

INTERPRETASI HASIL

Soal:

Diberikan sebuah file bernama “Online_Retail_Data.csv”. Analisislah data dalam file tersebut mulai dari total penjualan produk selama satu tahun, total penjualan setiap bulannya, total penerimaan (revenue) setiap bulannya, dan tren penjualan produk setiap bulan.

Jawaban:

- **Langkah 1: Impor Data**

1. Impor library yang diperlukan untuk analisis data dan visualisasi.
2. Baca data penjualan dari file CSV “Online_Retail_Data.csv” dan simpan ke dalam DataFrame.

- **Langkah 2 Data Cleaning: Bersihkan data dari**

1. Features yang data type nya tidak sesuai
2. Duplicate, missing value, atau outliers
3. Distribusi yang tidak masuk akal

```
12 # DATA CLEANSING
13
14 df_clean = df.copy()
15 # membuat kolom date
16 df_clean['date'] = pd.to_datetime(df_clean['order_date']).dt.date.astype('datetime64[ns]')
17 # menghapus semua baris tanpa customer_id
18 df_clean = df_clean[~df_clean['customer_id'].isna()]
19 # mengkonversi customer_id menjadi string
20 df_clean['customer_id'] = df_clean['customer_id'].astype(str)
21 # menghapus semua baris tanpa product_name
22 df_clean = df_clean[~df_clean['product_name'].isna()]
23 # membuat semua product_name berhuruf kecil
24 df_clean['product_name'] = df_clean['product_name'].str.lower()
25 # menghapus semua baris dengan product_code atau product_name test
26 df_clean = df_clean[(~df_clean['product_code'].str.lower().str.contains('test')) |
27                    (~df_clean['product_name'].str.contains('test '))]
28 # menghapus baris dengan status cancelled, yaitu yang order_id-nya diawali 'C'
29 df_clean = df_clean[df_clean['order_id'].str[:1]!='C']
30 # mengubah nilai quantity yang negatif menjadi positif karena nilai negatif tersebut hanya menandakan order ters
31 df_clean['quantity'] = df_clean['quantity'].abs()
32 # mengkonversi price_id menjadi number
33 df_clean['price'] = pd.to_numeric(df_clean['price'], errors='coerce')
34 # menghapus baris dengan price bernilai negatif
35 df_clean = df_clean[df_clean['price']>0]
36 # membuat nilai amount, yaitu perkalian antara quantity dan price
37 df_clean['amount'] = df_clean['quantity'] * df_clean['price']
38 # mengganti product_name dari product_code yang memiliki beberapa product_name dengan salah satu product_name-ny
39 most_freq_product_name = df_clean.groupby(['product_code', 'product_name'], as_index=False).agg(order_cnt=('order
40 most_freq_product_name['rank'] = most_freq_product_name.groupby('product_code')['order_cnt'].rank(method='first'
41 most_freq_product_name = most_freq_product_name[most_freq_product_name['rank']==1].drop(columns=['order_cnt', 'ra
42 df_clean = df_clean.merge(most_freq_product_name.rename(columns={'product_name': 'most_freq_product_name'}), how=
```

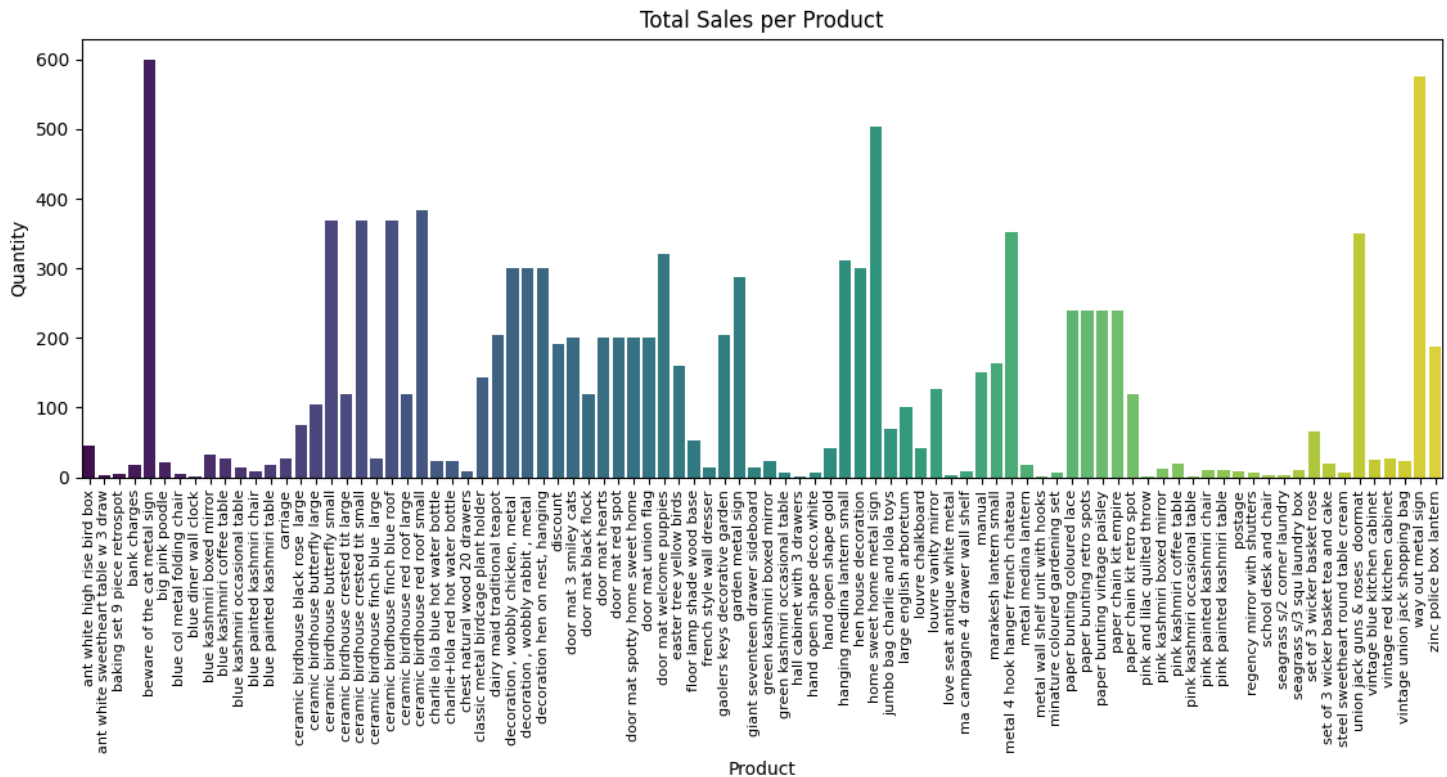
- **Langkah 3: Analisis dan Visualisasi**

1. Membuat variable yang berisi penghitungan total penjualan produk selama satu tahun, total penjualan setiap bulannya, total penerimaan (revenue) setiap bulannya, dan tren penjualan produk setiap bulan.
2. Visualisasikan setiap hasil dari variabel yang telah dibuat menggunakan grafik.

```
8 # Analisis sales by product
9 sales_by_product = df.groupby('product_name')['quantity'].sum().reset_index()
10 print(sales_by_product)
11
12 # Visualisasi sales by product
13 plt.figure(figsize=(16, 12))
14 sns.barplot(x='product_name', y='quantity', data=sales_by_product, palette='viridis')
15 plt.title('Total Sales per Product')
16 plt.xlabel('Product')
17 plt.xticks(rotation='vertical', size=8)
18 plt.ylabel('Quantity')
19 plt.show()
20
21 # Analisis sales by months
22 df['order_date'] = pd.to_datetime(df['order_date'])
23 df['month_date'] = df['order_date'].dt.month
24
25 sales_by_month = df.groupby('month_date')['quantity'].sum().reset_index()
26 print(sales_by_month)
27
28 # Visualisasi sales by Months
29 plt.figure(figsize=(10, 6))
30 plt.plot('month_date', 'quantity', data=sales_by_month, color='skyblue', linestyle='-')
31 plt.xlabel('Month')
32 plt.ylabel('Sales')
33 plt.title('Sales by Month')
34 plt.grid(True)
35 plt.show()
36
37 # Menghitung Total Revenue per Bulan
38 revenue_per_month= df.groupby('month_date')['amount'].sum().reset_index()
```

Hasil Analisis

➤ **Total Penjualan Produk**



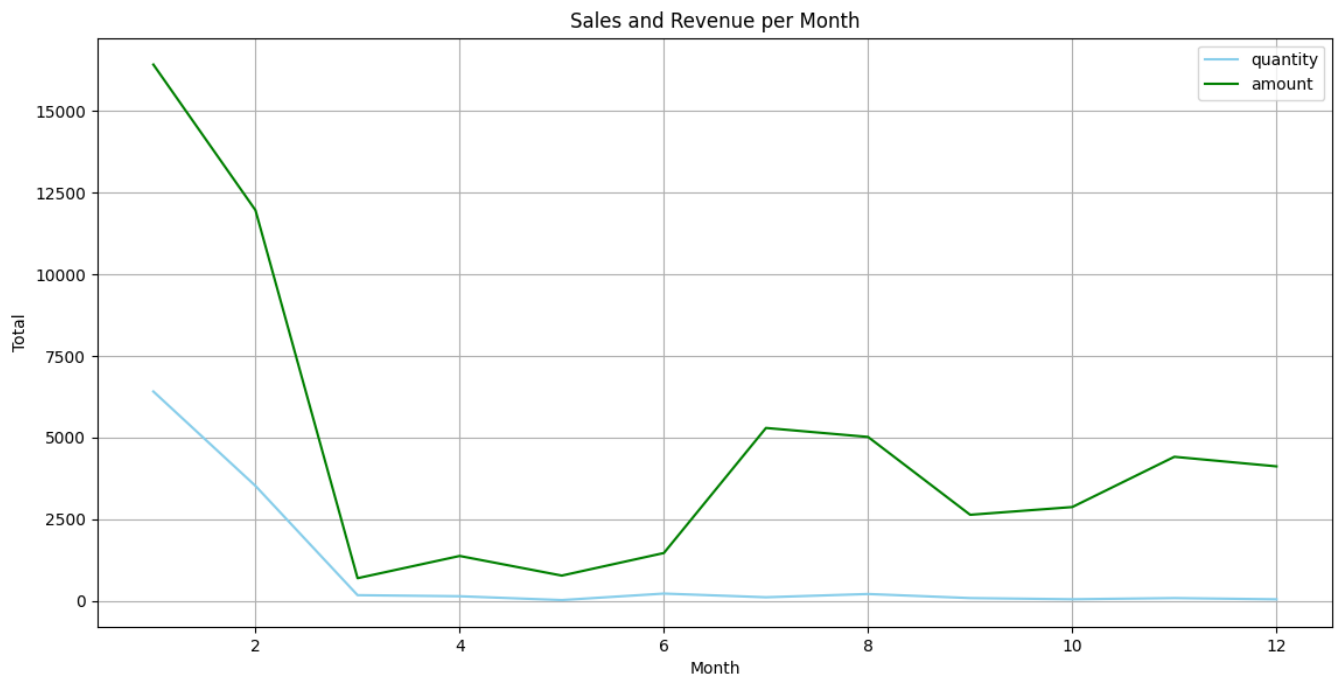
	product_name	quantity
0	ant white high rise bird box	45
1	ant white sweetheart table w 3 draw	3
2	baking set 9 piece retrospot	4
3	bank charges	18
4	beware of the cat metal sign	600
..
85	vintage blue kitchen cabinet	25
86	vintage red kitchen cabinet	27
87	vintage union jack shopping bag	24
88	way out metal sign	576
89	zinc police box lantern	187

Dari grafik dan hasil analisis diatas dapat dilihat bahwa penjualan produk tidak merata untuk setiap jenisnya. Produk dengan prnjualan tertinggi ialah *beware of the cat metal sign* dengan total penjualan dalam setahun sejumlah 600 unit. Peringkat

dua tertinggi ialah *way out metal sign* dan diikuti oleh *home sweet home metal sign*. Sedangkan produk dengan total penjualan terendah yang tidak melebihi 10 unit beberapa diantaranya ialah *ant white sweatheart table with 3 draw*, *baking set 9 piece retrospot*, *hall cabinet with 3 drawers*, *metal wall shelf unit with hooks*, *pink and lilac qualited throw*, *pink kashmiri occasional table*, dll.

➤ Total Penjualan Dan Total Revenue Setiap Bulan

Dalam tahun berjalan, total **penjualan terbanyak** dan **revenue tertinggi** terjadi di bulan Januari dengan total 6.414 unit dan dengan revenue \$16.426. Dan **penjualan paling sedikit** terjadi di bulan Oktober dengan total 51 unit namun dengan revenue \$2.875 yang berada pada peringkat 7. Sedangkan **revenue terendah** terjadi di bulan Maret dengan total \$696 dengan penjualan sebanyak 176 unit yang berada di peringkat 5. Perbedaan antara total penjualan dengan total revenue yang hubungannya *tidak* bersifat positif (ketika penjualan naik maka revenue akan naik) ini bisa terjadi disebabkan beberapa faktor salah satu diantaranya yaitu produk yang terjual pada bulan Oktober (penjualan paling sedikit) memiliki harga yang lebih tinggi dibandingkan dengan produk-produk yang terjual pada bulan Maret.



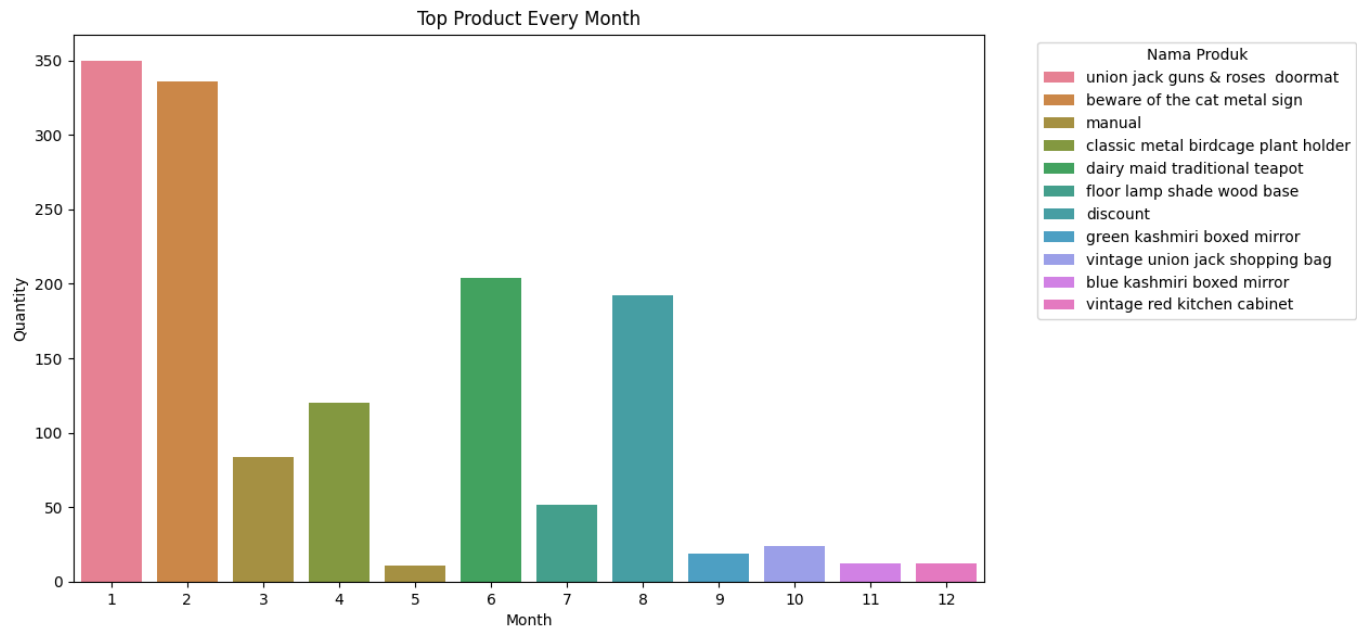
month_date	quantity
1	6414
2	3523
3	176
4	143
5	26
6	225
7	112
8	212
9	87
10	51
11	87
12	49

Total Penjualan

month_date	amount
1	16426.0
2	11965.0
3	696.0
4	1377.0
5	776.0
6	1467.0
7	5297.0
8	5022.0
9	2638.0
10	2875.0
11	4414.0
12	4121.0

Total Revenue

➤ **Tren Penjualan Produk Setiap Bulan**



month_date	product_name	quantity
1	union jack guns & roses doormat	350
2	beware of the cat metal sign	336
3	manual	84
4	classic metal birdcage plant holder	120
5	manual	11
6	dairy maid traditional teapot	204
7	floor lamp shade wood base	52
8	discount	192
9	green kashmiri boxed mirror	19
10	vintage union jack shopping bag	24
11	blue kashmiri boxed mirror	12
12	vintage red kitchen cabinet	12

Tren penjualan selama setahun setiap bulannya dapat dilihat pada grafik dan hasil analisi. Dapat disimpulkan bahwa produk dengan penjualan tertinggi setiap bulannya selalu berganti. Namun pada bulan Maret dan Mei penjualan tertinggi ialah produk “Manual”.