

Nama : Yulia Khairunnisa

SC ID : DS5B03

## 1. SQL

- Melihat dan menghitung asal negara *customer*

```
select country, count(country) from bigquery-public-data.thelook_ecommerce.users group by country
```

Pemilik *ecommerce* perlu untuk mengetahui negara apa saja dan ada berapa banyak *customer* di negara tersebut untuk melihat perkembangan *ecommerce*-nya di negara tersebut. Hal ini dapat membandingkan kinerja penjualan dan profitabilitas dari berbagai negara atau wilayah. Melihat dan menghitung asal negara ini dapat membantu mengidentifikasi peluang pertumbuhan dan mengevaluasi efektivitas strategi bisnis di setiap pasar. Berikut ini hasil yang ditampilkan:

country	count
Brasil	14552
Japan	2495
United States	22546
China	33985
South Korea	5318
Spain	4044

- Menampilkan barang yang dibeli oleh *customer*

```
1 select s.id, first_name, last_name, city, a.name from bigquery-public-data.thelook_ecommerce.users s
2 | inner join bigquery-public-data.thelook_ecommerce.products a on s.id=a.id
```

Dengan memiliki data tentang nama pelanggan dan barang yang dibeli, dapat memberikan pelayanan yang lebih personal. Ini dapat membantu merespons pertanyaan pelanggan dengan lebih baik dan menangani masalah atau pengembalian dengan lebih efisien. Selain juga dapat menganalisis riwayat pembelian pelanggan untuk memahami pola pembelian *customer*. Ini dapat membantu mengidentifikasi produk yang paling populer dan mengembangkan penjualan. Berikut ini hasil yang ditampilkan:

Row	id	first_name	last_name	city	name
1	15514	Phillip	Werner	Rio Branco	Bali Women's Double Support ...
2	16893	Eric	Cox	Rio Branco	Canvas Unisex 4.2 oz. Jersey D...
3	20778	Sara	Reeves	Rio Branco	Wrangler Men's Rugged Wear S...
4	3478	Robert	Alexander	Sena Madureira	Allegra K Ladies Scoop Neck Sleeveless Elastic Waist Mini Dress Multicolor XS
5	9094	Christina	Nash	Sena Madureira	Smartwool Women's PhD Outdoor Light Crew Socks - Aegean Blue M

- Menampilkan berapa banyak barang yang direturn dan dicancel

```
SELECT
p.name,
SUM(CASE WHEN oi.status = 'Cancelled' THEN 1 ELSE null END) AS Cancelled,
SUM(CASE WHEN oi.status = 'Returned' THEN 1 ELSE null END) AS Returned
FROM `bigquery-public-data.thelook_ecommerce.order_items` oi
LEFT JOIN `bigquery-public-data.thelook_ecommerce.products` p
ON oi.product_id = p.id
GROUP BY 1
ORDER BY 2 DESC;
```

Menampilkan barang yang direturn dan dicancel dapat digunakan untuk menganalisis pola dan memahami alasan umum mengapa pelanggan mengembalikan atau membatalkan produk. Analisis tersebut dapat membantu dalam perbaikan produk atau kebijakan lainnya. Berikut ini hasil yang ditampilkan:

name ▼	Cancelled ▼	Returned ▼
Puma Men's Socks	12	5
Fruit of the Loom Women's 6-P...	9	1
True Religion Men's Ricky Strai...	9	5
Wrangler Men's Premium Perfo...	8	8
Smith Optics Serpico Sunglass...	8	2

- Menampilkan jumlah pesanan *customer* dan status pengirimannya

```
select t1.first_name, last_name, t2.status, t3.num_of_item
from bigquery-public-data.thelook_ecommerce.users as t1
inner join
bigquery-public-data.thelook_ecommerce.order_items as t2 on t1.id=t2.id
inner join
bigquery-public-data.thelook_ecommerce.orders as t3 on t2.order_id=t3.order_id
```

Dengan melihat jumlah pembelian dan status pengirimannya dapat memahami sejauh mana seorang pelanggan aktif dalam berbelanja sekaligus dapat memberikan pelanggan informasi real-time tentang status pesanan mereka, termasuk apakah pesanan telah dikirim, dalam proses pengemasan, atau menunggu persetujuan. Selain itu, menggunakan data jumlah pembelian juga dapat mengidentifikasi produk yang paling populer di antara pelanggan. Ini dapat membantu dalam perencanaan inventaris dan pengembangan produk lebih lanjut. Berikut ini hasil yang ditampilkan:

first_name ▼	last_name ▼	status ▼	num_of_item ▼
Jose	Zuniga	Shipped	1
Michael	Bartlett	Shipped	1
Terri	Calderon	Complete	1
Bethany	Blake	Complete	1
Luis	Hall	Complete	1

- Menampilkan *traffic source* dari *customer* untuk mengunjungi *ecommerce*

```
SELECT
u.traffic_source,
COUNT(DISTINCT oi.user_id) total_customer
FROM `bigquery-public-data.thelook_ecommerce.order_items` oi
LEFT JOIN `bigquery-public-data.thelook_ecommerce.users` u
ON oi.user_id = u.id
WHERE oi.status NOT IN ('Cancelled', 'Returned')
GROUP BY 1
ORDER BY 2 DESC
```

Menampilkan *traffic source* terbanyak dari *customer* untuk mengunjungi *ecommerce* berfungsi untuk mengidentifikasi dari mana pengunjung datang ke web *ecommerce* tersebut. Ini memungkinkan untuk memahami apakah pengunjung berasal dari pencarian organik, tautan referensi, media sosial, iklan berbayar, atau sumber lainnya. Dengan informasi ini, dapat mengetahui *traffic source* yang paling efektif untuk mengoptimalkan dalam hal strategi pemasaran. Berikut ini hasil yang ditampilkan:

traffic_source ▼	total_customer ▼
Search	46425
Organic	10064
Facebook	3980
Email	3208
Display	2653

- Menampilkan barang yang belum terjual

```
select product_name, sold_at
from bigquery-public-data.thelook_ecommerce.inventory_items
where
sold_at is null
```

Menampilkan barang yang belum terjual dapat digunakan untuk mengidentifikasi barang yang mungkin memiliki masalah dalam hal desain, harga, atau daya tarik. Ini memberi kesempatan untuk mengevaluasi dan memperbaiki produk atau menghentikan penjualan produk yang tidak efektif. Selain itu, menampilkan hal tersebut dapat digunakan dalam merencanakan strategi pemasaran yang lebih efektif. Berikut ini hasil yang ditampilkan:

product_name	sold_at
Quiksilver Waterman Men's On ...	null
Quiksilver Waterman Men's On ...	null
Quiksilver Waterman Men's On ...	null
Quiksilver Waterman Men's On ...	null
Quiksilver Waterman Men's On ...	null
Quiksilver Waterman Men's On ...	null
Quiksilver Waterman Men's On ...	null
Quiksilver Waterman Men's On ...	null

- Menampilkan barang yang memiliki harga paling mahal

```
select * from bigquery-public-data.thelook_ecommerce.products
where
cost = ((select max(cost)
from
bigquery-public-data.thelook_ecommerce.products))
```

Menampilkan barang yang paling mahal dapat membantu mempromosikan produk-produk premium atau mewah yang mungkin menghasilkan margin keuntungan yang lebih tinggi dengan begitu dapat membantu meningkatkan pendapatan *e-commerce*. Berikut ini hasil yang ditampilkan:

id	cost	category	name
24899	557.1510021798...	Socks	Nike Jordan Retro 11 Bred Boo...

- Menampilkan barang yang memiliki harga paling murah

```
select * from bigquery-public-data.thelook_ecommerce.products
where
cost = (select min(cost)
from
bigquery-public-data.thelook_ecommerce.products)
```

Menampilkan barang yang paling murah membantu *customer* dalam melakukan perbandingan harga antara produk yang serupa atau sejenis. Ini memungkinkan mereka untuk memilih produk dengan harga yang sesuai dengan anggaran mereka. Selain itu, Jika pelanggan menemukan produk dengan harga terjangkau dan merasa puas dengan kualitas dan layanan, *customer* cenderung kembali untuk berbelanja lagi di kemudian hari. Berikut ini hasil yang ditampilkan:

id	cost	category	name
14235	0.00829999977...	Accessories	Indestructable Aluminum Alum...

## 2. Python

Berikut ini merupakan link google collab dari jawaban untuk soal python

<https://colab.research.google.com/drive/1yJ9xFNq6P72xXtn8gukSLHDr5pAldr8Q?usp=sharing>

### 3. Statistics

- **Probabilitas Dasar (*Basic Probability*)**

Probabilitas berasal dari kata *probably*, yang memiliki arti kemungkinan. Dalam konsep dasar statistik, probabilitas adalah mengukur peluang atau kemungkinan terjadinya suatu peristiwa. Peluang atau kemungkinan sering disebut dengan kata probabilitas yang secara umum merupakan peluang bahwa sesuatu akan terjadi. Probabilitas yang mengukur sejauh mana suatu peristiwa dapat terjadi atau tidak terjadi, biasanya dinyatakan dalam bentuk angka antara 0 dan 1. Dalam artian, 0 berarti peristiwa tersebut pasti tidak terjadi, dan 1 berarti pasti peristiwa tersebut terjadi. Semakin tinggi nilai probabilitas, semakin besar pula kemungkinan peristiwa tersebut terjadi. Probabilitas tidak hanya dapat dinyatakan pada nilai 0 sampai 1 saja, tetapi juga dalam bentuk persentase.

Konsep dasar probabilitas dapat diibaratkan dengan pengukuran kemungkinan bahwa suatu peristiwa tertentu akan terjadi. Namun, probabilitas menyajikan metode-metode yang berkaitan dengan ketidakpastian, dan berkaitan dengan peristiwa yang terjadi secara kebetulan. Contoh dari hal tersebut, yaitu kesalahan pengukuran, terjadinya kecelakaan, dan produksi cacat atau tidaknya suatu barang. Selain itu, permainan peluang seperti melempar koin, atau bermain kartu juga termasuk ke dalam contoh peristiwa yang terjadi secara kebetulan.

Dalam setiap kasusnya, biasanya memiliki pengetahuan tentang kemungkinan berbagai hasil yang mungkin, tetapi tidak dapat memprediksi dengan pasti hasil dari percobaan tertentu. Konsep dari probabilitas dapat diterapkan dalam berbagai bidang, termasuk ilmu statistik, ilmu komputer, hingga ilmu sosial. Probabilitas dapat digunakan untuk menjelaskan kesalahan pengukuran ilmiah atau kesalahan dalam suatu teknik sehingga dapat segera disadari kesalahannya.

- **Hipotesis 1 dan Hipotesis 2**

Hipotesis 1 dan 2 adalah pernyataan atau proposisi yang digunakan dalam metode ilmiah untuk menguji atau menjelaskan fenomena alam hingga fenomena sosial. Penjelasan dari hipotesis 1 dan hipotesis 2 adalah sebagai berikut:

1. Hipotesis 1

Hipotesis Nol ( $H_0$ ) disebut juga sebagai Hipotesis Statistik. Hipotesis 1 merupakan proposisi yang mengajukan suatu pernyataan atau prediksi yang akan diuji dalam sebuah penelitian atau eksperimen ilmiah. Hipotesis Nol ( $H_0$ ) berisikan pernyataan tentang nilai satu atau lebih parameter yang merupakan status saat ini dan biasanya tidak ditolak kecuali data sampel menyimpulkan dengan kuat bahwa hipotesis ini salah. Hipotesis Nol ( $H_0$ ) biasanya akan bersifat netral dan mengandaikan tidak adanya pengaruh atau hubungan antara variable yang sedang diuji.

2. Hipotesis 2

Hipotesis 2 juga dikenal sebagai hipotesis *alternative* ( $H_1$  atau  $H_a$ ). Hipotesis 2 adalah alternatif lain yang diajukan dalam sebuah penelitian atau eksperimen ilmiah yang merupakan kesimpulan sementara bahwa sudah dilakukan suatu penelitian tindakan dan hubungan antar variable yang sudah dipelajari dari teori-teori yang

berhubungan dengan masalah tersebut. Pada Hipotesis 2 berisikan pernyataan yang bertentangan dengan hipotesis nol. Hal ini menyatakan bahwa ada perbedaan atau efek yang signifikan antara *variable* atau kelompok yang sedang diuji.

- **Perbedaan antara distribusi diskrit dan distribusi kontinu dan contoh masing-masing distribusi, serta pemahaman sederhana tentang uji t**

- Distribusi diskrit adalah distribusi probabilitas yang terdiri dari nilai-nilai terbatas dan terpisah. Artinya, *variable* acak dalam distribusi ini hanya dapat mengambil sejumlah nilai yang terbatas, seperti bilangan bulat. Distribusi diskrit atau data diskrit memiliki data yang nilainya adalah bilangan asli, bukan bilangan berkoma maupun pecahan. Sedangkan, distribusi kontinu atau data kontinu adalah distribusi probabilitas yang terdiri dari *variable* acak yang dapat mengambil nilai dalam rentang yang kontinu, termasuk bilangan desimal.
- Contoh distribusi diskrit:  
Jumlah peserta UTBK tahun 2023. Variabel ini hanya dapat mengambil nilai berupa bilangan bulat positif, seperti 0, 1, 2, dan seterusnya.
- Contoh distribusi kontinu:  
Tinggi badan mahasiswa kelas Statistik. Variabel ini dapat mengambil nilai-nilai seperti 151.5 cm, 167.1 cm, dan lainnya.
- Uji t-tes

Uji t adalah teknik statistik yang digunakan untuk menguji apakah terdapat perbedaan yang signifikan antara dua kelompok atau sampel dalam hal rata-rata atau perbedaan signifikan antara dua kelompok atau populasi. Uji t-tes mengasumsikan bahwa data yang diuji memiliki distribusi normal (atau mendekati normal) dan memiliki varian yang sama. Ini sering digunakan untuk menentukan apakah perbedaan yang diamati antara dua kelompok adalah hasil dari variasi alamiah atau apakah perbedaan tersebut benar-benar signifikan secara statistik.

Uji t ini dapat digunakan dalam berbagai konteks, seperti perbandingan rata-rata gaji antara dua kelompok pekerja, perbandingan rata-rata skor tes sebelum dan setelah intervensi, atau perbandingan efektivitas dua metode pengobatan. Selain itu, uji t juga sering digunakan dalam penelitian ilmiah untuk menguji hipotesis yang telah diajukan. Misalnya, peneliti mungkin menggunakan uji t untuk menguji apakah terdapat perbedaan signifikan dalam tingkat keberhasilan belajar antara kelompok yang menggunakan metode belajar tertentu dengan kelompok yang menggunakan metode belajar lainnya. Dalam melakukan Uji t, peneliti harus mempertimbangkan tingkat signifikansi yang diinginkan. Tingkat signifikansi menunjukkan seberapa besar kemungkinan terjadinya kesalahan dalam menolak hipotesis nol.