

E-Commerce Shipping Data

“Asklepios”

Awalsyah Rinanto Putra

Fathah Oscar

M Rizky Septiansyah

Hermawan Febrianto

Devi Puji Ayuningsih

Anggita Citanegara Lubis



Stage 1 - EDA

1. Descriptive Statistics

```
df.info()

<class 'pandas.core.frame.DataFrame'>
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
```

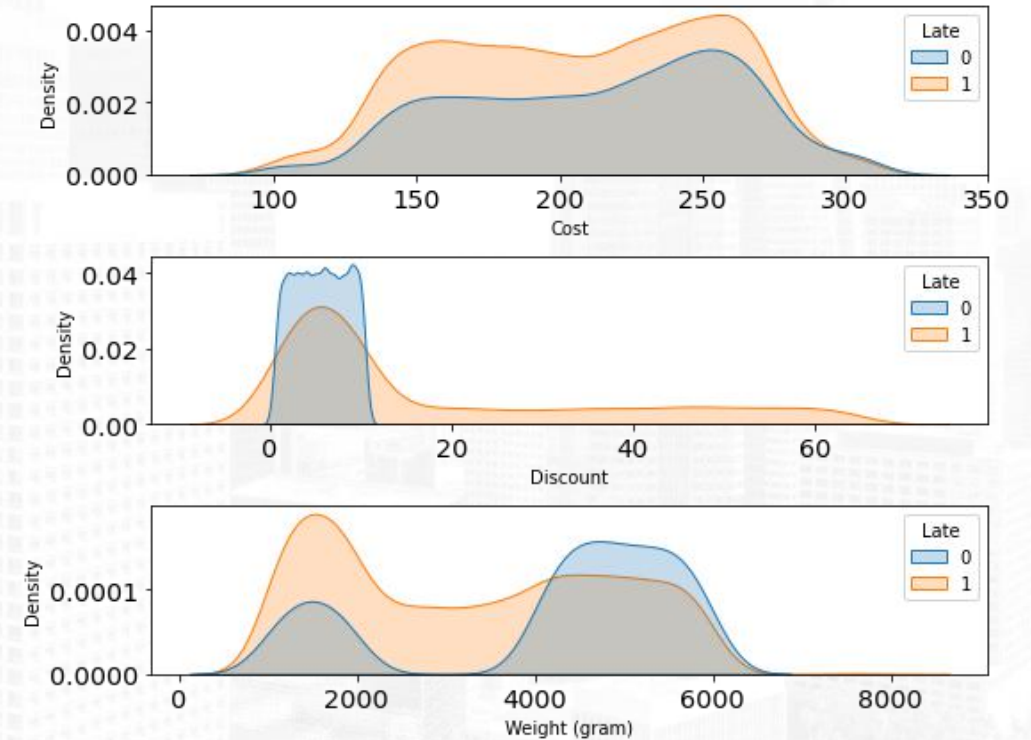
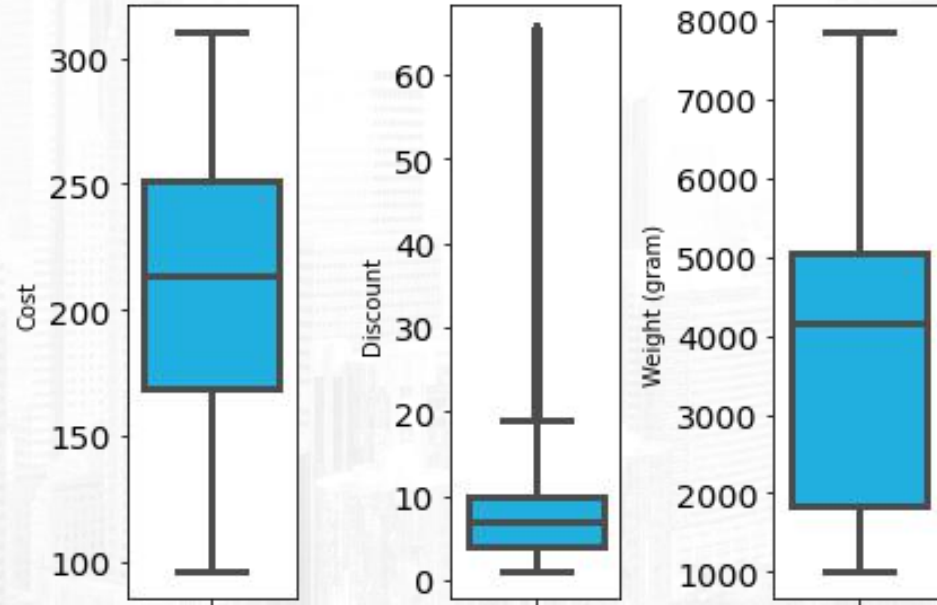
```
df_num = df[['Calls', 'Rating', 'Cost', 'Purchases', 'Discount', 'Weight (gram)']]
df_num.describe()
```

	Calls	Rating	Cost	Purchases	Discount	Weight (gram)
count	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000
mean	4.054459	2.990545	210.196836	3.567597	13.373216	3634.016729
std	1.141490	1.413603	48.063272	1.522860	16.205527	1635.377251
min	2.000000	1.000000	96.000000	2.000000	1.000000	1001.000000
25%	3.000000	2.000000	169.000000	3.000000	4.000000	1839.500000
50%	4.000000	3.000000	214.000000	3.000000	7.000000	4149.000000
75%	5.000000	4.000000	251.000000	4.000000	10.000000	5050.000000
max	7.000000	5.000000	310.000000	10.000000	65.000000	7846.000000

Hasil Observasi :

- Semua tipe data, nama kolom dan isinya sudah sesuai
- Tidak ada kolom yang memiliki missing value
- Jika dilihat dari nilai max, variabel Purchases dan Discount kemungkinan memiliki nilai outlier

2. Univariate Analysis



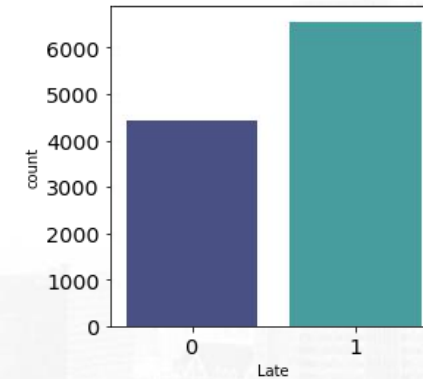
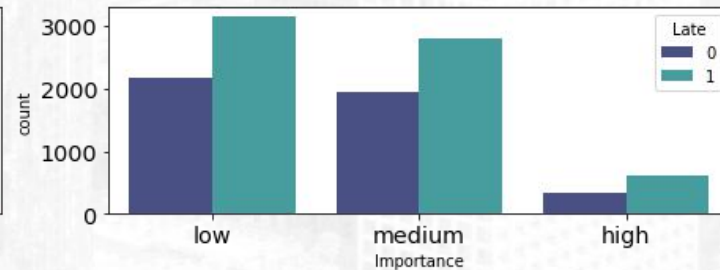
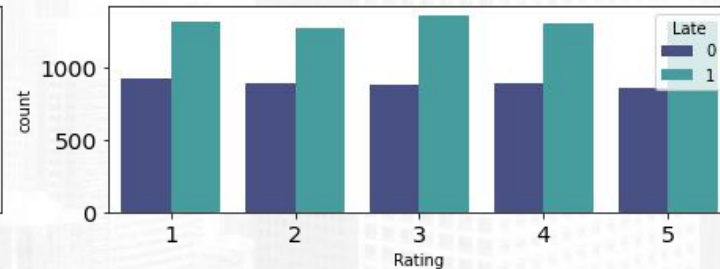
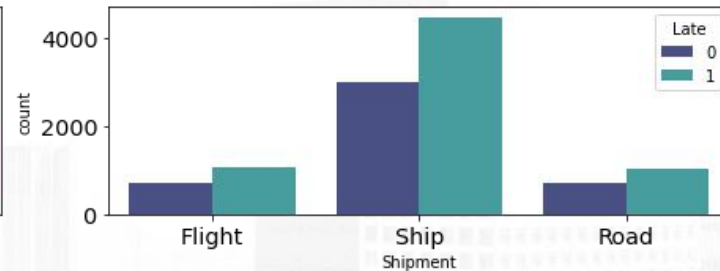
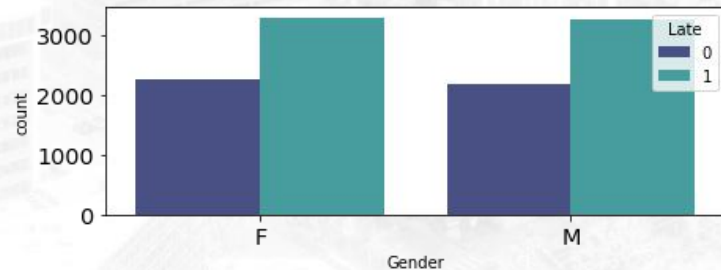
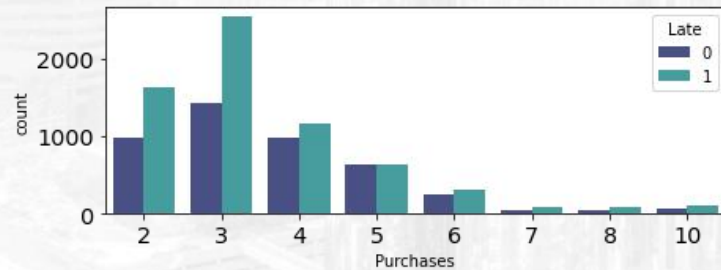
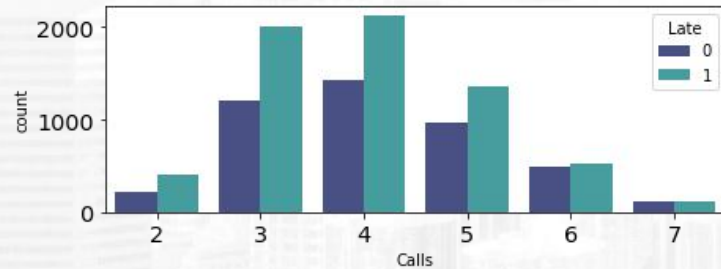
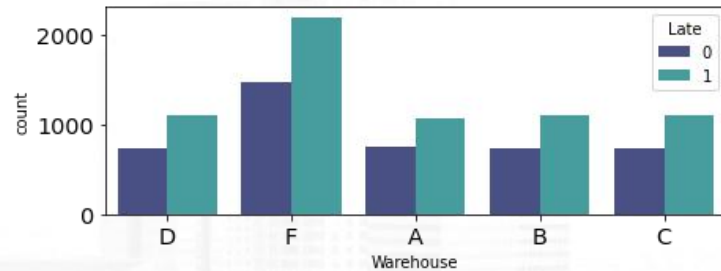
Hasil Observasi :

- Variabel Purchases dan Discount memiliki outlier dan membentuk pola positively Skewed
- Variabel Cost memiliki distribusi yang paling mendekati distribusi normal

```
# Menghitung nilai skew
df[numericals].skew(axis=0, skipna=True)
```

```
Cost          -0.157117
Discount       1.798929
Weight (gram) -0.249747
dtype: float64
```

2. Univariate Analysis



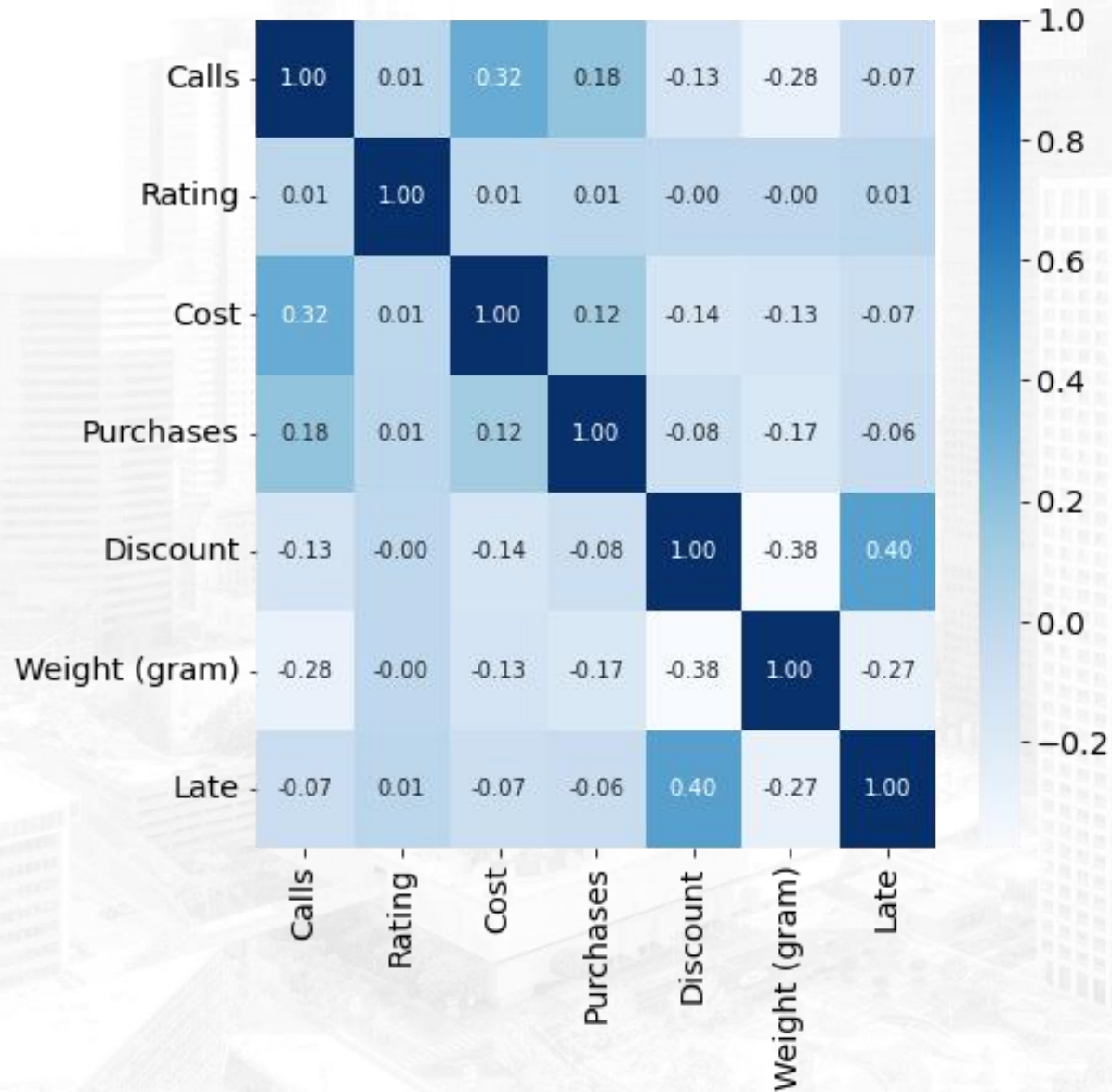
Hasil Observasi :

- Warehouse block yang paling banyak digunakan adalah Warehouse F
- Shipment mode yang dominan adalah pengiriman dengan kapal
- Jumlah sampel yang mengalami keterlambatan pengiriman lebih banyak

Apa yang harus di-follow up saat data pre-processing?

- Remove outlier
- Melakukan OHE pada feature Warehouse dan Shipment
- Melakukan label encoding pada feature Gender dan Product Importance
- Melakukan standardisasi pada feature Cost dan Weight
- Melakukan log transformation pada Discount dan Purchases

3. Multivariate Analysis



Hasil Observasi :

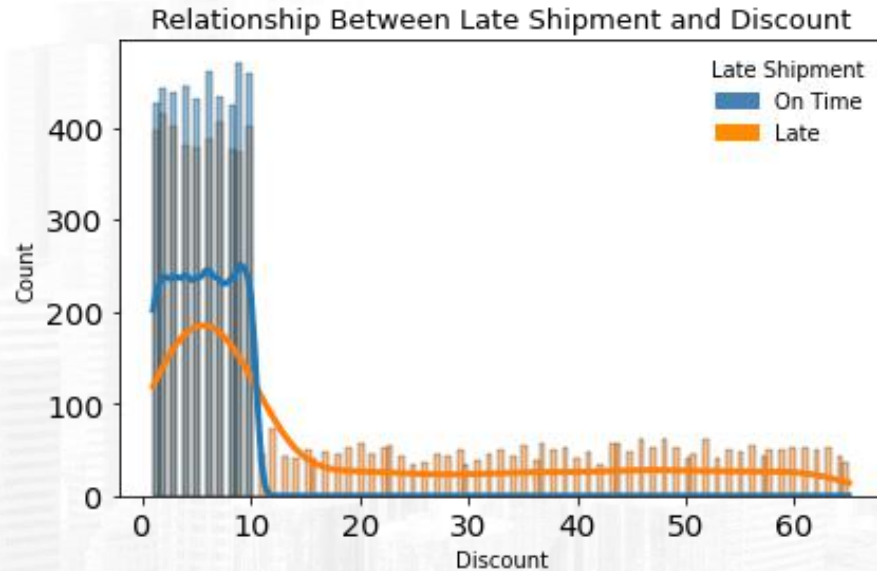
A. Bagaimana korelasi antara masing-masing feature dan label. Kira-kira feature mana saja yang paling relevan dan harus dipertahankan?

- Discount dan Weight

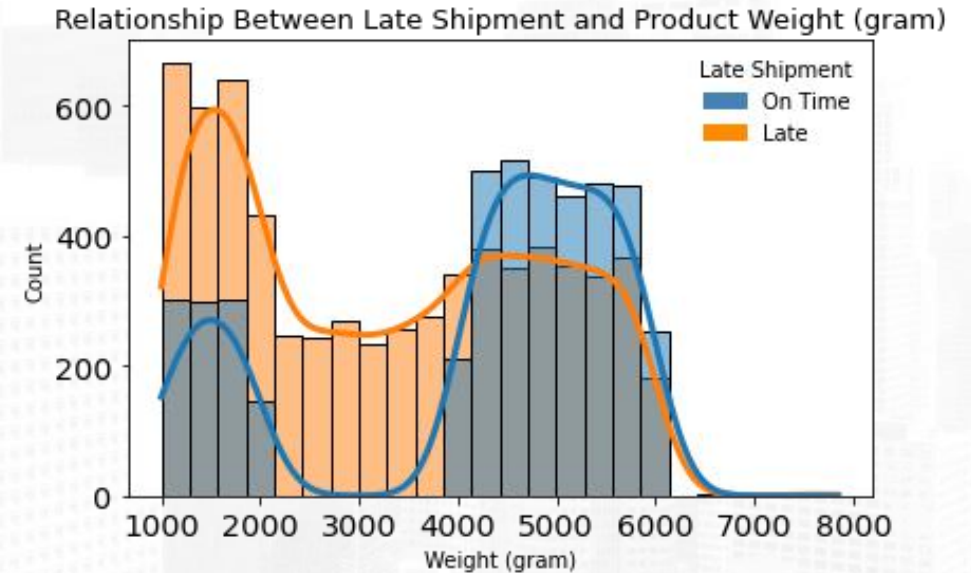
B. Bagaimana korelasi antar-feature, apakah ada pola yang menarik? Apa yang perlu dilakukan terhadap feature itu?

- Weight terhadap Calls
- Weight terhadap Discount
- Cost terhadap Calls

4. Business Insight

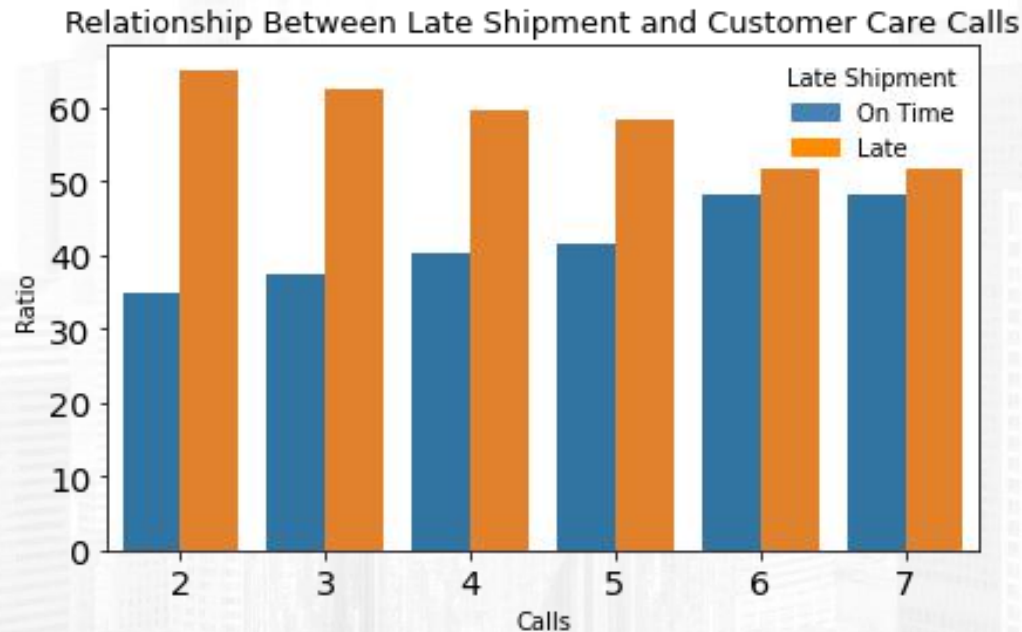


Untuk pembelian produk dengan discount diatas 10% banyak mengalami keterlambatan pengiriman. Dalam hal ini pihak e-commerce perlu memberikan notifikasi keterlambatan pengiriman kepada customer ketika melakukan pembelian dengan menggunakan discount yang besar yang memungkinkan produk yang dipesan tidak terkirim tepat waktu.



Barang dengan berat 2-4 Kg mengalami keterlambatan pengiriman. Dalam hal ini, pihak e-commerce perlu memberikan notifikasi keterlambatan pengiriman kepada customer yang membeli produk di rentang berat produk 2-4 kg sebelum customer melakukan transaksi.

4. Business Insight



Berdasarkan data jumlah keterlambatan pengiriman menurun dengan meningkatnya jumlah telepon yang diterima oleh customer care.

Perusahaan perlu mencari informasi mengenai isi telepon customer kepada customer care (siapa penelpon, isi telepon), sehingga bisa menentukan korelasi dengan jumlah keterlambatan pengiriman.

Rekomendasi dengan Asumsi:

1. Bila di asumsikan bahwa pelanggan menelpon untuk melakukan konfirmasi pemesanan, maka bisa dilakukan proses konfirmasi pemesanan dari pelanggan memberikan pengingat kepada penjual untuk segera melakukan konfirmasi ketersediaan barang dan kesiapan pengiriman kepada bagian pengiriman atau kurir. (bisa dengan aplikasi atau ditambahkan pada petugas tertentu).
2. Bila di asumsikan bahwa penjual menelpon untuk melakukan konfirmasi kesediaan pesanan dan barang siap di kirim, maka bisa dilakukan proses otomatisasi saat penjual konfirmasi kesediaan barang, langsung barang disiapkan untuk di kirim pada hari yang sama dan mengirimkan konfirmasi untuk kurir mengirimkan.