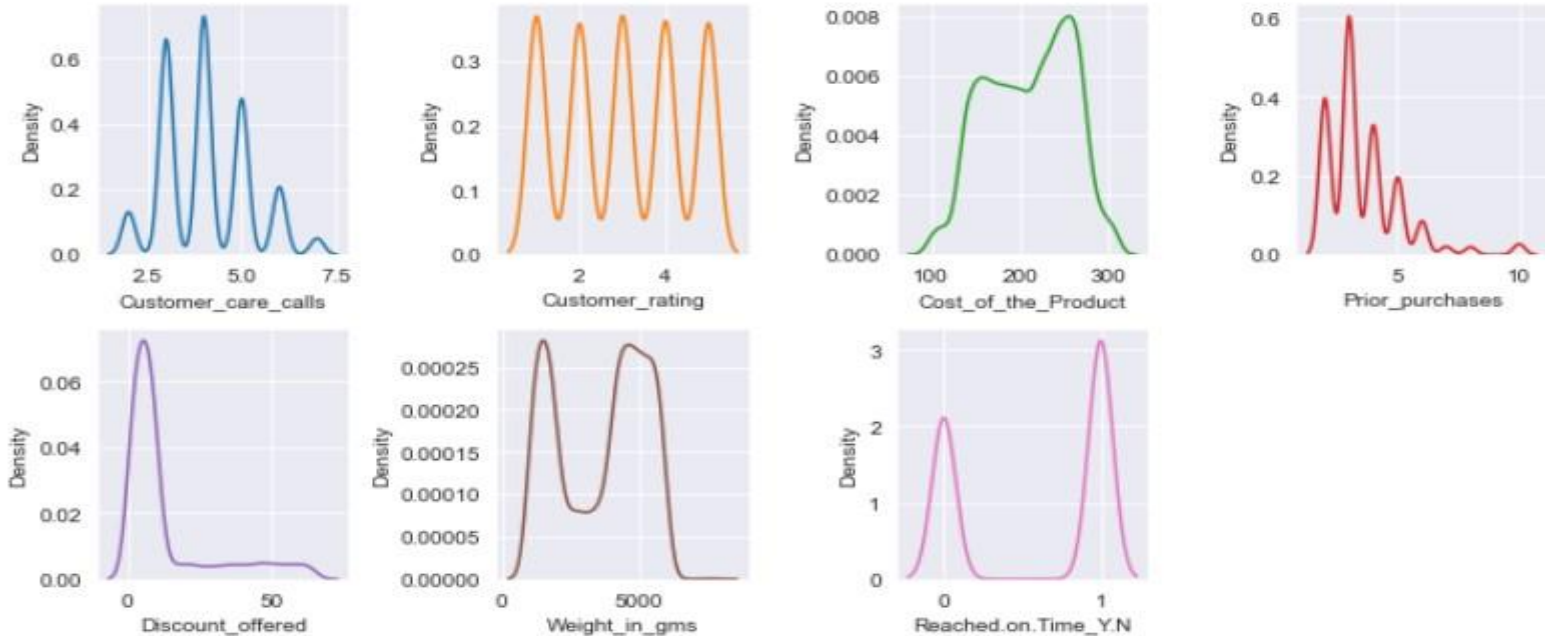
- **Nilai max dan min** dari setiap feature, secara umum terlihat **normal**. Pada 'Discount_offered' dan 'Prior_purchases' perlu diperhatikan karena nilai maxnya sangat jauh dari Q3nya
- Semua feature terkecuali 'Discount_offered' dan 'Weight_in_gms' terdistribusi normal dilihat dari mean dan median yang bernilai dekat
- Feature '**Discount_offered**' dan '**Weight_in_gms**' distribusinya **kemungkinan skewed**

Univariate Analysis (Boxplot)



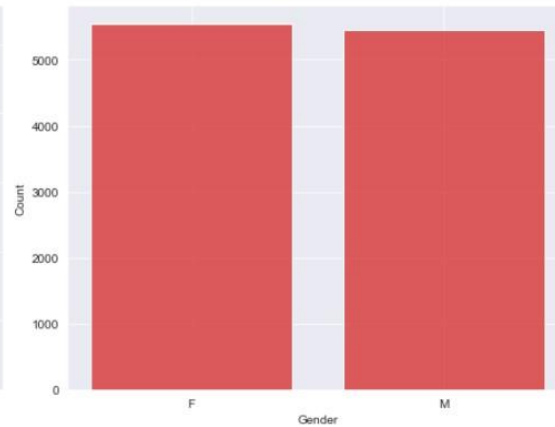
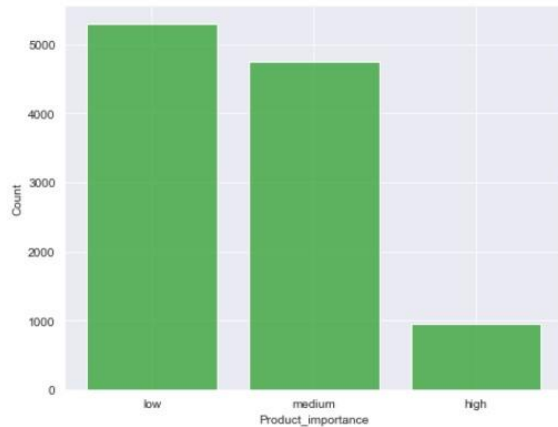
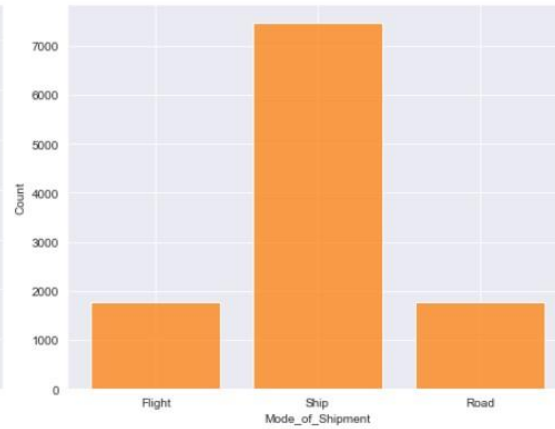
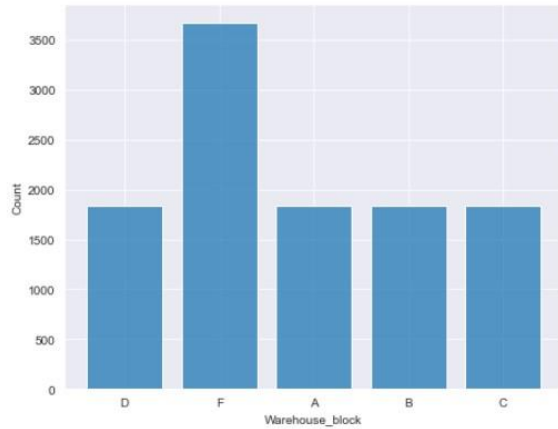
- Dapat dilihat pada plot 'Discount_offered' dan 'Prior_purchases' terdapat **outlier** dan cukup jauh dari nilai distribusi normalnya
- Feature lainnya terlihat tidak memiliki **outlier**

Univariate Analysis (Distplot)



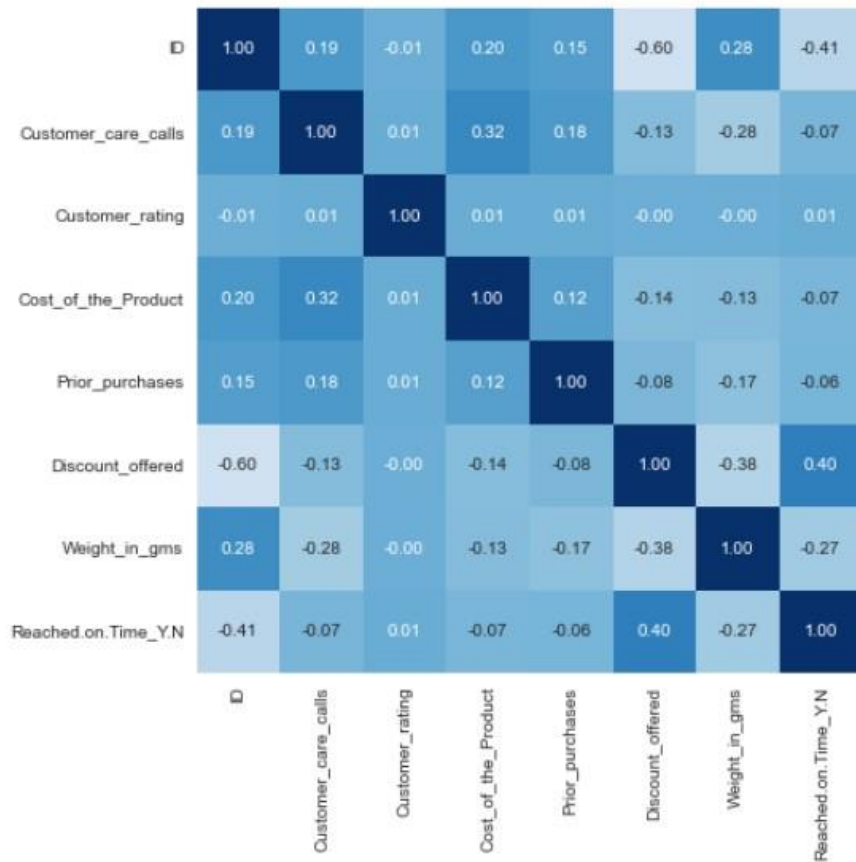
- Terdapat 2 feature yaitu 'Prior_purchase' dan 'Discount_offered' terdistribusi **negatively skewed**
- Feature 'Reached.on.Time_Y.N' memiliki distribusi **bimodal**

Univariate Analysis (Countplot)



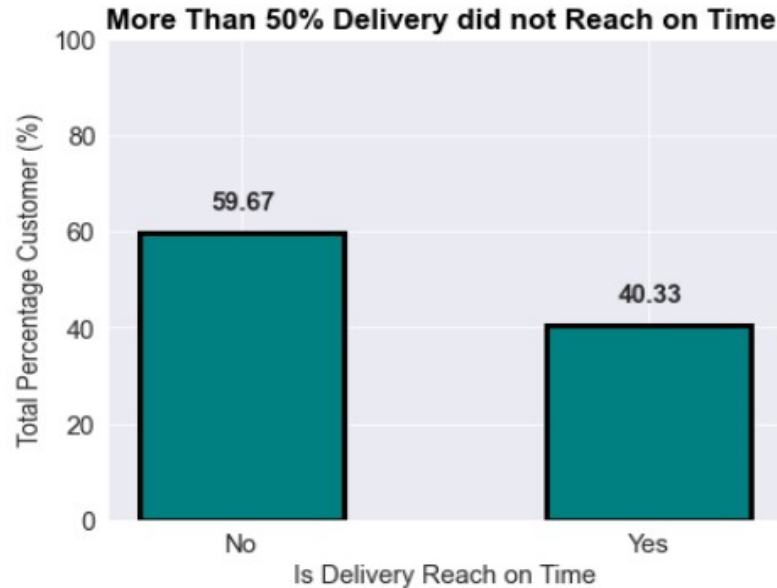
- **Warehouse** : Block A, B, C, D memiliki nilai yang serupa sedangkan block F mendominasi hingga 2x lipat nilai block lainnya.
- **Shipment** : Transportasi udara dan darat memiliki nilai yang serupa sedangkan transportasi laut(menggunakan kapal) mendominasi sekitar 4x lipat lebih banyak.
- **Importance** : Secara keseluruhan mayoritas **Importance** produk ada pada **low** dan **medium**, sedangkan **High importances** menjadi minoritas.
- **Genders** : Kedua gender memiliki nilai yang serupa.

Multivariate Analysis



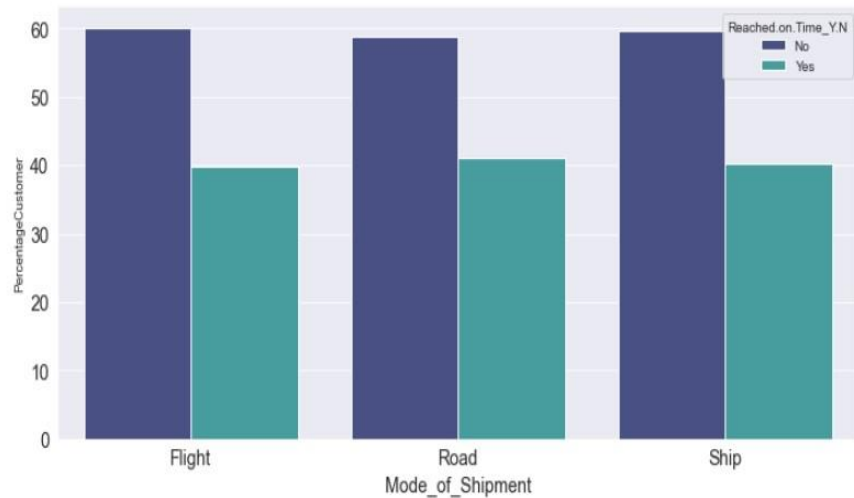
Dari hasil heatmap :

- Tidak ada 2 feature yang redundan (>0.7)
- Terdapat korelasi positif antara feature 'Discount_offered' dan 'Reached.on.Time_Y.N' (0.40)
- Terdapat korelasi positif antara feature 'Customer_care_calls' dan 'Cost_of_the_Product' (0.32)
- Terdapat korelasi negatif antara feature 'Weight_in_gms' dan 'Discount_offered' (-0.38)
- Terdapat korelasi negatif antara feature 'Reached.on.Time_Y.N' dan 'Weight_in_gms' (-0.27)



Reach on time

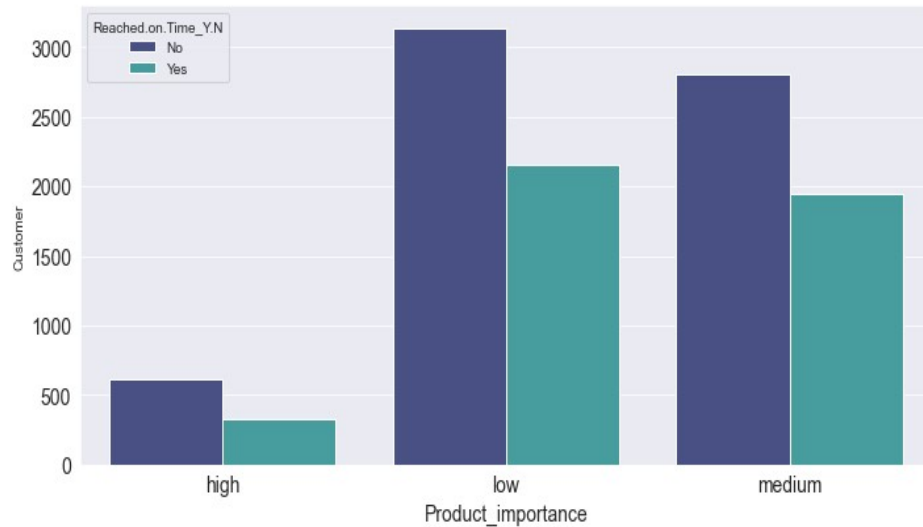
- Berdasarkan grafik di atas, dapat dilihat bahwa pengiriman yang mengalami keterlambatan lebih dari 50% yaitu sekitar **59,67%**



	Mode_of_Shipment	Reached.on.Time_YN	Customer	TotalCustomer	PercentageCustomer
0	Flight	No	1069	1777	60.16
1	Flight	Yes	708	1777	39.84
2	Road	No	1035	1760	58.81
3	Road	Yes	725	1760	41.19
4	Ship	No	4459	7462	59.76
5	Ship	Yes	3003	7462	40.24

Mode Pengiriman

- Banyak pelanggan memilih mode pengiriman dengan kapal. Sekitar **59,76%** pengirimannya mengalami keterlambatan
- Semua jenis pengiriman mengalami keterlambatan sekitar **58%-60%**.



	Product_importance	Reached.on.Time_Y.N	Customer	TotalCustomer	PercentageCustomer
0	high	No	616	948	60.16
1	high	Yes	332	948	39.84
2	low	No	3140	5297	58.81
3	low	Yes	2157	5297	41.19
4	medium	No	2807	4754	59.76
5	medium	Yes	1947	4754	40.24

Product Importance

- Keterlambatan terjadi pada semua jenis produk (low, medium, high)
- Keterlambatan pada setiap jenis produknya sekitar **58%-60%**

- GITHUB KELOMPOK 5 :

<https://github.com/aldiwachid/E-Commerce-Shipping-Data-.git>