

Analyzing eCommerce Business Performance with SQL

Created by Ammar Khaq Baasir



About Me

Saya merupakan seorang freshgraduate dari universitas muhammadiyah malang program studi teknik elektro. saya berhasil menyelesaikan pendidikan S1 saya dengan meraih penghargaan sebagai lulusan terbaik ke-1 periode 1 2024 dengan raihan IPK 3,76. sebagai seorang freshgraduate yang memiliki ketertarikan berkarir sebagai seorang data scientist maupun data analyst. saya memiliki beberapa kemampuan seperti pemahaman tentang statistik dan matematika, python programming, SQL, penerapan algorithms machine learning maupun deep learning.



Table of contents

01

**About
Project**

02

**Advantages
of Analyzing**

03

**Demo &
Insight
Business**





01

About Project

Business Problem

Sebuah perusahaan e-Commerce telah beroperasi selama beberapa tahun, menjual berbagai kategori produk seperti parfume, housewares, aksesoris elektronik dan lain sebagainya. Namun, selama beberapa bulan terakhir, mereka mulai mengalami penurunan dalam penjualan dan pertumbuhan yang lambat. Awalnya, perusahaan menyadari bahwa mereka kehilangan pangsa pasar kepada pesaing mereka dan mendapati bahwa pertumbuhan customers daily mereka telah menurun. Selain itu, mereka juga melihat bahwa tingkat konversi dari pengunjung situs web menjadi pelanggan aktif telah menurun secara signifikan.

Para pemimpin perusahaan memutuskan untuk menginvestasikan waktu dan sumber daya untuk memahami lebih lanjut penyebab di balik masalah ini. Mereka memulai dengan mengumpulkan data dari berbagai sumber, termasuk analisis customers, data transaksi, payments method, dan lain – lain.



• Analyzing eCommerce Business Performance with SQL •

Analyzing eCommerce Business Performance merupakan proses analisis, evaluasi dan penelitian yang dilakukan untuk memahami seberapa baik sebuah bisnis e-Commerce berkinerja. Proses ini melibatkan pengumpulan, analisis, dan interpretasi data terkait dengan berbagai metrik dan factor yang mempengaruhi kinerja bisnis online, seperti penjualan, daily active user, keuntungan, biaya operasional, dan lain-lain.

Analyzing eCommerce Business Performance sering melibatkan penggunaan berbagai alat dan metode, termasuk analisis data, model prediktif, analisis tren, survei pelanggan, dan lain – lain. Hasil dari analisis ini dapat digunakan untuk membuat rencana tindakan yang lebih baik, memprioritaskan inisiatif bisnis, dan mengukur kesuksesan dari strategi yang diimplementasikan.



SQL & Why Use SQL?

SQL (Structured Query Language) adalah bahasa pemrograman yang digunakan untuk mengelola dan mengakses database relasional. SQL digunakan untuk melakukan berbagai operasi pada database, seperti pengambilan data (query), pembaruan data, penghapusan data, serta pengelolaan struktur database seperti membuat tabel, mengubah tabel, dan menghapus tabel.

Terdapat beberapa alasan mengapa menggunakan SQL :

1. **Fleksibilitas** : SQL memungkinkan pengguna melakukan berbagai operasi pada database, mulai dari mengambil data yang spesifik hingga mengubah struktur database.
2. **Performa** : Dengan pengoptimalan yang tepat, SQL dapat memberikan kinerja yang baik pada query yang kompleks. DBMS modern memiliki optimasi query yang canggih untuk memastikan kinerja yang efisien.
3. **Keamanan** : SQL menyediakan mekanisme otentikasi dan izin akses yang memungkinkan administrator database untuk mengontrol siapa yang dapat mengakses data dan operasi apa yang diizinkan.



Seaborn & Matplotlib

Seaborn dan Matplotlib merupakan dua library populer dalam bahasa pemrograman python yang digunakan untuk visualisasi data.

Matplotlib adalah library dasar yang kuat untuk membuat plot dalam Python, sedangkan Seaborn merupakan library perpustakaan yang dibangun di atasnya yang menyediakan antarmuka yang lebih mudah digunakan dan plot yang lebih menarik secara estetik untuk visualisasi data statistik. Seringkali, seaborn digunakan untuk membuat plot yang lebih kompleks dan estetik, sementara Matplotlib digunakan untuk penyesuaian plot yang sangat spesifik.



CSV & Why Use CSV?

CSV (Comma-Separated Values) merupakan format file yang digunakan untuk menyimpan data dalam bentuk table, dimana setiap baris data dipisahkan oleh karakter koma (,) dan setiap kolom data dipisahkan oleh karakter pemisah, yang biasanya adalah titik koma (;) atau tab. Format CSV sering digunakan untuk pertukaran data antara aplikasi yang berbeda.

Beikut merupakan beberapa alasan menggunakan CSV :

1. **Kemudahan Penggunaan** : CSV adalah format yang sederhana dan mudah dipahami. Karena itu, hamper semua aplikasi database mendukung import dan ekspor data dalam format CSV.
2. **Ringan** : File CSV memiliki ukuran yang relative kecil karena tidak menyimpan format styling atau struktur yang kompleks seperti pada format dokumen lainnya.
3. **Fleksibilitas** : CSV dapat digunakan untuk menyimpan berbagai jenis data, termasuk teks, angka, tanggal, dan format data lainnya. Ini memungkinkan CSV dalam berbagai konteks, mulai dari menyimpan daftar kontak hingga data penjualan.



Environment & Tools For The Project

Language :



Tools :



Library :

matplotlib



seaborn



02

Advantages of Analyzing

Advantages of Analyzing

Analisis merupakan proses sistematis untuk memahami, menguraikan, memeriksa, dan menafsirkan informasi atau data untuk mendapatkan wawasan yang berguna atau mengambil keputusan yang tepat. Dalam berbagai konteks, analisis dapat merujuk pada berbagai metode atau Teknik yang digunakan untuk memeriksa dan memahami data.

Analisis data memberikan banyak keuntungan bagi bisnis, diantaranya :

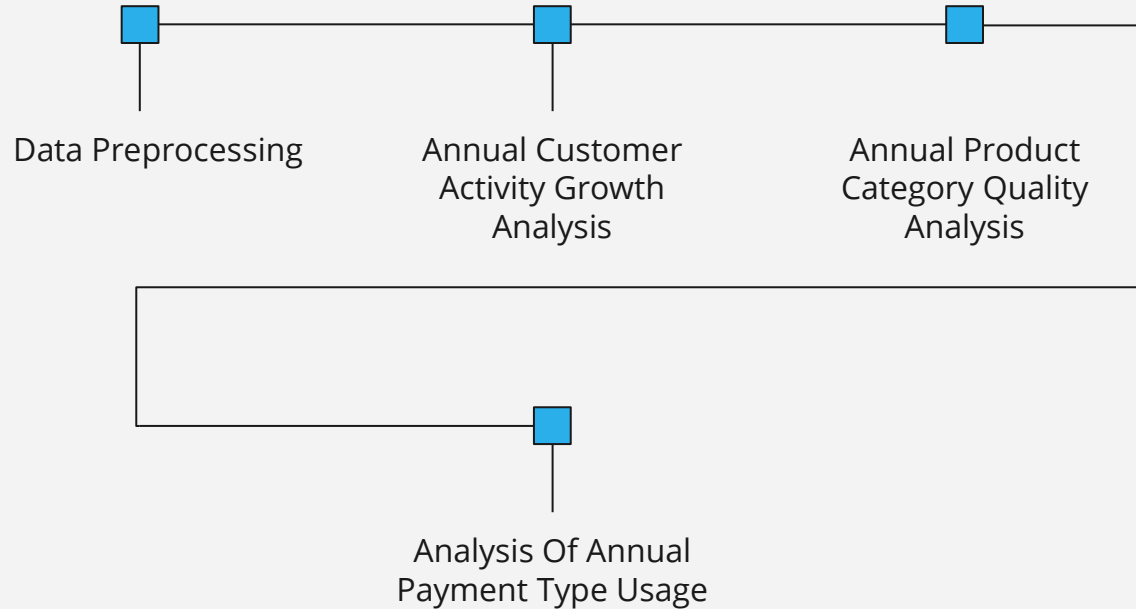
1. **Pengambilan Keputusan yang lebih baik** : Analisis data membantu dalam memahami tren, pola, dan perilaku pelanggan serta pasar. Dengan pemahaman ini, bisnis dapat membuat keputusan yang lebih baik dan lebih tepat, seperti dalam perencanaan produk, strategi pemasaran, dan pengelolaan rantai pasokan.
2. **Pengenalan Peluang Baru** : Analisis data membantu dalam mengidentifikasi peluang bisnis baru. Dengan memahami data pasar dan perilaku pelanggan, bisnis dapat menemukan segmen pasar yang baru, produk atau layanan yang diminati, dan cara untuk memperluas basis pelanggan.
3. **Peningkatan Pengalaman Customer** : Dengan memahami data pengguna, baik dari interaksi langsung maupun feedback, ini dapat mengarah pada desain produk yang lebih intuitif, situs web yang lebih responsif, dan layanan pelanggan yang lebih baik.
4. **Personalisasi Pengalaman Customer** : Analisis data memungkinkan bisnis untuk memahami preferensi dan kebutuhan individual pelanggan. Dengan demikian, mereka dapat menyajikan pengalaman yang lebih personal dan relevan, yang meningkatkan loyalitas pelanggan dan retensi.



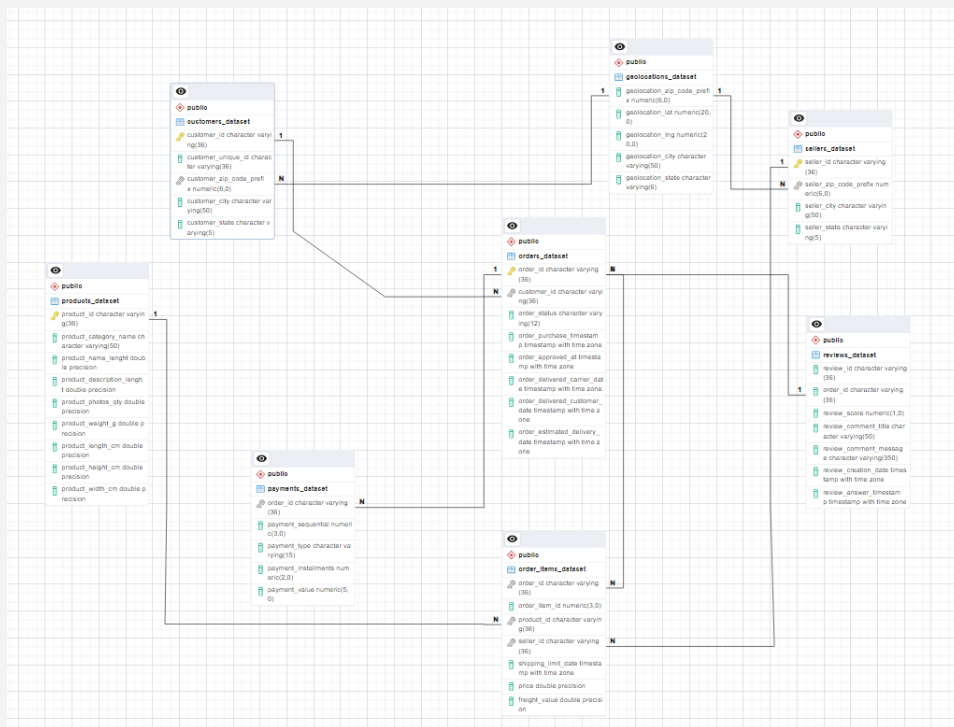
03

Demo & Insight Business

Demo Project Timeline



01. Data Preprocessing

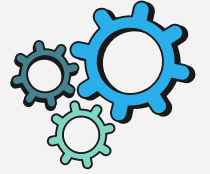


Data Insight :

1. Persebaran **Gender Customers** yang melakukan **clicked on ads** maupun **tidak**, cenderung **memiliki proporsi yang hampir sama**. Ini menandakan **tidak adanya dominasi gender** dalam feature male, feature ini masuk dalam kriteria **feature yang dapat digunakan sebagai feature prediksi**. Akan tetapi perlu dilakukan analisis lebih lanjut mengenai feature male ini, apakah memiliki pengaruh yang kuat terhadap kemungkinan customers melakukan clicked on ads.
2. Kolom target (**Clicked on Ad**) memiliki **proporsi** atau **jumlah customers yang sama** antara **customers "yes"** atau **"no"** dalam **clicked on ads**. Ini menandakan bahwa data yang digunakan merupakan data yang memiliki jumlah **class seimbang** atau **balance**, data yang **balance** akan memudahkan **machine learning** mempelajari semua feature yang diberikan pada **saat pembelajaran** dilakukan, hal ini karena **machine learning** akan **bersifat adil** dan **tidak fokus** hanya pada **class** yang memiliki jumlah lebih banyak.

Akses File Query [Disini](#)

01. Data Preprocessing

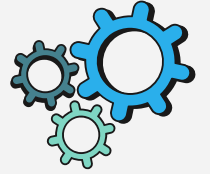


```
WITH master_table AS(
  SELECT
    database_customers.customer_id,
    database_customers.customer_unique_id,
    database_order.order_id,
    database_order.order_status,
    database_order.order_purchase_timestamp AS tanggal_order,
    database_order_item.order_item_id AS jumlah_order,
    database_order_item.price,
    database_order_item.freight_value,
    database_product.product_id,
    database_product.product_category_name
  FROM
    public.customers_dataset AS database_customers
  JOIN
    public.orders_dataset AS database_order
  ON
    database_order.customer_id = database_customers.customer_id
  JOIN
    public.order_items_dataset AS database_order_item
  ON
    database_order_item.order_id = database_order.order_id
  JOIN
    public.products_dataset AS database_product
  ON
    database_product.product_id = database_order_item.product_id
),
```

Gambar disamping merupakan Query Table Master dengan menggunakan CTE. CTE (Common Table Expression) merupakan sebuah query dalam bahasa SQL yang memungkinkan pembuatan temporary result set yang dapat digunakan dalam query yang lebih kompleks. CTE membantu dalam menyederhanakan dan memperjelas sintaks SQL, serta memungkinkan penulisan query yang lebih terstruktur dan mudah dimengerti. CTE memiliki beberapa kelebihan dalam penggunaannya yakni :

1. Kode Mudah Dibaca.
2. Kemudahan Dalam Pengelompokan Data.
3. Penggunaan Kode Berulang – Ulang.
4. Optimasi Kinerja

02. Annual Customer Activity Growth Analysis

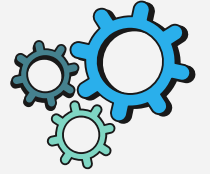


	tahun numeric	monthly_active_user bigint	new_customers bigint	total_repeat_order bigint	average_freq_order numeric
1	2016	312	309	3	1.265
2	2017	44579	43225	1354	1.196
3	2018	53775	52565	1210	1.199

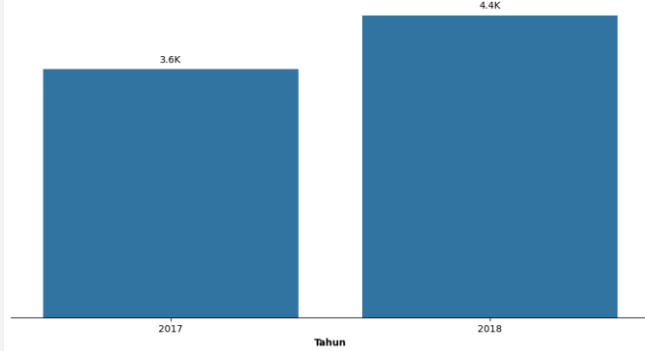
Gambar disamping merupakan Hasil Query Annual Customer Activity Growth Analysis. Ini merupakan sebuah analisis data yang digunakan untuk mencari informasi aktivitas pertumbuhan pelanggan dari tahun ke tahun dalam suatu periode tertentu. Analisis ini memiliki tujuan untuk memahami trend pertumbuhan pelanggan, pola perilaku order dan faktor – faktor lainnya. Analisis yang dilakukan dalam kasus ini adalah melihat beberapa point – point penting yakni :

1. Rata – Rata Monthly Active User (MAU) per tahun.
2. Total Customers baru per tahun.
3. Jumlah Customers yang melakukan repeat order per tahun.
4. Rata – Rata Frequency order untuk setiap tahun.

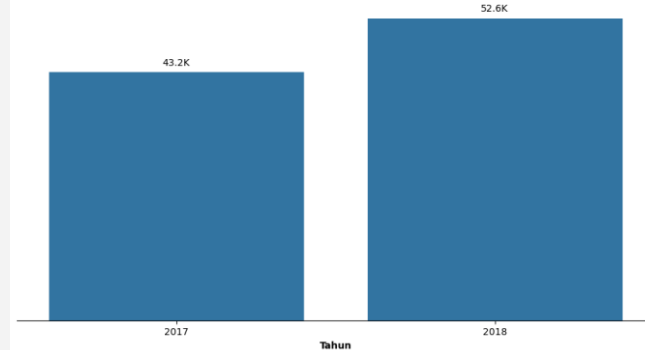
02. Annual Customer Activity Growth Analysis



Average Monthly Active User (MAU)



Total New Customer

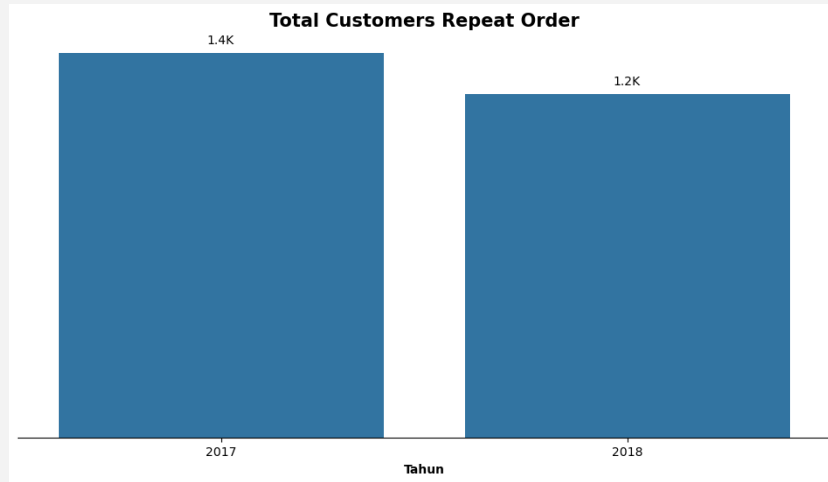
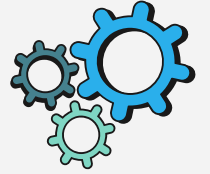


INSIGHT BISNIS

1. Dikarenakan data yang digunakan pada tahun 2016 dimulai pada **september 2016** dan menyebabkan **hasil analisis yang sangat besar perbedaannya** dengan tahun 2017 dan 2018 maka **data yang digunakan pada analisis ini hanya tahun 2017 dan 2018**.
2. Jumlah pengguna aktif (Monthly Active User) dan Total New Customer pada setiap tahun **mengalami peningkatan**. ini menandakan bahwa layanan ecommerce kami memiliki pertumbuhan market yang cukup besar
3. Dari analisis yang telah dilakukan dalam jangka waktu 2 tahun terakhir Ecommerce mampu mendapatkan **total users baru sebanyak 95.790 User**. Ini merupakan informasi yang sangat baik, dimana company memiliki prospek pertumbuhan yang sangat baik, dimana perusahaan dapat menarik minat para customers untuk mencoba menggunakan layanan dari ecommerce kami.

Akses File Query [Disini](#)

02. Annual Customer Activity Growth Analysis



INSIGHT BISNIS

1. Dalam 2 tahun terakhir, terjadi penurunan jumlah customer repeat order. Dimana pada tahun 2018 jumlah customer yang melakukan repeat order berkurang sebesar +- 1000 customer. Ini merupakan sebuah informasi yang buruk, dimana customer cenderung tidak memiliki keinginan untuk melakukan pembelian berulang kali pada layanan ecommerce kami. Kasus ini perlu dilakukan analisis lebih lanjut mengenai apa penyebab utama customer pada tahun 2018 tidak melakukan repeat order. Apabila permasalahan ini tidak segera ditangani maka dikhawatirkan penurunan customer repeat order akan mengalami peningkatan pada setiap tahunnya dan ini akan memberikan dampak cukup signifikan terhadap pendapatan perusahaan.

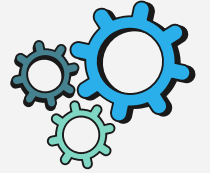
03. Annual Product Category Quality Analysis




	tahun numeric	product_category_revenue character varying (50)	total_revenue_product numeric	total_revenue_year numeric	product_category_canceled character varying (50)	total_canceled_product numeric	total_canceled_year numeric
1	2016	furniture_decor	6899.35	46653.74	health_beauty	3	17
2	2016	furniture_decor	6899.35	46653.74	auto	3	17
3	2016	furniture_decor	6899.35	46653.74	toys	3	17
4	2017	bed_bath_table	580949.20	6921535.24	housewares	43	343
5	2018	health_beauty	866810.34	8451584.77	furniture_decor	38	344

Business Insight :

1. **Annual Product Category Quality Analysis** merupakan proses evaluasi dan analisis kualitas produk dalam kategori tertentu selama periode tahunan. Analisis ini bertujuan untuk memahami product category apa saja yang memiliki performa terbaik dan terburuk dalam kurun waktu tahunan. Informasi dari tabel atau analisis ini nantinya dapat digunakan sebagai rujukan penyusunan strategi oleh layanan ecommerce kami, dimana tentunya strategi ini bertujuan untuk meningkatkan revenue perusahaan.
2. Untuk **product category** yang memiliki **revenue tertinggi** dalam periode tahunan, ini **dapat dijadikan** sebagai **produk unggulan ecommerce**, dimana perusahaan dapat melakukan **pemberian promo – promo** seperti pemberian diskon, pengurangan biaya ongkir, dan lain – lain untuk **dapat meningkatkan revenue perusahaan** lebih besar lagi pada tahun – tahun berikutnya.
3. Untuk **product category canceled order** terbesar dalam periode tahunan, maka perlu dilakukan **evaluasi ataupun kebijakan tertentu** yang dapat mengurangi total canceled order suatu product. Perusahaan perlu memberikan **pengawasan ketat** terhadap product tersebut, karena kita ingin para customer dapat **memiliki experience yang baik** ketika melakukan order melalui layanan ecommerce kami. Tentunya apabila customer selalu **mendapatkan experience** yang baik, maka ini dapat **meningkatkan peluang** suatu **customer melakuakn repeat order** diwaktu berikutnya.

03. Annual Product Category Quality Analysis

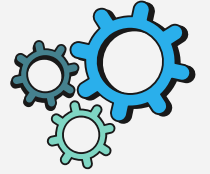





Tahun	Top Product	Total Revenue
2016	 Furniture Decor	\$6899.35
2017	 Bed Bath Table	\$580949.20
2018	 Health Beauty	\$866810.34

Business Insight :

1. Dalam kurun waktu 3 tahun terakhir total revenue Perusahaan berdasarkan product category mengalami trend positif yakni mengalami peningkatan revenue setiap tahunnya.
2. Top product category yang memberikan revenue paling banyak pada perusahaan setiap tahunnya selalu berubah – ubah. Dimana pada 2 tahun yakni 2016 dan 2017 product yang paling top berupa produk benda mati atau benda fisik. Sedangkan pada 2018 product yang paling top merupakan product obat – obatan, terutama obat kecantikan.

03. Annual Product Category Quality Analysis

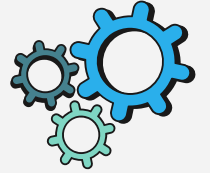


Tahun	Top Product	Total Canceled
2016	 Health Beauty, Auto, Toys	3
2017	 Housewares	43
2018	 Furniture Decor	38

Business Insight :

1. Dalam kurun waktu 3 tahun terakhir total canceled order berdasarkan product category mengalami peningkatan setiap tahunnya.
2. Top product category yang mendapatkan canceled order paling banyak setiap tahunnya selalu berubah - ubah. Terjadi peningkatan total canceled order yang cukup signifikan dari tahun 2016 ke tahun 2017 dan pada tahun berikutnya 2018 mengalami sedikit penurunan. Hal ini perlu dilakukan analisis lebih lanjut untuk mengetahui kenapa total canceled order mengalami peningkatan khususnya pada tahun 2017.

04. Analysis of Annual Payment Type Usage

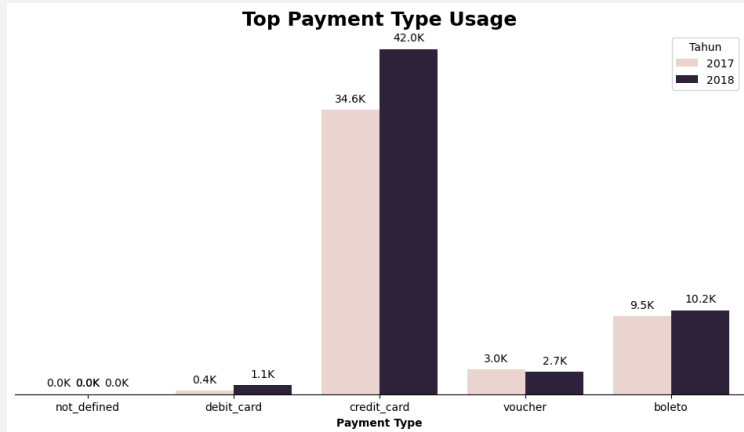


Business Insight

Pada analysis of Annual Payment Type Usage akan dicari sebuah informasi mengenai top payment type apa yang paling banyak digunakan customer pada setiap tahunnya. Informasi yang didapatkan dari analisis ini dapat digunakan sebagai sebuah strategi yang dapat meningkatkan jumlah transaksi maupun revenue perusahaan. Dengan informasi top payment type kita bisa melakukan treatment pada customer dengan cara memberikan promo – promo berupa diskon belanja apabila menggunakan type pembayaran tertentu, pastinya ini akan meningkatkan keinginan customer untuk dapat bertransaksi di layanan ecommerce kami.

	payment_type character varying (15) 🔒	tahun_2016 numeric 🔒	tahun_2017 numeric 🔒	tahun_2018 numeric 🔒
1	not_defined	0	0	3
2	debit_card	2	422	1105
3	credit_card	258	34568	41969
4	voucher	23	3027	2725
5	boleto	63	9508	10213

04. Analysis of Annual Payment Type Usage



Business Insight :

1. Dikarenakan data yang digunakan pada tahun 2016 dimulai pada **september 2016** dan menyebabkan **hasil analisis yang sangat besar perbedaannya** dengan tahun 2017 dan 2018 maka **data yang digunakan pada analisis ini hanya tahun 2017 dan 2018**.
2. Pada 2 tahun terakhir, customer yang **melakukan transaksi** pada layanan ecommerce kami paling banyak menggunakan **payment type credit card**. Pembayaran dengan **credit card** selalu **menjadi pilihan utama** oleh customer dari pada payment type jenis lainnya. Penggunaan credit card oleh customer **mengalami peningkatan sebesar 20%** dari tahun 2017 ke tahun 2018.
3. Terdapat hal unik pada **payment type debit card**, Dimana tipe ini memiliki **peningkatan jumlah penggunaan sebesar 175%** dalam kurun waktu 1 tahun. Hal ini mungkin terjadi karena beberapa hal seperti terdapat pemberian promo pada customer atau pelaksanaan kerja sama dengan pihak debit card yang membuat customer menjadi menggunakan tipe ini dalam transaksi mereka di ecommerce kami. Akan tetapi, hal ini perlu **dilakukan analisis lebih lanjut** untuk memastikan kenapa hal ini bisa terjadi.



Thanks!

Do you have any questions?

ammarbaasir19@gmail.com

<https://www.linkedin.com/in/ammar-baasir-973873234/>