



# **CAFE DATASET : ANALYSIS REPORT**



Syifa Gusriani Rohman

Mei 2025



# LATAR BELAKANG

Presentasi ini merupakan part kedua dari proyek latihan berjudul **“Cafe Dataset: Cleaning and Analysis”**, dimana ini merupakan penjelasan analisis yang didapatkan dari hasil pembersihan dataset cafe sebelumnya. Disini saya menganalisis sepenuhnya menggunakan Microsoft Excel. Saya berupaya semaksimal mungkin agar proses eksplorasi data berjalan bersih dan netral (bebas bias), sesuai kondisi asli dataset.

# TUJUAN PROYEK

- Melatih kemampuan membersihkan dan menganalisis data.
- Melatih penggunaan Microsoft Excel sebagai alat utama pembersihan dan analisis data.
- Menggali insight dengan menjawab serangkaian pertanyaan yang mengeksplorasi karakteristik dataset.

# DATASET YANG DIGUNAKAN

- Dataset : Cafe Sales - Dirty Data for Cleaning Training
- Periode : 2023
- Jumlah Data:
  - Kotor : 10000 baris
  - Bersih : 9515 baris
- Sumber : Kaggle ([Link](#))

## Catatan:

- Dataset ini merupakan dataset sintetis. Karena sumber aslinya tidak mencantumkan lokasi geografis, saya mengasumsikan data ini berasal dari wilayah Amerika Serikat, melihat dari harga produk yang umumnya berada di bawah angka 10, kemungkinan besar dalam satuan dolar (USD).
- Dari hasil pembersihan sebelumnya, pada beberapa kolom (Item, Payment, dan Location) saya memutuskan membiarkan kolom bernilai "ERROR", "UNKNOWN", dan mengisi nilai null menjadi "AMBIGOUS (Nama Kolom)"

# PENJELASAN KOLOM

Kolom	Penjelasan
Transaction ID	Id transaksi
Item	Produk
Quantity	Banyaknya produk yang dipesan
Price	Harga

Kolom	Penjelasan
Total Spent	Jumlah harga
Payment	Metode pembayaran
Location	Tempat makan (in-store atau takeaway)
Transaction Date	Tanggal transaksi

# PERTANYAAN BISNIS

1. Berapa total pendapatan yang dihasilkan selama tahun 2023, dan bagaimana distribusinya per bulan?
2. Bulan apa yang memiliki pendapatan tertinggi/terendah? Produk apa yang paling banyak terjual di bulan tersebut?
3. Hari apa yang memiliki rata-rata transaksi tertinggi? Pelanggan cenderung membeli di hari kerja atau libur (weekend atau tanggal merah)?
4. Metode pembayaran apa yang paling banyak digunakan? Apakah metode tertentu memiliki rata-rata transaksi lebih tinggi?
5. Produk mana yang memberikan kontribusi pendapatan tertinggi dan yang paling banyak terjual?
6. Bagaimana hubungan harga terhadap banyaknya produk yang terjual?
7. Manakah yang lebih dominan: transaksi in-store atau takeaway? Apakah salah satu metode memiliki rata-rata belanja lebih tinggi?

# PERTANYAAN 1

Berapa total pendapatan yang dihasilkan selama tahun 2023, dan bagaimana distribusinya per bulan?



Total pendapatan Café tahun 2023 mencapai **\$84.868,50**. Tren bulanan menunjukkan fluktuasi: misalnya, pendapatan Januari \$7.254 turun ke \$6.644 di Februari, lalu naik ke \$7.216 di Maret. Secara umum terlihat pola zigzag, dengan penurunan berturut-turut dari Maret hingga Mei sebelum kembali meningkat di Juni.

## PERTANYAAN 2

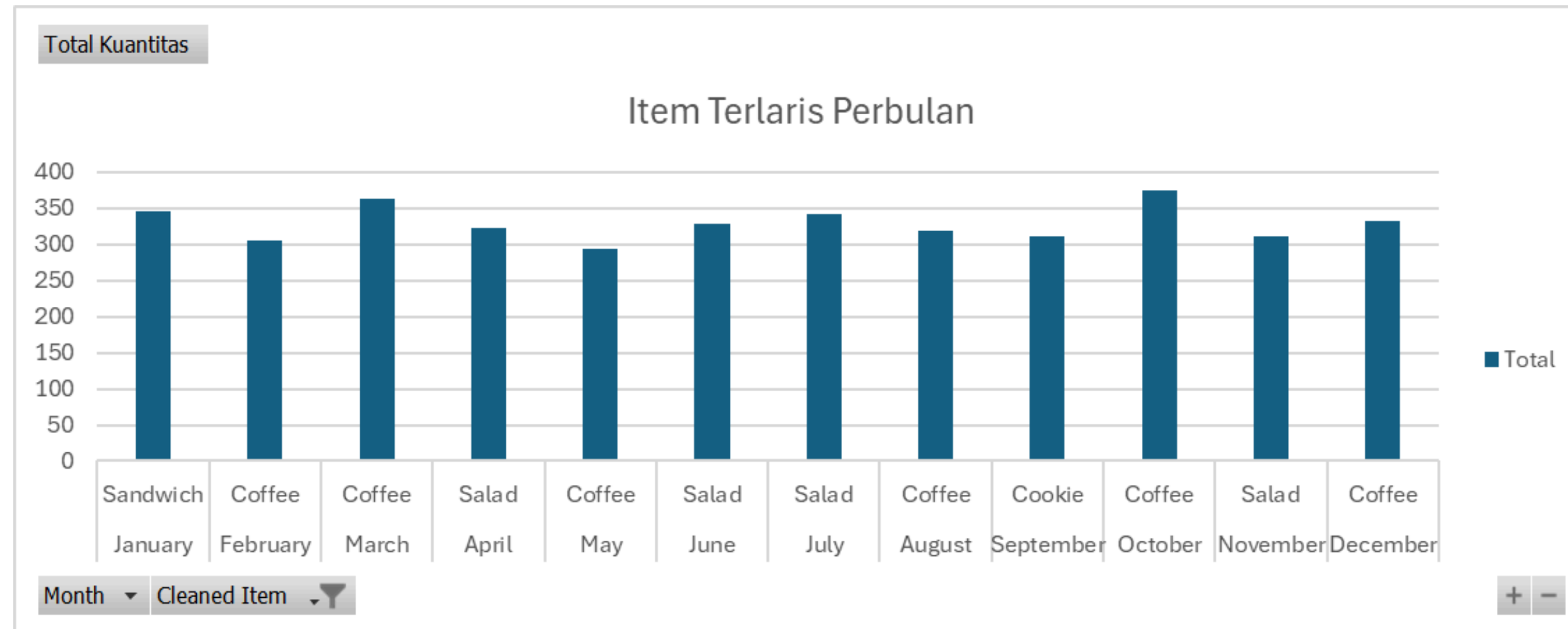
Bulan apa yang memiliki pendapatan tertinggi/terendah? Produk apa yang paling banyak terjual di bulan tersebut?



Dapat dilihat dari grafik disamping, **Juni** menjadi bulan dengan pendapatan tertinggi. Disusul **Oktober** dan **Januari**., Sementara **Februari**, **September**, dan **Juli** mencatat pendapatan terendah.



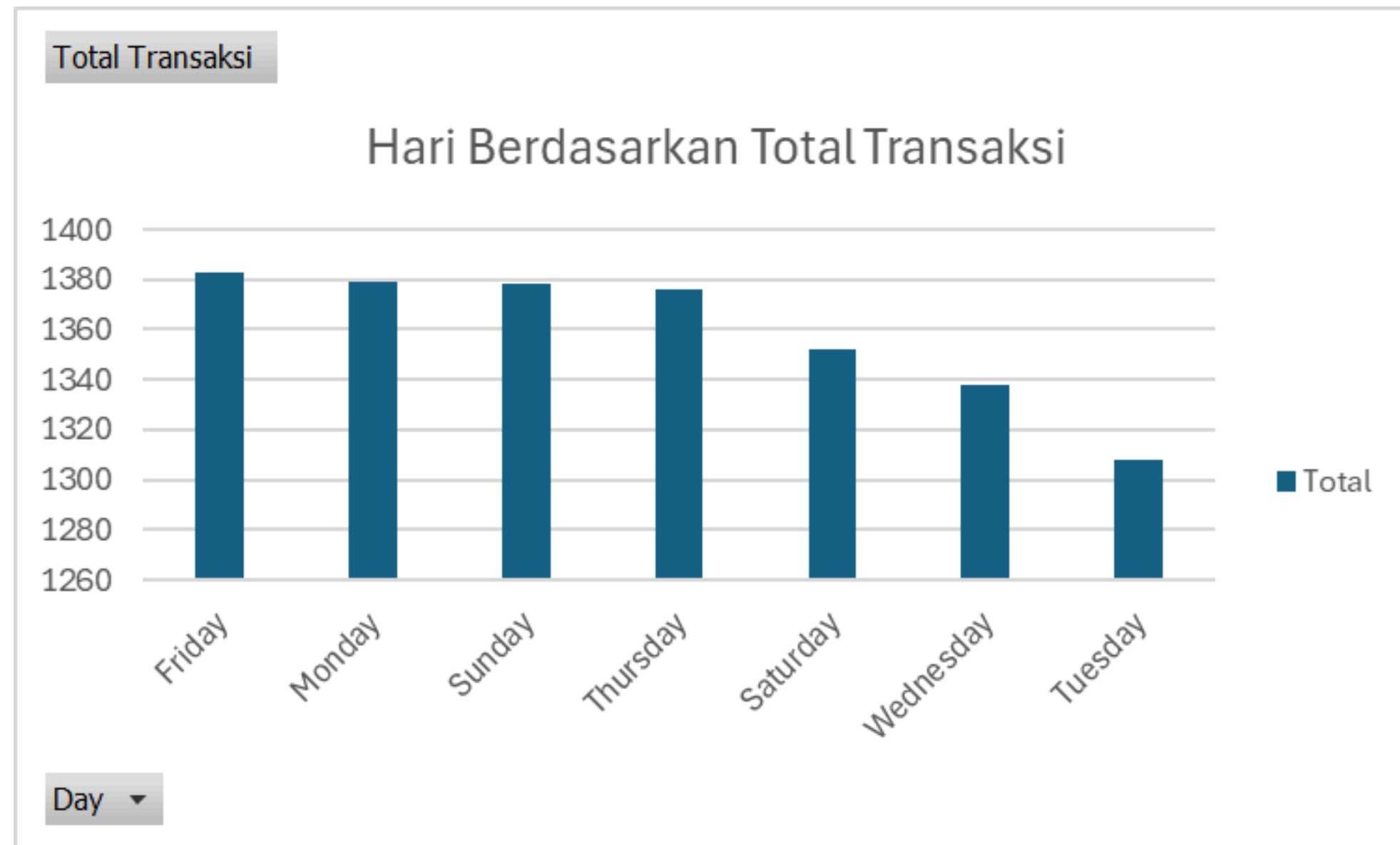
## PERTANYAAN 2



Beberapa produk konsisten menjadi andalan penjualan, seperti **Coffee** (nomor 1 di 6 bulan [Februari, Maret, Mei, Agustus, Oktober, dan Desember]) dan **Salad** (nomor 1 di 4 bulan [April, Juni, Juli, dan November]). Sandwich dan Cookie hanya muncul sekali (Januari & September).

# PERTANYAAN 3

Hari apa yang memiliki total transaksi tertinggi? Pelanggan cenderung membeli di hari kerja atau libur (weekend atau tanggal merah)?



Jumat mencatat total transaksi tertinggi, diikuti oleh Senin dan Minggu. Namun, perbedaan antara empat hari teratas tidak signifikan (selisihnya maksimal tujuh transaksi) seperti terlihat pada grafik. Perbedaan yang lebih nyata baru muncul pada posisi kelima hingga ketujuh (Sabtu, Rabu, dan Selasa).

## PERTANYAAN 3

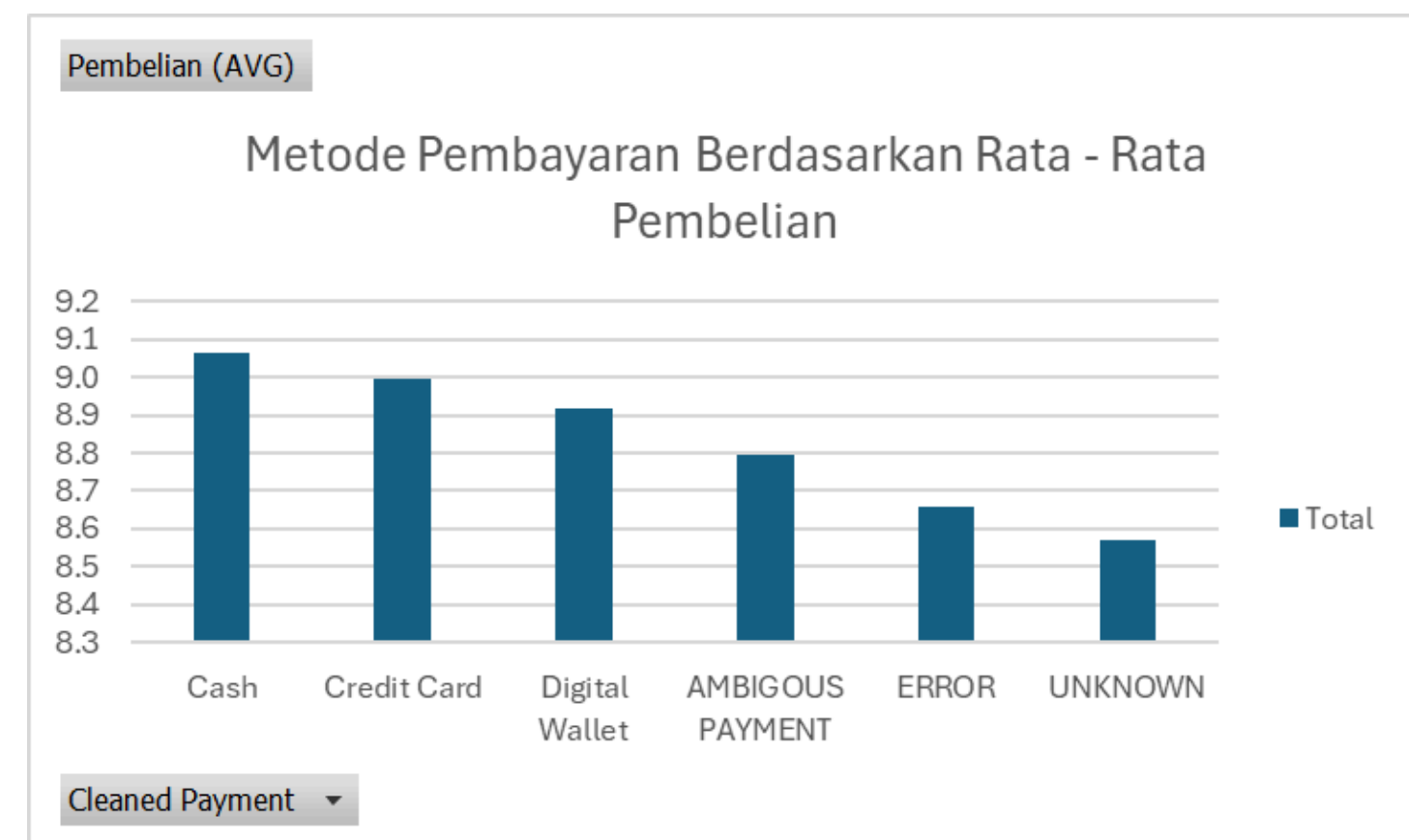
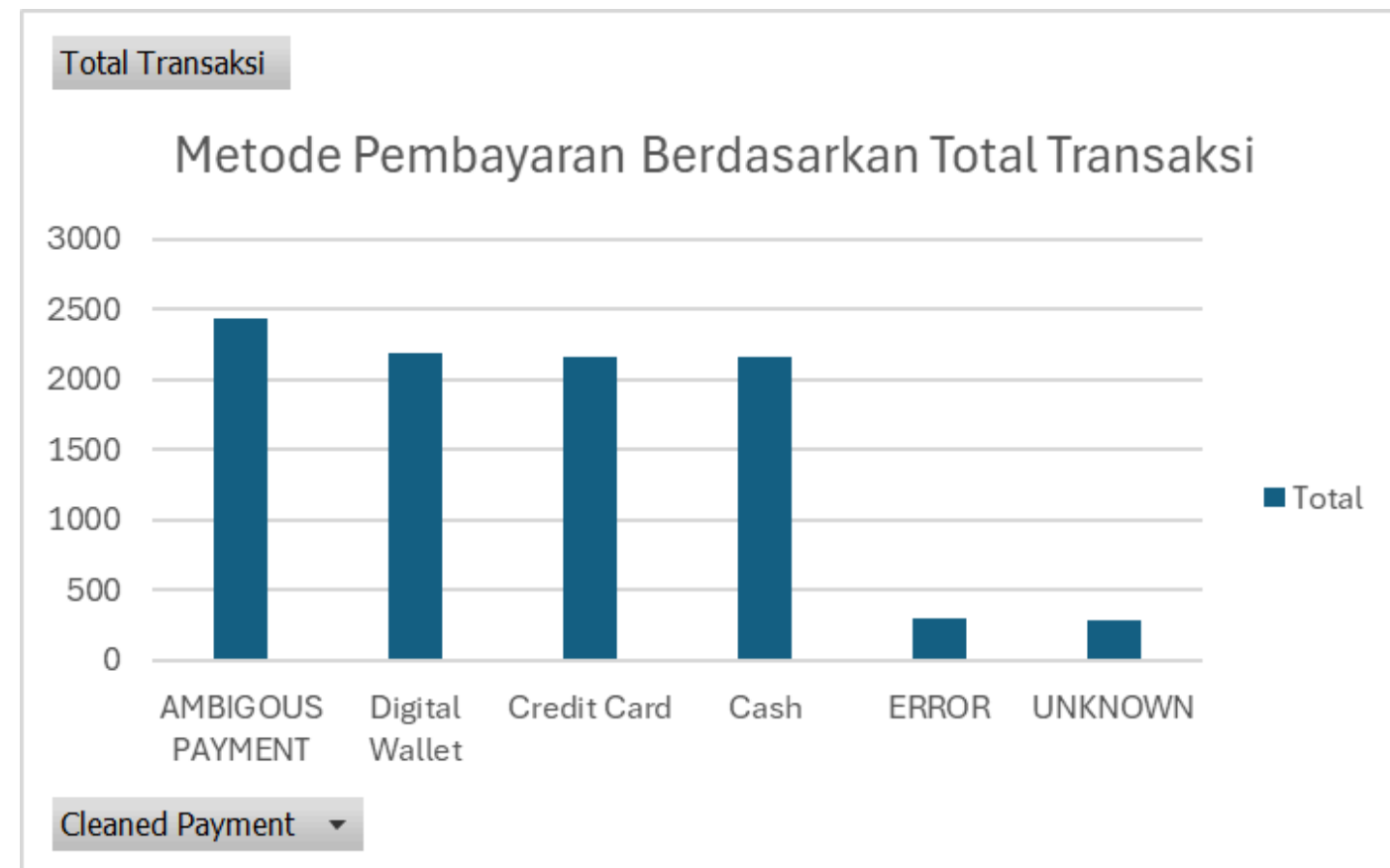
Hari apa yang memiliki total transaksi tertinggi? Pelanggan cenderung membeli di hari kerja atau libur (weekend atau tanggal merah)?

Jenis Hari	Total Transaksi	Total Hari	Rata - Rata Transaksi
Weekday	6538	250	25.152
Weekend	2730	105	26
Holiday	246	10	24.6

Jika dilihat berdasarkan tipe hari (weekday, weekend, atau hari libur nasional), hari kerja (weekday) memimpin jumlah transaksi dibandingkan dua kategori lainnya. Dengan demikian, pelanggan cenderung berbelanja pada hari kerja biasa.

# PERTANYAAN 4

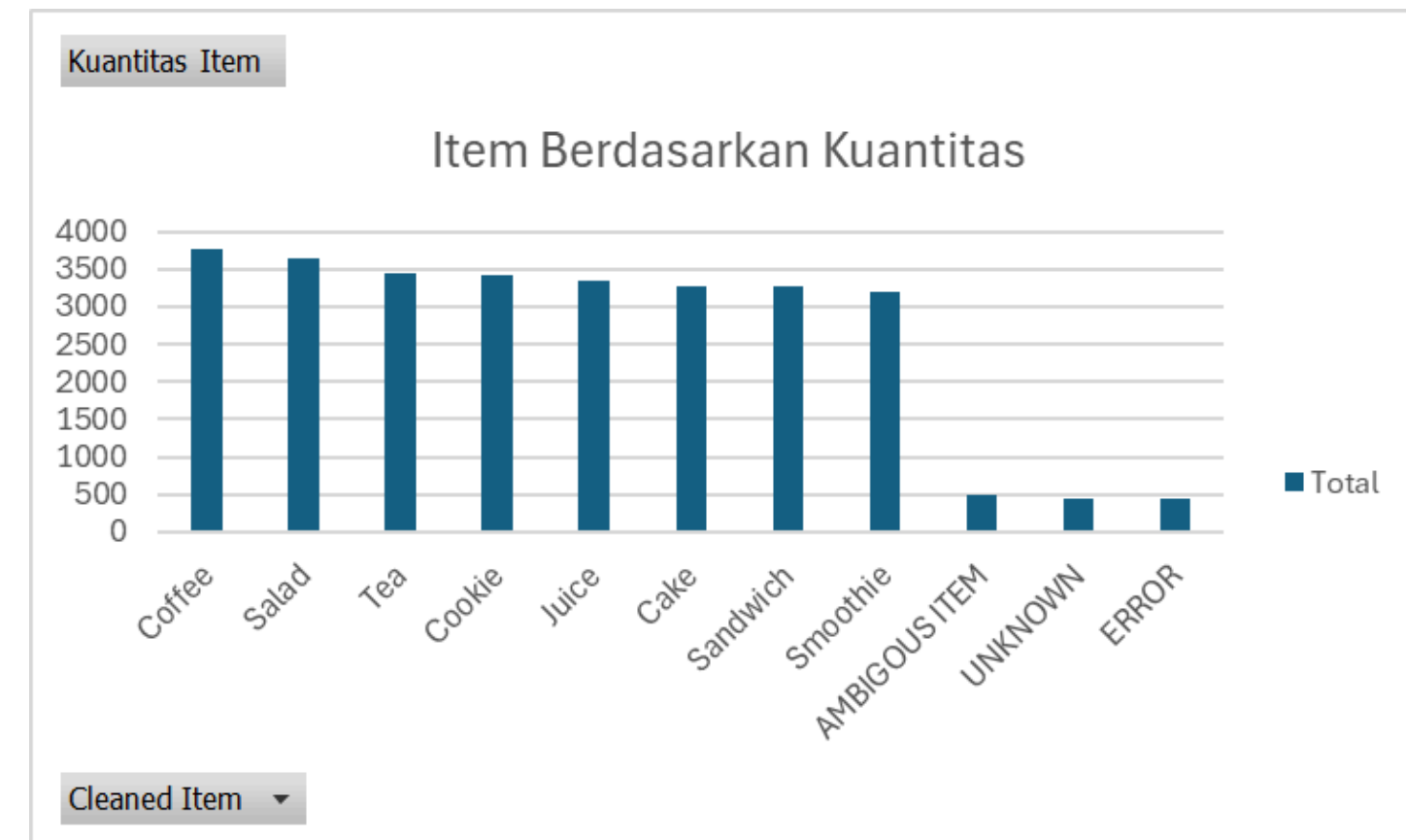
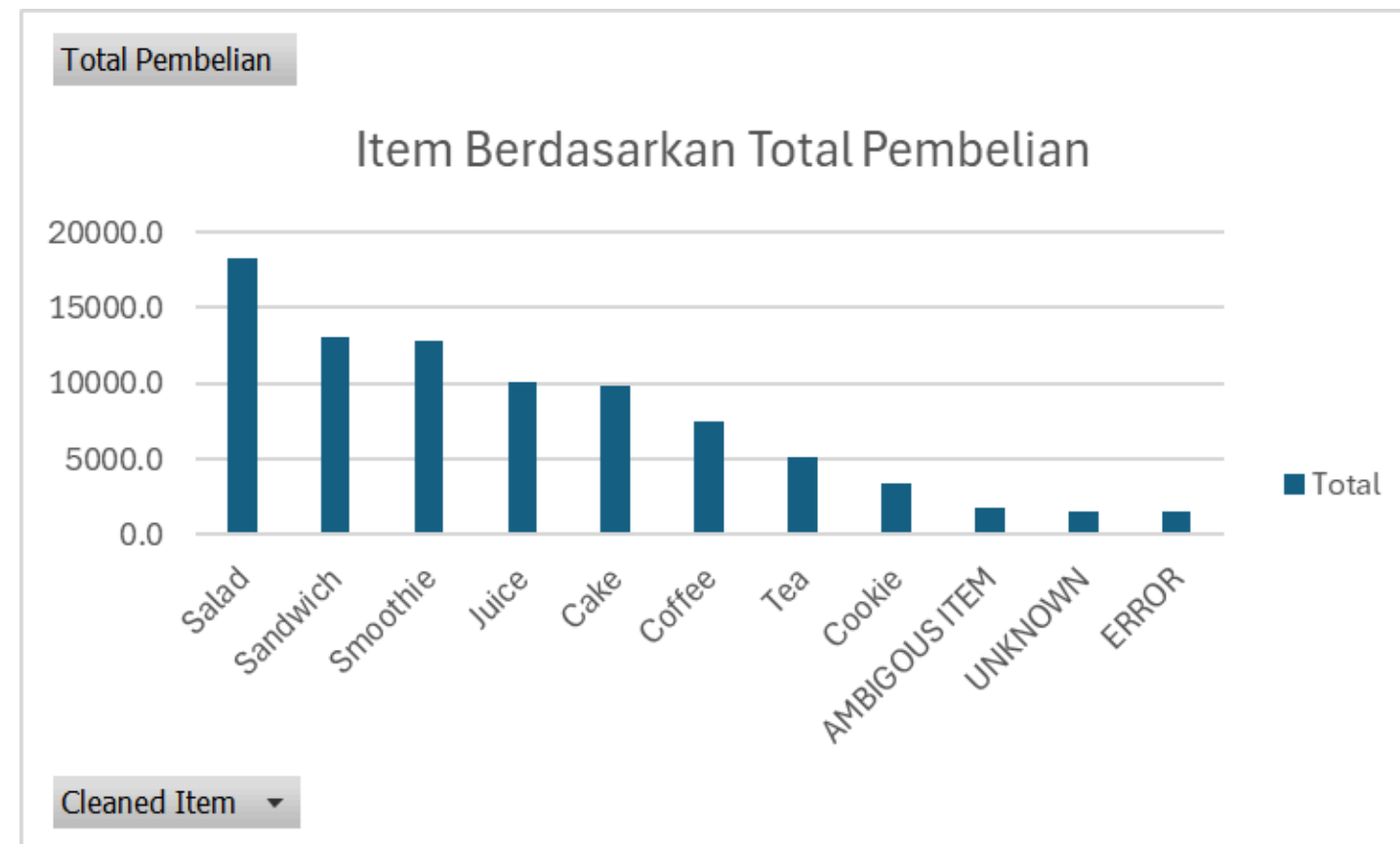
Metode pembayaran apa yang paling banyak digunakan? Apakah metode tertentu memiliki rata-rata transaksi lebih tinggi?



Sebagai catatan, saya tidak mengisi nilai kosong pada kolom metode pembayaran (diganti menjadi AMBIGUOUS PAYMENT) untuk menghindari bias. Berdasarkan data, transaksi terbanyak tercatat pada kategori AMBIGUOUS PAYMENT (nilai kosong), sehingga sulit mengidentifikasi metode pembayaran yang paling sering digunakan (total transaksi 2430). Namun, jika dilihat dari rata-rata nilai transaksi per metode, Cash menempati posisi pertama, diikuti oleh Credit Card dan Digital Wallet.

# PERTANYAAN 5

Produk mana yang memberikan kontribusi pendapatan tertinggi dan yang paling banyak terjual?



Diantara semua produk (tidak termasuk data ambigu), **Salad** menjadi produk dengan kontribusi pemasukkan tertinggi, disusul **Sandwich** dan **Smoothie**. Berbeda dengan pemasukkan tertinggi, untuk produk terlaris, yang pertama ditempati **Coffee**, disusul **Salad** dan **Tea**. Dari 450 transaksi ambigu: ada 234 transaksi (harga \$3) berpotensi menambah kuantitas Cake atau Juice dan 216 transaksi (harga 4) mungkin termasuk Sandwich atau Smoothie. Jika diklasifikasikan, peringkat 4–8 bisa berubah, tetapi **Coffee**, **Salad**, dan **Tea** tetap dominan sebagai top 3 terlaris.

## PERTANYAAN 6

Bagaimana hubungan harga terhadap banyaknya produk yang terjual?

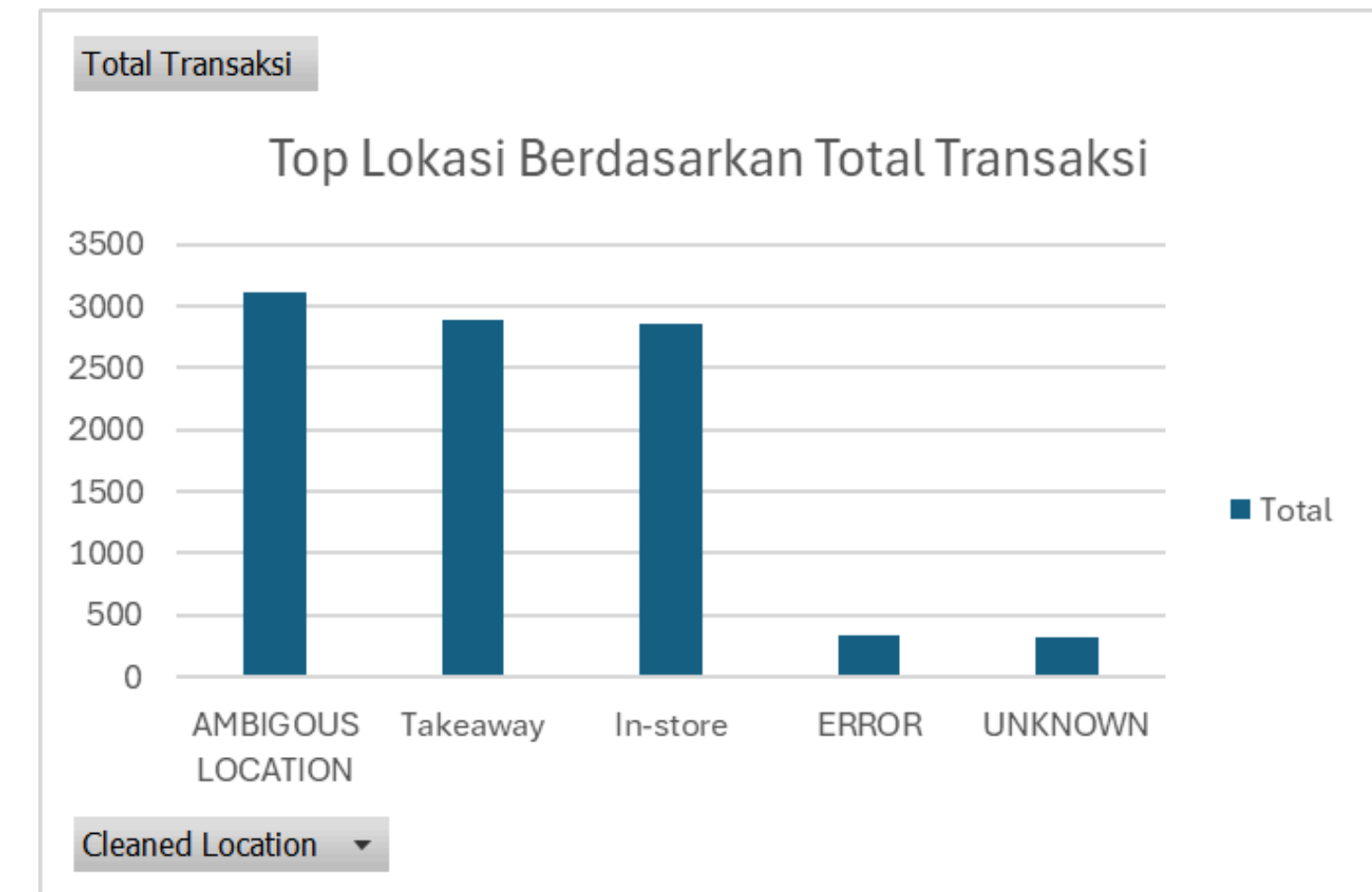
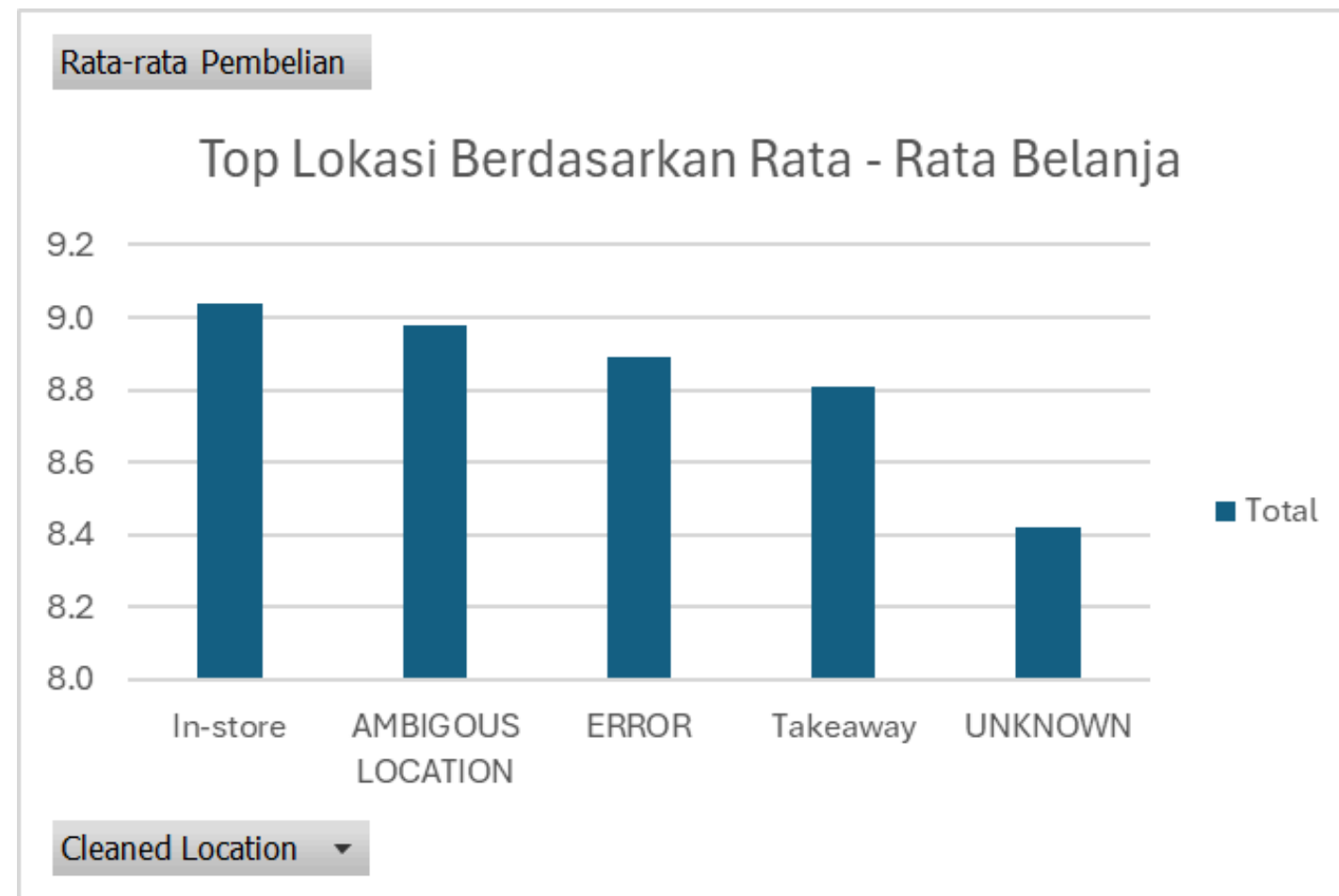
Item	Kuantitas	Harga (\$)
Coffee	3766	2
Salad	3644	5
Tea	3441	1.5
Cookie	3427	1

Item	Kuantitas	Harga (\$)
Juice	3341	3
Cake	3278	3
Sandwich	3267	4
Smoothie	3194	4

Sebagai catatan, karena analisis ini menilai korelasi antara kuantitas terjual dan harga, saya mengecualikan tiga item ambigu (AMBIGUOUS ITEM, UNKNOWN, dan ERROR). Berdasarkan perhitungan, diperoleh **nilai korelasi  $r = -0,17037$** , yang menunjukkan bahwa ada kecenderungan kecil (namun tidak kuat) bahwa penjualan menurun saat harga naik. Hal ini juga tercermin dari fakta bahwa Salad, meski berharga tertinggi, tetap menjadi salah satu produk terlaris.

# PERTANYAAN 7

Manakah yang lebih dominan: transaksi in-store atau takeaway? Apakah salah satu metode memiliki rata-rata belanja lebih tinggi?



Sebagai catatan, karena ini hanya dummy project untuk pembersihan data dan analisis, saya sengaja tidak mengisi nilai null pada kolom lokasi transaksi agar tidak menimbulkan bias. Akibat banyaknya data yang tidak terdefinisi, baik kategori **in-store** maupun **takeaway** menjadi agak ambigu (terutama top lokasi berdasarkan transaksi, dimenangkan AMBIGUOUS LOCATION yang bernilai 3110) dan perlu inspeksi lebih lanjut. Namun, dari data yang tersedia saat ini, **takeaway** tampak lebih sering dipilih pelanggan, sedangkan **in-store** mencatat rata-rata nilai transaksi tertinggi.



# KESIMPULAN

- **Tren pendapatan tahunan Café bersifat fluktuatif**, dimana **Juni** menjadi bulan puncak tertinggi (\$7k+), sedangkan Februari dan September mencatat pendapatan terendah.
- **Coffee dan Salad adalah produk unggulan yang mendominasi penjualan sepanjang tahun.** Coffee menjadi produk terlaris di 6 bulan berbeda, sementara Salad tidak hanya laris tetapi juga menjadi kontributor pendapatan tertinggi, meskipun harganya relatif mahal.
- **Transaksi paling aktif terjadi pada hari kerja** (weekday), dengan Jumat sebagai hari puncak. Namun, perbedaan jumlah transaksi antartop 4 hari (Jumat, Senin, Minggu, Kamis) sangat tipis, menunjukkan kebiasaan pelanggan yang konsisten sepanjang minggu.
- **Data ambigu (450 transaksi) berpotensi mengubah peringkat produk menengah seperti Juice, Cake, Sandwich, dan Smoothie.** Meskipun tidak mengganggu dominasi Coffee dan Salad, alokasi data ini perlu diverifikasi untuk akurasi analisis produk terlaris peringkat 4–8.



# KESIMPULAN

- Terdapat korelasi negatif lemah antara harga dan penjualan ( $r = -0,17$ ), yang menunjukkan pengaruh kecil kecenderungan penurunan penjualan saat harga naik. Hal ini tercermin dari produk Salad, yang karena tetap laris meski harganya tinggi.
- Data terkait metode pembayaran dan lokasi belum dapat dianalisis secara akurat karena sebagian besar didominasi oleh kategori **AMBIGUOUS PAYMENT** dan **AMBIGUOUS LOCATION (nilai kosong)**. Meskipun dua teratas menunjukkan nilai yang valid, hasil analisis secara keseluruhan belum bisa dianggap representatif.

# POLA PELANGGAN

- Pelanggan konsisten membeli Coffee dan Salad sepanjang tahun, dengan Salad sebagai pendorong pendapatan utama.
- Aktivitas transaksi paling tinggi di hari kerja, terutama Jumat, tetapi perbedaan antartop 4 hari sangat kecil, menunjukkan konsistensi jumlah pelanggan perharinya.
- Harga tidak signifikan memengaruhi penjualan untuk produk unggulan.
- Data ambigu menghambat analisis mendetail tentang preferensi produk menengah, metode pembayaran, dan lokasi transaksi.

# STRATEGI BISNIS

- Selalu pastikan konsistensi kualitas, terutama produk unggulan (Coffee dan Salad). Fokus pada standar rasa, penyajian, dan kesegaran bahan).
- Manfaatkan paket bundling produk unggulan + produk menengah untuk dorong penjualan di bulan rendah (Februari, September dan Juli).
  - Contoh paket:
    - "Weekday Coffee Combo": Coffee + Sandwich (diskon 15%).
    - "Salad Boost": Salad + Smoothie (harga spesial).
- Perbaiki Data Ambigu:
  - Kasir wajib isi metode pembayaran (cash/kartu) & lokasi (dine-in/takeaway) sebelum transaksi.
  - Audit 50 transaksi ambigu per minggu.
  - Perbaiki sistematisasi sistem database, dan selalu back up data (untuk mengatasi terjadinya error sistem).