

ASSIGNMENT CHAPTER 1

SQL FOUNDATION

Nama : Agnes Monica Puspitaningtyas

Kompi : 6

SC ID : SC60619

Analisis data memiliki peran yang penting untuk mengevaluasi kebijakan bisnis yang akan dijalankan. Tujuan dari analisis data adalah untuk menemukan informasi penting dan relevan dari data untuk mengambil keputusan terbaik dan informatif. Sebagai *data analyst*, kita diharuskan untuk fokus pada tujuan, yaitu membantu perusahaan membuat keputusan berdasarkan data yang ada. Untuk dapat menganalisis data, diperlukan sebuah *tools* untuk membantu pekerjaan seorang *data analyst*, yaitu SQL. Dalam hal ini, saya akan menganalisis sebuah dataset *e-commerce* The Look untuk mengetahui performa penjualan dan identifikasi pelanggan beserta perilakunya dalam melakukan transaksi.

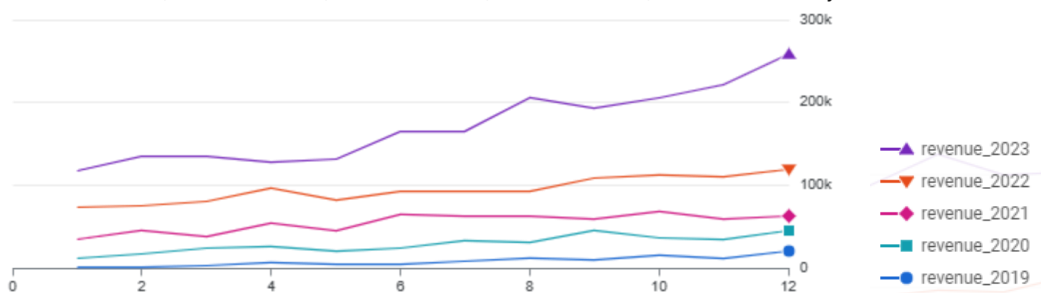
The Look merupakan situs *E-commerce* fiksi yang diciptakan oleh Looker team. Dataset ini memuat data dan tabel berupa:

Nama Tabel	Nama Kolom
users	id, first_name, last_name, email, age, gender, state, street_address, postal_code, city, country, latitude, longitude, traffic source, created_at
order_items	id, order_id, user_id, product_id, inventory_item_id, status, created_at, shipped_at, delivered_at, returned_at, sale_price
distribution_center	id, name, latitude, longitude
inventory_items	id, product_id, created_at, sold_at, cost, product_category, product_name, product_brand, product_retail_price, product_department, product_sku, product_distribution_center_id
products	id, cost, category, name, brand, retail_price, department, sku, distribution_center_id
orders	order_id, user_id, status, gender, created_at, returned_at, shipped_at, delivered_at, num_of_item
events	id, user_id, sequence_number, session_id, created_at, ip_address, city, state, postal_code, browser, traffic_source, url, event_type

Berikut *insight* yang didapatkan:

1. Bagaimana pendapatan *e-commerce* pada setiap tahun?

Revenue_2019, Revenue_2020, Revenue_2021, Revenue_2022, Revenue_2023 by m



- Pendapatan pada tahun dari tahun 2019 hingga 2023 terus mengalami peningkatan. Hal ini menunjukkan hasil yang signifikan perusahaan dalam mengembangkan bisnisnya sehingga mengalami kemajuan yang pesat.
- Pendapatan melonjak pada bulan Juli dan Desember. Hal ini dikarenakan Juli merupakan musim panas di beberapa wilayah sehingga orang cenderung lebih banyak berbelanja. Desember merupakan akhir tahun dimana banyak orang sedang menikmati liburan Natal dan tahun baru sehingga orang cenderung berbelanja pula pada bulan ini. Selain itu, promosi dan diskon menjadi faktor konsumen menjadi lebih sering melakukan transaksi.

Call to action: perusahaan perlu menyesuaikan strategi pemasaran dengan memberikan promosi dan diskon, meningkatkan persediaan barang, dan meningkatkan interaksi dengan pelanggan seperti melakukan *live* agar membangun koneksi yang lebih dekat dengan audiens, serta membuat konten yang bersifat menarik dan persuasif untuk dibagikan ke media sosial.

Query:

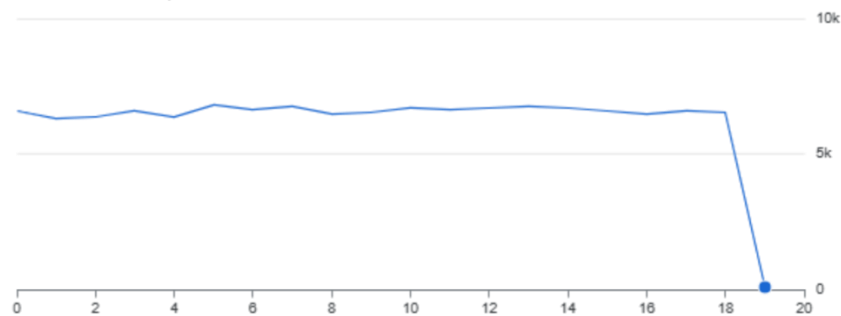
```
SELECT
  EXTRACT(MONTH FROM oi.Created_at) AS months,
  ROUND(SUM(CASE WHEN EXTRACT(YEAR FROM oi.created_at) = 2019 THEN
    oi.sale_price * o.num_of_item ELSE 0 END), 2) AS Revenue_2019,
  ROUND(SUM(CASE WHEN EXTRACT(YEAR FROM oi.created_at) = 2020 THEN
    oi.sale_price * o.num_of_item ELSE 0 END), 2) AS Revenue_2020,
  ROUND(SUM(CASE WHEN EXTRACT(YEAR FROM oi.created_at) = 2021 THEN
    oi.sale_price * o.num_of_item ELSE 0 END), 2) AS Revenue_2021,
  ROUND(SUM(CASE WHEN EXTRACT(YEAR FROM oi.created_at) = 2022 THEN
    oi.sale_price * o.num_of_item ELSE 0 END), 2) AS Revenue_2022,
  ROUND(SUM(CASE WHEN EXTRACT(YEAR FROM oi.created_at) = 2023 THEN
    oi.sale_price * o.num_of_item ELSE 0 END), 2) AS Revenue_2023
FROM bigquery-public-data.thelook_ecommerce.order_items oi
INNER JOIN `bigquery-public-data.thelook_ecommerce.orders` AS o ON
  oi.order_id = o.order_id
WHERE oi.status = 'Complete'
  AND EXTRACT(YEAR FROM oi.Created_at) BETWEEN 2019 AND 2023
GROUP BY months
ORDER BY months
```

Output:

Row	months	Revenue_2019	Revenue_2020	Revenue_2021	Revenue_2022	Revenue_2023
1	1	820.33	11275.97	34473.73	72540.16	116974.52
2	2	1377.85	16826.67	44910.84	75621.35	134878.38
3	3	2651.67	23643.49	37752.9	81114.13	135044.7
4	4	5821.84	24720.82	53270.24	97037.92	127951.58
5	5	5077.66	19414.86	44541.16	82879.09	130966.53
6	6	4307.24	23741.38	64401.89	91940.59	165557.68
7	7	7353.81	31856.47	61859.04	92437.75	165530.46
8	8	12199.55	30278.14	62593.37	91936.47	205159.79
9	9	9066.42	44422.29	58575.2	108770.58	193054.74
10	10	14719.68	36220.56	68326.44	112368.86	204843.62
11	11	11434.68	34505.45	59219.0	110550.17	222196.73
12	12	20658.14	44394.44	62537.35	118392.58	259327.06

2. Kapan waktu pemesanan yang paling sibuk?

Order_Count by Order_Hour



- Waktu pemesanan rata-rata konstan, tetapi pelanggan paling banyak melakukan transaksi pada saat jam 5 pagi sebanyak 6.779 pesanan dan 1 siang sebanyak 6.772 pesanan. Sementara itu, waktu paling sedikit bertransaksi adalah pukul 7 malam.
- Hal dikarenakan kebiasaan pelanggan yang lebih aktif berbelanja pada saat pagi hari sebelum memulai hari kerja atau jam istirahat makan siang. Sementara itu, pada pukul 7 malam sebagian besar orang sedang beristirahat setelah hari kerja sehingga aktivitas pembelian *online* cenderung menurun.

Call to action: Perusahaan dapat mengoptimalkan strategi promosi pada jam-jam puncak, seperti memberikan diskon/promosi dan melakukan *live* pada jam tersebut.

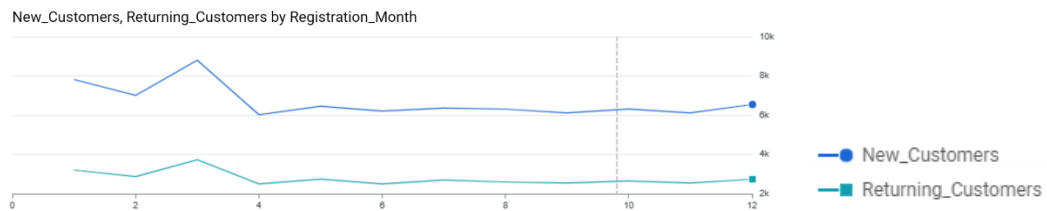
- Query:

```
SELECT
  EXTRACT(HOUR FROM created_at) AS Order_Hour,
  COUNT(order_id) AS Order_Count
FROM bigquery-public-data.thelook_ecommerce.orders
GROUP BY order_hour
ORDER BY order_count DESC
```

- Output:

Row	Order_Hour	Order_Count
1	5	6779
2	13	6772
3	7	6737
4	12	6679
5	14	6676

3. Bagaimana tren pelanggan baru yang melakukan transaksi di bulan yang sama?



- Bulan dengan registrasi pengguna paling banyak berada pada bulan Maret, yaitu sebanyak 7.800 pengguna dan yang melakukan transaksi sebanyak 3.707 pelanggan. Sementara itu, bulan dengan registrasi pengguna paling sedikit yaitu bulan April sebanyak 6.140 pengguna dan yang melakukan transaksi sebanyak 2.497 pelanggan.
- Hal ini dikarenakan pada bulan Maret terdapat promosi secara besar-besaran untuk pengguna yang baru melakukan transaksi. Ketika seluruh kebutuhannya serta hasrat untuk berbelanja sudah terpenuhi pada bulan tersebut, animo pelanggan turun pada bulan setelahnya karena pada bulan April perusahaan tidak lagi melakukan promosi serupa.

Call to action: perusahaan dapat terus memanfaatkan strategi promosi pada bulan Maret untuk menarik lebih banyak pelanggan baru. Namun, perusahaan perlu mengadakan promosi yang berkelanjutan pada bulan setelahnya agar dapat meningkatkan aktivitas dan transaksi pelanggan. Perusahaan perlu membangun komunikasi dan melakukan penawaran yang disesuaikan dengan preferensi pelanggan.

- Query:

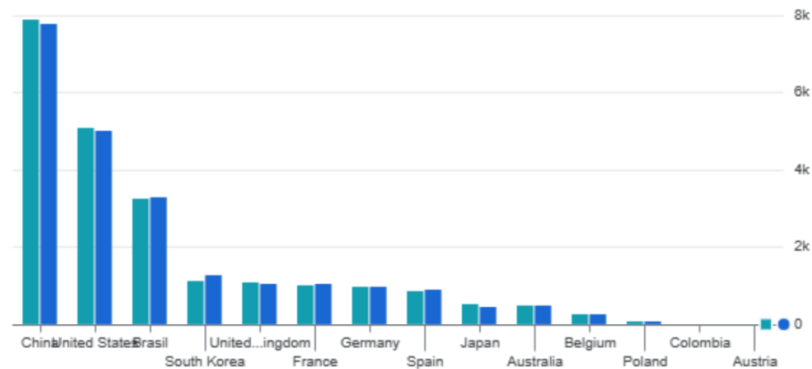
```
SELECT
  EXTRACT(MONTH FROM u.created_at) AS Registration_Month,
  COUNT(DISTINCT u.id) AS New_Customers,
  COUNT(DISTINCT CASE WHEN EXTRACT(MONTH FROM o.created_at) = EXTRACT(MONTH FROM
delivered_at) THEN user_id END) AS Returning_Customers
FROM bigquery-public-data.thelook_ecommerce.orders AS o
INNER JOIN bigquery-public-data.thelook_ecommerce.users AS u
ON o.user_id = u.id
GROUP BY Registration_Month
ORDER BY Registration_Month
```

- Output:

Row	Registration_Month	New_Customers	Returning_Customer
1	1	7800	3186
2	2	7025	2876
3	3	8811	3707
4	4	6025	2497
5	5	6425	2707
6	6	6221	2500
7	7	6335	2662
8	8	6315	2577
9	9	6141	2529
10	10	6299	2619
11	11	6140	2542
12	12	6559	2747

4. Bagaimana profil pelanggan pada setiap negara?

Male, Female by Country



- Pelanggan paling banyak berasal dari negara China dengan total sebanyak 15.665 pelanggan yang terdiri dari 7.887 pelanggan laki-laki dan 7.778 pelanggan perempuan. Sementara itu, pelanggan paling sedikit berasal dari negara Austria dengan total sebanyak 2 pelanggan laki-laki.
- Secara keseluruhan, proporsi pelanggan antara laki-laki dan wanita pada setiap negara hampir sama.

Call to action: Perusahaan dapat mengembangkan strategi pemasaran yang unik dan menarik berdasarkan preferensi dan kebutuhan antara laki-laki dan perempuan.

Query:

```
SELECT
  DISTINCT Country,
  SUM(CASE WHEN u.gender = 'M' THEN 1 else null END) AS Male,
  SUM(CASE WHEN u.gender = 'F' THEN 1 else null END) as Female,
  COUNT(oi.user_id) AS Total_Pelanggan
FROM bigquery-public-data.thelook_ecommerce.users AS u
INNER JOIN bigquery-public-data.thelook_ecommerce.order_items AS oi
ON u.id = oi.user_id
WHERE oi.status = 'Complete'
GROUP BY Country
```

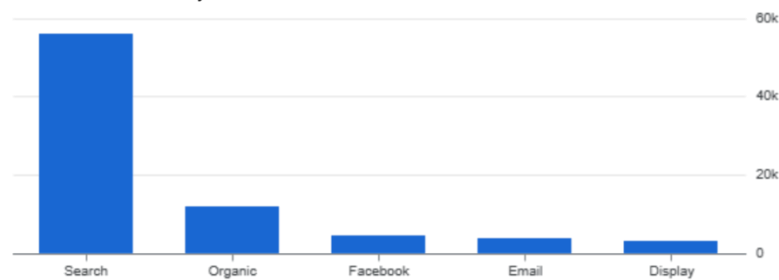
ORDER BY Total_Pelanggan DESC

Output:

Row	Country	Male	Female	Total_Customer
1	China	7887	7778	15665
2	United States	5071	5002	10073
3	Brasil	3245	3298	6543
4	South Korea	1137	1261	2398
5	United Kingdom	1069	1042	2111
6	France	1024	1044	2068
7	Germany	987	962	1949
8	Spain	876	889	1765
9	Japan	519	461	980
10	Australia	502	470	972
11	Belgium	254	273	527
12	Poland	78	57	135
13	Colombia	1	4	5
14	Austria	2	null	2

5. Apa sumber *traffic* yang lebih banyak diakses oleh pelanggan?

Total_Customers by Traffic_Source



- Pelanggan paling banyak datang ke situs web melalui mesin pencarian, seperti Google, Mozilla Firefox, Bing, dll sebanyak 56.009 pelanggan. Kemudian, diikuti oleh organik, yang merupakan pencarian yang muncul secara alami di mesin pencari sebanyak 12.080, diikuti oleh Facebook sebanyak 4.751, Email sebanyak 4.003, dan *display* (iklan) sebanyak 3.223.
- Hal ini dikarenakan karena mesin pencarian menjadi alat utama yang digunakan oleh banyak orang untuk mencari informasi dan produk secara *online* sehingga membuat situs web lebih mudah ditemukan dan diakses oleh pengguna yang ingin mencari informasi.

Call to action: Perusahaan mengoptimalkan situs web untuk mesin pencari, meningkatkan kehadiran di Facebook, dan mulai menyusun strategi pemasaran Email secara efektif untuk memaksimalkan *traffic*.

Query:

```
SELECT
  u.traffic_source AS Traffic_Source,
  COUNT(DISTINCT oi.user_id) AS Total_Customers,
FROM bigquery-public-data.thelook_ecommerce.users AS u
```

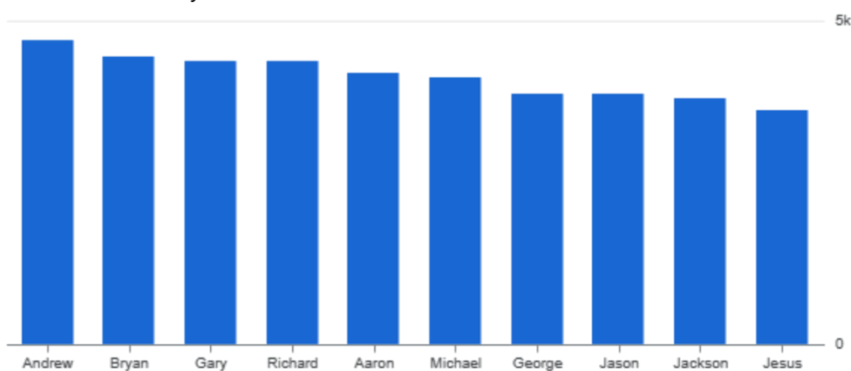
```
INNER JOIN bigquery-public-data.thelook_ecommerce.order_items AS oi
ON u.id = oi.user_id
GROUP BY Traffic_Source
ORDER BY Total_Customers DESC
```

- Output:

Row	Traffic_Source	Total_Customers
1	Search	56009
2	Organic	12080
3	Facebook	4751
4	Email	4033
5	Display	3223

6. Siapa saja pelanggan yang melakukan pembelian paling besar?

Total_Purchase by First_Name



- Pelanggan yang melakukan pembelian paling besar adalah Andrew Wilson yang berada di negara Poland dengan total pembayaran sebesar 4.702,3.
- Andrew Wilson memiliki total pembayaran paling besar karena melakukan pembelian produk dalam jumlah besar atau membeli barang dengan harga yang tinggi.

Call to action: Agar pelanggan menjadi loyal, perusahaan perlu memberikan layanan pelanggan layanan yang efisien, mempertahankan produk yang berkualitas, serta menawarkan program loyalitas. Selain itu, diskon dengan nominal yang lebih tinggi bisa dipertimbangkan untuk diberikan kepada pelanggan yang mencapai minimal pembayaran tertentu.

- Query:

```
SELECT
  DISTINCT u.id AS user_id,
  u.first_name AS First_Name,
  u.last_name AS Last_Name,
  u.country AS Country,
  ROUND(SUM(o.num_of_item * oi.sale_price),1) AS Total_Purchase
FROM bigquery-public-data.thelook_ecommerce.users as u
INNER JOIN bigquery-public-data.thelook_ecommerce.orders as o
ON u.id = o.user_id
```

```

INNER JOIN bigquery-public-data.thelook_ecommerce.order_items as oi
ON o.order_id = oi.order_id
WHERE o.status = 'Complete'
GROUP BY user_id, First_Name, Last_Name, Country
ORDER BY Total_Purchase DESC
LIMIT 10

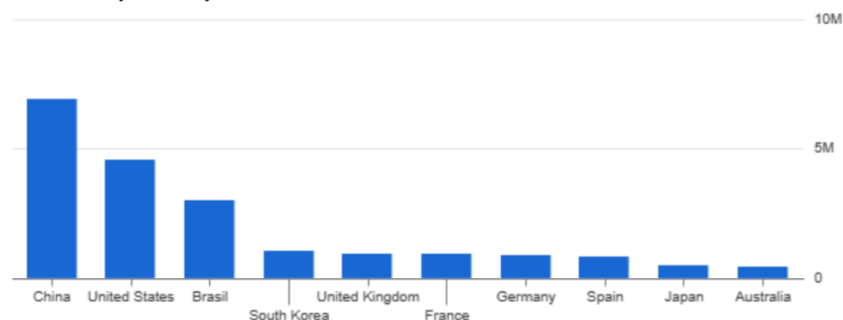
```

- Output:

Row	User_id	First_Name	Last_Name	Country	Total_Purchase
1	167	Andrew	Wilson	Poland	4702.3
2	69992	Bryan	Smith	United States	4454.0
3	31671	Gary	Miller	China	4373.5
4	97258	Richard	Dixon	China	4371.9
5	34311	Aaron	Carr	Japan	4204.9
6	44753	Michael	Hanson	United States	4124.0
7	86222	George	Vega	Brasil	3880.0
8	53837	Jason	Vazquez	United States	3873.0
9	11478	Jackson	Nelson	Brasil	3800.9
10	17531	Jesus	Rogers	China	3606.2

7. Negara apa yang memberikan pendapatan paling besar?

Revenue by Country



- China memberikan pendapatan untuk *e-commerce* paling besar dengan total pendapatan sebesar 6.936.191,9, diikuti oleh United States dan Brasil.
- Hal ini dikarenakan China memiliki populasi yang sangat besar, pertumbuhan pasar dan ekonomi yang pesat, dan infrastruktur teknologi yang canggih sehingga banyak warga China yang melakukan transaksi melalui *e-commerce*.

Call to action: Perusahaan dapat meningkatkan pendapatan *e-commerce* dengan melakukan ekspansi global, terus berinovasi pada teknologi, dan melakukan penguatan kemitraan sembari memantau perubahan pasar dan kebutuhan konsumen supaya negara-negara yang menghasilkan pendapatan besar dapat meningkatkan transaksinya terhadap *e-commerce*

- Query:

```

SELECT
DISTINCT u.country as Country,

```



```

ROUND(SUM(oi.sale_price * o.num_of_item),1) AS Revenue
FROM bigquery-public-data.thelook_ecommerce.users as u
INNER JOIN bigquery-public-data.thelook_ecommerce.order_items as oi
ON u.id = oi.user_id
INNER JOIN bigquery-public-data.thelook_ecommerce.orders as o
ON oi.order_id = o.order_id
GROUP BY Country
ORDER BY Revenue DESC
LIMIT 10

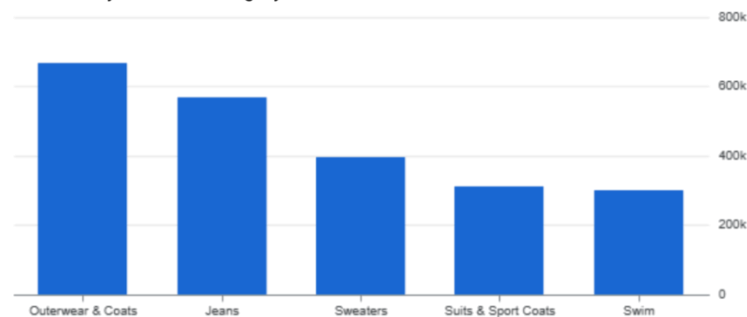
```

- Output:

Row	Country	Revenue
1	China	6936191.9
2	United States	4604762.6
3	Brasil	2992657.7
4	South Korea	1042489.7
5	United Kingdom	963673.0
6	France	948590.9
7	Germany	876007.1
8	Spain	828651.2
9	Japan	495621.6
10	Australia	445432.2

8. Apa kategori produk yang menghasilkan pendapatan paling besar?

Revenue by Product_Category



- Kategori yang memberikan pendapatan paling besar adalah *Outwear & Coats*, diikuti oleh *Jeans*, *Sweaters*, *Suit & Sport Coats*, dan *Swim*.
- Kategori *Outwear & Coats* kemungkinan besar terkait dengan musim tertentu, seperti musim dingin atau hujan sehingga konsumen membutuhkan penggunaan mantel atau pakaian luar.

Call to action: Untuk mempertahankan keuntungan ini, perusahaan dapat meningkatkan stok dan strategi promosi atau diskon ketika musim hujan atau musim dingin akan/sedang berlangsung.

- Query:

```

SELECT
category AS Product_Category,
SUM(num_of_item) AS Quantity,
ROUND(SUM(sale_price * num_of_item),1) AS Revenue

```

```

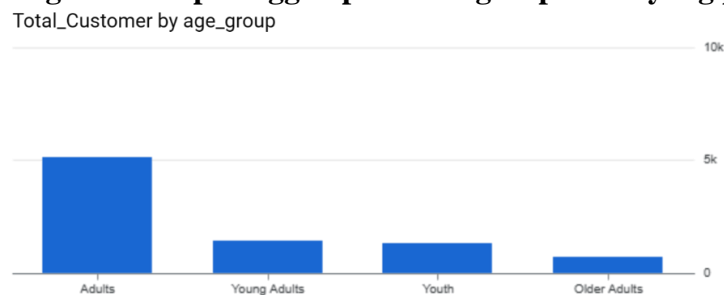
FROM bigquery-public-data.thelook_ecommerce.order_items AS oi
INNER JOIN bigquery-public-data.thelook_ecommerce.orders AS o
ON oi.order_id = o.order_id
INNER JOIN bigquery-public-data.thelook_ecommerce.products AS p
ON p.id = oi.product_id
WHERE oi.status = 'Complete'
GROUP BY Category
ORDER BY Revenue DESC
LIMIT 5

```

- Output:

Row	Product_Category	Quantity	Revenue
1	Outerwear & Coats	4533	666336.5
2	Jeans	5976	569032.6
3	Sweaters	5201	397493.3
4	Suits & Sport Coats	2436	310475.9
5	Swim	5341	301809.4

9. Bagaimana kategori umur pelanggan pada kategori produk yang paling laris?



- Pada kategori produk yang paling laris, yaitu *Outerwear & Coats*, pelanggan terbanyak berada pada kelompok umur *Adults* sebanyak 5.120 pelanggan, diikuti kelompok umur *Young Adults* sebanyak 1.433 pelanggan, kemudian kelompok umur *Youth* sebanyak 1.302 pelanggan, dan kelompok umur *Older Adults* sebanyak 345 pelanggan.
- Pakaian *outerwear & coats* menjadi populer di kalangan umur yang lebih dewasa, hal ini dikarenakan kebutuhan yang lebih besar untuk pakaian *outer* atau *fashion*. Sementara itu untuk kelompok umur *adults* dan *young adults*, kemungkinan mereka lebih peka terhadap tren mode dan gaya, mereka cenderung membeli produk *outerwear & coats* untuk menjaga penampilan mereka di musim dingin.

Call to action: Perusahaan dapat mengarahkan strategi pemasaran dan promosi untuk lebih tepat sasaran, seperti di kelompok umur *adults* dan *young adults* dan memenuhi kebutuhan pelanggan yang dominan pada kategori tersebut.

- Query:

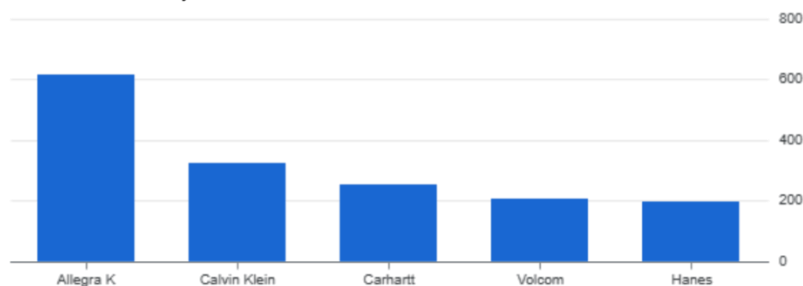
```
SELECT
CASE
  WHEN u.age < 12 THEN 'Children'
  WHEN u.age BETWEEN 12 AND 20 THEN 'Youth'
  WHEN u.age BETWEEN 21 AND 30 THEN 'Young Adults'
  WHEN u.age BETWEEN 31 AND 65 THEN 'Adults'
  WHEN u.age > 65 THEN 'Older Adults'
END AS age_group,
COUNT(DISTINCT u.id) AS Total_Pelanggan
FROM bigquery-public-data.thelook_ecommerce.users AS u
INNER JOIN bigquery-public-data.thelook_ecommerce.order_items AS oi
ON u.id = oi.user_id
INNER JOIN bigquery-public-data.thelook_ecommerce.inventory_items AS ii
ON ii.product_id = oi.product_id
WHERE ii.product_category = 'Outerwear & Coats'
GROUP BY age_group
ORDER BY Total_Pelanggan DESC
```

- Output:

Row	age_group	Total_Customer
1	Adults	5120
2	Young Adults	1433
3	Youth	1302
4	Older Adults	735

10. Brand apa yang paling sering dikembalikan oleh pelanggan?

Total_Returned by Brand



- Brand yang sering dikembalikan oleh pelanggan adalah *brand clothing*, seperti Allegra K sebanyak 615 unit, diikuti oleh Calvin Klein sebanyak 324 unit, dan Carhartt sebanyak 252 unit.
- Hal ini dikarenakan pakaian yang tiba tidak sesuai dengan ukuran yang diinginkan oleh pelanggan maupun kualitas produk yang tidak sesuai dengan ekspektasi pelanggan.

Call to action: Perusahaan dapat meninjau kembali kualitas produk yang dihasilkan, dan meningkatkan detail deskripsi produk agar barang lebih sesuai dan mengurangi tingkat pengembalian produk.

- Query:

```
SELECT
  DISTINCT p.brand AS Brand,
  COUNT(o.num_of_item) AS Total_Returned
FROM bigquery-public-data.thelook_ecommerce.products as p
INNER JOIN bigquery-public-data.thelook_ecommerce.order_items as oi
ON p.id = oi.product_id
INNER JOIN bigquery-public-data.thelook_ecommerce.orders as o
ON oi.order_id = o.order_id
WHERE o.status = 'Returned'
GROUP BY Brand
ORDER BY Total_Returned DESC
LIMIT 5
```

- Output:

Row	Brand ▼	Total_Returned ▼
1	Allegra K	615
2	Calvin Klein	324
3	Carhartt	252
4	Volcom	209
5	Hanes	199