**LAPORAN PRAKTIKUM**

**Praktikum Eksplorasi dan Visualisasi Data**

**Explorasi Data Analisis dan Visualisasi**

Dosen Pengampu : Edi Satriyanto S.Si., M.Si



**Disusun Oleh :**

**Dinda Ayu Permatasari – 3323600012**

**POLITEKNIK ELEKTRONIKA NEGERI SURABAYA** **PROGRAM STUDI SAINS DATA TERAPAN**

**TAHUN 2024/2025**

1. **Explorasi Data Analisis**
2. **Memahami Struktur Data**
3. **Import Library**

|  |
| --- |
| **Script** |
|  |
| **Analisis** |
| Pada Langkah pertama, kita menyiapkan library yang diperlukan, dalam hal ini memakai library pandas, seaborn, matplotlib dan numpy untuk kebutuhan analysis data dan visualisasi data. |

1. **Loading Data**

|  |
| --- |
| **Script** |
|  |
| **Analisis** |
| Selanjutnya, saya memuat data dari file Excel (Tokokita.xlsx) ke dalam objek DataFrame Pandas yang disimpan dalam variabel df. |

1. **Menampilkan Sample Data**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Lalu saya menggunakan fungsi head() untuk menampilkan 5 baris pertama dari dataset yang sudah saya load dari langkah sebelumnya. Dari output diatas kita bisa melihat apa saja kolom, baris dan isi dari dataset yang kita miliki yaitu ‘Tokokita.xlsx’. |

1. **Menampilkan Informasi Dataset**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Disini saya juga menggunakan fungsi info() untuk menampilkan informasi dari dataset ‘Tokokita.xlsx’. Kita bisa mengetahui jumlah datanya (ada 801 baris dan 12 kolom) serta tipe data untuk setiap kolomnya. |

1. **Memeriksa Nilai yang Hilang (Missing Values) dan Data Duplikat**
2. **Identifikasi jumlah nilai yang hilang di setiap kolom**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Kemudian untuk proses eksplorasi data analisis selanjutnya, yaitu memeriksa apakah ada data yang hilang dalam dataset kita (Missing Value). Hal ini perlu diperhatikan sebelum kita menggunakan data untuk analisis lebih lanjut. Oleh karena itu, pada kode diatas saya memeriksa jumlah nilai yang hilang (NaN) di setiap kolom DataFrame df. Dan hasilnya menunjukkan bahwa kolom Region dan City memiliki nilai di kolom yang lengkap. Ini menunjukkan bahwa tidak ada data yang hilang dalam kolom ini, sehingga kolom ini bisa dianggap terisi sepenuhnya. Sedangkan hampir semua kolom lainnya memiliki nilai yang hilang. Ini mencakup banyak kolom, sehingga penanganan nilai hilang menjadi sangat penting. |

1. **Menangani nilai yang hilang**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Lalu untuk menangani nilai yang hilang, saya menggunakan dropna() untuk menghapus semua baris dalam DataFrame df yang mengandung nilai hilang (NaN). Dengan demikian, setelah baris-baris ini dihapus, tidak ada baris yang memiliki nilai hilang di DataFrame yang tersisa. Dan fungsi isnull().sum() untuk memeriksa jumlah nilai yang hilang di setiap kolom. Output menunjukkan bahwa semua kolom sekarang memiliki 0 nilai yang hilang. |

1. **Memeriksa data duplikat**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Terakhir saya menggunakan fungsi duplicated() untuk memeriksa apakah ada data yang duplikat dalam dataset. Dan hasil dari outputnya menunjukkan bahwa tidak ada baris yang memiliki nilai duplikat. |

1. **Memahami Distribusi Data**
2. **Descriptive Statistics**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Untuk memahami distribusi data yang kita miliki, bisa memakai fungsi describe() untuk memberikan ringkasan statistik deskriptif dari kolom numerik dalam DataFrame df. Didalamnya mencakup metrik-metrik seperti count, mean, min, max, serta kuartil (25%, 50%, 75%). |

1. **Menghapus Kolom Yang Tidak Perlu**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Kode drop() diatas untuk menghapus kolom-kolom tertentu dari DataFrame berfungsi untuk membersihkan data dan menyederhanakan analisis dengan hanya menyimpan kolom yang dianggap relevan. DataFrame yang tersisa kini siap untuk analisis lebih lanjut dan visualisasi data. |

1. **Mengubah Tipe Data**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Kode diatas digunakan untuk mengubah tipe data dari float64 ke int32 pada beberapa kolom numerik: 'Qty', 'Jual(Rp)', 'Beli(Rp)', 'Penjualan', dan 'Pembelian'. Ini penting karena dalam konteks tertentu, nilai numerik seperti kuantitas dan harga lebih cocok disimpan sebagai bilangan bulat (int) dibandingkan bilangan desimal (float), terutama jika data tidak memerlukan representasi desimal. |

1. **Analisis Korelasi Antar Variabel**
2. **Menghitung korelasi**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Kode diatas digunakan untuk menghitung dan menampilkan matriks korelasi untuk beberapa variabel numerik dalam dataset. Output menunjukkan bahwa terdapat hubungan yang sangat kuat antara harga jual, harga beli, penjualan, dan pembelian. Korelasi yang lebih rendah antara kuantitas dan variabel lainnya mengindikasikan bahwa jumlah barang yang terjual tidak selalu berkorelasi erat dengan harga atau total nilai transaksi. |

1. **Visualisasi korelasi (heatmap)**

|  |
| --- |
| **Script** |
|  |
| **Output** |
|  |
| **Analisis** |
| Visualisasi heatmap ini efektif dalam memberikan gambaran visual dari hubungan korelasi antar variabel. Warna yang lebih merah menunjukkan korelasi positif yang kuat, sedangkan warna biru menunjukkan korelasi negatif. Dalam heatmap ini, ada warna yang sangat merah antara Jual(Rp) dan Beli(Rp), Penjualan dan Pembelian, serta hubungan lainnya yang kuat. |

1. **Menyimpan Data Setelah Proses Pembersihan**

|  |
| --- |
| **Script** |
|  |
| **Analisis** |
| Karna proses eksplorasi data analisis sudah kita lakukan sebelumnya, maka kita perlu menyimpan file yang sudah dibersihkan tersebut dan membuat file baru Bernama ‘Tokokita\_clean.xlsx’. File ini akan dibuat di direktori kerja saat ini. File ini akan berisi data yang telah dibersihkan dan diformat sesuai dengan perubahan yang telah dilakukan pada DataFrame df. Menyimpan data yang telah dibersihkan sangat penting untuk menjaga hasil analisis dan menghindari melakukan pembersihan ulang setiap kali data tersebut diperlukan. |

1. **Visualisasi Data**
2. **Tren penjualan dan pembelian dalam per bulan**

|  |
| --- |
| **Visualisasi** |
|  |
| **Manfaat** |
| Dengan visualisasi tren penjualan dan pembelian per bulan, kita dapat memperoleh wawasan yang mendalam tentang bagaimana bisnis beroperasi sepanjang waktu. Visualisasi ini membantu dalam membuat keputusan yang lebih informasional dan memungkinkan penyesuaian strategi yang lebih baik untuk mencapai tujuan bisnis di masa depan.  Dari hasil visualisasi tersebut, terlihat bahwa penjualan menunjukkan tren yang lebih signifikan dibandingkan dengan pembelian sepanjang waktu. Hal ini dapat mengindikasikan bahwa ada peningkatan permintaan atau keberhasilan dalam strategi penjualan yang diterapkan. |

1. **Produk dengan penjualan dan pembelian terbanyak**

|  |
| --- |
| **Visualisasi** |
|  |
| **Manfaat** |
| Visualisasi ini memberikan pandangan yang jelas dan terperinci tentang produk mana yang paling berkontribusi terhadap penjualan dan pembelian, yang dapat membantu dalam pengambilan keputusan dan perencanaan bisnis yang lebih baik.  Produk seperti beras lalu diikuti dengan minuman, detergen, sabun, dan pewangi menunjukkan angka penjualan yang tinggi. Ini menunjukkan bahwa produk-produk ini sangat dibutuhkan dan populer di kalangan konsumen. Sedangkan snack, minuman, detergen, pewangi, dan roti muncul sebagai produk dengan pembelian terbanyak. Ini menunjukkan bahwa produk-produk ini sering dibeli dalam jumlah besar atau lebih sering dibandingkan produk lainnya.  Hal ini bisa dipertimbangkan agar perusahaan mempertimbangkan untuk meningkatkan stok dan memperluas variasi produk dalam kategori dengan penjualan tertinggi seperti beras dan minuman. Atau berdasarkan pola penjualan dan pembelian tersebut, perusahaan dapat merencanakan diversifikasi produk dan inovasi yang sesuai dengan tren pasar dan kebutuhan konsumen. |

1. **Distribusi penjualan dan pembelian berdasarkan kota**

|  |
| --- |
| **Visualisasi** |
|  |
| **Manfaat** |
| Visualisasi distribusi penjualan dan pembelian berdasarkan kota dapat memberikan wawasan yang mendalam mengenai performa pasar di berbagai lokasi. Dengan mengidentifikasi kota-kota dengan total penjualan dan pembelian tertinggi, perusahaan dapat merencanakan strategi pemasaran, pengelolaan stok, dan logistik yang lebih efisien. Selain itu, data ini dapat membantu dalam menyesuaikan penawaran produk sesuai dengan kebutuhan lokal, serta merencanakan ekspansi pasar dengan lebih strategis. |

1. **Distribusi penjualan berdasarkan sub-kategori**

|  |
| --- |
| **Visualisasi** |
|  |
| **Manfaat** |
| Visualisasi distribusi penjualan berdasarkan subkategori menggunakan pie chart memberikan gambaran jelas tentang proporsi kontribusi masing-masing subkategori terhadap total penjualan. Dengan cara ini, perusahaan dapat mengidentifikasi subkategori mana yang mendominasi penjualan dan mana yang memiliki kontribusi lebih kecil. Informasi ini sangat berharga untuk alokasi sumber daya, strategi promosi, dan pengelolaan inventaris.  Pada visualisasi diatas, subkategori kebutuhan pokok dan kebutuhan rumah tangga menyumbang sebagian besar penjualan, perusahaan dapat memfokuskan upaya pemasaran dan stok pada subkategori tersebut. Sebaliknya, subkategori dengan kontribusi kecil seperti jasa dan perlengkapan bayi mungkin memerlukan perhatian lebih untuk meningkatkan strategi penjualannya. |

1. **Tren penjualan per sub-kategori**

|  |
| --- |
| **Visualisasi** |
|  |
| **Manfaat** |
| Visualisasi tren penjualan per subkategori ini dapat memberikan wawasan mendalam mengenai bagaimana penjualan untuk setiap subkategori berubah seiring waktu. Dengan grafik ini, kita dapat melacak fluktuasi penjualan dalam setiap subkategori dari bulan ke bulan dan mengidentifikasi pola musiman, tren jangka panjang, atau perubahan mendadak dalam penjualan. Hal ini memungkinkan perusahaan untuk mengevaluasi performa subkategori secara lebih detail, mengidentifikasi peluang untuk peningkatan penjualan, dan menyesuaikan strategi pemasaran atau stok sesuai kebutuhan.  Seperti pada hasil visualisasi tren diatas, dapat kita lihat bahwa subkategori dari kebutuhan pokok lebih signifikan terutama pada sekitar bulan Januari-Mei tahun 2020. Hal ini dapat menjadi fokus untuk pengembangan lebih lanjut atau promosi khusus. Sebaliknya untuk subkategori yang menunjukkan tren penurunan penjualan, perusahaan dapat melakukan analisis lebih lanjut untuk memahami penyebabnya dan mengambil tindakan yang diperlukan. |

1. **Analisis rata-rata keuntungan per sub-kategori**

|  |
| --- |
| **Visualisasi** |
|  |
| **Manfaat** |
| Visualisasi rata-rata keuntungan per kategori memungkinkan identifikasi yang jelas mengenai kategori produk yang paling menguntungkan. Hal ini mempermudah pengambilan keputusan strategis, seperti memfokuskan usaha pada kategori yang memberikan keuntungan maksimal dan mengevaluasi kembali kategori yang kurang menguntungkan.  Selain itu, visualisasi ini membantu dalam perencanaan keuangan dan strategi bisnis dengan memberikan wawasan yang jelas tentang distribusi keuntungan di berbagai kategori produk. Dengan informasi ini, perusahaan dapat merancang strategi pemasaran yang lebih efektif dan melakukan penyesuaian operasional untuk meningkatkan profitabilitas. |

1. **Analisis tren keuntungan per bulan**

|  |
| --- |
| **Visualisasi** |
|  |
| **Manfaat** |
| Visualisasi tren keuntungan per bulan memberikan wawasan mendalam tentang fluktuasi keuntungan sepanjang waktu. Dengan grafik ini, kita dapat mengidentifikasi pola musiman, tren pertumbuhan atau penurunan, serta periode dengan keuntungan yang sangat tinggi atau rendah. Hal ini penting untuk memahami bagaimana keuntungan berubah dari waktu ke waktu dan dapat membantu dalam merencanakan strategi bisnis jangka panjang. |

1. **Analisis penjualan berdasarkan hari dalam seminggu**

|  |
| --- |
| **Visualisasi** |
|  |
| **Manfaat** |
| Visualisasi penjualan berdasarkan hari dalam seminggu membantu dalam memahami pola penjualan yang mungkin dipengaruhi oleh hari tertentu. Dengan informasi ini, perusahaan dapat mengidentifikasi hari-hari dengan penjualan tertinggi dan terendah, memungkinkan mereka untuk mengoptimalkan strategi pemasaran dan operasional sesuai dengan pola penjualan.  Pada visualisasi diatas menunjukkan bahwa penjualan tertinggi pada hari Senin dan Minggu, dimana kedua hari tersebut merupakan waktu-waktu yang signifikan untuk penjualan. Hal ini mengindikasikan adanya permintaan yang tinggi pada awal dan akhir minggu, yang dapat mempengaruhi strategi bisnis. Perusahaan harus mempertimbangkan untuk meningkatkan promosi dan persediaan produk pada hari-hari ini, serta memastikan layanan pelanggan yang optimal. |