



FASHION RETAIL SALES EXPLORATORY DATA ANALYSIS

Mini Project for Digital Skill Fair 39.0 by
dibimbingid

Table of Content

Business Understanding

Data Understanding

Exploratory Data Analysis

Data Visualization

Insight & Conclusion

Business Understanding

Latar belakang

Penjualan produk-produk fashion semakin meningkat seiring dengan bertambahnya permintaan pasar yang terus berkembang. Hal ini menunjukkan bahwa pelanggan semakin tertarik dan percaya terhadap produk yang saat ini sedang beredar. Selain itu, kemudahan dari penggunaan metode pembayaran melalui kartu kredit semakin meningkatkan jumlah pembelian produk-produk fashion.

Tujuan

Berdasarkan latar belakang tersebut, maka tujuan analisis proyek ini adalah:

1. Mengetahui distribusi penjualan yang terjadi
2. Mengetahui tren penjualan dari produk-produk fashion
3. Memperkirakan produk terlaris berdasarkan penjualan

Data Understanding

Dataset dalam proyek kali ini menggunakan Fashion Retail Sales Dataset yang berasal dari pencarian melalui website Kaggle. Dataset yang digunakan ini berisi 3.400 data terkait penjualan produk fashion pada ritel, yang juga memuat berbagai detail tentang pembelian pelanggan, termasuk detail barang, jumlah pembelian, peringkat, dan metode pembayaran. Dataset ini juga berisi 6 kolom dengan penjelasan semua kolom sebagai berikut:

- Customer Reference ID: ID untuk setiap pelanggan.
- Item Purchased : Nama barang yang dibeli oleh pelanggan.
- Purchase Amount (USD): Harga barang yang dibeli dalam USD.
- Date Purchase: Tanggal pembelian dilakukan (format: DD-MM-YYYY).
- Review Rating: Rating ulasan dari pelanggan.
- Payment Method: Metode pembayaran yang digunakan.

Read Data

```
✓ [306] # Import library yang dibutuhkan
          import pandas as pd
          import seaborn as sns
          import matplotlib.pyplot as plt
          import numpy as np
          import warnings
          warnings.filterwarnings("ignore")

✓ [307] # Membaca data dari upload file
          df = pd.read_csv('/content/Fashion_Retail_Sales.csv')
```

Sebelum dilakukan read data, terlebih dahulu melakukan import library yang dibutuhkan. Kemudian dilakukan pemanggilan file csv yang telah diupload.

Read Data

```
✓ 0d # Menampilkan 5 data teratas dan 5 data terakhir
df
```

	Customer Reference ID	Item Purchased	Purchase Amount (USD)	Date Purchase	Review Rating	Payment Method
0	4018	Handbag	4619.0	05-02-2023	NaN	Credit Card
1	4115	Tunic	2456.0	11-07-2023	2.0	Credit Card
2	4019	Tank Top	2102.0	23-03-2023	4.1	Cash
3	4097	Leggings	3126.0	15-03-2023	3.2	Cash
4	3997	Wallet	3003.0	27-11-2022	4.7	Cash
...
3395	4118	Shorts	37.0	09-07-2023	2.7	Cash
3396	4056	Slippers	176.0	16-03-2023	2.5	Cash
3397	3991	Onesie	NaN	03-10-2022	2.4	Credit Card
3398	4042	Sandals	128.0	01-01-2023	1.8	Cash
3399	4001	Backpack	115.0	22-03-2023	1.4	Cash

3400 rows × 6 columns

Melakukan perintah untuk menampilkan data pada 5 data teratas dan 5 data terakhir

Exploratory Data Analysis

```
✓ [309] # Menampilkan jumlah baris dan kolom  
0d df.shape
```

```
→ (3400, 6)
```

Dilakukan pengecekan jumlah baris dan kolom, serta jumlah nilai beserta tipe data untuk setiap kolom

```
✓ 0d ➔ # Menampilkan informasi data  
df.info()
```

```
→ <class 'pandas.core.frame.DataFrame'>  
RangeIndex: 3400 entries, 0 to 3399  
Data columns (total 6 columns):  
 #   Column           Non-Null Count  Dtype     
---  --  
 0   Customer Reference ID  3400 non-null  int64    
 1   Item Purchased        3400 non-null  object    
 2   Purchase Amount (USD) 2750 non-null  float64   
 3   Date Purchase         3400 non-null  object    
 4   Review Rating         3076 non-null  float64   
 5   Payment Method        3400 non-null  object    
 dtypes: float64(2), int64(1), object(3)  
memory usage: 159.5+ KB
```

Missing Value

```
[315] # Cek missing value  
df.isnull().sum()  
  
Customer Reference ID      0  
Item Purchased              0  
Purchase Amount (USD)     650  
Date Purchase                0  
Review Rating             324  
Payment Method                0  
  
dtype: int64
```

Dilakukan pengecekan missing value pada dataset. Ditemukan dua kolom yang memiliki missing value, yaitu Purchase Amount (USD) dan Review Rating, sehingga harus dilakukan handling missing value untuk kedua kolom.

```
[316] # Mengatasi missing value  
for column in df.columns:  
    if df[column].dtype == 'object':  
        # Jika kolom bertipe object, isi dengan mode  
        df[column].fillna(df[column].mode()[0], inplace=True)  
    else:  
        # Jika kolom bertipe numerik, isi dengan mean  
        df[column].fillna(df[column].mean(), inplace=True)
```

Missing Value

```
[317] # Cek missing value  
df.isnull().sum()
```

Customer Reference ID	0
Item Purchased	0
Purchase Amount (USD)	0
Date Purchase	0
Review Rating	0
Payment Method	0

```
[318] # Menampilkan informasi data  
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 3400 entries, 0 to 3399  
Data columns (total 6 columns):  
 #   Column           Non-Null Count  Dtype     
---  --  
 0   Customer Reference ID  3400 non-null   int64    
 1   Item Purchased        3400 non-null   object    
 2   Purchase Amount (USD) 3400 non-null   float64  
 3   Date Purchase         3400 non-null   datetime64[ns]  
 4   Review Rating         3400 non-null   float64  
 5   Payment Method        3400 non-null   object    
dtypes: datetime64[ns](1), float64(2), int64(1), object(2)  
memory usage: 159.5+ KB
```

Hasil setelah dilakukan handling missing value.
Terlihat bahwa missing value sudah 0 serta jumlah nilai pada semua kolom sudah berjumlah 3400.

Data Duplicate

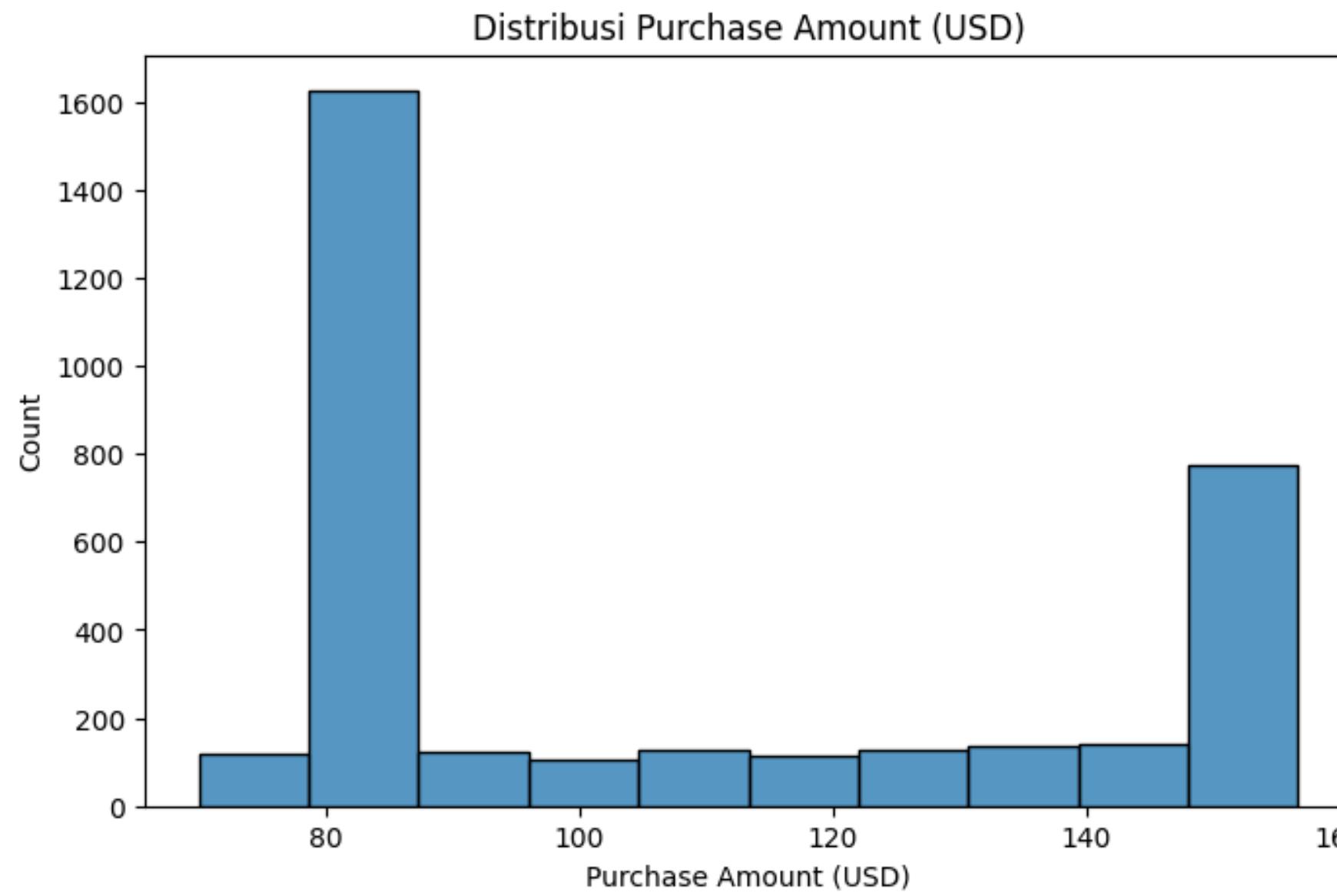
```
✓ [53] # Cek data duplikat  
0d duplicates = df.duplicated().sum()  
print(duplicates)
```

```
→ 0
```

Dilakukan pengecekan data yang duplikat.
Dari hasil yang ditampilkan menunjukkan bahwa data tidak ada yang duplikat, sehingga tidak perlu dilakukan handling data duplicate

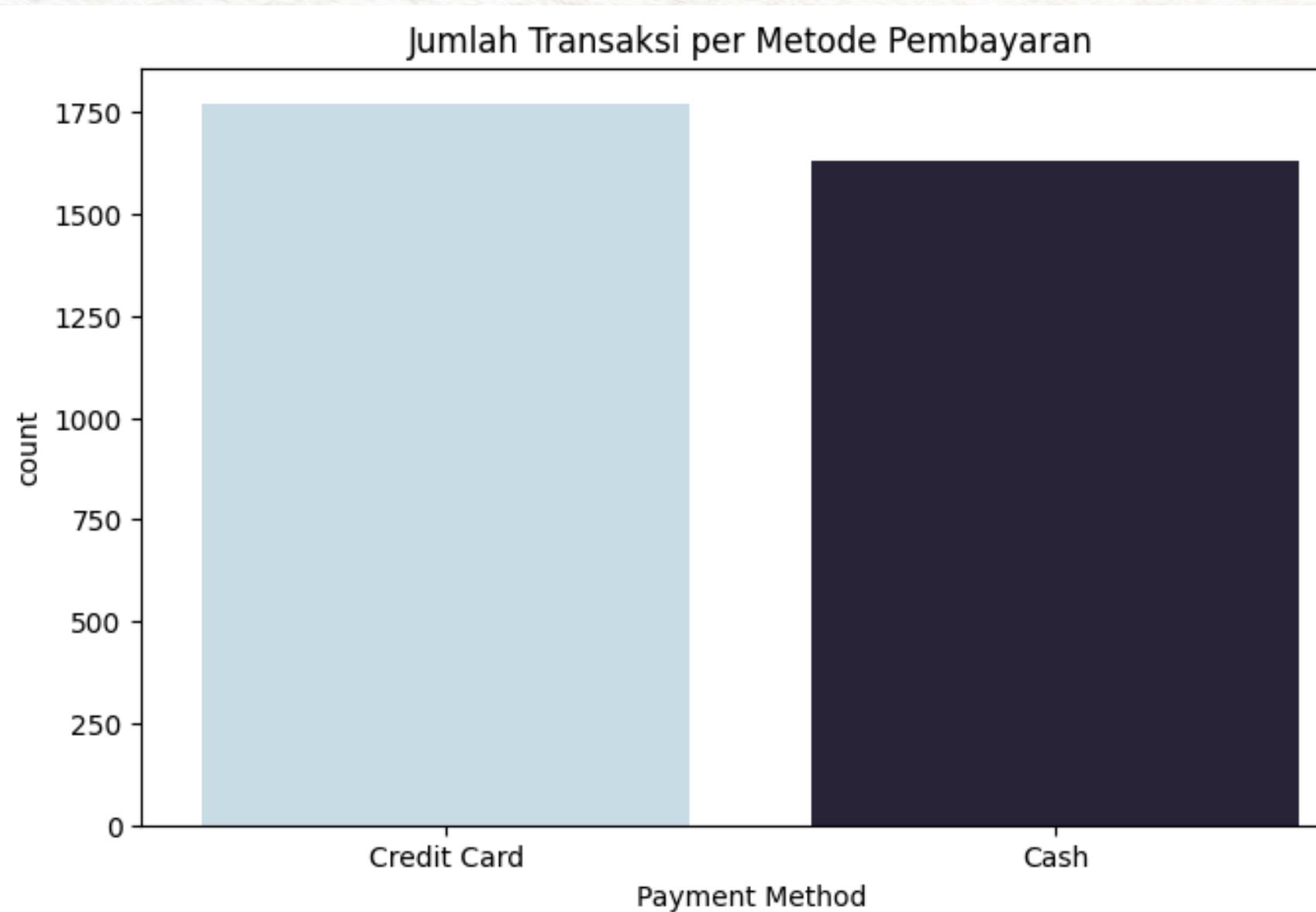
Data Visualization

Distribusi Harga Barang (USD)



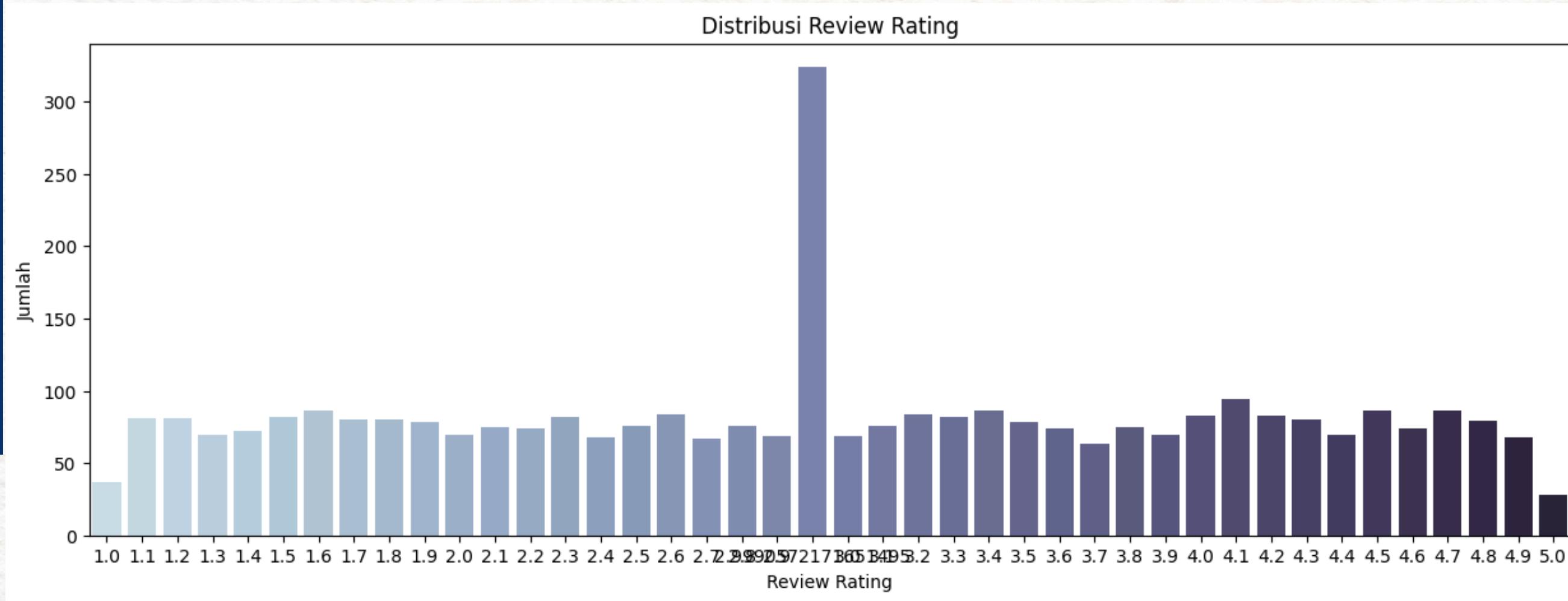
Hasil visualisasi data terkait distribusi harga barang-barang fashion. Terlihat bahwa barang fashion dengan harga yang berkisar di harga 80 USD menjadi yang tertinggi.

Jumlah Transaksi Metode Pembayaran



Hasil visualisasi data terkait jumlah transaksi dari metode pembayaran yang digunakan. Terlihat bahwa kebanyakan pelanggan memilih menggunakan credit card dibanding cash.

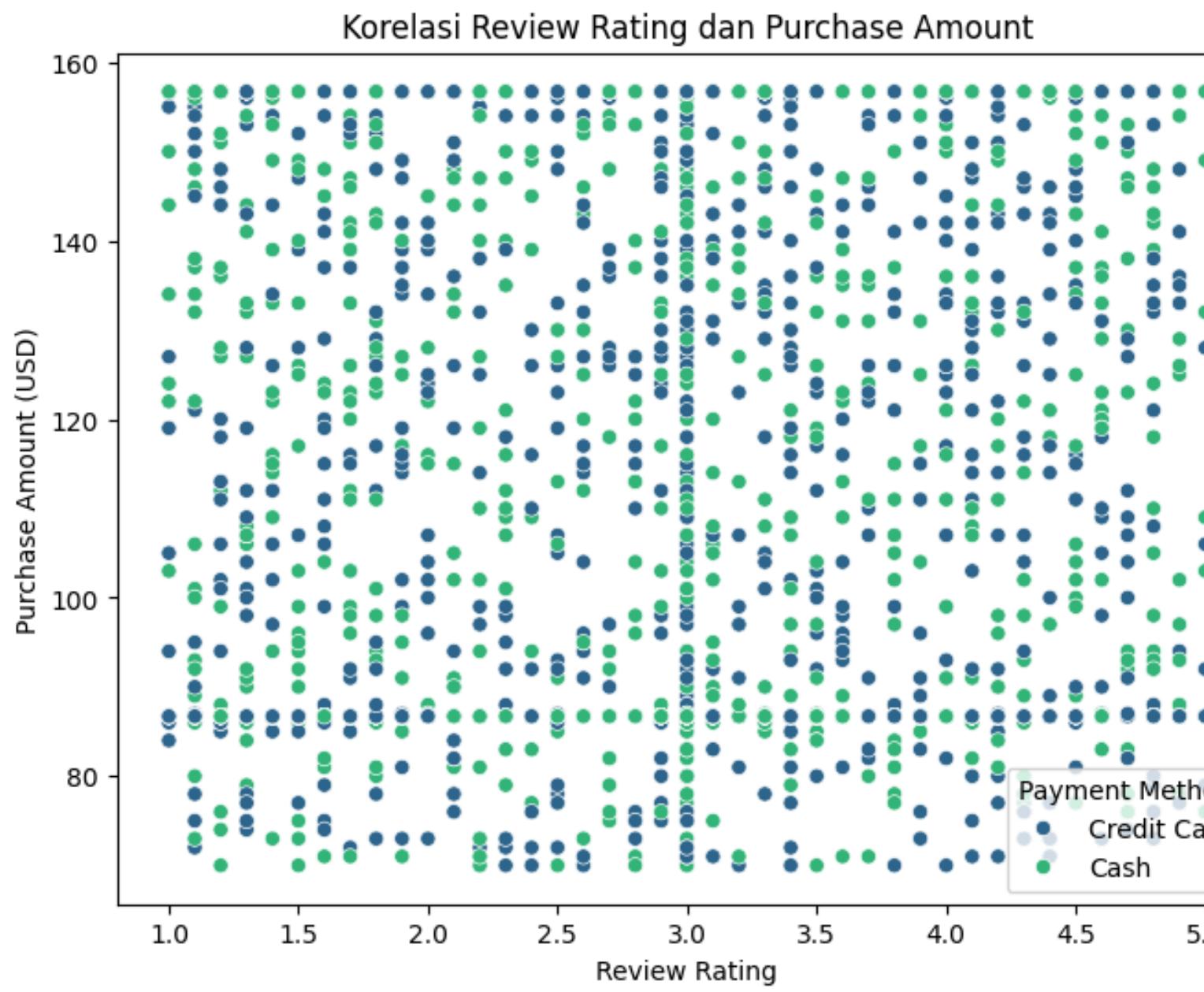
Distribusi Review Rating



Hasil visualisasi data terkait distribusi review rating.

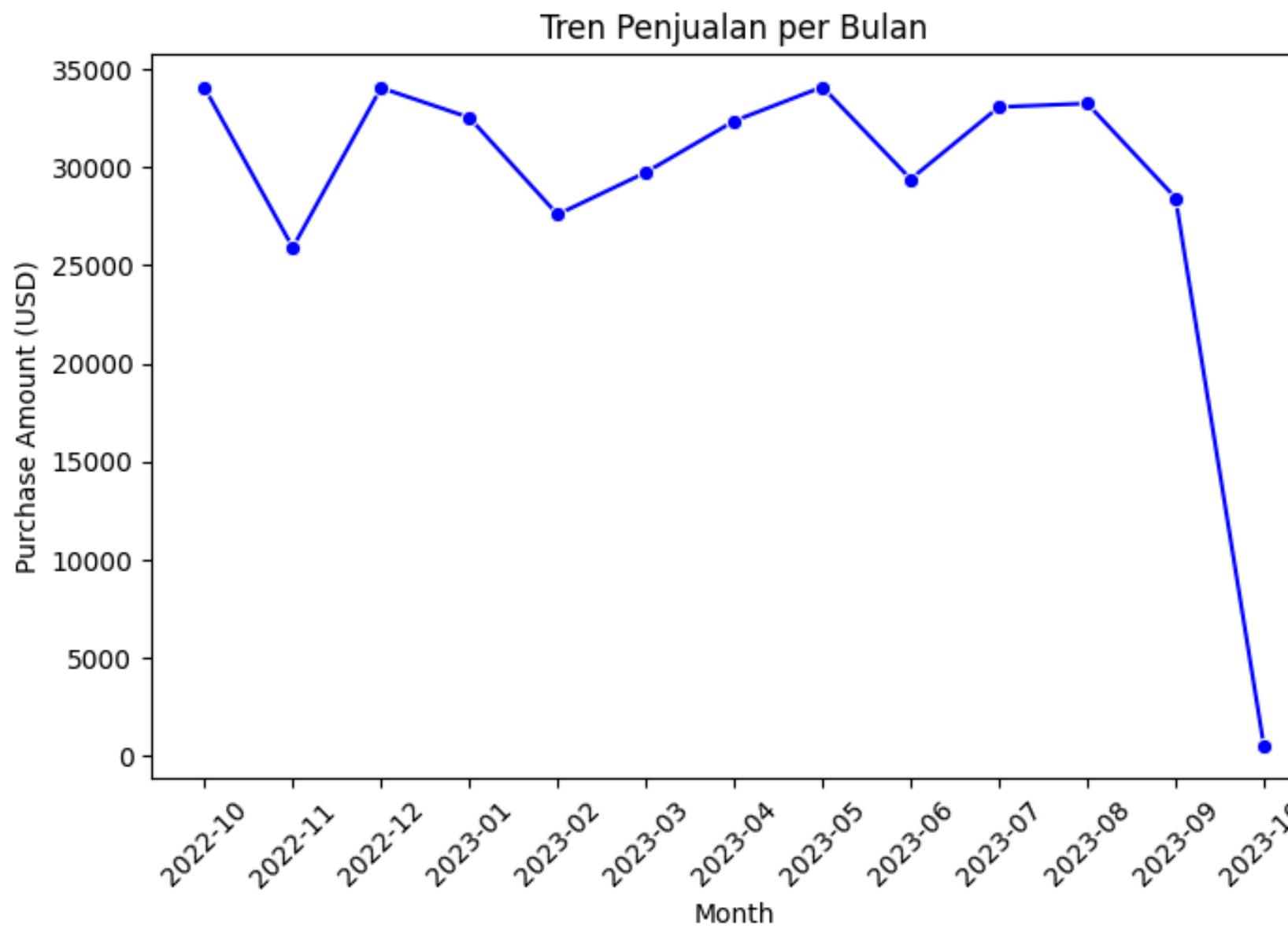
Terlihat bahwa kebanyakan pelanggan memberikan review rating pada rentang nilai 2 hingga 3. Menjadikan bahwa barang ataupun pelayanan cenderung biasa-biasa saja.

Korelasi Review Rating dan Harga Barang (USD)



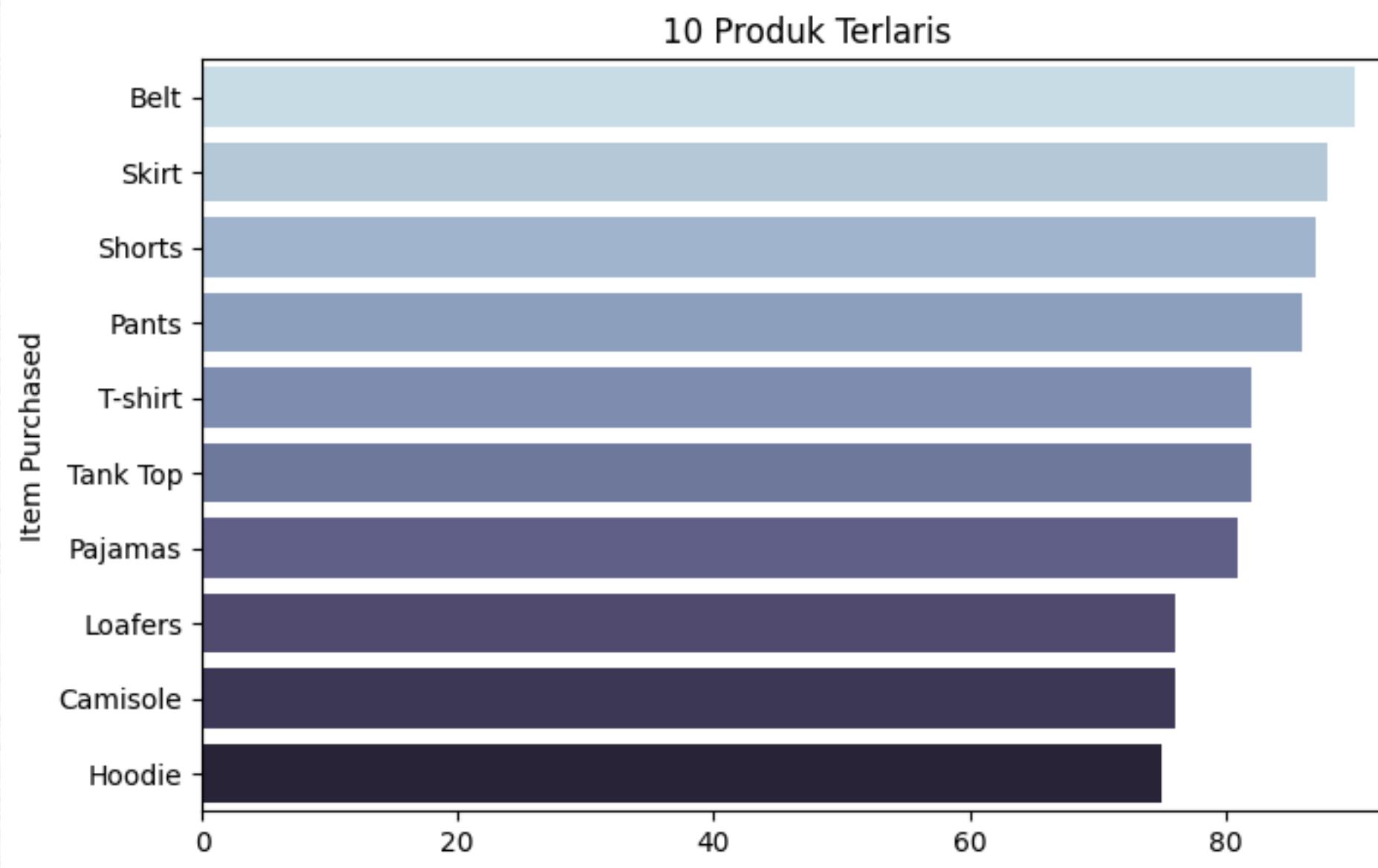
Hasil visualisasi data terkait hubungan antara harga barang, review rating, dan metode pembayaran. Terlihat bahwa kebanyakan review rating lebih berpusat di angka 3.0, tetapi harga barang tidak menjadi faktor hubungan dengan review rating. Pada metode pembayaran pun terlihat bahwa data tersebar, sehingga preferensi metode tidak melihat harga dari barang dan review rating.

Tren Penjualan per Bulan



Hasil visualisasi data terkait grafik tren penjualan per bulan menggunakan harga barang setiap transaksi. Terlihat bahwa tren penjualan terjadi kenaikan dan penurunan yang sering terjadi. Kenaikan penjualan terjadi sebanyak 3 kali, sedangkan penurunan penjualan terjadi sebanyak 4 kali.

Produk Terlaris



Hasil visualisasi data terkait produk-produk terlaris yang dibeli. Terlihat bahwa belt atau sabuk menjadi produk yang paling sering dibeli oleh pelanggan. Produk dengan angka lebih dari 80 ditempati oleh 7 produk terlaris dari 10 produk yang ditampilkan.

Insight & Conclusion

Insight

- Harga setiap barang yang ada cenderung bervariasi
- Metode pembayaran credit card lebih banyak digunakan pelanggan dibanding cash
- Review rating pelanggan sekitar nilai 2-3, mengindikasikan kepuasan pelanggan yang cenderung biasa-biasa saja
- Tren penjualan yang mengalami kenaikan dan penurunan bergantian yang hampir secara rutin terjadi
- Produk-produk terlaris kebanyakan terkait pakaian, mengindikasikan pelanggan lebih mencari produk terkait pakaian

Conclusion

Dapat disimpulkan bahwa meskipun penjualan beberapa kali mengalami penurunan, tetapi pelanggan masih tetap kembali lagi membeli hingga penjualan kembali naik. Berdasarkan penjualan tersebut, produk terkait pakaian menjadi produk terlaris, meskipun review rating untuk semua produk yang ada terkesan biasa saja.

About me

Fajar Alfiansyah Asikin



Brebes, Central Java



alfifajaras22@gmail.com



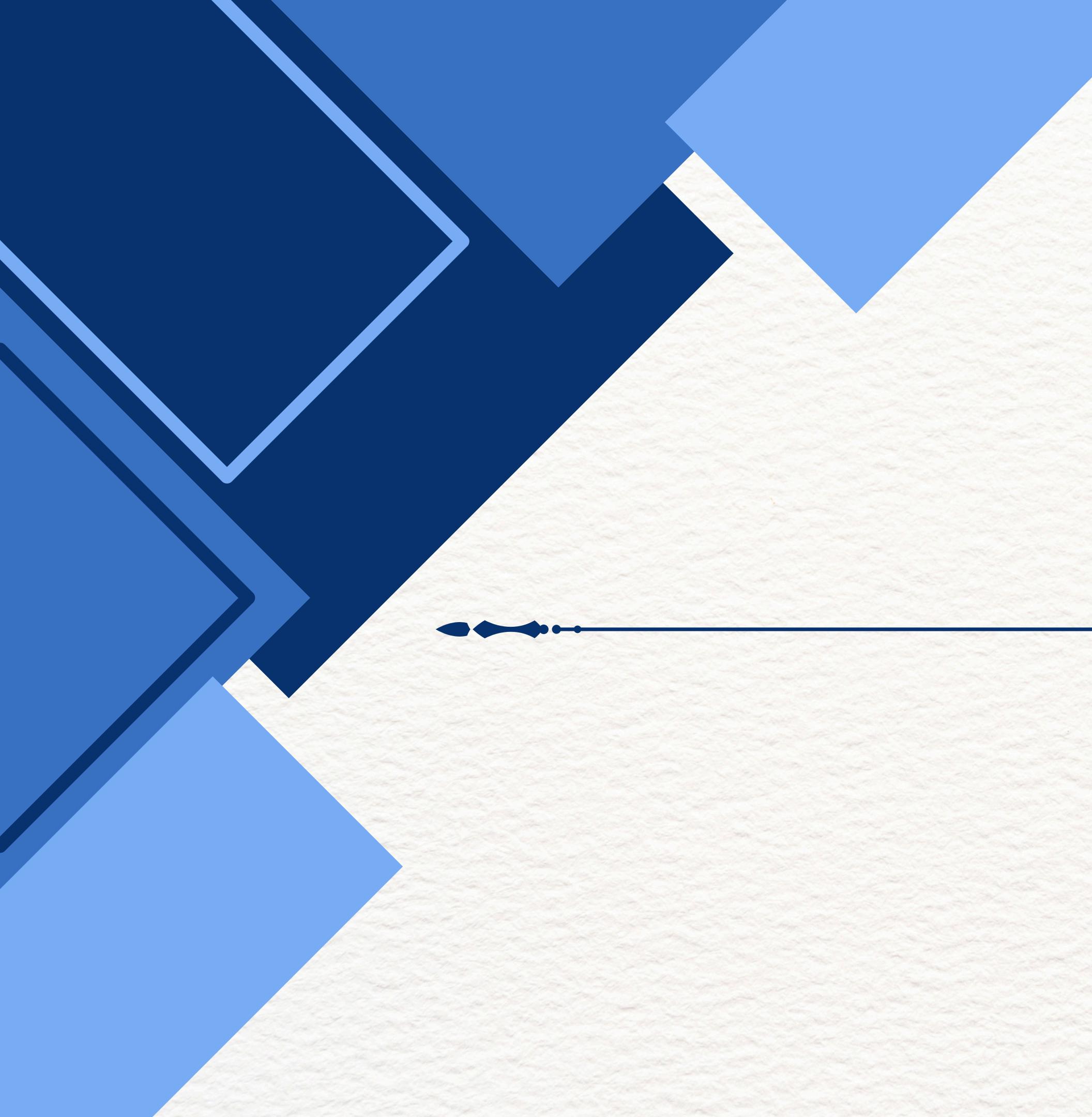
github.com/alfhyansalfi



Information Systems Student
at Telkom University Purwokerto



linkedin.com/in/fajaralfiansyahasikin

The background features abstract, overlapping blue geometric shapes. On the left, there are several large, dark blue arrows pointing diagonally upwards and to the right. In the bottom right corner, there is a dark blue square containing a smaller, lighter blue square.

THANK YOU

