



JUNIOR
DATA ANALYST

CAPSTONE
PROJECT

CAPSTONE PROJECT GUIDE

TUJUAN

Memberi kesempatan ke para peserta untuk menerapkan semua materi yang telah dipelajari selama ini, mulai dari **managing big data** dengan **SQL**, **basic programming in dataframe** dalam **data pre-processing** (data cleansing, data wrangling, data transformation, etc) dan menerapkan **ilmu statistik** pada **exploratory data analysis** (EDA) menggunakan Python, serta **visualisasi data dan storytelling** dari insight yang didapatkan menggunakan **Tableau**.

TOPIK

DATA PREPARATION



EDA



- 1) Critical thinking dalam merumuskan problem statement
- 2) Statistika deskriptif
- 3) Visualisasi

INTERPRETATION



- 1) Interpretasi pada setiap insight yang didapat

PRESENTATION & STORYTELLING



KEY PENILAIAN CAPSTONE PROJECT

1. Data Preparation
2. EDA
3. Interpretation
4. Presentation & Storytelling

	40%
	30%
	15%
	15%

Setiap tahapan harus unik. Diharapkan peserta memiliki keunikan masing-masing dan alasan mengambil tindakan tersebut. **Setiap tambahan insight di luar main problem akan mendapatkan extra credit.**



DEADLINE AND TIMELINE

Pengumpulan akan dilakukan terakhir pada hari **Rabu, tanggal 15 Desember 2021, 20:00 WIB.** Untuk pengumpulan setelah deadline akan dikenakan **sanksi pengurangan sebesar 10 poin.**

Konsultasi Notebook (Check Point 1):
Jum'at, 10 Desember 2021

Konsultasi Tableau (Check Point 2):
Selasa, 14 Desember 2021

Presentasi:
Kamis, 17 Desember 2021

SUBMISSION FORMAT

File yang diharapkan dari peserta adalah:

- 1 Jupyter Notebook (.ipynb)
- 2 Tableau workbook untuk mendukung presentation dan storytelling (.twbx)

PEMBAGIAN CASE

CASE 1: PENJUALAN COFFEE SHOP

Mentor: Adit

Assignee:

1. Ardan Zaki
2. Lazuardi Firdaus
3. Jihan Alfiyyah Munajat
4. Alex Negara
5. Ganes Aryudha

Note: Case ini merupakan **Individual Project**, sehingga dikerjakan perorangan.



CASE 2: RESTAURANT RECOMMENDATION

Mentor: Ayu

Assignee:

1. Muhammad Zaky Umadi
2. Indra Wahyu Pratama
3. Karas Chandra Gupta Khan
4. Fikri Irfan Adristi
5. Nyoto Wijaya

Note: Case ini merupakan **Individual Project**, sehingga dikerjakan perorangan.



CASE 3: RETAIL ONLINE SHOP

Mentor: Syahrul

Assignee:

1. Izzatul Mardiah Saini
2. Fazlinafi Aulia Asdin
3. Rahmat Taufiq Sigit
4. Muhammad Zuhdi Afi Abiyyi

Note: Case ini merupakan **Individual Project**, sehingga dikerjakan perorangan.



CASE 4: FOODS AND GOODS ORDERS IN BRAZIL

Mentor: Yusfi

Assignee:

1. Andrew Nangoy
2. Syadiyah Israita
3. Uly Aldini
4. Dwi Yuwono Kristanto

Note: Case ini merupakan **Individual Project**, sehingga dikerjakan perorangan.



CASE STUDY 1

DATASET: DATA PENJUALAN COFFEE SHOP

Diasumsikan Anda adalah calon dari salah satu tim data yang akan melamar pada perusahaan coffee shop. Namun, perusahaan tersebut memberi data sampel penjualan kopi hanya dalam lingkup rentang waktu tertentu. Data yang Anda dapatkan adalah data terkait profiling dari setiap pelanggan, transaksi berdasarkan product, dan inventory.



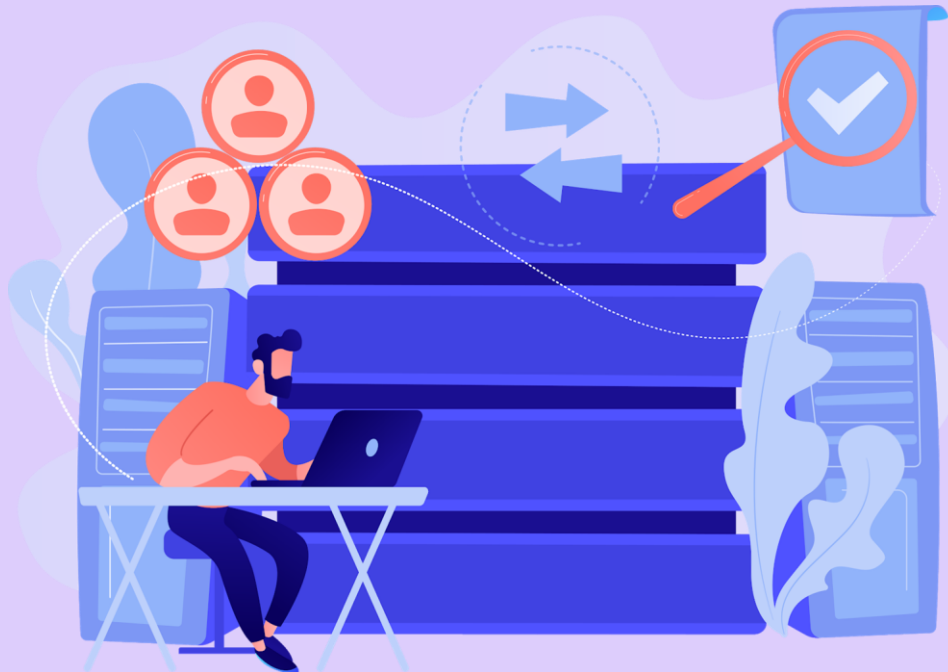
DATASET: DATA PENJUALAN COFFEE SHOP

Dari user perusahaan tersebut memberikan tugas, agar Anda sebagai tim data bisa menemukan keanehan, insight dari data, serta bisa memberikan rekomendasi terhadap perusahaan tersebut.



DATASET: DATA PENJUALAN COFFEE SHOP

Anda diwajibkan untuk menemukan hal-hal yang menarik dari data tersebut, seperti:



01

Profiling Customer

02

Transaksi Penjualan

03

Product Recommendation

04

Optimalisasi Inventory

DATA UNDERSTANDING | KETERANGAN DARI VARIABEL

Tabel	Variables	Ket
customer	customer_id	Customer id unik
	home_store	Nama store
	customer_first-name	Nama customer
	customer_email	Email customer
	customer_since	Tanggal bergabung
	loyalty_card_number	Nomor loyalty card
	birthdate	Tanggal lahir
	gender	Jenis kelamin
	birth_year	Tahun lahir
generation	birth_year	Tahun lahir
	generation	Tipe generasi
pastry inventory	sales_outlet_id	Outlet id unik
	transaction_date	Tanggal transaksi
	product_id	Produk id unik
	start_of_day	Target penjualan/stok harian per hari

Tabel	Variables	Ket
product	quantity_sold	Jumlah produk yang habis
	waste	Jumlah produk sisa
	% waste	% produk sisa
	product_id	Produk id unik
	product_group	Group produk - level 3
	product_category	Kategori produk - level 2
	product_type	Tipe produk - level 1
	product	Produk - level 0
	product_description	Deskripsi dari produk
	unit_of_measure	Ukuran size
	current_wholesale_price	Harga grosir
	current_retail_price	Harga jual
	tax_exempt_yn	Bebas pajak/tidak
	promo_yn	Dapat promo/tidak
	new_product_yn	Produk baru/tidak

DATA UNDERSTANDING | KETERANGAN DARI VARIABEL

Tabel	Variables	Ket
sales outlet	sales_outlet_id	Sales id unik
	sales_outlet_type	Tipe outlet
	store_square_feet	-
	store_address	Alamat store
	store_city	Kota store
	store_state_province	Provinsi store
	store_telephone	Nomor telepon store
	store_postal_code	Postal code store
	store_longitude	Longitude store
	store_latitude	Latitude store
	manager	Kode manager
	neighborhood	Lingkungan dari lokasi store
sales receipts	transaction_id	Transaksi id unik
	transaction_date	Tanggal transaksi
	transaction_time	Jam transaksi

Tabel	Variables	Ket
	sales_outlet_id	Sales outlet id unik
	staff_id	Staff id unik
	customer_id	Customer id unik
	instore_yn	Pembelian in store/tidak
	order	-
	line_item_id	-
	product_id	Product id unik
	quantity	Jumlah orderan
	line_item_amount	-
	unit_price	Harga satuan
	promo_item_yn	Dapat promo/tidak

HINTS!

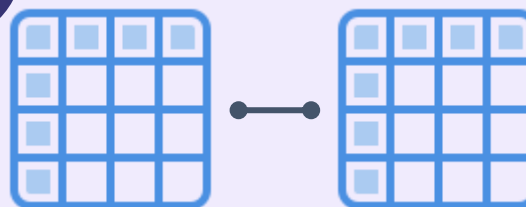
01



Lakukan **data cleansing** dari setiap tabel:

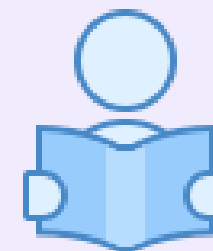
- Customer
- Generation
- Pastry Inventory
- Product
- Sales Outlet
- Sales Receipts

02



Perhatikan dalam **proses joining tabel**, apakah menghilangkan makna dari tujuan di level customer atau di level transaksi.

03



Perhatikan **data understanding** yang menyatakan unique id apakah benar-benar unik.

HINTS!

04



Listing **pertanyaan** yang menjadi **problem statement** agar bisa mendukung goal dari analisis data tersebut. Goal utama dalam analisis adalah sebagai berikut:

- Melakukan profiling customer agar bisa mendapatkan customer yang lebih tertarget, misal mendukung dalam pendistribusian produk di setiap store.
- Memberikan product recommendation agar bisa meningkatkan revenue.
- Mengoptimalkan stok atau mengurangi stok produk agar tidak banyak yang terbuang.

HINTS!

04



Contoh pertanyaan:

01

Dari segmentasi generation, customer mana yang memiliki kecintaan terhadap kopi?

02

Produk rekomendasi apa yang bisa menjadi produk terlaris dari masing-masing store?

HINTS!

05

Selamat bereksplorasi, tingkatkan rasa keingintahuan Anda terhadap insight apa yang akan Anda ambil dari data tersebut, tetap semangat ! :)



CASE STUDY 2

DATASET: RESTAURANT RECOMMENDATION

Anda bekerja sebagai Data Analyst di sebuah perusahaan restoran yang memiliki cabang perusahaan di berbagai wilayah. Anda ingin mengetahui kondisi perusahaan tersebut berdasarkan data sampel yang telah Anda peroleh dari atasan ketika rapat kemarin. Data sample tersebut berisi data pesanan, vendor, dan customer yang diambil dalam kurun waktu tertentu.



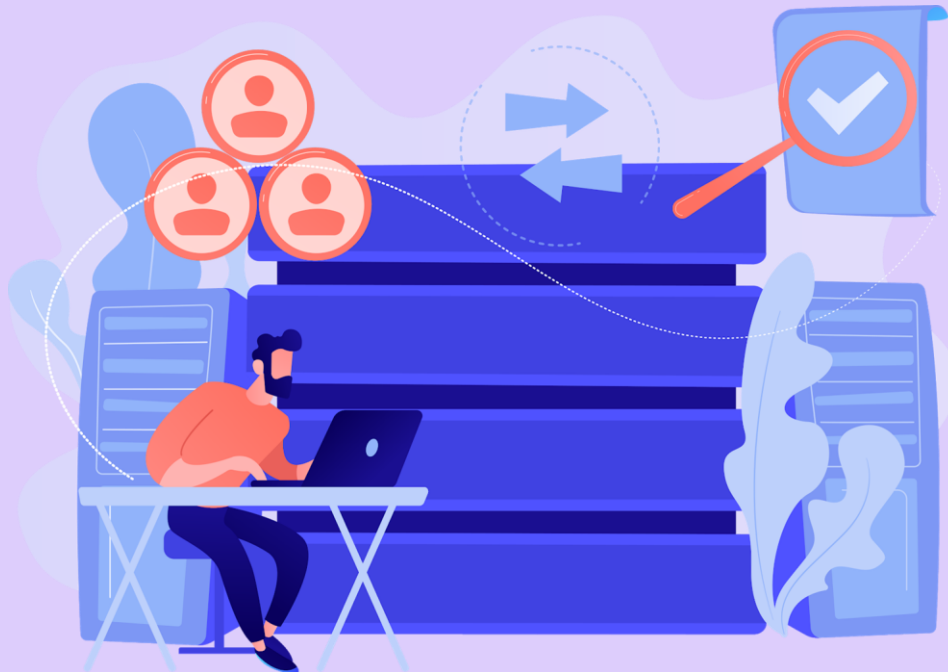
DATASET: RESTAURANT RECOMMENDATION

Mendengar hal itu, atasan Anda meminta untuk segera melaporkan hasil analisis data tersebut beserta insight apa saja yang bisa didapatkan. Anda juga diminta untuk memberikan rekomendasi strategi bisnis terhadap perusahaan tersebut.



DATASET: RESTAURANT RECOMMENDATION

Anda diwajibkan untuk menemukan hal-hal yang menarik dari data tersebut, seperti:



01 Profiling Customer

02 Transaksi Penjualan

03 Profiling Vendor

DATA UNDERSTANDING | KETERANGAN DARI VARIABEL

Tabel	Variables	Keterangan
Customer	customer_id	Identitas unik customer, digunakan untuk mengakses data locations dan orders
	gender	Jenis kelamin customer
	dob	Tahun lahir
	status and verified	Status akun
	Created_at and updated_at	Waktu ketika akun dibuat dan diupdate
Order	order_id	Identitas unik pesanan
	customer_id	Identitas unik customer (digunakan untuk menghubungkan dataset customer)
	item_count	Banyaknya item yang dipesan
	grand_total	Total biaya
	payment_mode	Metode pembayaran (kategorik)
	is_favorite	Item favorit dari vendor
	is_rated	Item yang dinilai
	driver_rating	Rating pengemudi
	deliverydistance	Jarak pengiriman
	created_at	Waktu pemesanan
Locations	vendor_id	Identitas unik vendor
	LOCATION_NUMBER	Nomor lokasi menentukan lokasi customer mana yang menjadi tujuan pengiriman
	customer_id	Identitas unik customer
	location_number	Nomor lokasi (kebanyakan customer memiliki nomor satu atau dua)
	Latitude and longitude	Menunjukkan lokasi (bukan lokasi sebenarnya karena lokasi telah disamarkan)

DATA UNDERSTANDING | KETERANGAN DARI VARIABEL

Tabel	Variables	Keterangan
Vendor	id	Identitas unik vendor
	authentication_id	Identitas otentikasi
	latitude	Lokasi pelanggan (garis lintang)
	longitude	Lokasi pelanggan (garis bujur)
	vendor_category_en	Kategori vendor (Restaurants dan Sweets & Bakes)
	delivery_charge	Biaya pengiriman
	serving_distance	Jarak pelayanan
	is_open	-
	OpeningTime	Waktu buka
	preparation_time	Waktu persiapan (menit)
	discount_percentage	Persentase diskon
	status	-
	verified	-
	rank	-
	vendor_rating	Rating dari vendor skala (0 s/d 5)
	vendor_tag	Menunjukkan label dari vendor
	vendor_tag_name	Menunjukkan nama label vendor
	created_at	Waktu dibuat
	updated_at	Waktu diupdate
	device_type	Tipe device yg digunakan

HINTS!

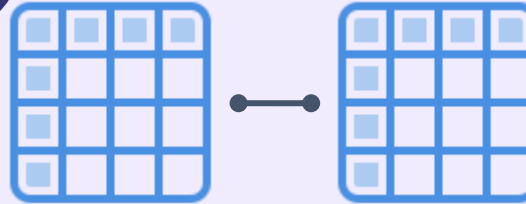
01



Lakukan **data cleansing** dari setiap tabel:

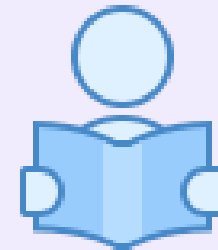
- Customer
- Vendor
- Order
- Locations

02



Perhatikan dalam **proses joining tabel**, apakah menghilangkan makna dari tujuan di level customer atau di level transaksi.

03



Perhatikan **data understanding** yang menyatakan unique id apakah benar-benar unik.

HINTS!

04



Listing **pertanyaan** yang menjadi **problem statement** agar bisa mendukung goal dari analisis data tersebut. Goal utama dalam analisis adalah sebagai berikut:

- Melakukan profiling customer agar bisa mendapatkan customer yang lebih tertarget, misal mendukung dalam pendistribusian produk.
- Memberikan product recommendation agar bisa meningkatkan revenue.
- Mengoptimalkan penggunaan Payment agar lebih tertarget pada customer.

HINTS!**04****Contoh pertanyaan:****01**

Dari segmentasi generation, customer mana yang banyak melakukan pesanan ke vendor?

02

Dari nama tag vendor, manakah nama tag yang paling banyak muncul?

HINTS!

05

Selamat bereksplorasi, tingkatkan rasa keingintahuan Anda terhadap insight apa yang akan Anda ambil dari data tersebut, tetap semangat ! :)



CASE STUDY 3

DATASET: RETAIL ONLINE SHOP

Anda bekerja di perusahaan besar yang memiliki banyak cabang di beberapa wilayah. User Anda memberikan Anda sampel data yang berisi informasi berupa deskripsi produk, distribusi produk, dan transaksi dari masing-masing produk. Saat ini, Anda sedang mempersiapkan untuk membuat laporan data transaksi.



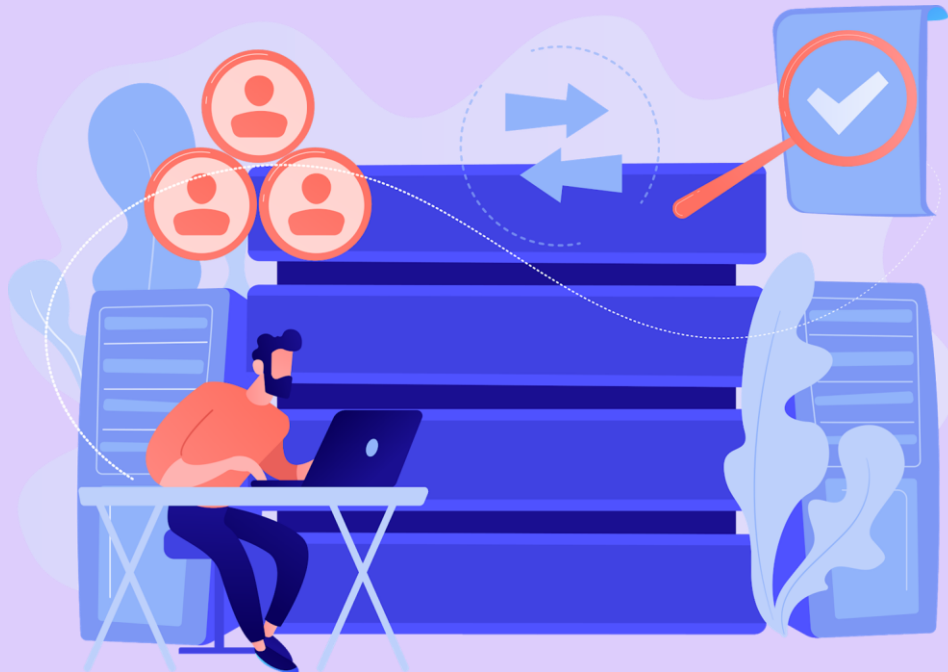
DATASET: RETAIL ONLINE SHOP

Dikarenakan distribusi barang berada di beberapa wilayah, Anda ingin mengetahui apakah terdapat keanehan dari data yang anda miliki. Kemudian Anda juga ingin mencari tahu insight dari data tersebut, agar Anda dapat memberikan rekomendasi kepada user perusahaan Anda.



DATASET: RETAIL ONLINE SHOP

Anda diwajibkan untuk menemukan hal-hal yang menarik dari data tersebut, seperti:



01

Profiling Company

02

Performa Penjualan

03

Product Recommendation

04

Distribusi Wilayah Penjualan

DATA UNDERSTANDING | KETERANGAN DARI VARIABEL

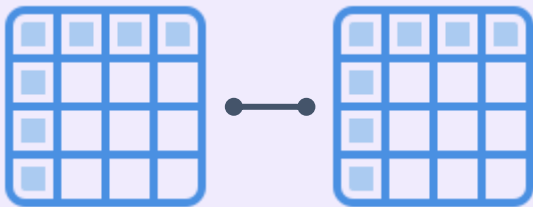
Tabel	Variables	Keterangan
bill_locations	loc_ID	ID unik dari lokasi transaksi
	bill_ID	ID unik transaksi
	address	Alamat lokasi transaksi
	latitude	Latitude lokasi transaksi
	longitude	Longitude lokasi transaksi
	town/city	Kabupaten/kota lokasi transaksi
	country	Negara lokasi transaksi
bills	bill_ID	ID unik transaksi
	invoice	Kode transaksi
	bill_date	Tanggal transaksi
bill_items	bill_item_ID	ID unik item setiap transaksi
	bill_ID	ID unik transaksi
	product_ID	ID unik produk
	quantity	Jumlah unit item yang dibeli
	selling_price	Harga jual per unit (dalam USD)
	total_sales	Total harga yang harus dibayarkan = quantity * selling_price (dalam USD)
bill_locations	loc_ID	ID unik dari lokasi transaksi
	bill_ID	ID unik transaksi
	address	Alamat lokasi transaksi
	latitude	Latitude lokasi transaksi

DATA UNDERSTANDING | KETERANGAN DARI VARIABEL

Tabel	Variables	Keterangan
main_categories	category_ID	ID unik dari jenis kategori
	category_name	Nama kategori
sub_categories	subcategory_ID	ID unik dari jenis subkategori
	subcategory_name	Nama sub kategori
brand_details	brand_ID	ID unik dari brand/merk
	brand_name	Nama brand/merk
	company	Nama perusahaan produk (Note : 1 nama brand dapat dimiliki oleh beberapa nama perusahaan)
products	product_ID	ID unik dari produk
	product_name	Nama produk
	brand_ID	ID dari brand/merk
	category_ID	ID dari jenis kategori
	subcategory_ID	ID dari jenis sub kategori

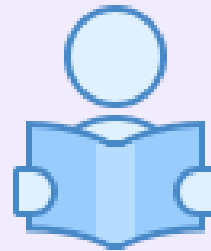
HINTS!

01



Lakukan **joining tabel** untuk mendapatkan informasi produk dan bagaimana produk tersebut dipasarkan.

02



Perhatikan baik-baik definisi variabel tiap data. Carilah **keywords variabel** untuk memudahkan Anda join tabel.

03



Perhatikan kondisi data Anda. Lakukanlah **data cleaning** sehingga tidak ada yang misrepresentasi.

HINTS!

04



Listing **pertanyaan** yang menjadi **problem statement** agar bisa mendukung goal dari analisis data tersebut. Goal utama dalam analisis adalah sebagai berikut:

- Melakukan profiling perusahaan, agar bisa mendapatkan informasi produk apa saja yang dimiliki perusahaan saat ini, serta apakah perlu diperluas variasi produknya.
- Memberikan product recommendation agar bisa meningkatkan revenue.
- Memberikan gambaran distribusi wilayah penjualan untuk mengetahui potensi pembukaan cabang toko baru.

HINTS!**04****Contoh pertanyaan:****01**

Produk apa yang paling digemari customer dan bagaimana distribusi wilayah penjualan produk tersebut?

02

Negara mana yang menjadi top contributor terhadap penjualan retail?

HINTS!

05

Selamat bereksplorasi, tingkatkan rasa keingintahuan Anda terhadap insight apa yang akan Anda ambil dari data tersebut, tetap semangat ! :)



CASE STUDY 4

DATASET: DELIVERY CENTER - FOOD & GOODS ORDERS IN BRAZIL

Anda bekerja di perusahaan tempat pusat pengiriman di Brazil. Dikarenakan mitra dan pelanggan sudah mulai bertambah, maka perlu dilakukan analisa terkait pengelolaan pengiriman barang selama ini. Beruntungnya, Anda telah memiliki record pengiriman mulai bulan Januari hingga April 2021. Untuk membuat ekosistem yang sehat antara penjual dan pembeli, Anda sebagai data analyst diminta untuk menganalisis data dan memberikan rekomendasi untuk keperluan pengaturan strateginya di masa mendatang.



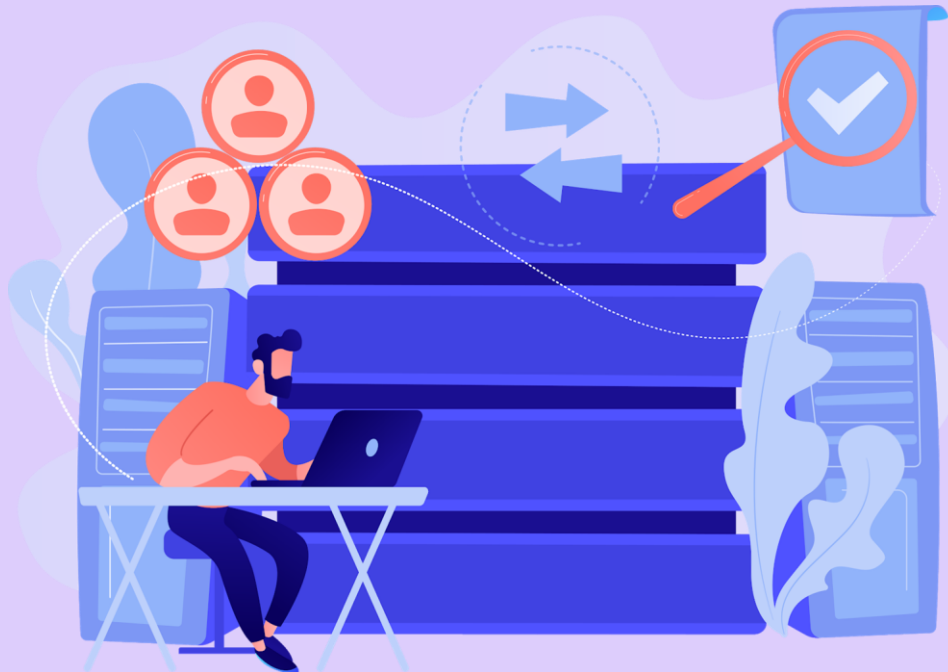
DATASET: DELIVERY CENTER - FOOD & GOODS ORDERS IN BRAZIL

Data yang Anda miliki saat ini berisi informasi mengenai channel, deliveries, drivers, hubs, orders, payments, dan stores. User perusahaan Anda ingin mengetahui bagaimana kondisi perusahaan saat ini, baik insight data maupun jika ada keanehan dari data.



DATASET: DELIVERY CENTER - FOOD & GOODS ORDERS IN BRAZIL

Anda diwajibkan untuk menemukan hal-hal yang menarik dari data tersebut, seperti:



01 Profiling Perusahaan

02 Performa Pengiriman

03 Customer Behavior

04 Performa Mitra

DATA UNDERSTANDING | KETERANGAN DARI VARIABEL

Tabel	Variables	Keterangan
channels	channel_id	ID channel
	channel_name	Nama channel
	channel_type	Jenis/tipe channel
deliveries	driver_id	ID pengiriman
	delivery_order_id	ID pengiriman pesanan
	delivery_id	ID pengirim yang melakukan pengiriman
	delivery_distance_meters	Jarak tempuh
drivers	delivery_status	Status pengiriman
	driver_id	ID orang yang mengirim
	driver_modal	Jenis/tipe pengiriman
hubs	hub_id	ID hub pusat pengiriman
	hub_name	Nama hub pusat pengiriman
	hub_city	Kota pusat pengiriman
	hub_state	Negara bagian/provinsi hub pusat pengiriman
	hub_latitude	Posisi lintang hub pusat pengiriman
	hub_longitude	Posisi busur hub pusat pengiriman

Tabel	Variables	Keterangan
payments	payment_id	ID pembayaran
	payment_order_id	ID pesanan terkait dengan pembayaran
	payment_amount	Jumlah pembayaran
	payment_fee	Jumlah diskon
	payment_method	Cara pembayaran
stores	payment_status	Status pembayaran
	store_id	ID toko/pedagang
	hub_id	ID hub pusat pesanan yang terkait dengan toko/pedagang
	store_name	Nama toko/pedagang
	store_segment	Segmen toko/pedagang
	store_plan_price	Nilai paket pedagang/toko
	store_latitude	Letak garis lintang toko
	store_longitude	Letak garis busur toko

DATA UNDERSTANDING | KETERANGAN DARI VARIABEL

Tabel	Variables	Keterangan
orders	order_id	ID order yang dilakukan
	store_id	ID penjual/toko
	channel_id	ID channel
	payment_order_id	ID pembayaran
	delivery_order_id	ID pengiriman pesanan
	order_status	Status pemesanan
	order_amount	Jumlah pemesanan
	order_delivery_fee	Pendapatan pengiriman atau jumlah harga yang dibayar
	order_delivery_cost	Biaya pengiriman
	order_created_hour	Waktu pemesanan dalam jam
	order_created_minuta	Waktu pemesanan dalam menit
	order_created_day	Waktu pemesanan dalam hari
	order_created_year	Waktu pemesanan dalam tahun
	order_moment_created	Waktu pemesanan

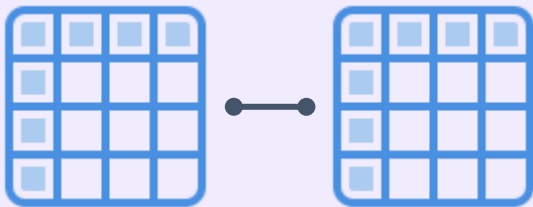
Tabel	Variables	Keterangan
	order_moment_accepted	Waktu pemesanan diterima
	order_moment_ready	Waktu pesanan siap diantarkan
	order_moment_collected	Waktu pesanan terkumpul
	order_moment_in_expedition	Waktu pesanan di ekspedisi
	order_moment_delivering	Waktu pesanan dalam pengiriman
	order_moment_delivered	Waktu ketika pesanan terkirim
	order_moment_finished	Waktu pesanan selesai dikirim
	order_metric_collected_time	Metrik waktu pengumpulan pesanan

DATA UNDERSTANDING | KETERANGAN DARI VARIABEL

Tabel	Variables	Keterangan
orders	order_metric_paused_time	Metrik waktu jeda pesanan
	order_metric_production_time	Metrik waktu produksi pesanan
	order_metric_walking_time	Metrik waktu berjalannya pesanan
	order_metric_expedition_speed_time	Metrik kecepatan waktu ekspedisi pesanan
	order_metric_transit_time	Metrik waktu transit pesanan
	order_metric_cycle_time	Metrik siklus waktu pesanan

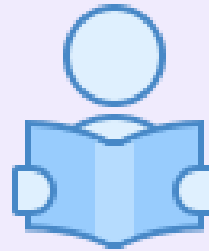
HINTS!

01



Lakukan **joining tabel** untuk mendapatkan informasi produk dan bagaimana produk tersebut dipasarkan.

02



Perhatikan baik-baik definisi variabel tiap data. Carilah **keywords variabel** dan pastikan bahwa items untuk variabel yang berisi informasi produk memiliki nilai yang unik.

03



Perhatikan kondisi data Anda. Lakukanlah **data cleaning** sehingga tidak ada yang misrepresentasi.

HINTS!

04



Listing **pertanyaan** yang menjadi **problem statement** agar bisa mendukung goal dari analisis data tersebut. Goal utama dalam analisis adalah sebagai berikut:

- Melakukan profiling perusahaan agar bisa mendapatkan customer, mitra, dan trend pengiriman saat ini.
- Memberikan recommendation agar bisa meningkatkan revenue dan efisiensi operasional serta menciptakan ekosistem yang sehat.

HINTS!**04****Contoh pertanyaan:****01**

Berapa jumlah customer dan mitra saat ini?

02

Apakah ada penurunan trend pengiriman dari waktu ke waktu?

HINTS!

05

Selamat bereksplorasi, tingkatkan rasa keingintahuan Anda terhadap insight apa yang akan Anda ambil dari data tersebut, tetap semangat ! :)





THANK YOU!