STAGE 1  
EXPLORATORY DATA ANALYSIS: AMAZON SALES DATASET

**\_\_\_**



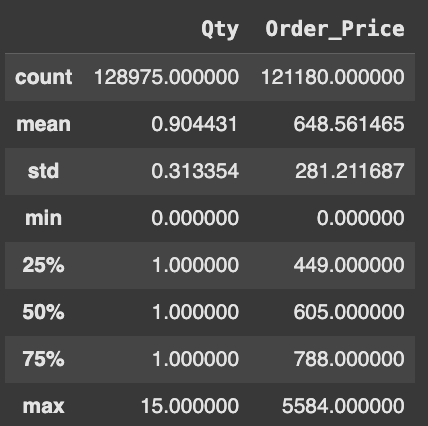
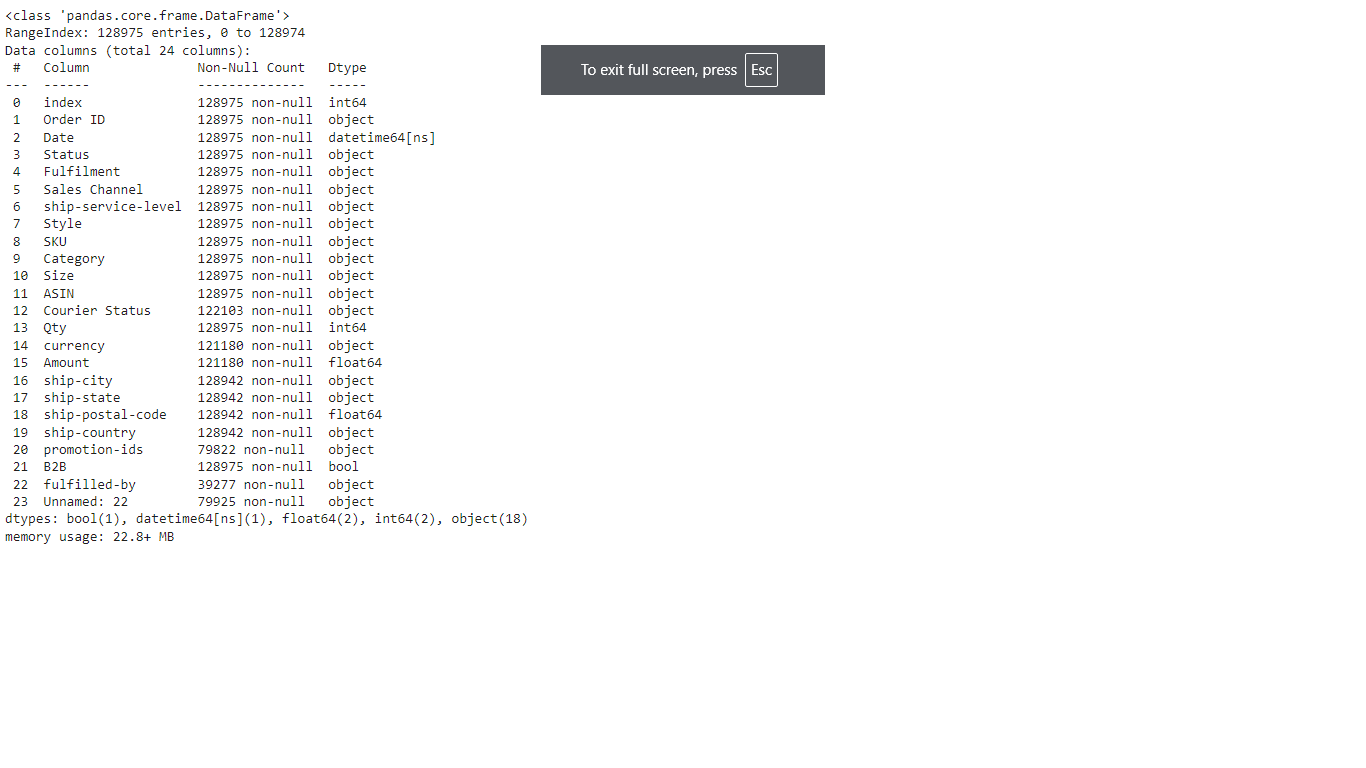
Oleh Kelompok 1- Last Voyage:

1. Mardio Edana Putra
2. Mochamad Rizqi
3. Muhammad Gusti Maulana Alfin
4. Nena Herdiani

# 

Dengan melihat jumlah fitur yang cukup banyak untuk mendukung pemodelan prediksi status paket, dan jumlah baris data yang cukup banyak yaitu sekitar 128ribu-an maka dataset Amazon Sale Report dirasa sudah cukup memadai untuk menghasilkan model machine learning yang cukup baik nantinya. Fitur-fitur yang relevan dan dianggap sangat berguna untuk dieksplor dan dilihat adalah **Order ID, Date, Fulfilment, Sales Chanel, ship-service-level, SKU, Category, Size, Qty, Amount, ship-city, ship-state, ship-postal-code, B2B, Courier Status, dan Status.**

# DESCRIPTIVE STATISTICS



A. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai?

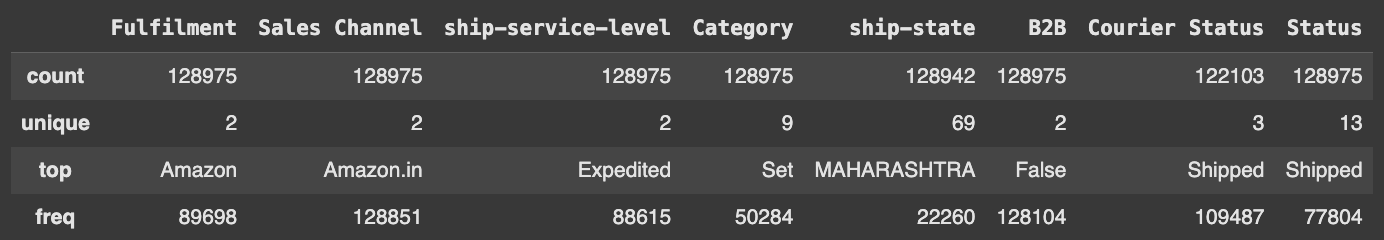
Berdasarkan output dari fungsi info(), tipe data kolom sudah terlihat sesuai, yaitu datetime, int64, float64, dan object. Tidak ada kolom dengan tipe data yang tidak sesuai dengan isinya. Sebagai contoh, kolom index menggunakan int64, kolom seperti Order ID, Status, dan Category menggunakan object, serta kolom numerik seperti Qty dan Amount menggunakan tipe float64. Jadi, tidak ada masalah dengan tipe data.

B. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?

Terdapat beberapa kolom yang memiliki nilai null, yaitu **Courier Status, Order\_Price, ship-city, ship-state, dan ship-postal-code.**

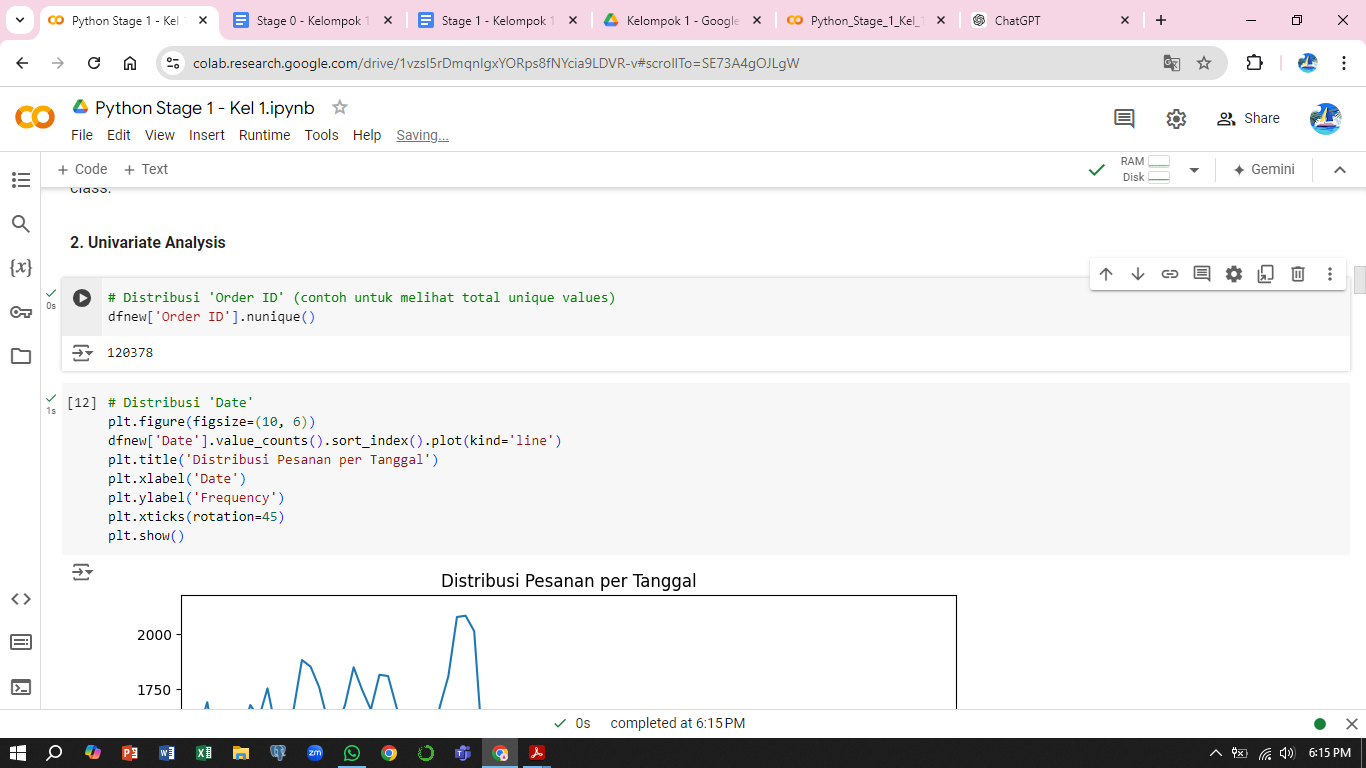
C. Apakah ada kolom yang memiliki nilai summary agak aneh (min/mean/median/max/unique/top/freq)?

Tidak terdapat perbedaan yang signifikan antara mean dan median pada kolom Order\_Price dan Qty sehingga dapat dikatakan bahwa kolom tersebut memiliki data **mendekati distribusi normal.** Selain itu, nilai minimum pada kolom Order\_Price yang bernilai 0 juga perlu diperhatikan, karena ini mungkin menunjukkan pesanan tanpa transaksi, diskon penuh, ataupun kesalahan input.

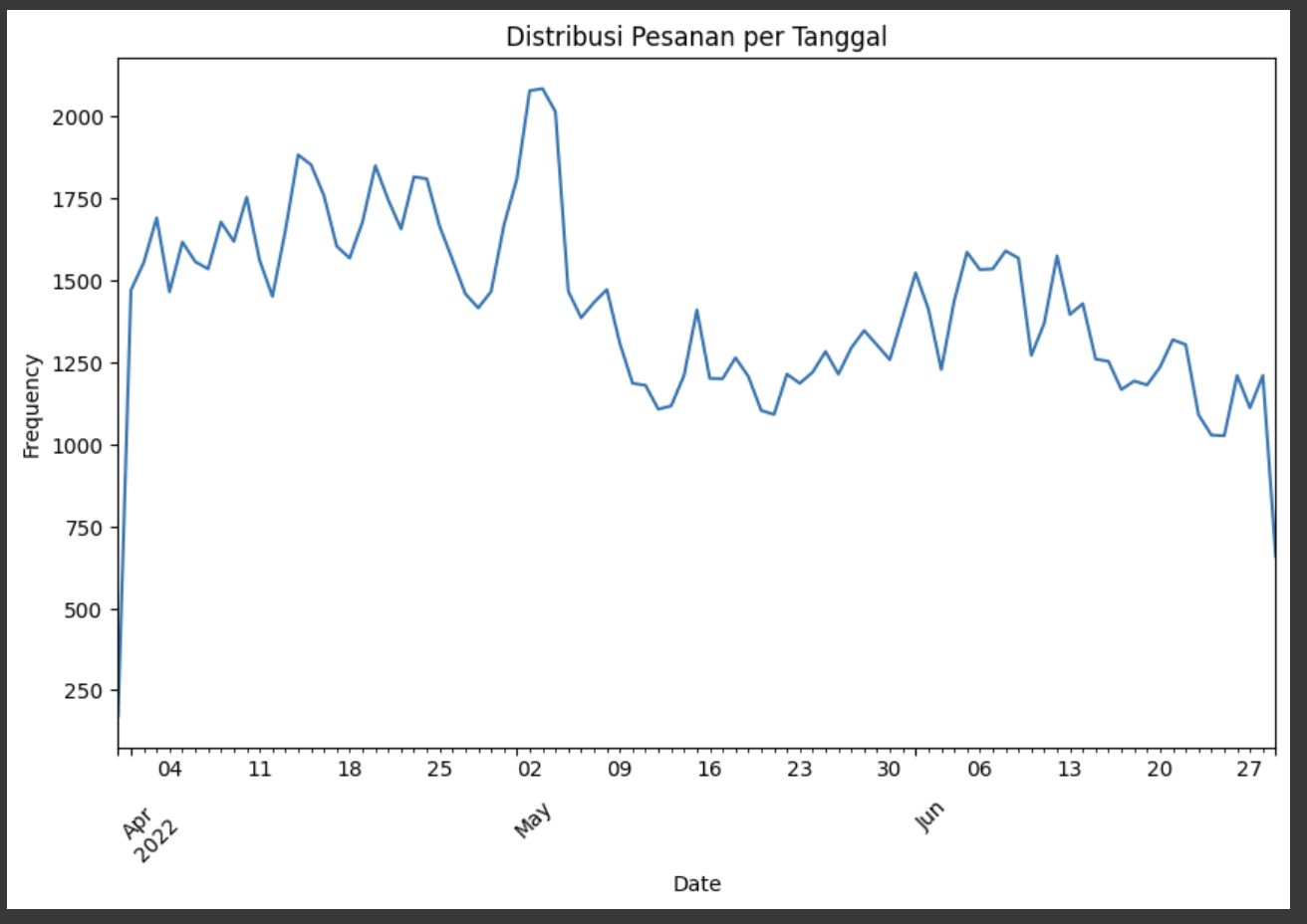


Jumlah unique values pada semua kolom kategorikal masih masuk akal, sebagai contoh untuk ship-state pengiriman terdapat 69 state. Kemudian untuk kolom target 'status' terdapat 13 jenis untuk status pengiriman pada paket yang ada. Untuk nilai yang sering muncul pada kolom Sales Channel dan B2B dengan jumlah nilai unik hanya 2 dan frekuensinya hampir mendekati jumlah baris mengindikasikan bahwa distribusi nilai terlalu timpang sehingga kemungkinan fitur tidak akan terlalu berguna dalam klasifikasi nantinya dan perlu dilakukan balance class.

# UNIVARIATE ANALYSIS



Dari gambar diatas dapat disimpulkan terdapat 120378 unique values untuk kolom Order ID dari 128975 baris data.



Grafik tersebut menunjukkan distribusi atau frekuensi pesanan per tanggal dari bulan April hingga akhir Juni 2022 yang dipetakan dari kolom ‘Date’. Berikut adalah beberapa poin analisis yang dapat diperoleh dari grafik tersebut:

1. **Tren Frekuensi Pesanan:**

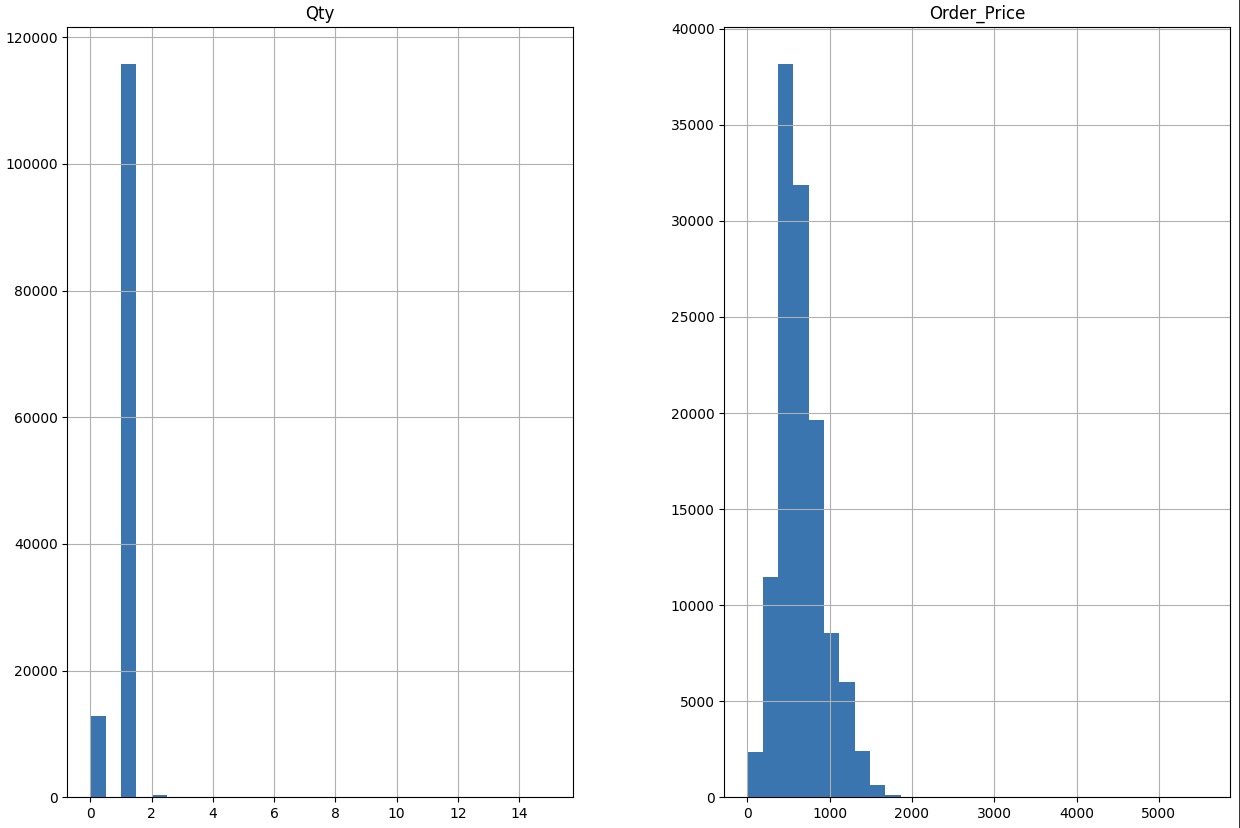
Frekuensi pesanan tampaknya relatif tinggi di awal bulan April, dengan jumlah pesanan yang fluktuatif tapi cukup konsisten hingga awal Mei. Setelah puncak yang terjadi di awal Mei, terjadi sedikit penurunan pada pertengahan hingga akhir bulan Mei. Pada bulan Juni, frekuensi pesanan cenderung mengalami fluktuasi yang lebih stabil namun sedikit menurun secara bertahap, terutama menjelang akhir bulan.

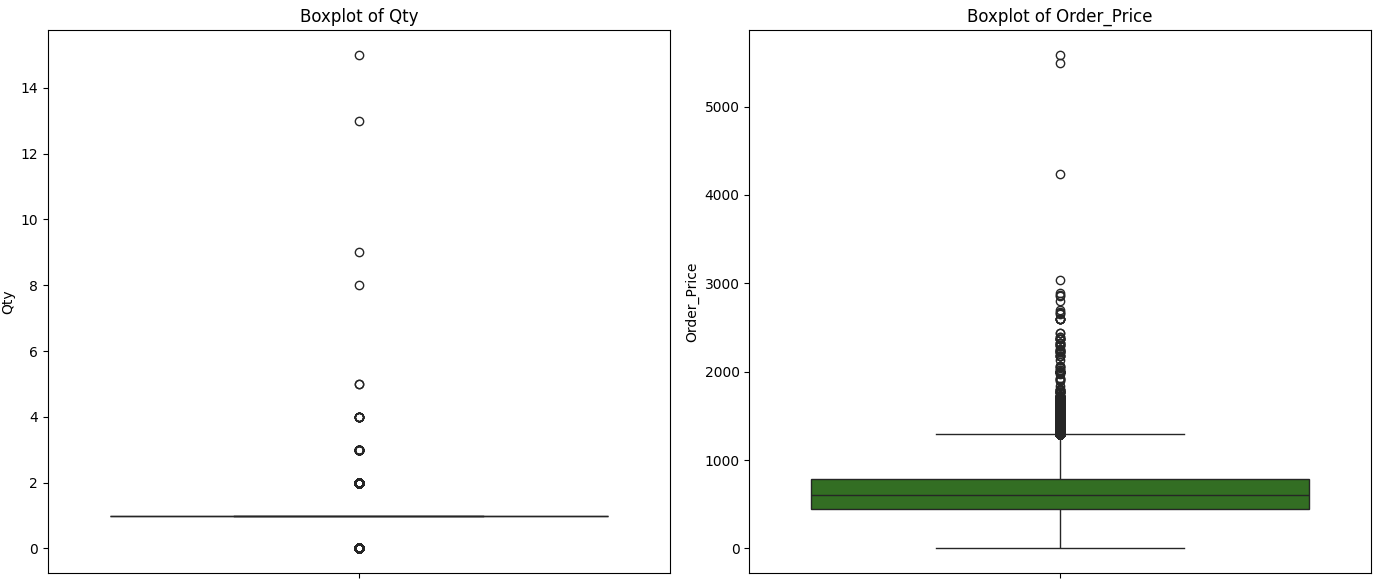
1. **Puncak dan Penurunan Frekuensi:**

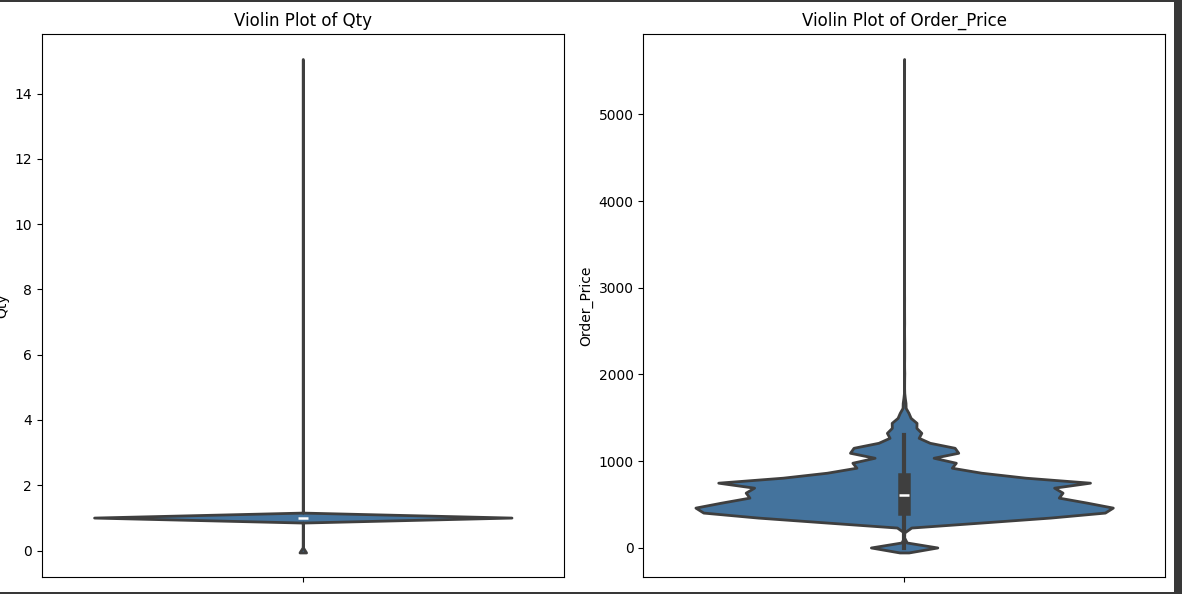
Ada puncak frekuensi pesanan yang cukup menonjol di awal Mei, yang mungkin menunjukkan periode promosi khusus atau peningkatan permintaan. Hal ini dapat terjadi pada periode liburan, promosi musiman, atau hari-hari besar. Setelah puncak tersebut, terjadi penurunan frekuensi pesanan. Ini bisa menjadi sinyal untuk melakukan analisis lebih lanjut mengenai penyebabnya, seperti berakhirnya promosi, atau periode pasca-hari besar ketika permintaan biasanya menurun.

# Frekuensi Menurun di Akhir Periode:

# Pada akhir Juni, frekuensi pesanan terlihat menurun tajam, bahkan hingga mencapai titik terendah. Penurunan ini bisa jadi karena adanya perubahan pola belanja konsumen, pergeseran musim, atau bahkan masalah eksternal seperti keterlambatan logistik atau stok yang terbatas.

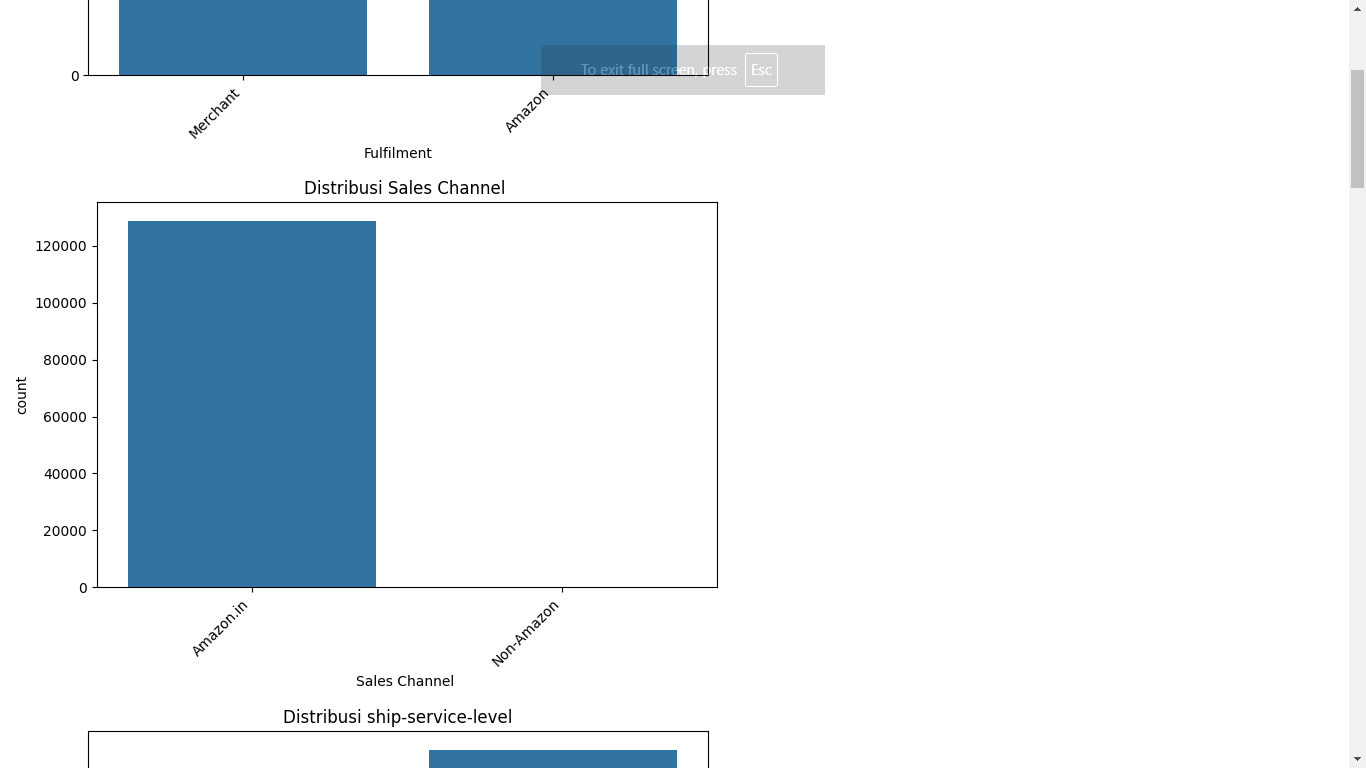
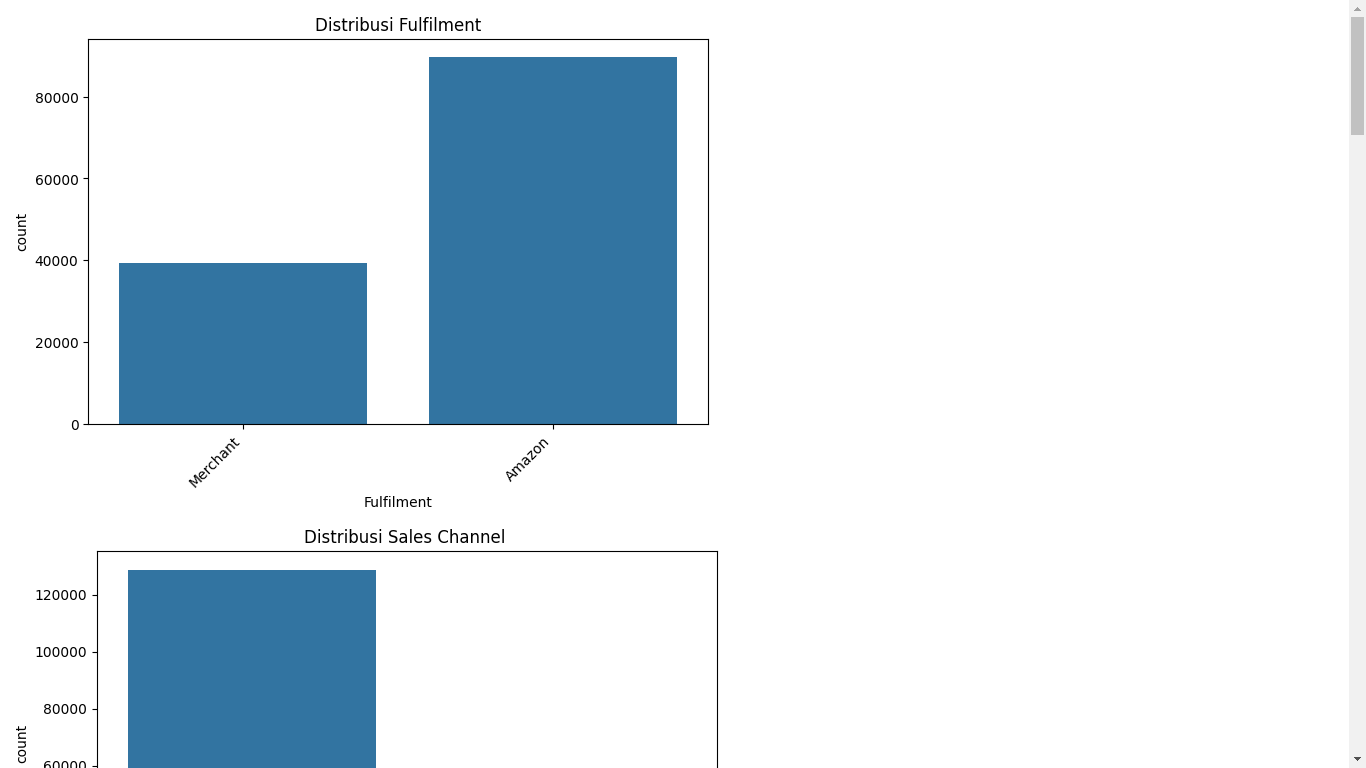


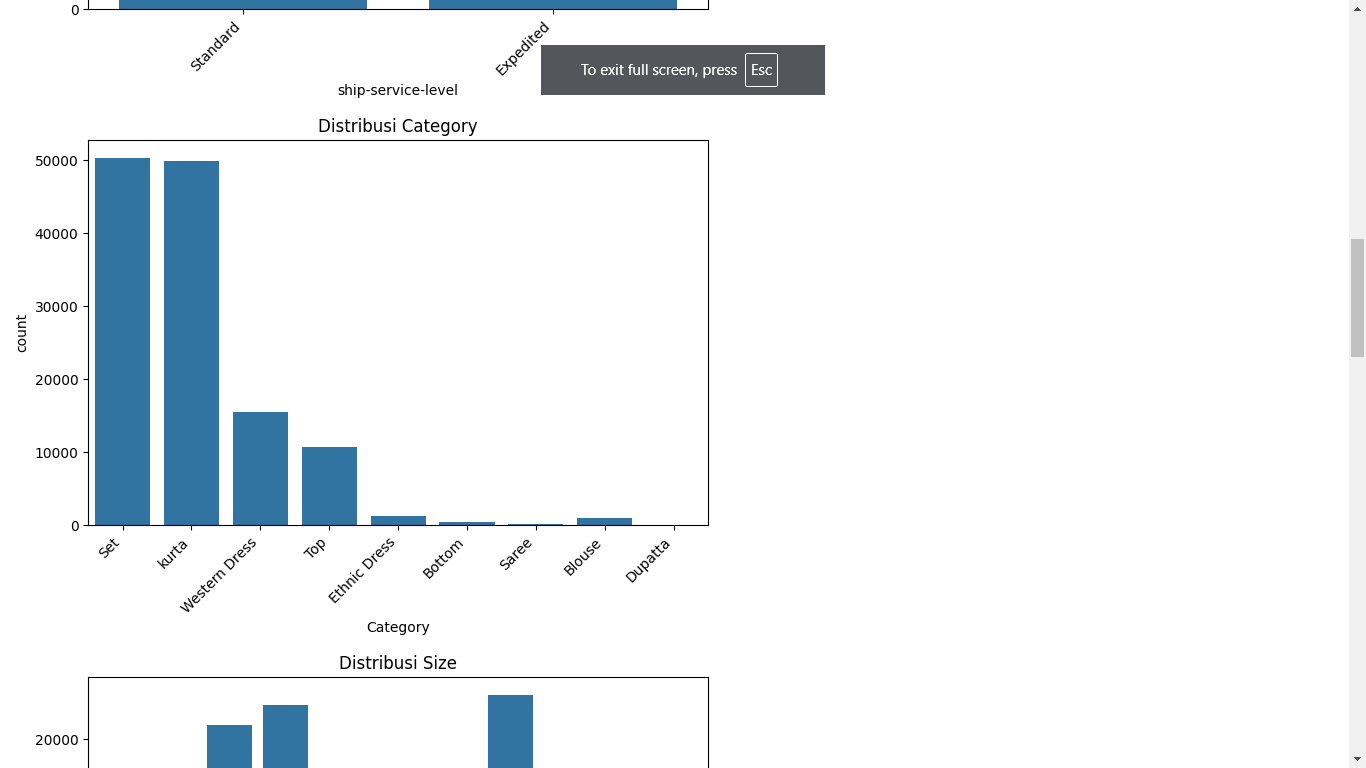
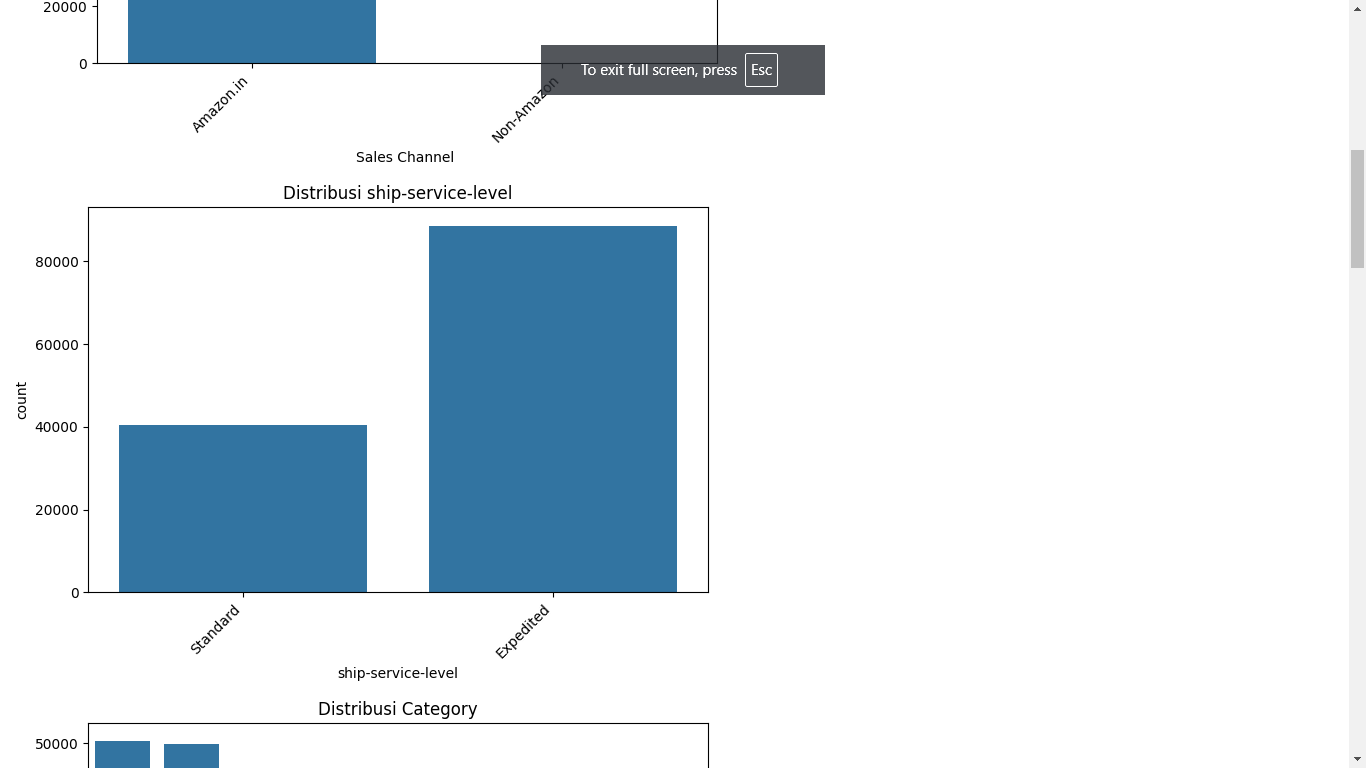


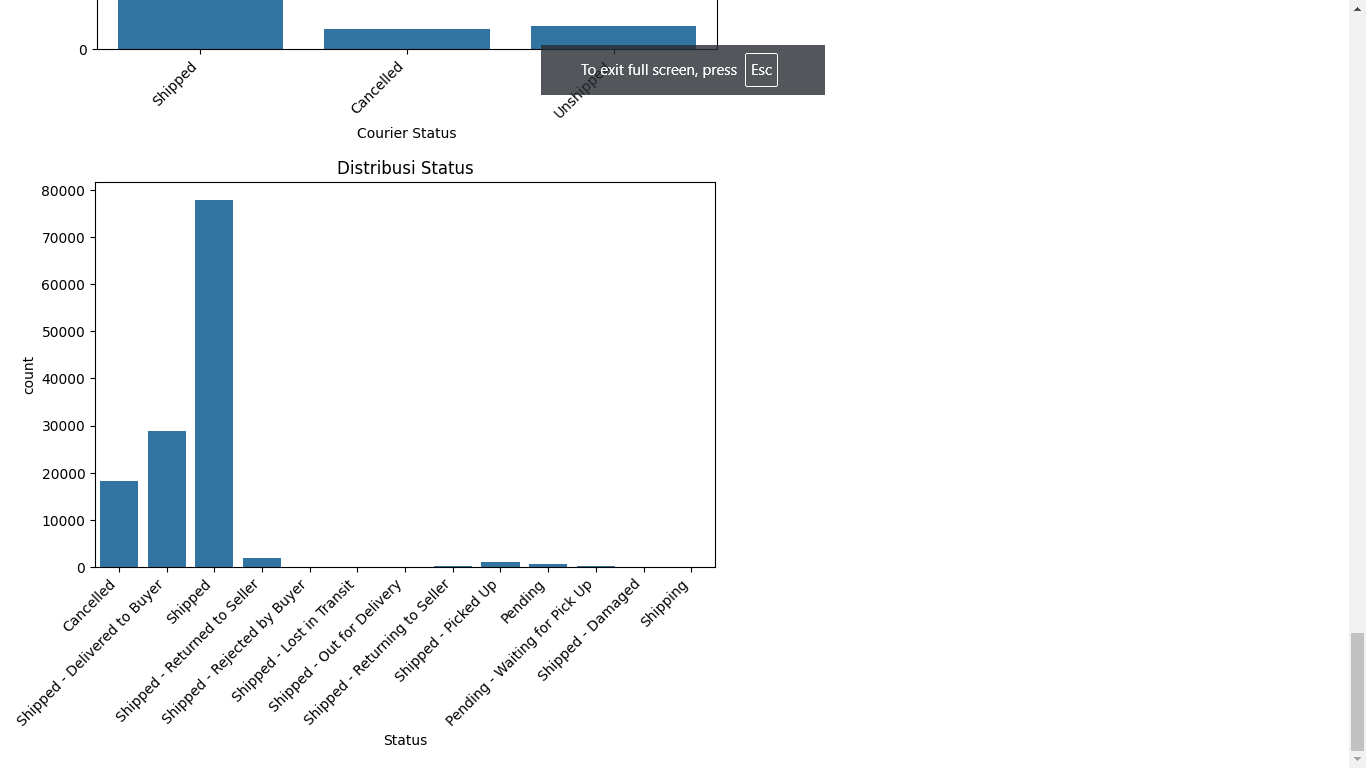
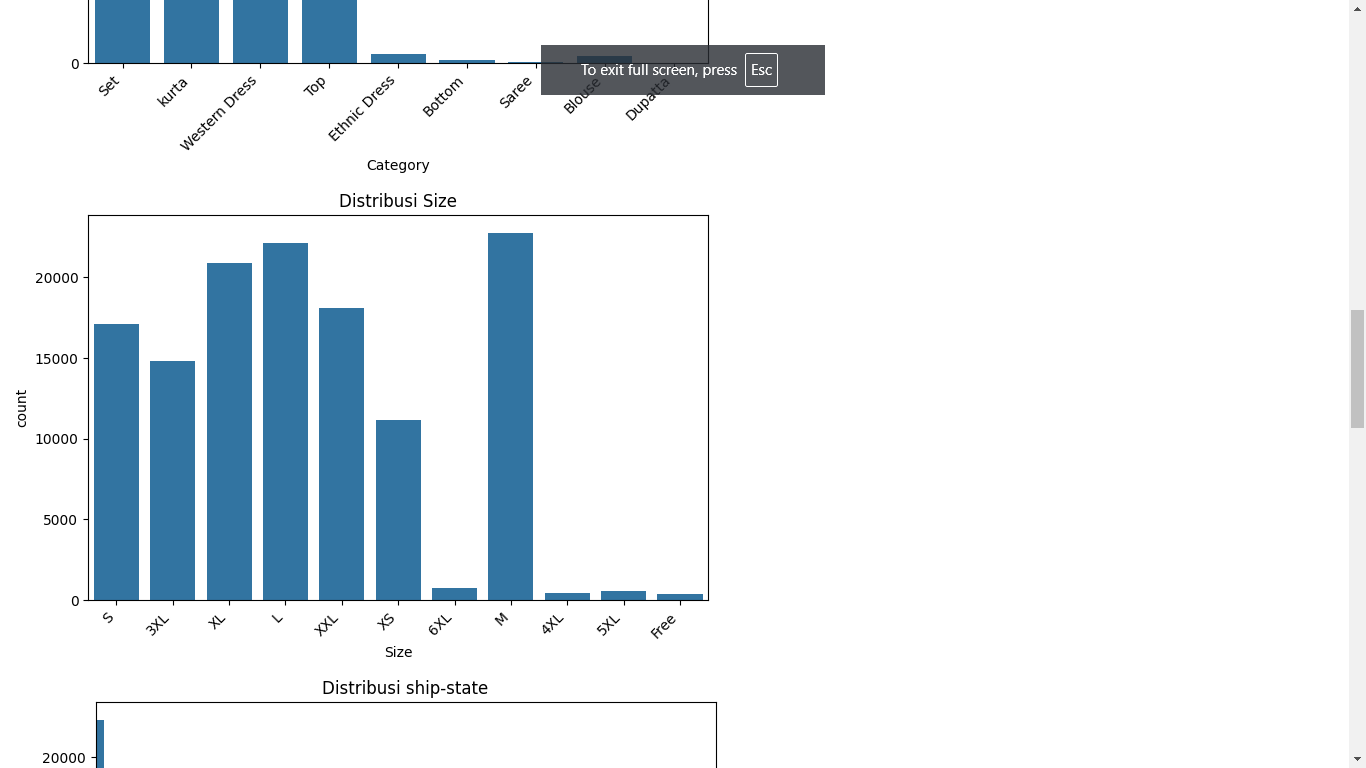


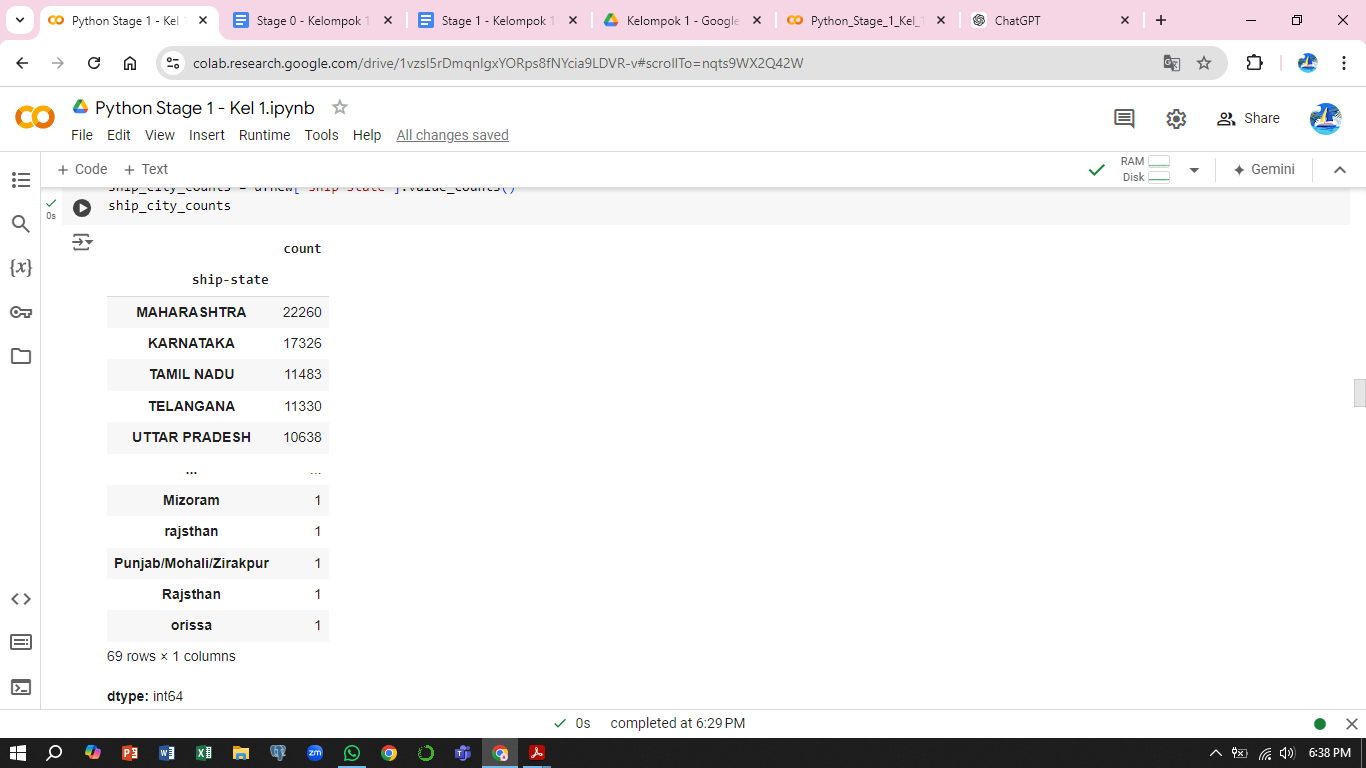
Berdasarkan hasil boxplot dan violinplot terlihat bahwa distribusi nilai pada kolom Qty/kuantitas adalah cenderung berkumpul di nilai 1, artinya jumlah barang yang dipesan dan dilakukan pengiriman hampir seluruhnya 1 pcs untuk setiap baris transaksi. Kemudian untuk distribusi nilai untuk amount/harga pesan cenderung berkumpul di nilai 1000-an dan bentuk distribusinya dapat dianggap mendekati distribusi normal walau agak sedikit skewed ke kanan dengan nilai mean sedikit lebih besar dibanding nilai mediannya.

Kemudian untuk data kategorikal berikut pemetaan countplot-nya.









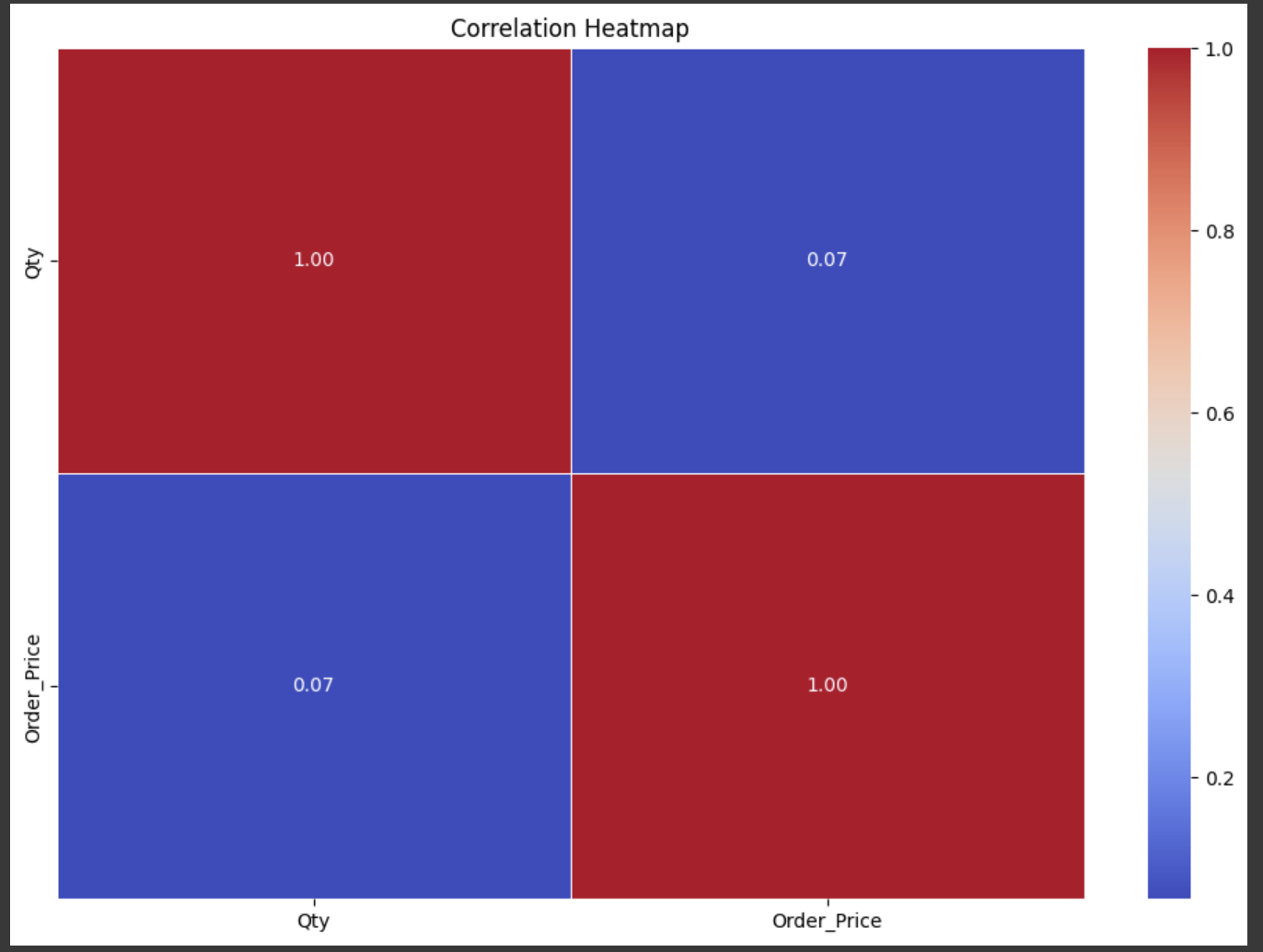
Untuk data kategorikal, kesimpulan yang dapat diambil adalah:

1. Proporsi pesanan yang dipenuhi langsung oleh Amazon lebih banyak dibandingkan dengan merchant atau tingkat fullfilment pesanan lebih dominan dipenuhi oleh Amazon.
2. Platform Amazon.in hampir mendominasi terhadap keseluruhan pesanan dalam perusahaan. Ini bisa berarti bahwa produk yang dijual kemungkinan besar lebih dikenal atau lebih efektif dipasarkan melalui Amazon dibandingkan saluran lain.
3. Sebagian besar transaksi adalah Business to Consumer (B2C), di mana perusahaan menjual produk langsung kepada konsumen akhir. Ini menunjukkan bahwa model bisnis yang diadopsi lebih fokus pada penjualan ritel kepada individu daripada menjual dalam jumlah besar kepada bisnis lain.
4. Set dan Kurta adalah dua kategori produk yang paling banyak dilakukan pemesanan dengan total sekitar hampir 10000 baris pemesanan.
5. Jenis pengiriman expedited atau lebih cepat memiliki jumlah yang lebih banyak dibandingkan dengan pengiriman standard.
6. Jenis pakaian M, L, dan XL adalah 3 jenis pakaian dengan jumlah pemesanan terbanyak dan 4XL, 5XL, 6XL, dan Free adalah jenis pakaian dengan jumlah pemesanan terendah.
7. State dengan pemesanan terbanyak adalah Maharashta, Karnataka, Tamil Nadu, Telangana, Uttar Pradesh dan state dengan pemesanan terendah adalah Mizoram, Rajhstan, dan Orisa.
8. Status yang paling dominan adalah Shipped (77,804) dan Shipped - Delivered to Buyer (28,769), kemudian status Cancelled (18,332) juga cukup signifikan, menunjukkan bahwa pembatalan adalah bagian yang tidak bisa diabaikan dalam alur pemesanan.
9. Status mungkin kedepannya dapat dikelompokkan menjadi kategori yang lebih sederhana: Delivered, Cancelled, Returned, Pending, dan Lost/Damaged. Hal ini akan membantu meningkatkan akurasi model, mempermudah analisis, dan dapat menyoroti permasalahan utama: Kategori seperti Returned atau Cancelled dapat menjadi fokus untuk menganalisis sebab-sebab utama kegagalan pengiriman atau kepuasan pelanggan.

# 

# MULTIVARIATE ANALYSIS

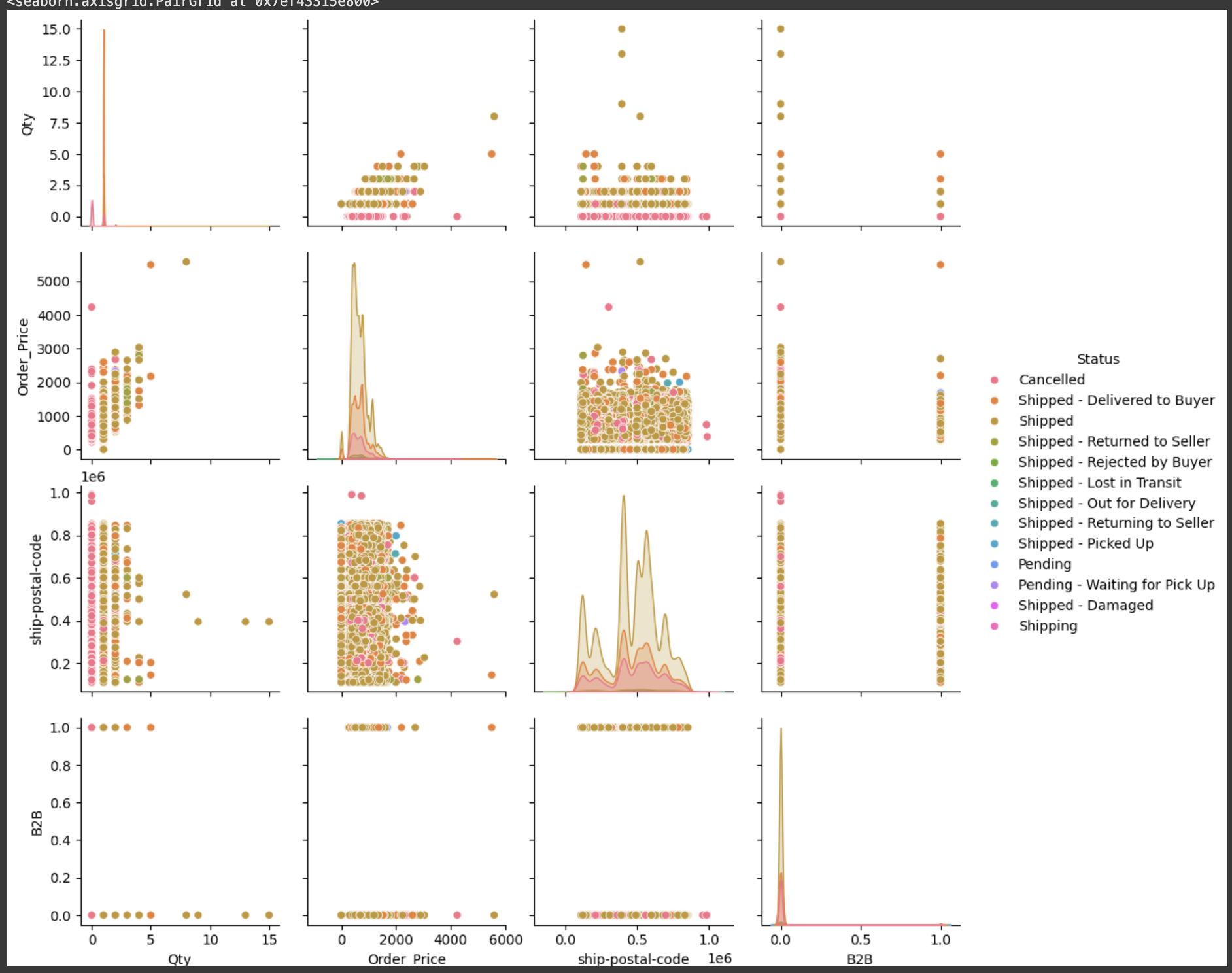
Correlation Heatmap untuk fitur numerikal:



### Interpretasi Korelasi:

* Korelasi Positif: Nilai 0.07 menunjukkan bahwa ketika jumlah item yang dibeli sedikit meningkat, harga pesanan cenderung meningkat sedikit juga. Namun, hubungan ini sangat lemah dan mungkin tidak signifikan secara praktis.
* Tidak Ada Hubungan Kuat: Nilai mendekati 0 menunjukkan bahwa tidak ada hubungan yang kuat antara Qty dan Order Price. Ini berarti perubahan pada salah satu variabel tidak banyak memengaruhi yang lain.

**Hubungan Status Pesanan dengan Variabel Numerik**



Berdasarkan status paket:

1. Cancelled:

Status ini menunjukkan bahwa pesanan dibatalkan. Dalam scatterplot, Cancelled tersebar di berbagai nilai Order\_Price dan Qty, tetapi bisa memiliki persebaran yang lebih rendah di kolom Qty, yang menunjukkan pesanan dengan jumlah item kecil. Ini mungkin karena pelanggan atau sistem membatalkan pesanan dengan jumlah kecil.

1. Shipped - Delivered to Buyer:

Status ini menunjukkan pesanan yang berhasil dikirim dan diterima oleh pembeli. Jika status ini memiliki persebaran yang lebih luas di nilai Order\_Price dan Qty, ini mengindikasikan bahwa pesanan yang dikirim dan diterima oleh pembeli bisa berupa pesanan kecil maupun besar. Hal ini menunjukkan kelancaran transaksi di berbagai kisaran harga dan jumlah.

1. Shipped - Returned to Seller dan Shipped - Rejected by Buyer:

Kedua status ini terkait dengan pesanan yang tidak berhasil diterima oleh pembeli dan dikembalikan ke penjual, baik karena pengembalian sukarela atau penolakan. Biasanya, pesanan dengan status ini mungkin memiliki Order\_Price yang lebih tinggi atau Qty yang berbeda dibanding pesanan yang berhasil diterima. Mungkin pembeli lebih cenderung mengembalikan barang dengan nilai atau jumlah yang lebih tinggi.

1. Shipped - Lost in Transit:

Ini adalah status untuk pesanan yang hilang selama pengiriman. Meskipun tidak terlihat jelas dalam scatterplot, biasanya status ini jarang muncul, dan dapat menjadi indikasi tantangan logistik yang perlu dianalisis lebih lanjut, terutama jika terjadi pada pesanan dengan Order\_Price atau Qty tertentu.

1. Shipped - Out for Delivery dan Shipped - Returning to Seller:

Kedua status ini menunjukkan pesanan yang masih dalam perjalanan, baik menuju pembeli atau dikembalikan ke penjual. Status ini mungkin tidak memiliki banyak variasi dalam Order\_Price atau Qty, karena status pengiriman ini mungkin relatif sama di berbagai jenis pesanan.

1. Pending, Pending - Waiting for Pick Up:

Status Pending menunjukkan pesanan yang belum diproses atau menunggu tahap selanjutnya, sementara Pending - Waiting for Pick Up mungkin menandakan pesanan yang sudah siap tetapi menunggu diambil oleh pengirim atau pembeli. Scatterplot menunjukkan bahwa banyak pesanan Pending berada pada kisaran Order\_Price yang lebih rendah, yang bisa menunjukkan bahwa pesanan yang masih menunggu biasanya adalah pesanan dengan harga yang tidak terlalu tinggi.

1. Shipping:

Status Shipping menunjukkan pesanan dalam proses pengiriman. Pesanan dalam status ini kemungkinan besar tersebar di berbagai nilai Order\_Price dan Qty, yang menunjukkan bahwa beragam jenis pesanan dapat berada dalam tahap pengiriman, baik yang bernilai rendah maupun tinggi.

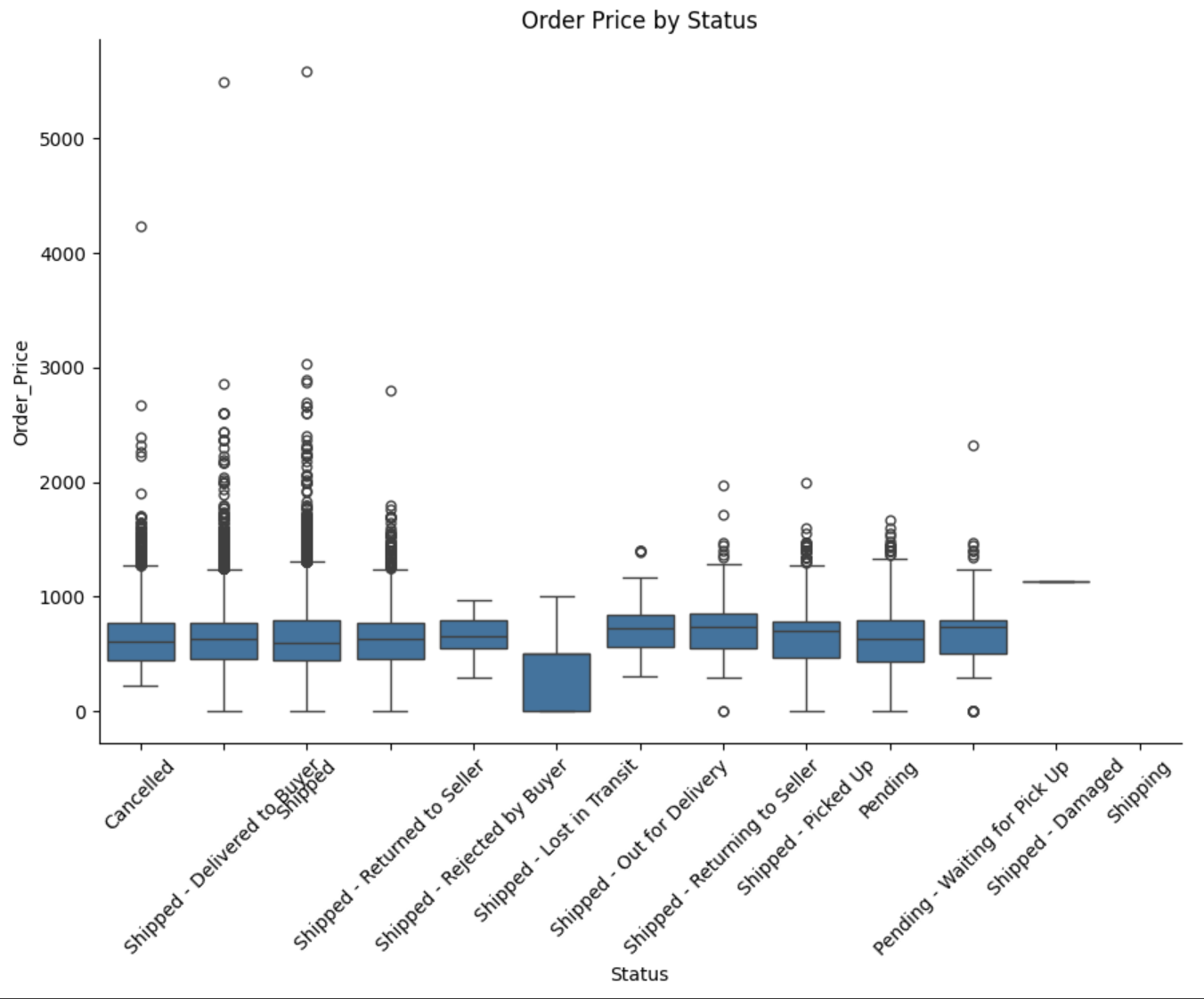
1. Shipped - Damaged:

Status ini menunjukkan pesanan yang mengalami kerusakan selama pengiriman. Scatterplot mungkin menunjukkan bahwa pesanan ini memiliki nilai Order\_Price atau Qty tertentu yang berbeda dari pesanan lainnya. Kerusakan selama pengiriman bisa terjadi pada pesanan yang bernilai tinggi atau barang yang mudah rusak.

Status pesanan dalam scatterplot matrix ini memberi gambaran mengenai kondisi atau tahapan yang berbeda dalam proses pemesanan dan pengiriman. Distribusi status dalam kaitannya dengan variabel seperti Order\_Price dan Qty dapat membantu kita mengidentifikasi pola, misalnya:

* Status yang cenderung muncul pada pesanan bernilai tinggi (misalnya, pengembalian atau kerusakan).
* Status yang sering muncul pada pesanan dengan jumlah item kecil (misalnya, pesanan Pending atau Cancelled).
* Status Shipped - Delivered to Buyer yang menunjukkan bahwa pesanan berhasil dikirim dan diterima di berbagai kisaran harga dan jumlah.

Kemudian berikut adalah contoh hasil analisis untuk distribusi data order\_price berdasarkan status pengiriman.



Dari visualisasi boxplot yang tampilkan, berikut adalah beberapa analisis dan insight yang dapat diperoleh mengenai distribusi Order\_Price berdasarkan Status:

1. Median Order Price:

Boxplot ini menunjukkan bahwa median Order\_Price untuk hampir semua status relatif konsisten, dengan sebagian besar berada di kisaran nilai yang serupa (sekitar beberapa ratus hingga di bawah seribu). Ini mengindikasikan bahwa rentang harga pesanan tidak banyak berbeda di antara status yang berbeda.

1. Variabilitas (IQR) dan Persebaran Data:

Rentang Antar-Kuartil (IQR): Beberapa status memiliki IQR yang lebih lebar, seperti Shipped - Returned to Seller dan Shipped - Out for Delivery, yang menunjukkan variasi yang lebih besar dalam harga pesanan di status ini. Status Shipped - Rejected by Buyer memiliki IQR yang paling sempit, menunjukkan variabilitas harga yang lebih rendah di antara pesanan dengan status ini.

Ekstremitas dan Outlier: Hampir semua status memiliki outlier pada Order\_Price yang relatif tinggi. Outlier ini mungkin adalah pesanan bernilai tinggi yang jarang terjadi, seperti pesanan besar atau barang mahal. Misalnya, status Cancelled, Shipped - Delivered to Buyer, dan Shipped - Returned to Seller memiliki beberapa outlier yang jauh di atas nilai maksimum untuk sebagian besar pesanan lainnya.

1. Perbedaan di Antara Status:

Status Shipped - Lost in Transit: Status ini memiliki median Order\_Price yang lebih rendah dibanding status lainnya, dengan IQR yang sempit. Ini mungkin menunjukkan bahwa pesanan yang hilang dalam pengiriman cenderung memiliki nilai yang lebih rendah.

Status Shipped - Damaged: Ada data dengan Order\_Price spesifik dan sempit (tidak ada persebaran yang terlihat), menunjukkan hanya ada sedikit atau satu pesanan yang masuk dalam kategori ini, atau data yang sangat konsisten untuk status ini.

Status Pending dan Pending - Waiting for Pick Up: Kedua status ini menunjukkan variasi yang lebih besar dalam Order\_Price, tetapi dengan banyak outlier di rentang harga yang lebih tinggi. Ini mungkin menandakan pesanan yang masih menunggu proses atau pengambilan mencakup berbagai jenis produk, mulai dari yang murah hingga yang mahal.

1. Insight untuk Pengambilan Keputusan:

Konsistensi di Order\_Price: Dengan distribusi harga pesanan yang relatif konsisten di sebagian besar status, ini menunjukkan bahwa berbagai tahap pesanan (seperti pengiriman, pengembalian, penolakan) cenderung terjadi pada kisaran harga yang mirip. Hal ini dapat berguna untuk memprediksi perilaku pesanan di masa mendatang.

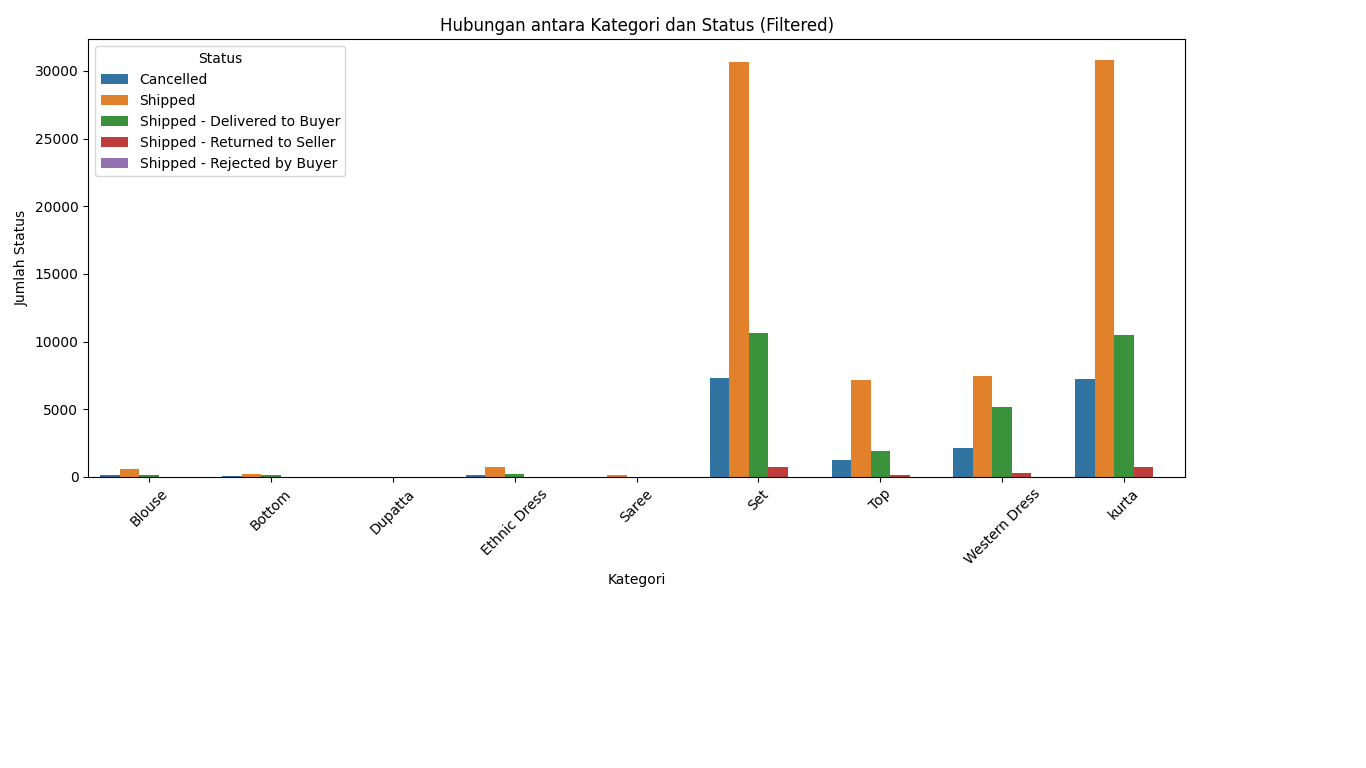
Pengelolaan Outlier: Banyaknya outlier di berbagai status menunjukkan adanya beberapa pesanan bernilai tinggi. Ini mungkin perlu diperhatikan untuk kebijakan khusus, seperti asuransi atau penanganan ekstra untuk pesanan bernilai tinggi.

Status Khusus: Status seperti Shipped - Lost in Transit dan Shipped - Damaged memiliki pola yang unik dan mungkin memerlukan kebijakan khusus atau perhatian tambahan dalam proses pengiriman.

Secara keseluruhan, visualisasi ini menunjukkan bagaimana Order\_Price terdistribusi di berbagai status pesanan dan memberikan gambaran tentang variasi harga di setiap tahap. Outlier yang muncul di beberapa status tertentu juga bisa menjadi fokus untuk perbaikan proses operasional.

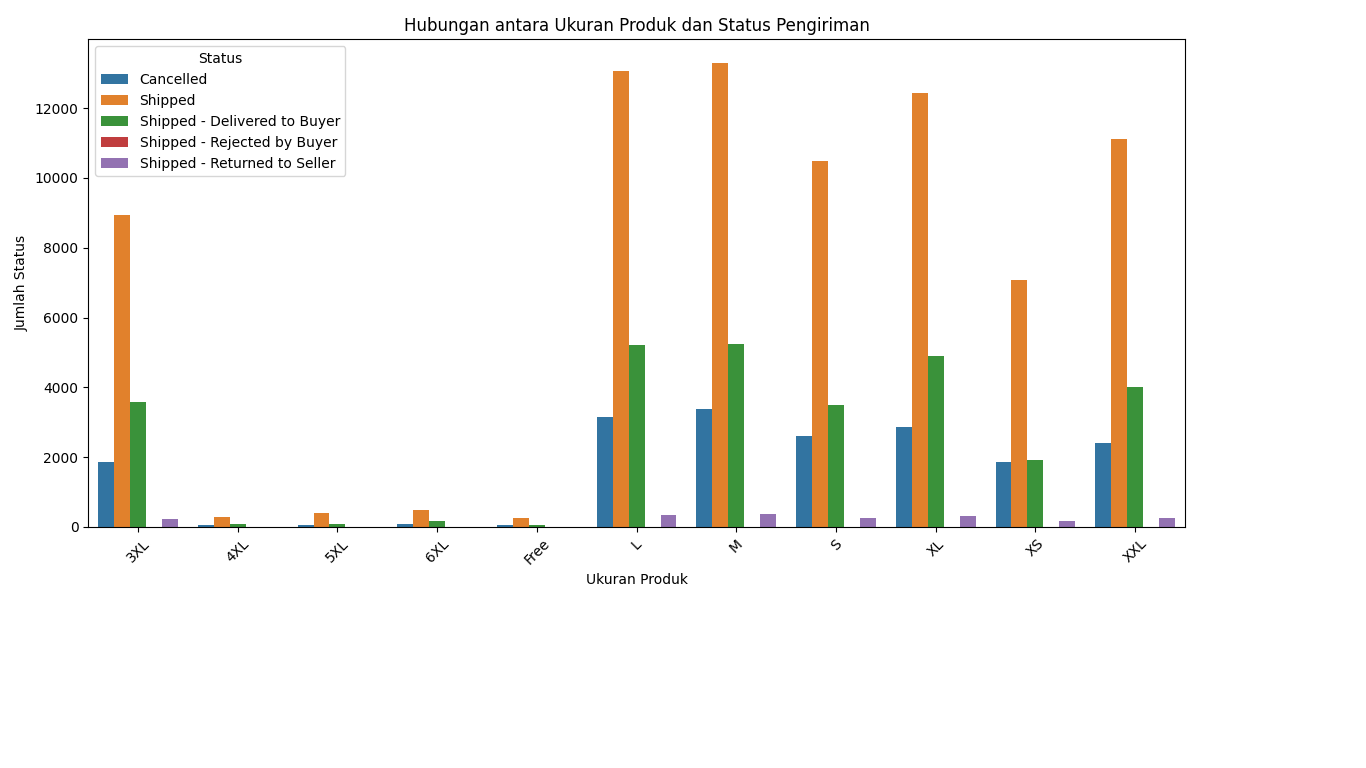
# BUSINESS INSIGHT & RECOMMENDATIONS

1. **Melihat distribusi status barang berdasarkan kategori barang yang dipesan.**



Dari gambar diatas terlihat bahwa dua kategori dengan jumlah pemesanan terbanyak, yaitu set dan kurta memiliki jumlah status barang yang sudah shipped dan diterima pembeli cukup banyak dibandingkan dengan jumlah barang yang dikembalikan ke pembeli karena ketidakpuasan tertentu atau faktor lainnya. Untuk kategori produk lainnya kurang lebih memiliki pola yang sama dimana barang yang diterima pembeli dan tidak dikembalikan atau ditolak pembeli lebih banyak dibandingkan barang dengan status dikembalikan ataupun ditolak.

1. **Melihat hubungan antara distribusi status barang berdasarkan ukuran pakaian/barang yang dipesan.**



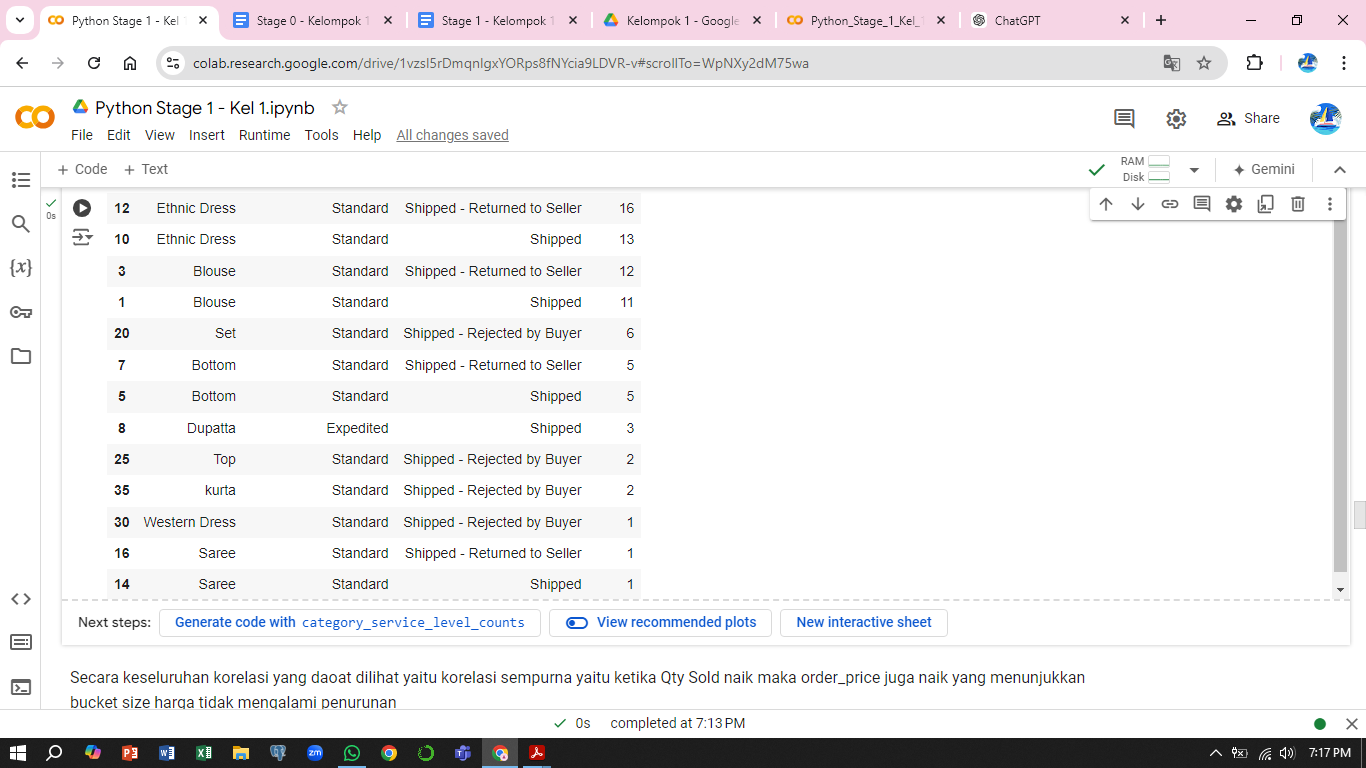
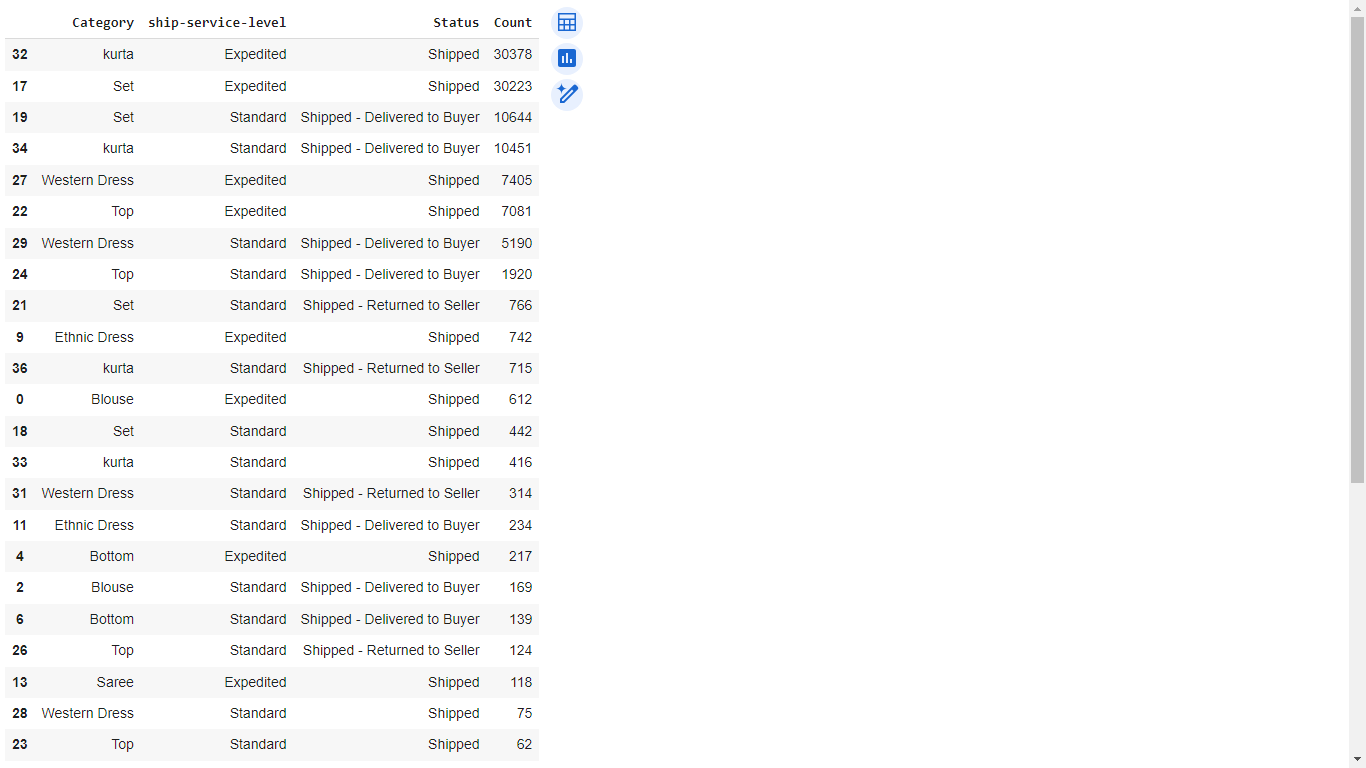
Sama dengan sebelumnya terlihat bahwa untuk ukuran produk yang memiliki jumlah pemesanan terbanyak seperti M, L, XL, XXL, S, dan XS memiliki jumlah status barang yang shipped dan diterima pembeli lebih banyak dibandingkan dengan jumlah status barang yang dikembalikan ataupun ditolak.

Distribusi yang menunjukkan lebih banyak status Shipped dan Shipped - Delivered to Buyer dibandingkan dengan status Shipped - Returned to Seller dan Shipped - Rejected by Buyer mencerminkan efektivitas pengiriman, kepuasan pelanggan yang tinggi, dan rendahnya masalah dengan pengembalian atau penolakan. Ini adalah indikasi positif untuk bisnis dan dapat menunjukkan bahwa strategi pengiriman dan manajemen produk berjalan dengan baik. Namun perlu diperhatikan bahwa tetap pentingnya untuk meneliti produk-produk baik dari aspek kategori dan ukuran yang masih memiliki

Dari dua visualisasi data diatas dapat dibangun rekomendasi bisnis sebagai berikut:

* **Fokus pada Kategori Populer:**
  + **Set** dan **Kurta** menunjukkan angka pengiriman yang tinggi dan tingkat pengembalian yang rendah. Perkuat strategi pemasaran dan promosi untuk kedua kategori ini, termasuk penawaran khusus atau diskon, untuk mendorong lebih banyak pembelian.
* **Ketersediaan Ukuran:**
  + Ukuran produk seperti M, L, XL, XXL, S, dan XS memiliki permintaan yang tinggi. Pastikan stok untuk ukuran-ukuran ini selalu mencukupi dan pertimbangkan untuk meningkatkan variasi dalam ukuran, termasuk ukuran tambahan, untuk menarik lebih banyak pelanggan dan mengurangi kemungkinan kehabisan stok.
* **Kualitas Pengemasan dan Pengiriman:**
  + Dengan tingkat pengembalian yang rendah, penting untuk mempertahankan kualitas. Evaluasi dan tingkatkan metode pengemasan untuk melindungi barang selama pengiriman, serta perkuat kemitraan dengan penyedia logistik untuk memastikan pengiriman tepat waktu dan aman. Ini dapat meningkatkan kepuasan pelanggan dan mengurangi risiko pengembalian.

1. **Melihat jumlah status barang yang dikelompokkan berdasarkan kategori produk dan jenis pengirimannya.**



Dari data di atas berikut kesimpulan yang dapat diambil:

### Kesimpulan Berdasarkan Output

* **Kinerja Pengiriman yang Baik:**

Kategori "kurta" dan "Set" menunjukkan performa pengiriman yang sangat baik, masing-masing dengan jumlah 30378 dan 30223 paket yang berhasil dikirim (Shipped) menggunakan metode pengiriman Expedited. Ini menunjukkan bahwa kategori-kategori ini mungkin sangat populer atau permintaan tinggi, sehingga pengiriman cepat sangat dibutuhkan.

* **Keberhasilan dalam Pengiriman yang Diterima:**

Kategori "Set" dan "kurta" juga memiliki jumlah yang signifikan dalam status "Shipped - Delivered to Buyer", masing-masing dengan 10644 dan 10451. Ini menunjukkan bahwa banyak dari produk ini berhasil sampai ke pelanggan tanpa masalah.

* **Pengembalian yang Terkendali:**

Meskipun ada beberapa produk yang mengalami status "Shipped - Returned to Seller", jumlahnya relatif rendah dibandingkan dengan jumlah yang berhasil dikirim. Contohnya, 766 untuk kategori "Set" dan 715 untuk "kurta". Ini menunjukkan bahwa pengembalian berada dalam tingkat yang terkendali dan mungkin tidak menjadi masalah besar saat ini.

* **Variasi dalam Status:**

Kategori-kategori lain seperti "Western Dress", "Top", dan "Ethnic Dress" juga menunjukkan performa yang baik dalam pengiriman, dengan sejumlah besar paket yang berhasil dikirim. Namun, mereka juga menunjukkan beberapa kasus pengembalian yang perlu dipantau.

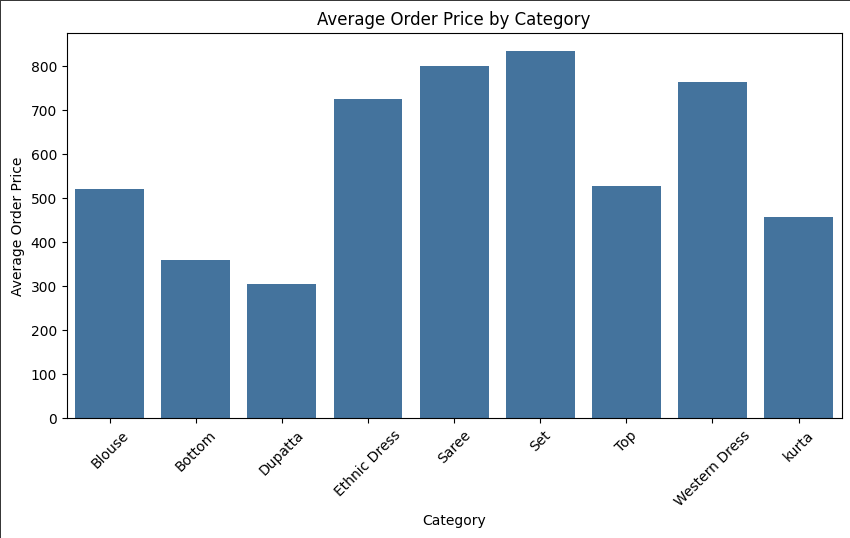
* **Minimnya Penolakan:**

Status "Shipped - Rejected by Buyer" cukup rendah, dengan hanya 6 untuk kategori "Set" dan sangat sedikit untuk kategori lainnya. Ini menunjukkan bahwa pelanggan tidak sering menolak paket yang mereka terima, yang merupakan indikator positif terhadap kepuasan pelanggan.

* **Kesiapan untuk Meningkatkan Pengiriman:**

Metode Expedited sangat digunakan untuk kategori tertentu, menunjukkan preferensi konsumen terhadap pengiriman yang lebih cepat. Ini bisa menjadi area fokus untuk meningkatkan penjualan atau menarik lebih banyak pelanggan.

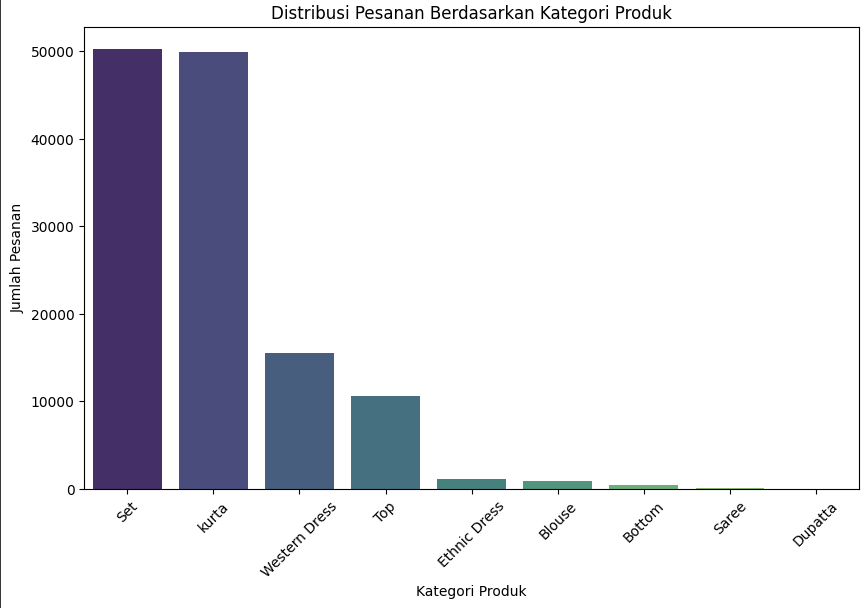
1. **Melihat average order price per kategori produk**



Berdasarkan gambar diatas dapat dilihat bahwa:

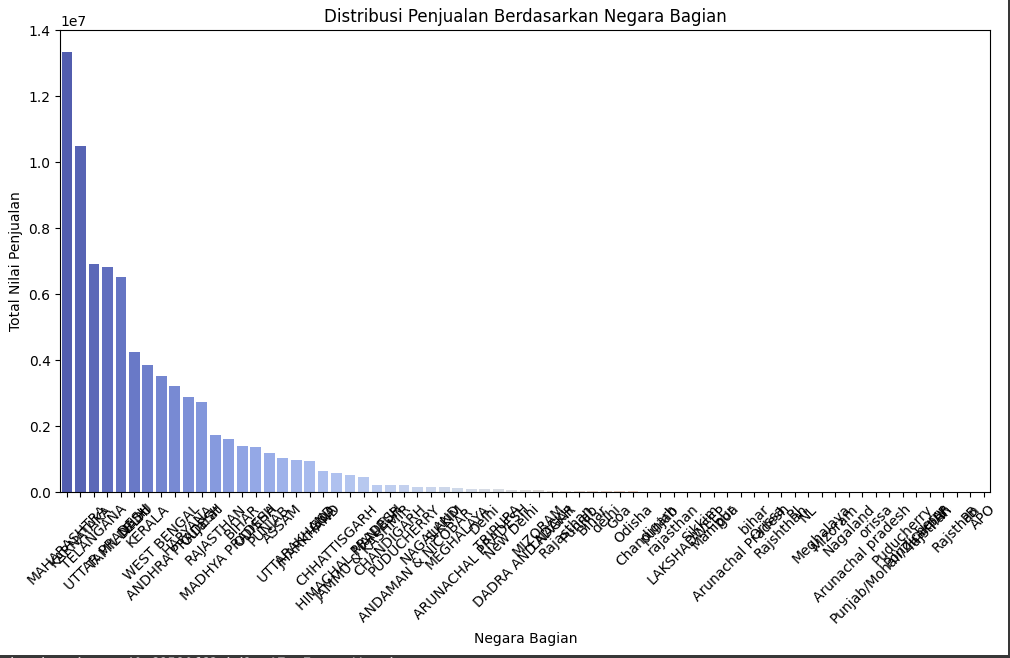
* Kategori **Set**, **Saree**, dan **Western Dress** mungkin lebih diminati oleh pelanggan, yang terlihat dari harga rata-rata yang lebih tinggi. Ini bisa menunjukkan bahwa pelanggan bersedia membayar lebih untuk kategori ini, kemungkinan karena nilai estetika, gaya, atau kepraktisan produk.
* Kategori **Dupatta** dan **Bottom** yang memiliki rata-rata harga lebih rendah dapat menunjukkan bahwa produk di kategori ini mungkin memiliki lebih banyak pesaing atau lebih sedikit diferensiasi, sehingga menyebabkan harga yang lebih rendah.
* Beberapa rekomendasi bisnis yang dapat diberikan adalah:
* Fokus pada promosi dan pemasaran untuk kategori **Set**, **Saree**, dan **Western Dress**. Buat kampanye yang menonjolkan keunikan dan kualitas produk dalam kategori ini. Misalnya, gunakan influencer atau model untuk menunjukkan cara mengenakan produk dalam konteks gaya hidup yang menarik.
* Tinjau dan evaluasi kategori **Dupatta** dan **Bottom** untuk mengidentifikasi potensi untuk meningkatkan daya tarik mereka. Pertimbangkan untuk memperkenalkan variasi desain, kualitas bahan yang lebih baik, atau fitur tambahan yang dapat meningkatkan nilai persepsi produk.
* Untuk kategori **Dupatta** dan **Bottom**, pertimbangkan untuk menawarkan diskon atau bundling dengan produk dari kategori yang lebih populer. Ini dapat menarik lebih banyak pelanggan dan meningkatkan penjualan keseluruhan untuk kategori yang kurang berkinerja.

1. **Pola data masuk distribusi pesanan barang berdasarkan kategori produk**



* Dari tabel distribusi pesanan berdasarkan kategori produk, ditemukan bahwa kategori set dan kurta memliki jumlah kurang lebih 50000 dari total pesanan. Ini menunjukan bahwa katagori set dan kurta adalah kategori yang lebih menarik bagi pelanggan, yang mungkin dapat dijadikan fokus untuk promosi lebih lanjut. Dengan mengumpulan dan menganalisis data pesanan berdasarkan kategori produk, dapat mengambil keputusan yang lebih baik dan strategis untuk meningkatkan kinerja bisnis.
* Dengan mengalokasikan produk yang lebih relevan dan banyak dicari, serta mengurangi stok barang yang kurang diminati bisa membantu efisiensi operasional.

1. **Pola distribusi berdasarkan negara bagian**



* Dari tabel distribusi berdasarkan negara bagian, Maharashtra dan Karnataka memiliki jumlah pesanan tertinggi, bisa meningkatkan stategi bisnis dengan dengan melakukan promosi produk sehingga meningkatkan kepuasaan pelanggan.
* Dan melakukan analisis mendalam mengapa beberapa negara bagian seperti Rajsthan, AR, PB, memiliki jumlah pesanan terkecil, dengan pelacakan umpan balik pelanggan, jika terdapat pesanan disebabkan pengiriman waktu yang lama, bisa menggunakan mitra pengiriman yang lebih cepat di wilayah tersebut

# GIT REPOSITORY