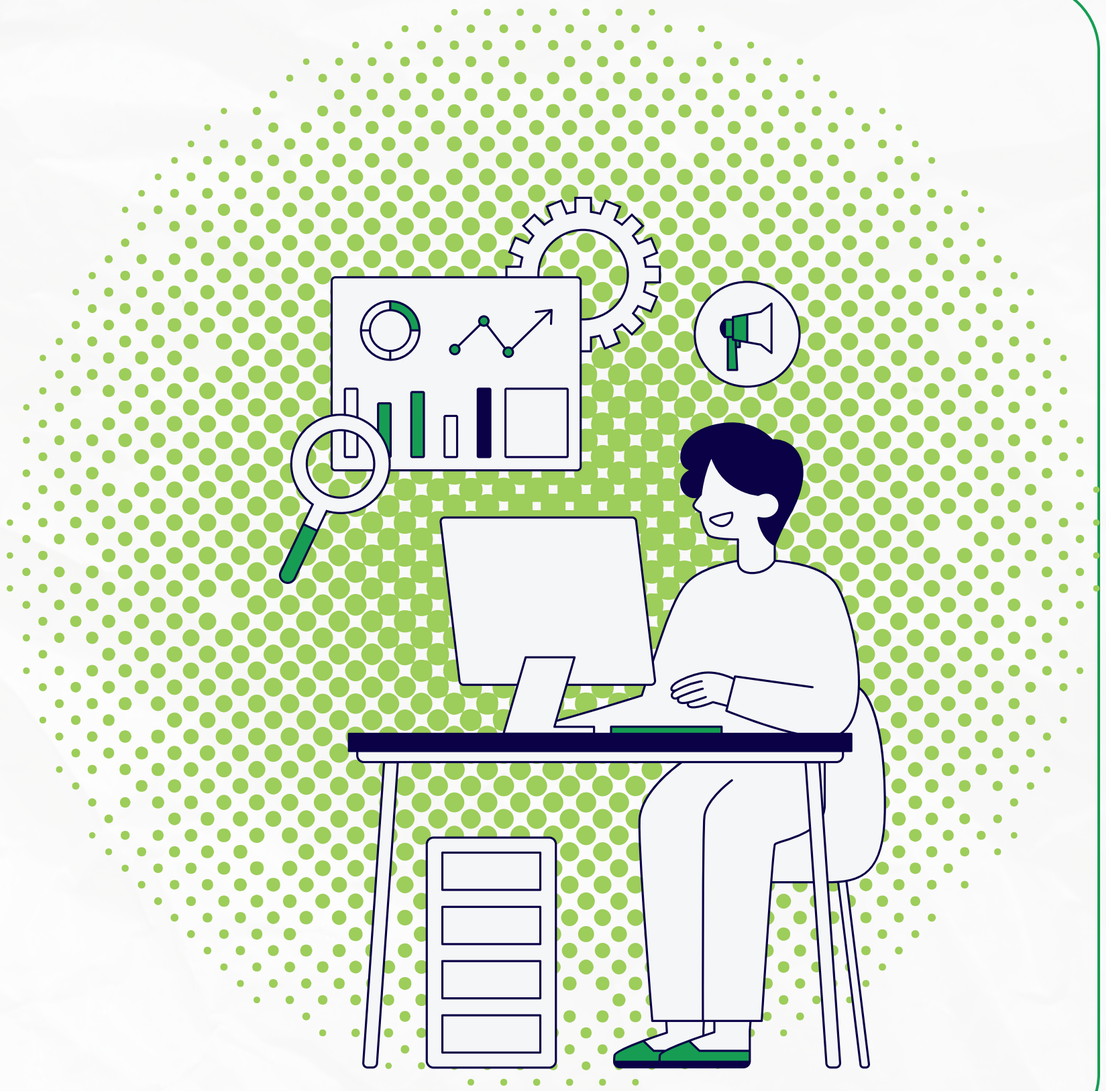
EXPLORATORY DATA ANALYSIS: SUPERSTORE DATASET

Agil Firli Gunawan | EDA Temudataku



Latar Belakang

SuperStore adalah perusahaan retail yang menjual berbagai produk mulai dari Furniture, Office Supplies, hingga Technology. Seiring pertumbuhan bisnis, penting bagi perusahaan untuk **memahami data penjualannya** secara mendalam. Dengan melakukan Exploratory Data Analysis (EDA), kita dapat menggali **informasi** penting terkait tren penjualan, performa produk, serta efisiensi bisnis secara keseluruhan.

Dengan melakukan Exploratory Data Analysis (EDA), kita dapat:

- Menggali tren dan pola penjualan
- Menemukan masalah tersembunyi seperti anomali dan outlier
- Memberikan dasar yang kuat untuk pengambilan keputusan bisnis yang lebih tepat
- Menemukan insghit keputusan yang tepat

Rumusan Masalah

1. Bagaimana pola dan tren penjualan berdasarkan kategori produk?
2. Bagaimana keuntungan yang diperoleh dari tiap kategori produk?
3. Bagaimana distribusi kuantitas produk yang terjual?
4. Bagaimana hubungan antara sales, profit, dan quantity?
5. Apakah terdapat outlier signifikan pada data profit?

Proses EDA

1 **Data Collection**

Mengimpor dan memahami dataset dengan menginput data yang sudah ditemukan dataset-nya

2 **Statistik Deskriptif**

Menggambarkan karakteristik data seperti ukuran pemusatan, penyebaran, dan distribusi sebelum di lakukan data cleaning.

3 **Data Manipulation**

Membersihkan dan mentransformasi data agar siap dianalisis lalu di buat grafik keputusan dengan visualisasi data

4 **Data Visualization**

Menyajikan data dalam bentuk grafik untuk menemukan pola dan interpretasi lalu insight yang relevan.

EDA Tools

EDA dilakukan di **Google Colab** sebuah platform berbasis cloud yang memungkinkan eksekusi Python secara interaktif, menggunakan library **Pandas** untuk manipulasi dan analisis data, **NumPy** untuk perhitungan numerik, serta **Matplotlib** dan **Seaborn** untuk membuat visualisasi data yang informatif dan estetis

Data Collection

Data Collection adalah tahap awal dalam EDA di mana data dikumpulkan dari sumber yang relevan. Pada proyek ini, data diimpor ke dalam lingkungan Google Colab menggunakan library Pandas, yang memungkinkan pembacaan file seperti .xlsx untuk dianalisis lebih lanjut.

Statistik Deskriptif

Statistik Deskriptif digunakan untuk menganalisis dan merangkum karakteristik utama data. Pada tahap ini, kita menghitung **ukuran pemusatan** seperti mean, median, dan modus, serta **ukuran penyebaran** seperti range, variansi, standar deviasi, dan interquartile range (IQR). Selain itu, **distribusi data** juga diperiksa menggunakan skewness dan kurtosis untuk memahami bentuk dan kecenderungan data.

Data Manipulation

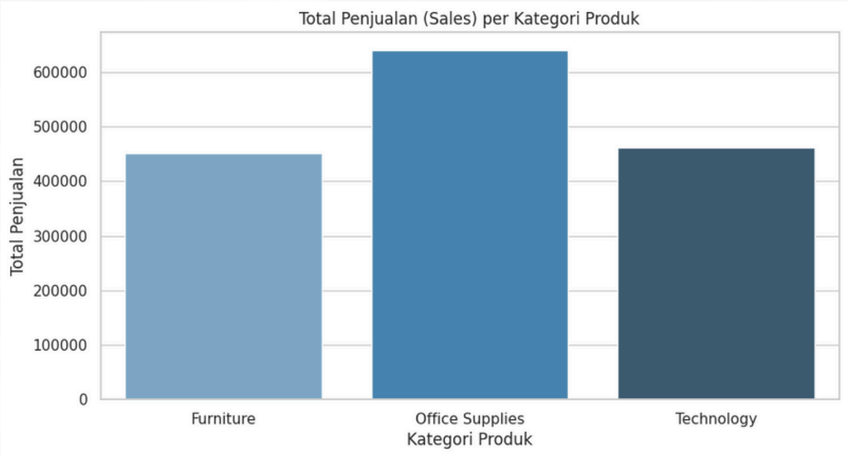
Data Manipulation adalah proses pembersihan dan transformasi data agar siap untuk analisis lebih lanjut. Ada **Data Cleaning** Tahap ini meliputi penanganan missing values (imputasi atau penghapusan), penghapusan data duplikat, koreksi inkonsistensi, serta deteksi dan penanganan outlier. Selain itu, dilakukan **transformasi data** seperti agregasi dan normalisasi untuk memudahkan pemodelan dan visualisasi.

Data Visualization

Data Visualisasi digunakan untuk menggambarkan pola, tren, dan hubungan dalam data secara visual agar lebih mudah dipahami. Visualisasi ini membantu menjawab rumusan masalah seperti pola dan tren penjualan berdasarkan kategori produk, performa keuntungan tiap kategori, distribusi kuantitas produk yang terjual, hubungan antara sales, profit, dan quantity, serta identifikasi outlier yang signifikan pada data profit. Dengan grafik dan diagram, insight penting dapat diambil untuk pengambilan keputusan bisnis.

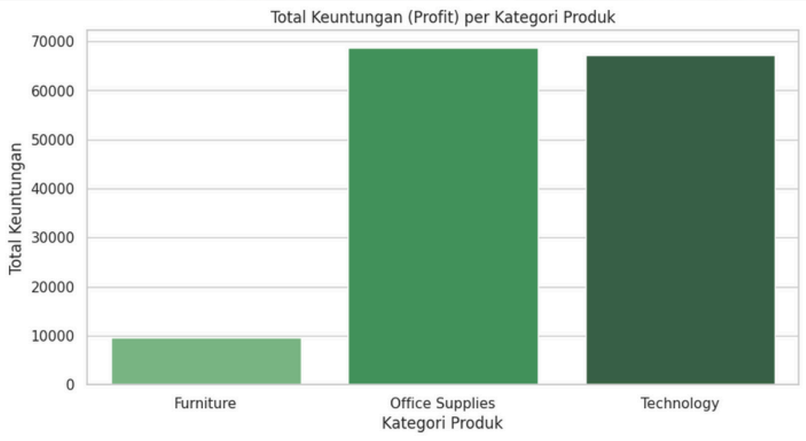
Data Visualization

1



Visualisasi menunjukkan bahwa kategori Office Supplies mendominasi penjualan, diikuti oleh Furniture dan Technology, menandakan fokus utama bisnis ada pada produk tersebut.

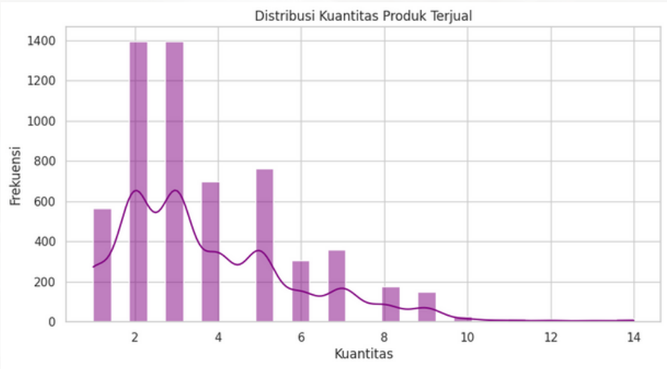
2



Meskipun Office Supplies menghasilkan keuntungan terbesar, kategori Technology juga menunjukkan performa profit yang hampir sebanding, sedangkan Furniture memiliki margin keuntungan yang lebih rendah dibanding penjualan.

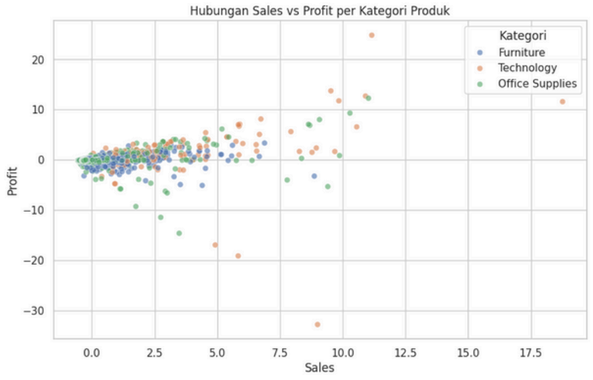
Data Visualization

3



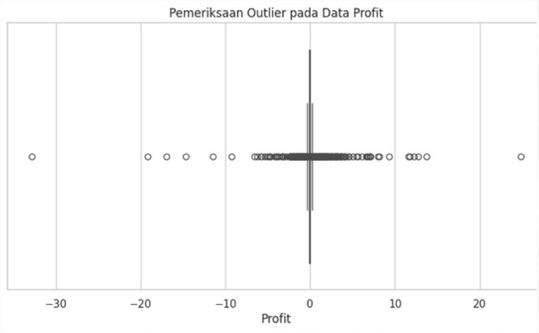
Distribusi quantity menunjukkan sebagian besar transaksi melibatkan pembelian dalam jumlah kecil, dengan beberapa outlier pembelian besar yang perlu diperhatikan.

4



Terdapat korelasi positif antara sales dan profit, serta quantity produk yang terjual, yang mengindikasikan bahwa peningkatan penjualan berkontribusi langsung pada keuntungan.

5



Visualisasi boxplot mengidentifikasi beberapa outlier profit yang sangat tinggi dan rendah, yang dapat menjadi fokus untuk analisis lebih lanjut terkait penyebabnya.

Kesimpulan & Rekomendasi

Analisis data menunjukkan pola penjualan yang dominan pada kategori Office Supplies dan Technology dengan profit yang signifikan, hubungan positif antara sales, profit, dan quantity, serta adanya beberapa outlier yang perlu diperhatikan untuk menjaga kualitas data dan pengambilan keputusan bisnis.

Perkuat strategi penjualan pada kategori utama, lakukan pemantauan dan penanganan outlier secara berkala, serta optimalkan proses pembersihan dan transformasi data untuk meningkatkan akurasi analisis di masa mendatang.

Thank You