

Analyzing eCommerce Business Performance with SQL



Created by:

Nadiana Yemin Mukti

muktinadiana@gmail.com

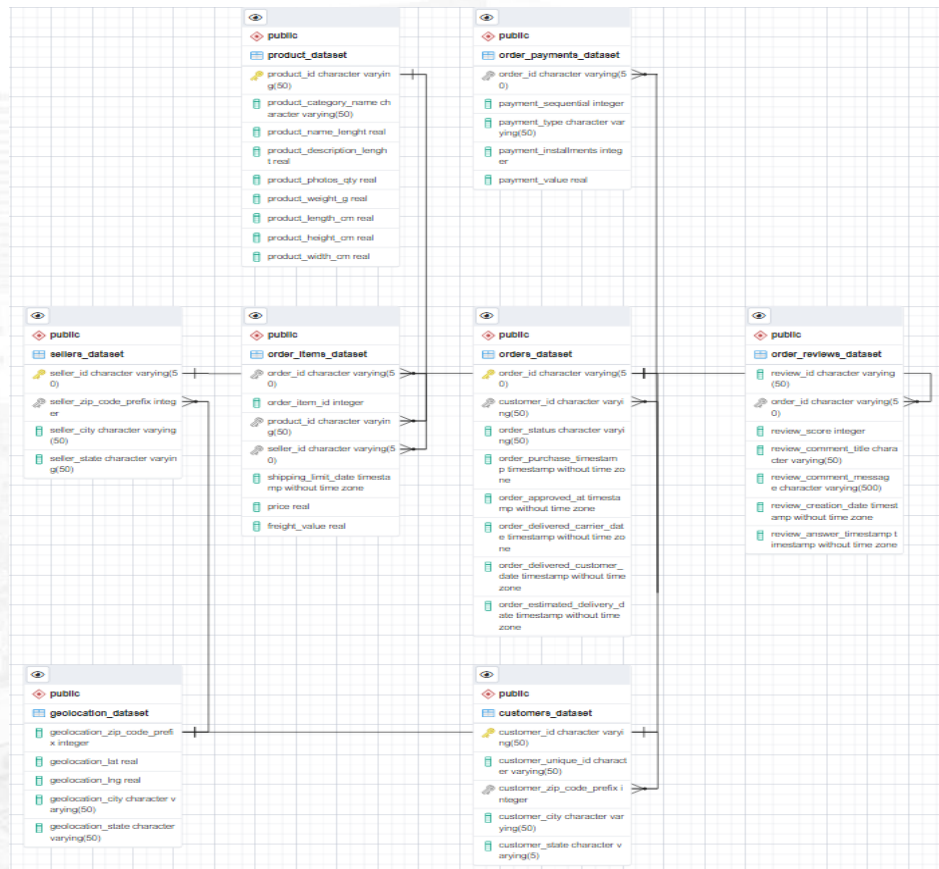
www.linkedin.com/in/nadiana-yemin-mukti-316a91149

“Seorang lulusan Universitas Pertamina Jakarta pada Program Studi Teknik Perminyakan, juga lulus dari Bootcamp Data Science Rakamin Academy, mempelajari PostgreSQL, Python, Data Visualisasi, Statistika, dan Machine Learning. Saat ini tengah melaksanakan Volunteering di SDN Sukamulya 2 untuk menganalisis data siswa, membuat machine learning Clustering mengenai potensi siswa berdasar nilai, membuat database menggunakan PostgreSQL, dan membuat sistem penilaian serta analisis hasil ujian siswa menggunakan google drive, docs, spreadsheet.”

“Dalam suatu perusahaan mengukur performa bisnis sangatlah penting untuk melacak, memantau, dan menilai keberhasilan atau kegagalan dari berbagai proses bisnis. Oleh karena itu, dalam paper ini akan menganalisa performa bisnis untuk sebuah perusahaan eCommerce, dengan memperhitungkan beberapa metrik bisnis yaitu pertumbuhan pelanggan, kualitas produk, dan tipe pembayaran.”

Data Preparation







ERD



- Pada tahap data preparation, merupakan proses persiapan data yang akan dipakai, agar data siap dan layak untuk digunakan.
- Pelaksanaannya mulai proses:
 - import data external ke dalam database dengan menggunakan query CREATE TABLE (nama_tabel tipe_data),
 - melaksanakan pengecekan tipe data dan mengganti kolom shipping_limit_date (tabel order_items_dataset), review_creation_date dan review_answer_timestamp (tabel order_reviews_dataset), order_purchase_timestamp (tabel orders_dataset), order_approved_at (tabel orders_dataset), order_delivered_carrier_date (tabel orders_dataset), order_delivered_customer_date (tabel orders_dataset), dan order_estimated_delivery_date (tabel orders_dataset) yang bertipe VARCHAR menjadi TIMESTAMP WITHOUT TIME ZONE menggunakan query ALTER TABLE nama_tabel ALTER COLUMN nama_kolom TYPE tipe_data_tujuan USING nama_kolom::tipe_data_tujuan,
 - setelah itu membuang sebuah kolom bernama unnamed (tabel product_dataset) dengan query ALTER TABLE nama_tabel DROP COLUMN nama_kolom,
 - kemudian membuat primary key customer_id (tabel customer_dataset), order_id (tabel orders_dataset), product_id (tabel product_dataset), seller_id (tabel sellers_dataset), dan foreign key order_id (tabel order_items_dataset), product_id (tabel order_items_dataset), seller_id (tabel order_items_dataset), order_id (tabel order_payments_dataset), order_id (tabel order_reviews_dataset), customer_id (tabel orders_dataset), menggunakan query ALTER TABLE nama_tabel ADD PRIMARY KEY (nama_kolom) dan ALTER TABLE nama_tabel_1 ADD FOREIGN KEY (nama_kolom_1) REFERENCES nama_tabel_2(nama_kolom_2), dan foreign key customer_zip_code_prefix (tabel customers_dataset), dan seller_zip_code_prefix (tabel sellers_dataset) pada ERD, dan
 - terakhir adalah memunculkan ERD

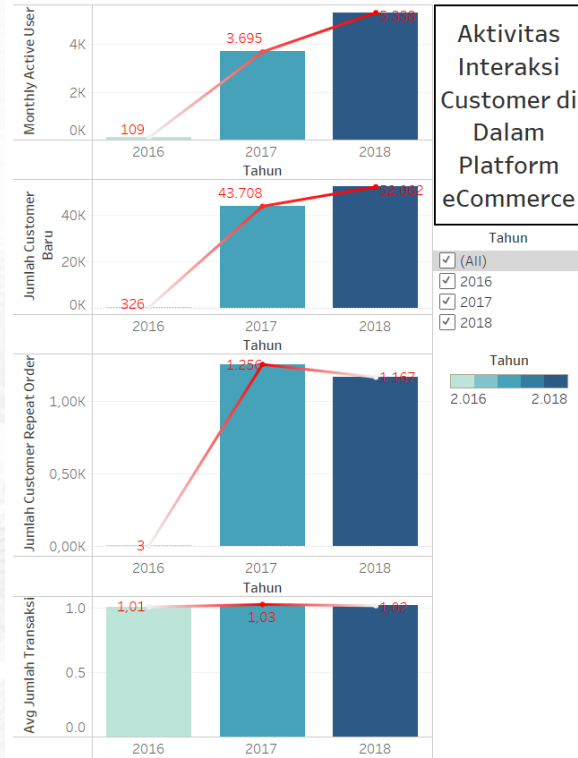
Annual Customer Activity Growth Analysis

- Master Table

	 tahun numeric 	monthly_active_user numeric 	jumlah_customer_baru bigint 	jumlah_customer_repeat_order bigint 	avg_jumlah_transaksi numeric 
1	2016	108.67	326	3	1.01
2	2017	3694.83	43708	1256	1.03
3	2018	5338.20	52062	1167	1.02

Annual Customer Activity Growth Analysis

● Visualisasi Data



Pada Monthly Active User dan Jumlah Customer Baru, tahun berbanding lurus dengan jumlah. Artinya, semakin bertambah tahun, jumlah customer semakin bertambah.

- Monthly Active User Meningkat sebanyak 3.290% atau 30 kali lipat dari tahun 2016 sampai 2017 dan kembali meningkat sebanyak 44% dari 2017 sampai tahun 2018.

- Jumlah Customer Baru Meningkat sebanyak 13.307% atau 133 kali lipat hingga tahun 2017, dan kembali meningkat sebanyak 19% hingga tahun 2018.

Pada Jumlah Customer Repeat Order, mengalami kenaikan pada tahun 2016 sampai 2017 sebanyak 41.567% atau sekitar 416 kali lipat, dan menurun ketika tahun 2018 sebanyak 6,64%.

Rata-rata Jumlah Transaksi menunjukkan nilai cenderung konstan, yaitu 1 transaksi pertahun.

Berdasarkan hasil Analisa grafik Annual Customer Activity Growth, didapatkan kesimpulan bahwa pertumbuhan pesat dalam hal jumlah pelanggan terjadi pada tahun 2016 – 2017, namun cenderung tidak bisa mempertahankan kesetiaan. Hal ini dapat dilihat dari menurunnya Customer Repeat Order pada tahun 2018, dan Avg Jumlah Transaksi yang cenderung konstan meskipun dengan penambahan jumlah customer yang besar.

Query selengkapnya dapat dilihat [disini](#)

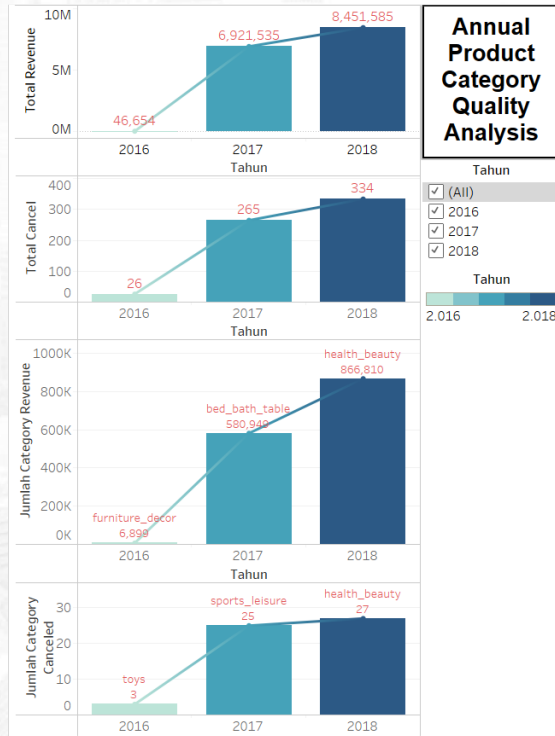
Annual Product Category Quality Analysis

- Master Table

	tahun numeric	total_revenue numeric	total_cancel bigint	top_category_revenue character varying (50)	jumlah_category_revenue numeric	top_category_canceled character varying (50)	jumlah_category_canceled bigint
1	2016	46653.74	26	furniture_decor	6899.35	toys	3
2	2017	6921535.24	265	bed_bath_table	580949	sports_leisure	25
3	2018	8451584.77	334	health_beauty	866810	health_beauty	27

Annual Product Category Quality Analysis

- Visualisasi Data



Pada revenue, jumlah cancel, revenue pada kategori terlaris, dan jumlah cancel pada kategori paling banyak cancel menunjukkan grafik yang meningkat. Jika dikaitkan dengan hasil analisa pertumbuhan customer, dapat disimpulkan bahwa semakin tinggi aktivitas customer maka akan menyebabkan meningkatnya aktivitas transaksi, baik transaksi yang berhasil dan menambah revenue, maupun transaksi yang gagal terjadi (cancel).

Pertumbuhan revenue 2016 ke 2017 14.736% dan 2017 ke 2018 22%.
Pertumbuhan transaksi yang dicancel dari 2016 ke 2017 919% dan 26% dari 2017 ke 2018.

Kategori yang paling banyak berkontribusi pada pertambahan revenue adalah furniture_decor dengan total 6.899.000 (15% dari total revenue 2016) pada tahun 2016, 2017 adalah bad_bath_table sebanyak 580.949.000 (8% dari total revenue 2017), 8.320% lebih banyak dari 2016, dan pada 2018 health_beauty sebanyak 866.810.000 (10% dari total revenue 2018), 49% lebih banyak dari 2017.

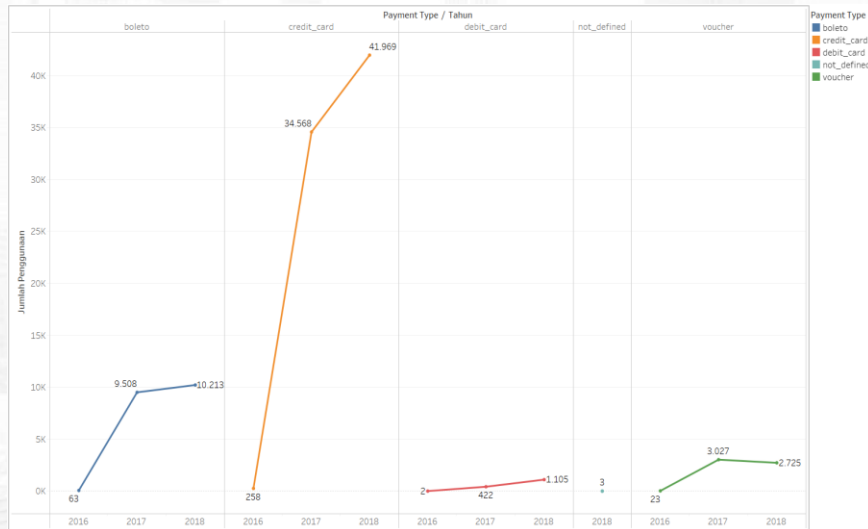
Kategori dengan jumlah cancel terbanyak adalah toys sebanyak 3 kali (12% dari total cancel) pada tahun 2016, sports_leisure sebanyak 25 (9% dari total cancel), lebih banyak 733% pada tahun 2017 dari 2016, health_beauty 27(8% dari total cancel), 8% lebih banyak pada tahun 2018.

Analysis of Annual Payment Type Usage

- Master Table

	tahun numeric	jumlah_boleto numeric	jumlah_credit_card numeric	jumlah_debit_card numeric	jumlah_not_defined numeric	jumlah_voucher numeric
1	2016	63	258	2	0	23
2	2017	9508	34568	422	0	3027
3	2018	10213	41969	1105	3	2725

- Visualisasi Data



Secara umum mayoritas, tipe pembayaran mendapatkan kenaikan setiap tahun, kecuali voucher yang menurun pada 2018 dari tahun sebelumnya.

Kenaikan signifikan terlihat pada tahun 2017 di seluruh tipe pembayaran, kecuali pada pembayaran yang tidak terdefinisi yang tidak terdapat data.

Pembayaran boleto meningkat sebanyak 14.992% pada 2017 dan 7% pada 2018.

Pembayaran credit_card meningkat 13.298% tahun 2017 dan 21% tahun 2018.

Pembayaran debit_card meningkat dari tahun 2016 sebanyak 21000% sampai tahun 2017 dan meningkat lagi sebanyak 162% sampai 2018.

Pembayaran voucher meningkat 13.061% sampai tahun 2017 dan menurun 10% pada tahun 2018.

Tipe pembayaran yang popularitasnya paling signifikan bertumbuh adalah penggunaan debit_card yang terus bertumbuh pesat mulai tahun 2016. Sedangkan yang paling lambat adalah voucher, sehingga perlu dilakukan Analisa lanjutan untuk memperbaiki penyebabnya.

Sedangkan tipe pembayaran yang popularitasnya paling besar adalah credit_card, dan yang paling tidak populer adalah debit_card.