

E-commerce Analysis

Bootcamp Data Analyst with SQL &
Python using Google Platform

Luthfiyah Astutiningtyas



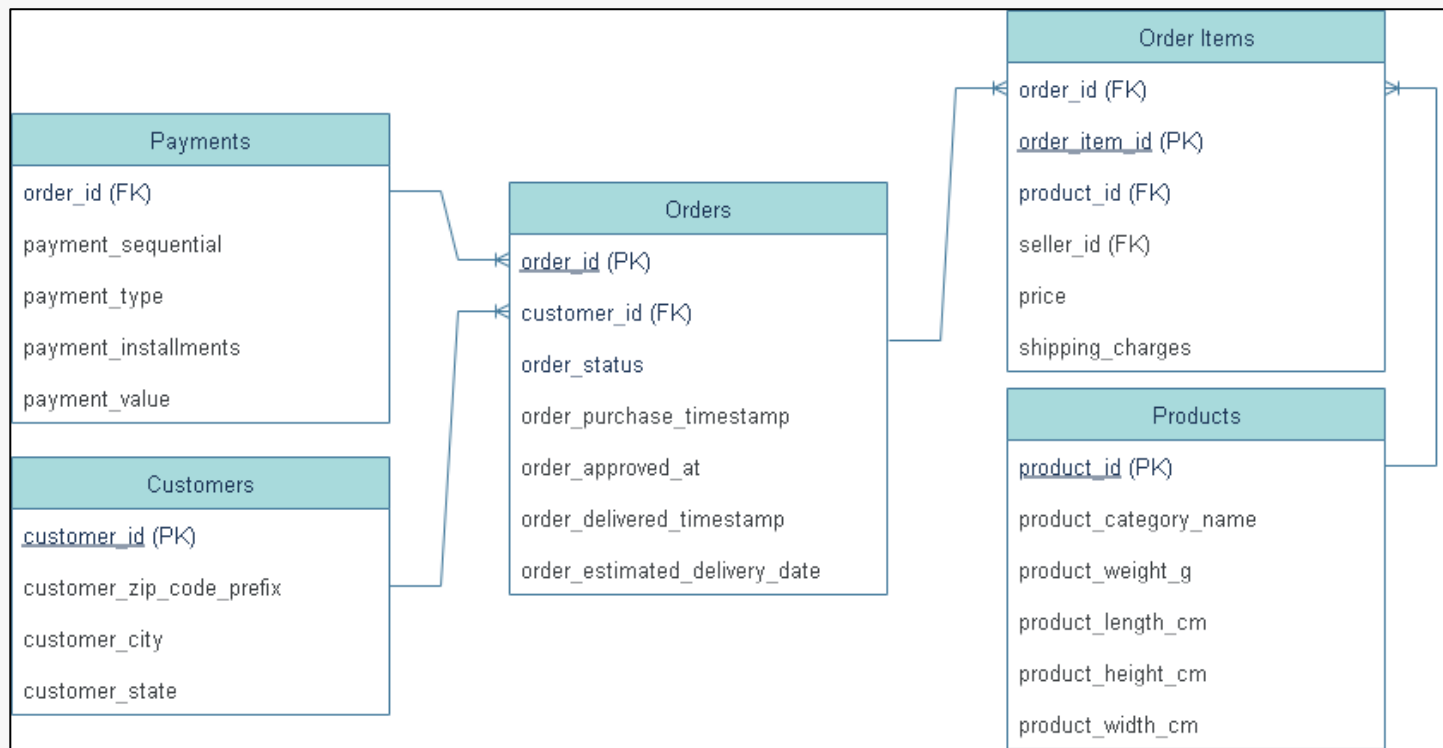
Dataset ini berisi informasi lengkap mengenai transaksi e-commerce, mencakup detail pesanan (orders), pelanggan (customers), produk (products), penjual (sellers), metode pembayaran (payments), serta proses pengiriman. Dengan struktur data yang saling terhubung, dataset ini memungkinkan analisis menyeluruh terhadap alur belanja pelanggan, mulai dari pemesanan, persetujuan, pembayaran, hingga pengiriman produk. Data ini dapat digunakan untuk memahami perilaku pelanggan, tren penjualan, distribusi pembayaran, serta performa pengiriman.

Dari sisi bisnis, tantangan utama yang dapat diidentifikasi adalah bagaimana memastikan pesanan pelanggan diproses dan dikirim tepat waktu, meminimalkan pembatalan atau penundaan pesanan, serta mengoptimalkan nilai transaksi. Keterlambatan pengiriman dan tingginya jumlah pesanan yang dibatalkan dapat menurunkan kepuasan pelanggan dan berdampak pada loyalitas jangka panjang. Dengan menganalisis data ini, perusahaan dapat menemukan pola keterlambatan, mengukur kontribusi penjual terhadap total penjualan, serta mengidentifikasi peluang peningkatan efisiensi operasional dan pengalaman pelanggan.

Dataset yang digunakan

1. **Orders Table:** Tabel ini berisi informasi utama mengenai pesanan (order) dari pelanggan. Informasi lengkap mengenai pemesanan: mulai dari ID order, status, waktu pembelian, persetujuan, hingga perkiraan dan waktu pengiriman.
2. **Order Items Table:** Tabel ini berisi detail setiap item di dalam sebuah pesanan. Detail setiap item dalam order: mencakup produk, seller, harga, dan biaya pengiriman.
3. **Customers Table:** Tabel ini berisi informasi pelanggan. Data pelanggan seperti kota dan provinsi, berguna untuk analisis demografis dan segmentasi.
4. **Payments Table:** Tabel ini berisi informasi pembayaran untuk setiap pesanan. Rincian metode pembayaran, jumlah cicilan, dan nilai transaksi—penting untuk melihat preferensi dan pola bayar.
5. **Products Table:** Tabel ini berisi informasi produk. Data produk lengkap dengan kategori dan dimensi fisik (berat dan ukuran).

Diagram ERD (Entity-Relationship Diagram)



Langkah Pengerjaan

1. Mengumpulkan Data

✓ Extract Data from Kaggle

```
!pip install kaggle
from google.colab import files

files.upload() # upload kaggle.json dari akun Kaggle
!mkdir -p ~/.kaggle
!cp kaggle.json ~/.kaggle/
!chmod 600 ~/.kaggle/kaggle.json
!kaggle datasets download -d bytadit/ecommerce-order-dataset
!unzip "ecommerce-order-dataset.zip" "Ecommerce Order Dataset/train/*" -d dataset
!ls dataset/"Ecommerce Order Dataset"/train
```

Data berasal dari Kaggle:

<https://www.kaggle.com/datasets/bytadit/ecommerce-order-dataset>

Python code dapat diakses [disini](#)

Langkah Pengerjaan

2. Membersihkan Data (*Data Cleaning / Preprocessing*)

Missing Value

```
#Cek missing value
customers.isnull().sum()
```

	0
customer_id	0
customer_zip_code_prefix	0
customer_city	0
customer_state	0

dtype: int64

Duplikat Data

```
# Cek jumlah baris duplikat
print("Jumlah duplikat:", customers.duplicated().sum())

# Tampilkan baris duplikat (kalau ada)
customers[customers.duplicated()]
```

Jumlah duplikat: 0

customer_id	customer_zip_code_prefix	customer_city	customer_state
-------------	--------------------------	---------------	----------------

Pada tahapan ini, dilakukan pembersihan data untuk mengatasi missing value, duplikat data dan inkonsisten tipe data pada masing-masing tabel.

Python code dapat diakses [disini](#)

Langkah Pengerjaan

3. Data Integration (Join/Merge)

```
# Merge step-by-step
df = orders.merge(order_items, on='order_id', how='left')
df = df.merge(customers, on='customer_id', how='left')
df = df.merge(products, on='product_id', how='left')
df = df.merge(payments, on='order_id', how='left')

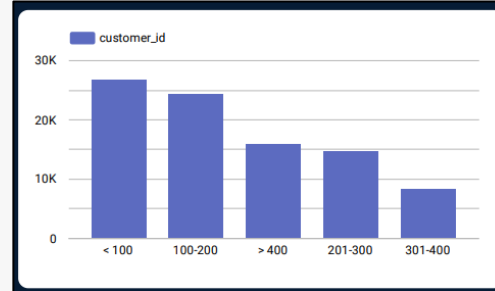
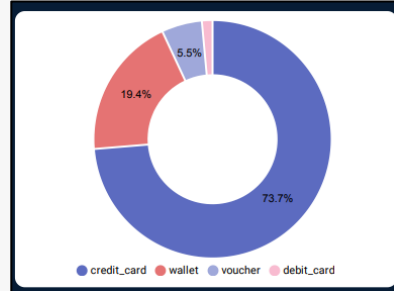
# Hasil final
print("Jumlah baris setelah merge:", len(df))
df
```

Tahapan ini dilakukan left join pada Orders Table, Order Items Table, Customers Table, Products Table dan Payment Table.

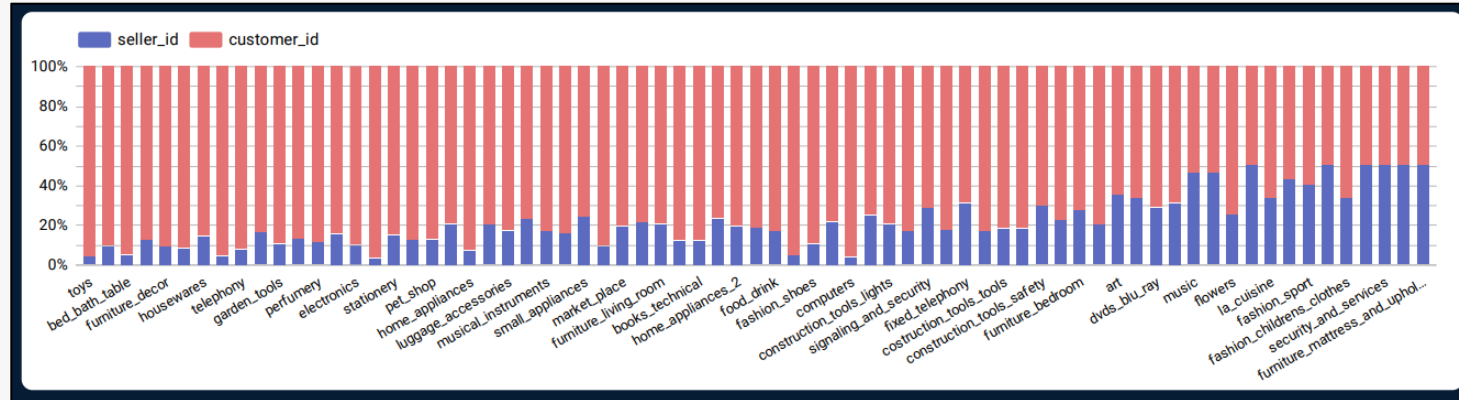
Python code dapat diakses [disini](#)

Langkah Pengerjaan

4. Eksplorasi Data (EDA - Exploratory Data Analysis)

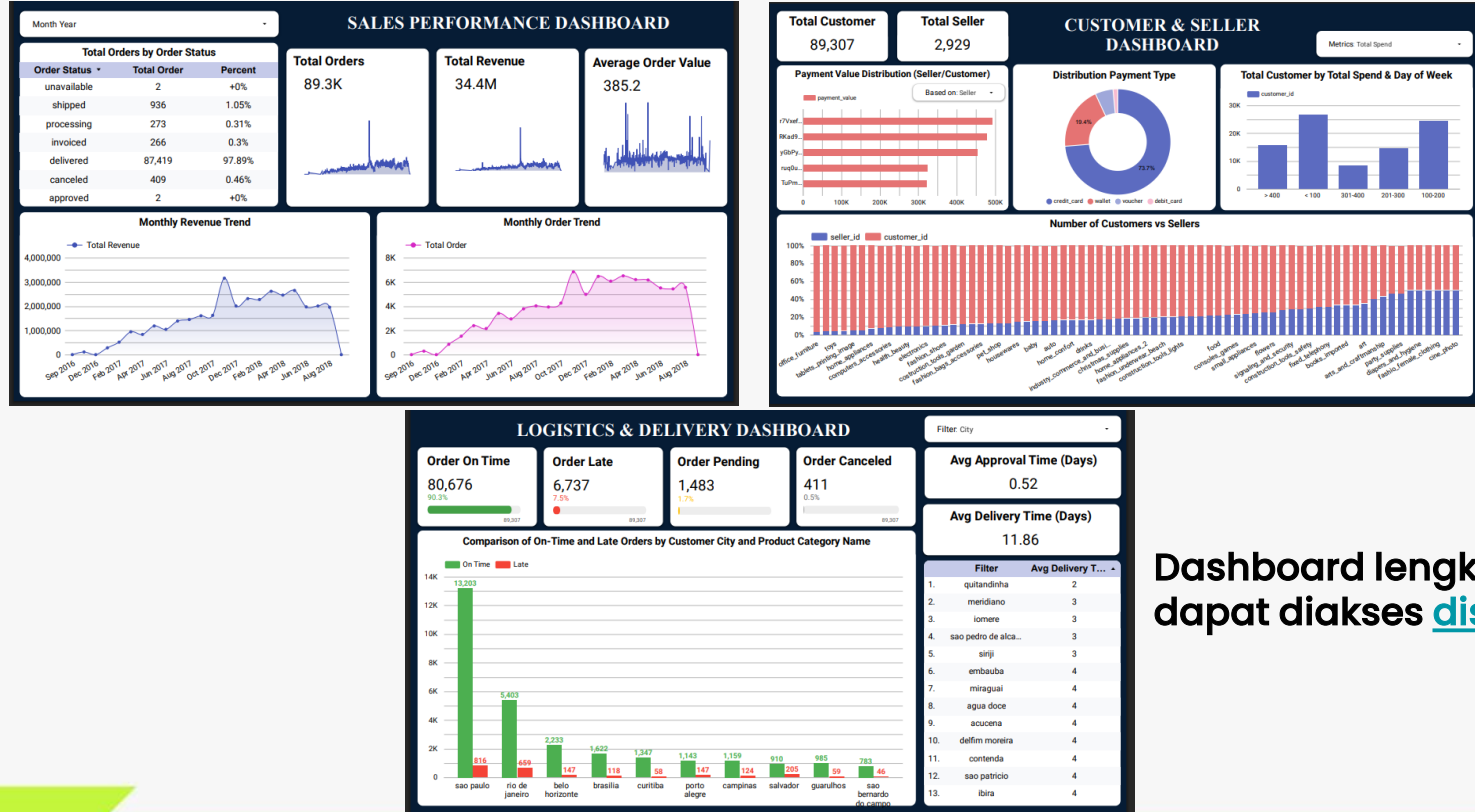


Tahapan ini dilakukan untuk menggali insight dari data yang ada. Untuk memudahkan, maka dilakukan visualisasi data.



Python code dapat diakses [disini](#)

5. Pembuatan Dashboard Interaktif



Dashboard lengkap
dapat diakses [disini](#)

Business Recommendation

Berdasarkan hasil analisis, perusahaan sebaiknya fokus pada peningkatan efisiensi proses persetujuan dan pengiriman order. Hal ini dapat dilakukan dengan mempercepat validasi pembayaran, memastikan ketersediaan stok, serta memperkuat koordinasi dengan seller agar estimasi pengiriman lebih akurat. Selain itu, perlu dibuat sistem monitoring untuk mendeteksi potensi keterlambatan sejak awal, sehingga tindakan preventif bisa segera diambil sebelum berdampak pada kepuasan pelanggan.

Dari sisi revenue, perusahaan dapat meningkatkan strategi segmentasi pelanggan dan optimasi kategori produk. Analisis transaksi dapat digunakan untuk mengidentifikasi produk dengan margin tinggi dan pelanggan dengan nilai belanja signifikan, lalu diarahkan pada promosi yang lebih personal. Untuk menjaga keberlanjutan, perusahaan juga sebaiknya membangun dashboard performa seller yang transparan, agar penjual dengan tingkat keterlambatan tinggi dapat diberi peringatan sebelumnya. Dengan langkah ini, pengalaman belanja pelanggan dapat lebih konsisten, kepuasan meningkat, dan revenue dapat terus tumbuh.

Dataset:

<https://www.kaggle.com/datasets/bytadit/ecommerce-order-dataset>

Python Code:

<https://colab.research.google.com/drive/1tCBnq9vSqGKKOlQPESBoZ-qgqfQJlhSB?usp=sharing>

Dashboard Interaktif:

https://lookerstudio.google.com/s/qBzlolw5l_E

Thanks!