**Analyzing eCommerce Business Performance with SQL**

**Overview** :

Selamat datang di Mini Project: *Analyzing eCommerce Business Performance with SQL*. Pada Mini Project ini, Anda berperan sebagai anggota dari tim Data Analytics di sebuah perusahaan eCommerce. Perusahaan ini merupakan salah satu marketplace terbesar di Amerika Selatan yang menghubungkan pelaku bisnis mikro dengan para pelanggannya. Sebagai anggota dari tim Data Analytics, Anda mendapatkan tanggung jawab untuk menganalisis tiga aspek yang berkaitan dengan performa bisnis perusahaan tersebut. Ketiga aspek tersebut di antaranya adalah pertumbuhan pelanggan, kualitas produk dan tipe pembayaran. Anda akan mengolah data yang sudah disediakan untuk membuat laporan performa bisnis pada ketiga aspek tersebut.

**Skill level** : Beginner

**Keterampilan yang dibutuhkan** :

* Business understanding
* Analytical thinking
* SQL Reporting

**Learning output** :

* Mengenal dan memahami beberapa metrik bisnis pada industri eCommerce.
* Mampu memanfaatkan database PostgreSQL dari data mentah dengan format CSV.
* Mampu menerjemahkan pertanyaan bisnis ke dalam query SQL.

**Estimasi waktu** : 2-4 jam

**Dataset** : https://drive.google.com/file/d/1m3QI\_d3gZUZRGg4aJYkZzoXQpX9WPmIn/view?usp=sharing

**Tools yang digunakan** : PostgreSQL

**Tugas 1**

**Data Preparation**

**Deskripsi singkat** :

Sebelum memulai pemrosesan data, tahap paling awal yang harus dilakukan adalah mempersiapkan data mentah menjadi data yang terstruktur dan siap diolah. Kamu harus memasukkan data dengan format csv ke dalam tabel-tabel database menggunakan PostgreSQL. Kemudian membuat entity relationship antar tabel-tabel tersebut. Kamu akan diminta untuk melakukan instalasi software terkait dengan panduan yang sudah disiapkan.

**Tujuan tugas**  : Membuat data mentah menjadi data yang siap diolah

**Kemampuan yang akan dikuasai** :

1. Importing data csv ke database
2. Querying alter statement dalam membuat primary key dan foreign key

**Langkah - langkah** :

1. Download file "Dataset.rar", kemudian ekstrak file tersebut.
2. Membuat database baru beserta tabel-tabelnya untuk data yang sudah disiapkan

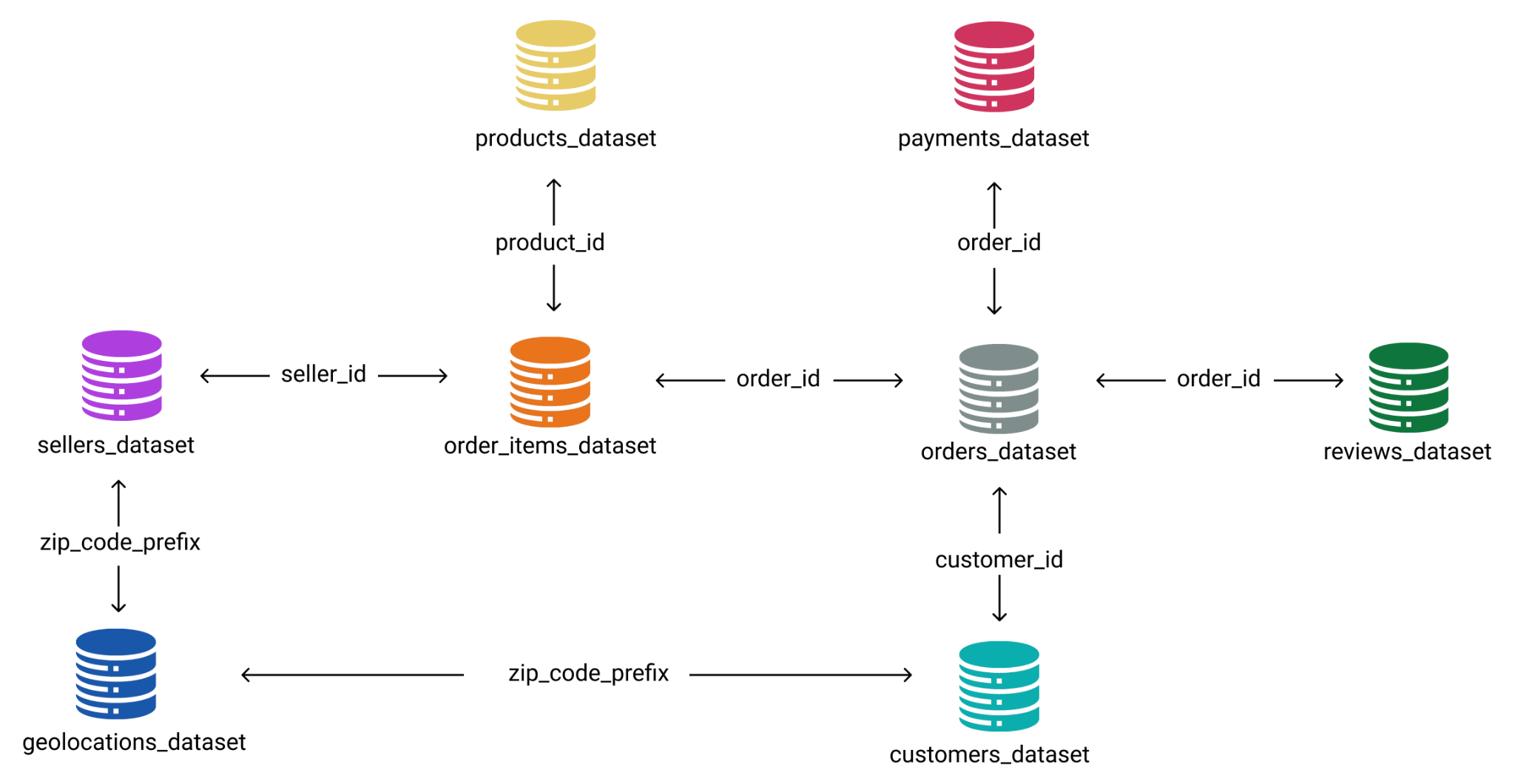
*Hint*: Perhatikan tipe data setiap kolom

1. Importing data csv ke dalam database

*Hint*: Perhatikan path penyimpanan dataset

1. Membuat entity relationship antar tabel, berdasarkan skema pada file "Data Relationship.png". Kemudian export Entity Relationship Diagram (ERD) dalam bentuk gambar.

*Hint*: Perhatikan kesesuaian tipe data dan penamaan kolom antar tabel yang saling berhubungan



**Resource**  :

1. <https://www.postgresql.org/download/> (panduan install postgreSQL)
2. [Import CSV File Into PostgreSQL Table (postgresqltutorial.com)](https://www.postgresqltutorial.com/import-csv-file-into-posgresql-table/)
3. [PostgreSQL Tutorial](https://docs.google.com/document/d/1LiyTCvTZifdDbAPLso5LOnIx6VFSU_wAYMid_WQEW4c/edit?usp=sharing)

**Hasil** :

1. Gambar entity relationship diagram
2. Kesimpulan tugas 1 (apa saja yang dilakukan dan bagaimana melakukannya)

**Template** : [Tugas 1 - Template](https://docs.google.com/presentation/d/14Hwmzlc29GDlqhULItdNmasa1ordyn1B3oWcVb0qbKA/edit?usp=sharing)

**Tugas 2**

**Annual Customer Activity Growth Analysis**

**Deskripsi singkat** :

Salah satu metrik yang digunakan untuk mengukur performa bisnis eCommerce adalah aktivitas customer yang berinteraksi di dalam platform eCommerce tersebut. Dalam Tugas 2 ini, Anda akan menganalisis beberapa metrik yang berhubungan dengan aktivitas customer seperti jumlah customer aktif, jumlah customer baru, jumlah customer yang melakukan repeat order dan juga rata-rata transaksi yang dilakukan customer setiap tahun.

**Tujuan tugas**  :

Melihat dan menganalisis apakah performa bisnis eCommerce dari sisi aktivitas customer mengalami pertumbuhan, stagnan atau bahkan mengalami penurunan dalam kurun waktu tahunan.

**Kemampuan yang akan dikuasai** :

1. Kemampuan memahami partisi data bertipe timestamp untuk mengekstrak komponen timestamp
2. Kemampuan memahami cara melakukan agregasi group by
3. Kemampuan memahami penggunaan subquery

**Langkah - langkah** :

1. Menampilkan rata-rata jumlah customer aktif bulanan (monthly active user) untuk setiap tahun  
   *Hint*: Perhatikan kesesuaian format tanggal
2. Menampilkan jumlah customer baru pada masing-masing tahun  
   *Hint*: Pelanggan baru adalah pelanggan yang melakukan order pertama kali
3. Menampilkan jumlah customer yang melakukan pembelian lebih dari satu kali (repeat order) pada masing-masing tahun*Hint*: Pelanggan yang melakukan repeat order adalah pelanggan yang melakukan order lebih dari 1 kali
4. Menampilkan rata-rata jumlah order yang dilakukan customer untuk masing-masing tahun*Hint*: Hitung frekuensi order (berapa kali order) untuk masing-masing customer terlebih dahulu
5. Menggabungkan keempat metrik yang telah berhasil ditampilkan menjadi satu tampilan tabel*Hint*: Lakukan pembuatan tabel sementara terhadap subtask-subtask sebelumnya terlebih dahulu

**Resource**  :

1. <https://www.postgresqltutorial.com/postgresql-date_part/>
2. <https://www.postgresqltutorial.com/postgresql-group-by/>
3. <https://www.postgresqltutorial.com/postgresql-subquery/>
4. <https://www.postgresql.org/docs/9.1/queries-with.html>
5. <https://www.geeksforgeeks.org/postgresql-joins/>

**Hasil** :

1. Rata-rata Monthly Active User (MAU) per tahun
2. Total customer baru per tahun
3. Jumlah customer yang melakukan repeat order per tahun
4. Rata-rata frekuensi order untuk setiap tahun
5. Master tabel yang berisi semua informasi di atas

**Template** : [Tugas 2 - Template](https://docs.google.com/presentation/d/1RIJF3H5PwCawGjjeCQapUrKD0Tb1qaNDp2LU6MxN1F0/edit?usp=sharing)

**Tugas 3**

**Annual Product Category Quality Analysis**

**Deskripsi singkat** :

Performa bisnis eCommerce tentunya sangat berkaitan erat dengan produk-produk yang tersedia di dalamnya. Menganalisis kualitas dari produk dalam eCommerce dapat memberikan keputusan untuk mengembangkan bisnis dengan lebih baik. Dalam Tugas 3 ini, Anda akan menganalisis performa dari masing-masing kategori produk yang ada dan bagaimana kaitannya dengan pendapatan perusahaan.

**Tujuan tugas**  :

Melihat dan menganalisis kategori-kategori produk yang memberikan dampak positif (dilihat dari metrik pendapatan/revenue) dan juga kategori produk yang memberikan dampak negatif (dilihat dari metrik jumlah cancel order) bagi perusahaan setiap tahunnya.

**Kemampuan yang akan dikuasai** :

1. Kemampuan memahami teknik membuat tabel dari query result
2. Kemampuan memahami teknik window function dan juga fungsi RANK

**Langkah - langkah** :

1. Membuat tabel yang berisi informasi pendapatan/revenue perusahaan total untuk masing-masing tahun*Hint*: Revenue adalah harga barang dan juga biaya kirim. Pastikan juga melakukan filtering terhadap order status yang tepat untuk menghitung pendapatan
2. Membuat tabel yang berisi informasi jumlah cancel order total untuk masing-masing tahun*Hint*: Perhatikan filtering terhadap order status yang tepat untuk menghitung jumlah cancel order
3. Membuat tabel yang berisi nama kategori produk yang memberikan pendapatan total tertinggi untuk masing-masing tahun*Hint*: Perhatikan penggunaan window function dan juga filtering yang dilakukan
4. Membuat tabel yang berisi nama kategori produk yang memiliki jumlah cancel order terbanyak untuk masing-masing tahun*Hint*: Perhatikan penggunaan window function dan juga filtering yang dilakukan
5. Menggabungkan informasi-informasi yang telah didapatkan ke dalam satu tampilan tabel*Hint*: Perhatikan teknik join yang dilakukan serta kolom-kolom yang dipilih

**Resource**  :

1. <https://www.postgresql.org/docs/8.2/sql-createtableas.html>
2. <https://www.postgresql.org/docs/9.1/tutorial-window.html>

**Hasil** :

1. Tabel revenue per tahun
2. Tabel jumlah cancel order per tahun
3. Tabel top kategori yang menghasilkan revenue terbesar per tahun
4. Tabel kategori yang mengalami cancel order terbanyak per tahun
5. Master tabel yang berisi informasi-informasi di atas

**Template** : [Tugas 3 - Template](https://docs.google.com/presentation/d/1FOEYZjc-oSH6FuY18fynFq-JVIm_Uql8W6v3YF4UOhY/edit?usp=sharing)

**Tugas 4**

**Annual Payment Type Usage Analysis**

**Deskripsi singkat** :

Bisnis eCommerce umumnya menyediakan sistem pembayaran berbasis open-payment yang memungkinkan customer untuk memilih berbagai macam tipe pembayaran yang tersedia. Menganalisis performa dari tipe pembayaran yang ada dapat memberikan insight untuk menciptakan strategic partnership dengan perusahaan penyedia jasa pembayaran dengan lebih baik. Dalam Tugas 3 ini, Anda akan menganalisis tipe-tipe pembayaran yang tersedia dan melihat tren perubahan yang terjadi selama beberapa tahun terakhir.

**Tujuan tugas**  :

Melihat dan menganalisis tipe pembayaran yang menjadi favorit customer secara all time dan juga menganalisis perubahan tren yang terjadi dari tahun ke tahun untuk masing-masing tipe pembayaran yang ada.

**Kemampuan yang akan dikuasai** :

1. Kemampuan mengartikan masalah bisnis ke dalam bahasa SQL
2. Kemampuan dalam menggunakan CASE WHEN sebagai salah satu bantuan untuk mendapatkan pivot table

**Langkah - langkah** :

1. Menampilkan jumlah penggunaan masing-masing tipe pembayaran secara all time diurutkan dari yang terfavorit *Hint*: Perhatikan struktur (kolom-kolom apa saja) dari tabel akhir yang ingin didapatkan
2. Menampilkan detail informasi jumlah penggunaan masing-masing tipe pembayaran untuk setiap tahun*Hint*: Perhatikan struktur (kolom-kolom apa saja) dari tabel akhir yang ingin didapatkan

**Resource**  :

1. <https://www.postgresqltutorial.com/postgresql-case/>

**Hasil** :

1. Jumlah penggunaan masing-masing jenis pembayaran all time
2. Detail jumlah penggunaan masing-masing jenis pembayaran untuk setiap tahun
3. Satu tabel summary jumlah penggunaan tipe-tipe pembayaran untuk masing-masing tahun.

**Template** : [Tugas 4 - Template](https://docs.google.com/presentation/d/1tqW9ZE4JEMVuy765CsvVO6OKl_qYfX3SZkTf_XCdjqw/edit?usp=sharing)