

Clothing Sales Analyst and Visualization

1. Data Cleansing

a. Membuat klasifikasi dari kategori

```
#cek sample
df.sample(5)
```

	product_title	Cat_product	shipping_city	Prov	referrer_source	orders	gross_sales	discounts	returns	net_sales	shipping	total_sales	net_quantity
199	Talia Dress Woodrose	Dress	Batam	Kepulauan Riau	Social	1	475000000	0	0	475000000	0	475000000	1
89	Klamby Monogram Brooch Rose Gold	Brooch	Kota Padang	Sumatera Barat	Direct	1	385000000	-6581190	0	378418810	0	378418810	1
37	Yasmin Tunic Desert	Tunic	Tangerang	Banten	Search	1	575000000	0	0	575000000	0	575000000	1
207	Yara Pleated Skirt Navy	Skirt	Kota Adm. Jakarta Selatan	DKI Jakarta	Social	1	649000000	0	0	649000000	0	649000000	1
60	Klara Dress Almond	Dress	Kota Jambi	Jambi	Direct	1	475000000	0	0	475000000	0	475000000	1

Menambahkan kolom Category Product dengan cara mengambil kata ke-dua dari kolom product title. Serta menambahkan kolom provinsi berdasarkan nama kota dari kolom shipping city. Agar membantu proses analisa dengan mengklasifikasi kolom yang bersifat kategori.

b. Menangani data yang hilang

```
df=df.replace('Unknown', np.nan)

[130] #mengecek apakah ada kolom yang kosong
df.isna().sum()

product_title      0
Cat_product        0
shipping_city      0
Prov               0
referrer_source     2
orders             0
gross_sales        0
discounts          0
returns           0
net_sales          0
shipping           0
total_sales        0
net_quantity       0
dtype: int64
```

Terdapat value “Unknown” di kolom referrer_source yang dapat di artikan sebagai Missing Value, yang selanjutnya akan di tangani dengan menggunakan modus.

c. Duplikat data

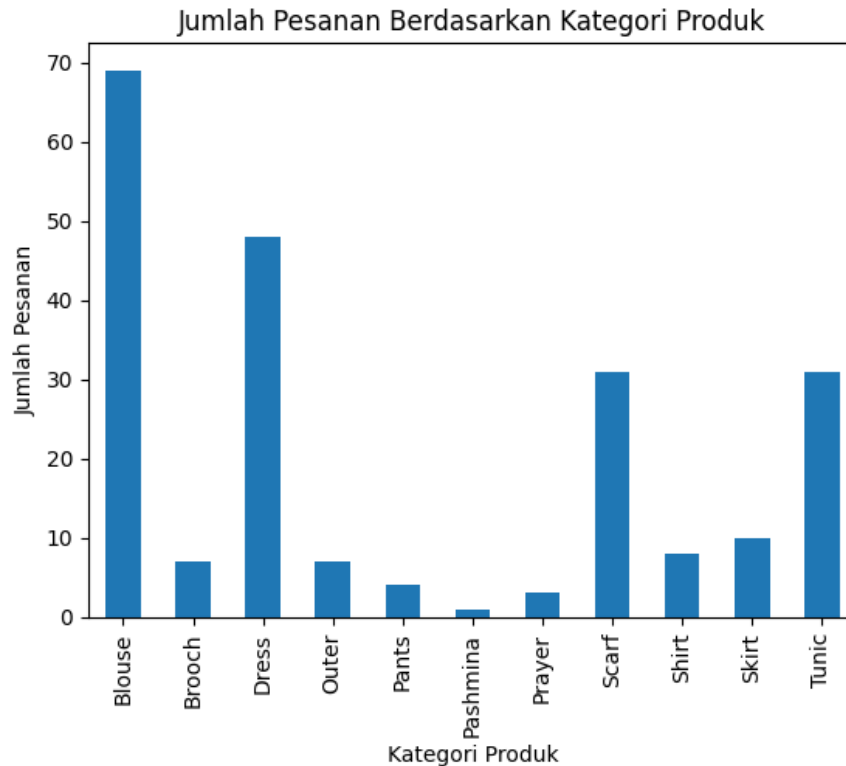
```
[132] #cek duplikat data
duplikat = df.duplicated()
jumlah_duplikat = duplikat.sum()
jumlah_duplikat

0
```

Tidak terdapat data yang duplikat sehingga tidak perlu adanya tindak lanjutan.

2. Visualisasi dan Analisa

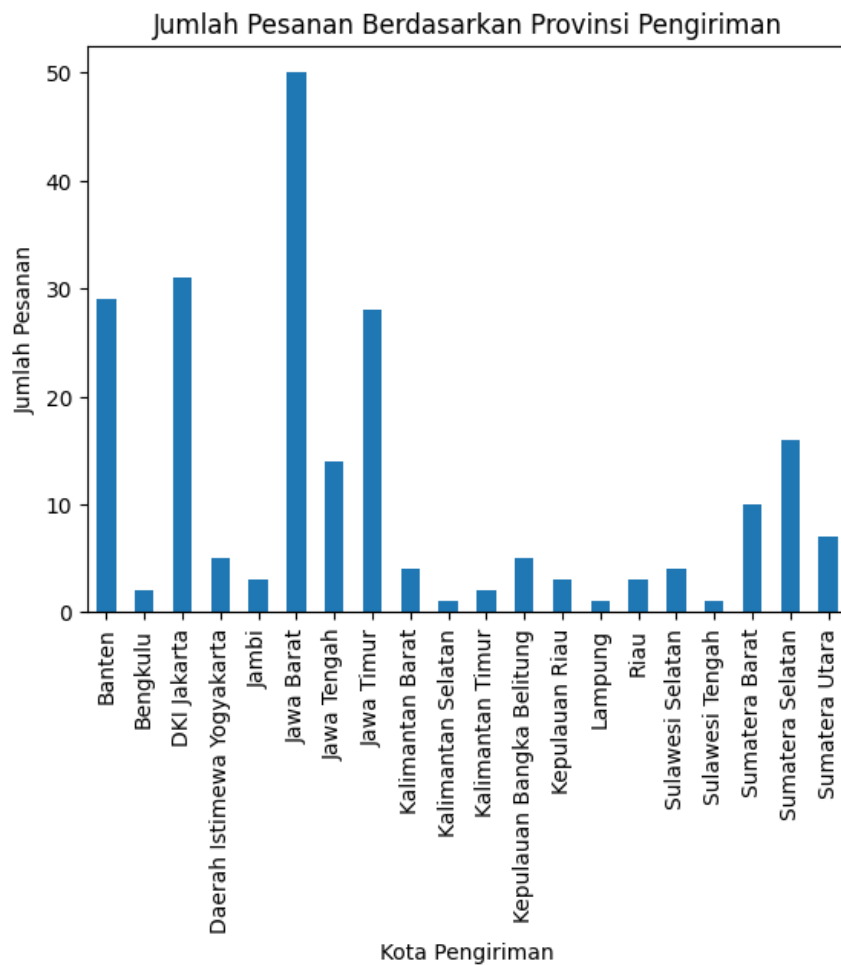
a. Frekuensi Kategori Produk



Kategori dengan jumlah pesanan tertinggi adalah kategori Blouse di ikuti dengan Dress, dan kategori dengan pembelian terendah adalah Pashmina.

Kita dapat memberikan promo khusus untuk pembelian lebih dari 1 atau 2 agar produk yang sudah banyak peminatnya ini terjual lebih banyak kuantitasnya. Produk dengan pembelian terendah di berikan promo walaupun hanya membeli satu agar dapat menarik pembeli dan lebih di promosikan agar masyarakat lebih aware dengan produk ini.

b. Frekuensi Kategori Provinsi

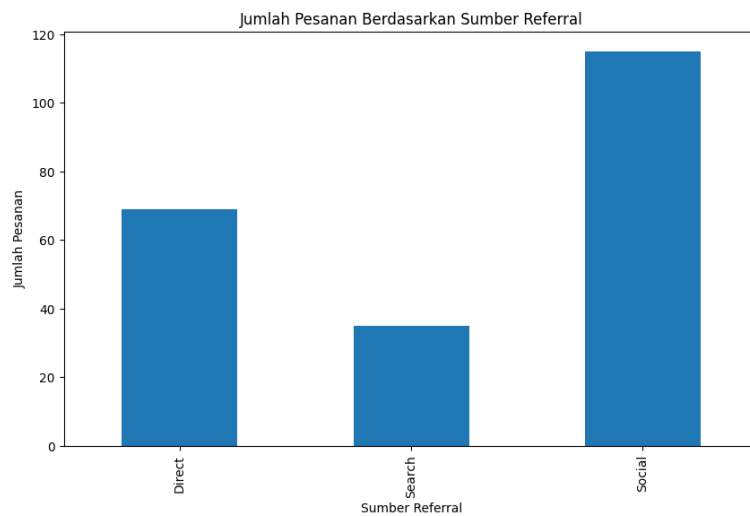


Provinsi dengan pemesanan tertinggi ada di kota Jawa Barat dan provinsi dengan pemesanan terendah adalah Sulawesi tengah, Lampung, dan Kalimantan Selatan.

Ibu kota Jakarta berada di posisi kedua dengan jumlah penjualan hamper setengahnya dari Jawa Barat, seharusnya Ibu kota Jakarta merupakan penjualan tertinggi di karenakan tingkat penduduk dan pusat perbelanjaan lebih banyak di Jakarta.

Kita dappat meningkatkan promosi, iklan, atau kampanye tentang produk kita agar dapat meningkatkan brand awareness brand kita ke masyarakat Ibu kota Jakarta.

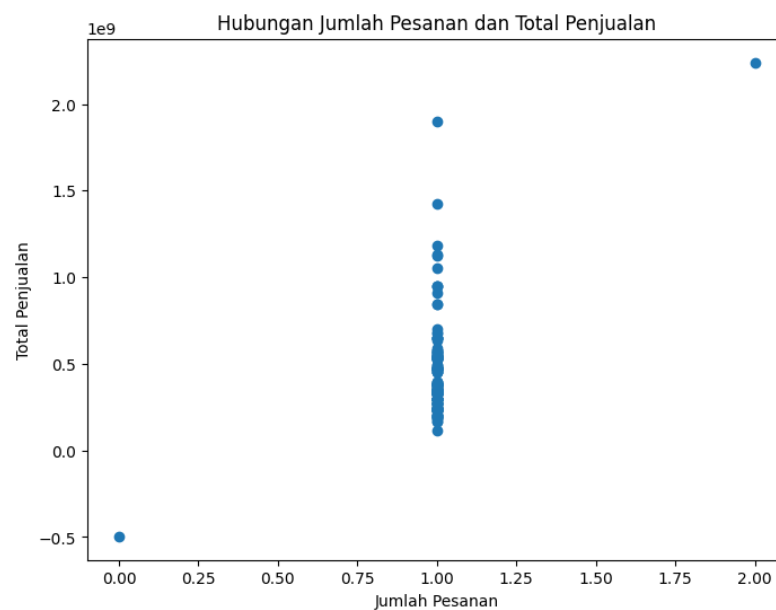
c. Frekuensi Kategori Referensi



Penjualan terbanyak bersumber dari referensi sosial dan yang terendah adalah referensi search. Dapat disimpulkan Produk kita lebih di kenal dari mulut ke mulut.

Karena kurangnya brand awareness brand kita di masyarakat, kita dapat melakukan iklan dan kampanye untuk meningkatkan brand awareness serta kepercayaan ke masyarakat tau dan tertarik dengan brand kita.

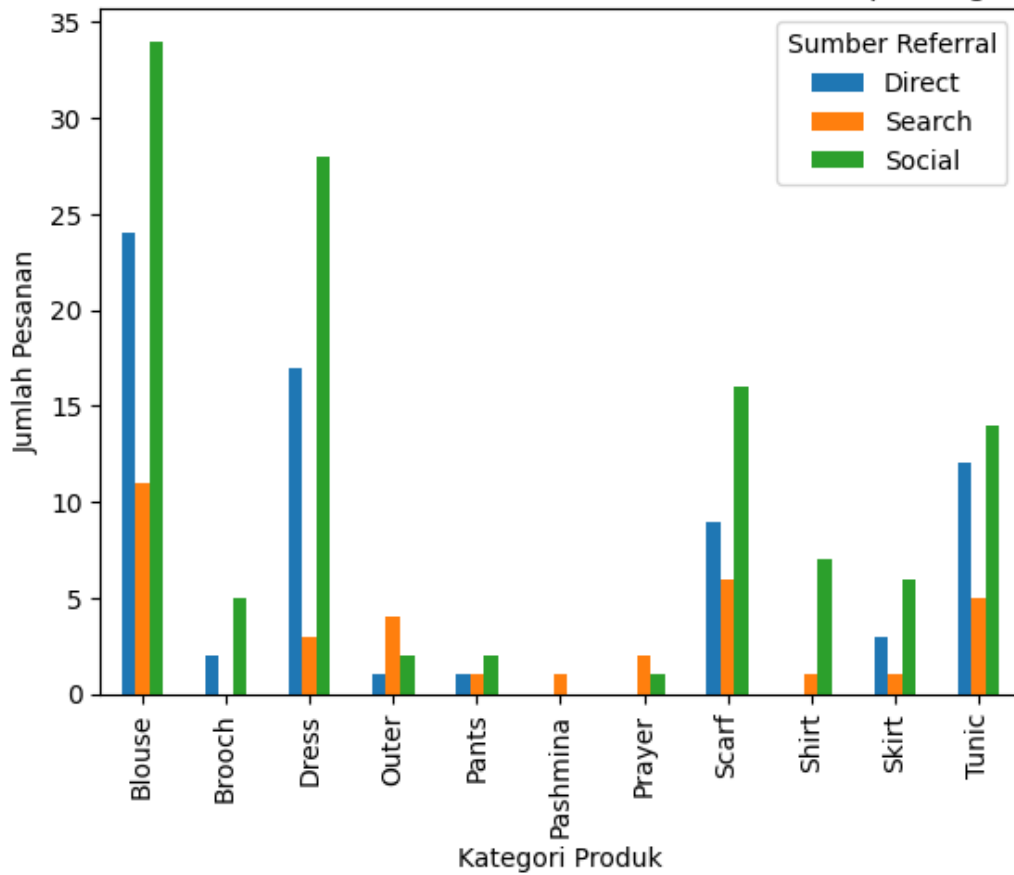
d. Relasi antar Jumlah Pesanan (qty) dan Total penjualan



Tidak ada relasi atau tidak dapat ditarik kesimpulan karena hampir semua penjualan hanya menjual 1 barang, dan yang menjual 2 barang hanya sedikit.

e. Frekuensi Berdasarkan Sumber Referral untuk Setiap Kategori Produk

Jumlah Pesanan Berdasarkan Sumber Referral untuk Setiap Kategori Produk



Dari plot di atas ada hal yang menarik, yaitu produk brooch tidak ada penjualan yang berasal dari search, dapat di tarik kesimpulan kurangnya brand awareness di masyarakat terhadap produk Brooch.