

ASSIGNMENT CHAPTER 1 - FOUNDATION FOR DATA SCIENCE

SQL FOUNDATION

Nama : Rizqon Maulana
SCID : SC61224
Kompi : 12

Studi Case 1

Kita akan melamar menjadi seorang data scientist di suatu perusahaan e-commerce yaitu The Look e-Commerce. Sebagai ujian agar kita bisa diterima di perusahaan, HRD meminta kita untuk menganalisis sebaran dari pengguna The Look e-Commerce. HRD memberikan kita suatu data perusahaan yang dapat kita akses melalui laman berikut ini:

<https://console.cloud.google.com/marketplace/product/bigquery-public-data/thelook-ecommerce>

HRD meminta kita menganalisis data tersebut menggunakan Structured Query Language (SQL)!

Solution

Pertama, kita harus mengetahui berapa jumlah pengguna The Look e-Commerce di seluruh dunia hingga saat ini. Kita dapat mengetahuinya dengan mengeksekusi query berikut ini:

```
SELECT COUNT(id) AS jumlah_pengguna  
FROM `bigquery-public-data.thelook_ecommerce.users`
```

Jika dieksekusi, maka query tersebut akan menghasilkan output sebagai berikut:

Row	jumlah_pengguna
1	100000

Dari sini kita tahu bahwa jumlah pengguna The Look e-Commerce di dunia adalah 100000 pengguna.

Dari data tersebut, tersebar dari mana sajakah pengguna The Look e-commerce? Kita dapat mulai menganalisis dari tingkat negara. Misalnya kita ingin melihat 5 negara dengan pengguna The Look e-Commerce terbesar, kita dapat mengeksekusi query berikut:

```
SELECT country AS negara, COUNT(id) AS jumlah_pengguna  
FROM `bigquery-public-data.thelook_ecommerce.users`  
GROUP BY negara  
ORDER BY jumlah_pengguna DESC  
LIMIT 5
```

Jika dieksekusi, maka query tersebut akan menghasilkan output sebagai berikut:

Row	negara	jumlah_pengguna
1	China	34009
2	United States	22468
3	Brasil	14568
4	South Korea	5331
5	France	4673

Dari sini kita tahu bahwa 5 negara dengan jumlah pengguna The Look e-Commerce terbesar di dunia adalah China, United States, Brasil, South Korea, dan France. Negara China menjadi negara dengan jumlah pengguna The Look e-Commerce terbesar di dunia dengan jumlah 34009 pengguna, atau sekitar 34% dari seluruh pengguna.

Kita telah mengetahui bahwa berdasarkan negara, China menjadi negara dengan jumlah pengguna The Look e-Commerce terbesar di dunia. Namun kita ingin menganalisis dengan lebih detil, sekarang kita akan menganalisis sebaran pengguna The Look e-Commerce di Negara China. Kita ingin mengetahui 5 kota di China dengan pengguna The Look e-Commerce terbesar. Hal ini dapat dilakukan dengan mengeksekusi query berikut:

```
SELECT city AS kota, COUNT(id) AS jumlah_pengguna
FROM `bigquery-public-data.thelook_ecommerce.users`
WHERE country = "China"
GROUP BY kota
ORDER BY jumlah_pengguna DESC
LIMIT 5
```

Jika dieksekusi, maka query tersebut akan menghasilkan output sebagai berikut:

Row	kota	jumlah_pengguna
1	Shanghai	2484
2	Beijing	2090
3	Shenzhen	1329
4	Dongguan	970
5	Chengdu	775

Dari sini kita tahu bahwa 5 kota dengan jumlah pengguna The Look e-Commerce terbesar di China adalah Shanghai, Beijing, Shenzhen, Dongguan, dan Chengdu.

Sepertinya akan menarik jika kita membandingkan 5 kota dengan jumlah pengguna The Look e-Commerce terbesar di China dan di dunia. Apakah kota-kota dari China akan mendominasi? Langsung saja kita cari tahu bersama dengan mengeksekusi query berikut ini:

```
SELECT city as kota, COUNT(id) AS jumlah_pengguna
FROM `bigquery-public-data.thelook_ecommerce.users`
GROUP BY kota
```

ORDER BY jumlah_pengguna DESC

LIMIT 5

Jika dieksekusi, maka query tersebut akan menghasilkan output sebagai berikut:

Row	kota	jumlah_pengguna
1	Shanghai	2484
2	Beijing	2090
3	Seoul	1479
4	Shenzhen	1329
5	null	990

Dari sini kita tahu bahwa jika dibandingkan dengan kota di seluruh dunia pun, kota-kota di China seperti Shanghai, Beijing, dan Shenzhen masih mendominasi dalam hal jumlah pengguna The Look e-Commerce.

Kita telah berhasil menganalisis sebaran jumlah pengguna The Look e-Commerce di dunia. Dari jumlah pengguna The Look e-Commerce di dunia yang berjumlah 100000 pengguna, 34% di antaranya berasal dari negara China. Sebagian besar pengguna The Look e-Commerce di China tersebar di 5 kota, yaitu Shanghai, Beijing, Shenzhen, Dongguan, dan Chengdu. Jika dibandingkan dengan kota di seluruh dunia pun, kota-kota di China seperti Shanghai, Beijing, dan Shenzhen masih mendominasi dalam hal jumlah pengguna The Look e-Commerce. Kita bisa menyarankan kepada perusahaan The Look e-Commerce menggencarkan promosi di negara China, terutama di kota Shanghai, Beijing, dan Shenzhen.

Studi Case 2

Pekerjaan kita sebelumnya ternyata cukup menyenangkan HRD, namun HRD masih belum bisa memutuskan apakah kita diterima bekerja sebagai seorang data scientist di The Look e-Commerce atau tidak, karena juga masih ada kandidat lain. Keadaan kita dengan kandidat lain bisa di bilang imbang atau seri. Maka dari itu, HRD memberikan ujian kembali untuk menentukan kandidat yang akan diterima. Sebelumnya kita telah menganalisis sebaran pengguna The Look e-Commerce dan kita mendapati bahwa negara China menjadi pengguna The Look e-Commerce terbesar di dunia, utamanya di kota-kota seperti Shanghai, Beijing, dan Shenzhen. Kali ini HRD ingin kita untuk menganalisis perihal penjualan, apakah jumlah pengguna sebanding dengan jumlah transaksi yang dilakukan. Jika negara China menjadi pengguna The Look e-Commerce terbesar di dunia, harusnya China jugalah yang memiliki jumlah transaksi terbesar. Namun apakah fakta lapangan berkata demikian? Itu menjadi tugas kita untuk membuktikannya.

Solution

Untuk mengetahui tentang jumlah transaksi, kita perlu mengakses table “order_items”. Untuk menentukan jumlah transaksi, hal pertama yang perlu kita cermati adalah status dari suatu transaksi. lebih jelasnya dapat kita lihat dengan mengeksekusi query berikut:

```
SELECT DISTINCT status
```

```
FROM `bigquery-public-data.thelook_ecommerce.order_items`
```

Jika dieksekusi, maka query tersebut akan menghasilkan output sebagai berikut:

Row	status
1	Cancelled
2	Complete
3	Processing
4	Returned
5	Shipped

Dari sini kita tahu terdapat 4 status transaksi, yaitu Cancelled, Complete, Processing, Returned, dan Shipped. Complete adalah transaksi yang telah selesai, kita memperoleh pendapatan. Cancelled adalah transaksi yang dibatalkan, tentunya pada status ini kita tidak memperoleh pendapatan. Processing adalah transaksi yang sedang diproses, kita bisa memperoleh ataupun tidak memperoleh pendapatan. Returned adalah transaksi yang telah selesai namun pembeli mengembalikan barang yang dibeli, disini kita mungkin peroleh mengganti rugi sehingga bisa di bilang kita tidak memperoleh pendapatan. Shipped adalah transaksi yang sedang dalam proses pengiriman, kita bisa memperoleh pendapatan jika transaksi selesai, namun bisa juga kita perlu mengganti rugi jika pembeli melakukan return. Disini kita harus menyepakati bahwa untuk menghitung jumlah transaksi, kita hanya akan menyertakan transaksi yang berstatus "Complete".

Mengelaborasi dengan kasus sebelumnya, disini kita ingin mengetahui jumlah transaksi yang dilakukan serta total pendapatan yang diperoleh dari berbagai negara. kita dapat mengetahuinya dengan mengeksekusi query berikut ini:

```
SELECT users.country as negara, count(order_items.sale_price) as
jumlah_transaksi, sum(order_items.sale_price) as total_pendapatan
FROM `bigquery-public-data.thelook_ecommerce.order_items` as order_items
INNER JOIN `bigquery-public-data.thelook_ecommerce.users` as users
ON order_items.user_id = users.id
WHERE order_items.status = "Complete"
GROUP BY negara
ORDER BY total_pendapatan DESC, jumlah_transaksi DESC
LIMIT 5
```

Jika dieksekusi, maka query tersebut akan menghasilkan output sebagai berikut:

Row	negara	jumlah_transaksi	total_pendapatan
1	China	15656	946935.9408802...
2	United States	10119	592155.1204013...
3	Brasil	6657	394289.1504597...
4	South Korea	2370	143345.1099982...
5	United Kingdom	2207	136042.3501014...

Hasilnya tidak cukup mengejutkan karena dari urutan pertama hingga keempat, negara dengan jumlah transaksi terbesar juga merupakan negara dengan jumlah pengguna The Look e-Commerce terbesar, yaitu China, United States, Brasil, dan South Korea. Hanya berbeda di urutan kelima di mana walaupun France

adalah negara dengan pengguna terbesar nomor 5, namun dari sisi penjualan United Kingdom lebih unggul.

Bagaimana jika kita menganalisis jumlah transaksi yang dilakukan serta total pendapatan yang diperoleh dari berbagai kota di dunia? Apakah kota-kota dari China tetap mendominasi? Kita akan mengetahui hasilnya dengan mengeksekusi query berikut:

```
SELECT users.city as kota, count(order_items.sale_price) as jumlah_transaksi,
sum(order_items.sale_price) as total_pendapatan
FROM `bigquery-public-data.thelook_ecommerce.order_items` as order_items
INNER JOIN `bigquery-public-data.thelook_ecommerce.users` as users
  ON order_items.user_id = users.id
WHERE order_items.status = "Complete"
GROUP BY kota
ORDER BY total_pendapatan DESC, jumlah_transaksi DESC
LIMIT 5
```

Jika dieksekusi, maka query tersebut akan menghasilkan output sebagai berikut:

Row	kota	jumlah_transaksi	total_pendapatan
1	Shanghai	1107	64088.64007973...
2	Beijing	1016	57707.05004525...
3	Shenzhen	654	41583.59999322...
4	Seoul	600	37346.06002068...
5	Dongguan	461	29114.70001077...

Hasilnya juga tidak terlalu mengejutkan di mana kota-kota dari China tetap menjadi negara dengan penjualan terbanyak (terbukti dengan jumlah transaksi dan total pendapatan).

Sejauh ini, fakta di lapangan masih sesuai dengan yang seharusnya di mana jumlah pengguna sebanding dengan jumlah penjualan. Lalu mungkin terbersit pertanyaan, barang dengan brand apa yang paling banyak diburu oleh para pembeli (menarik perhatian pembeli untuk melakukan transaksi)? Kita dapat mengetahuinya dengan mengeksekusi query berikut ini:

```
SELECT products.brand as brand_barang, count(order_items.sale_price) as
jumlah_transaksi, sum(order_items.sale_price) as total_pendapatan
FROM `bigquery-public-data.thelook_ecommerce.order_items` as order_items
INNER JOIN `bigquery-public-data.thelook_ecommerce.products` as products
  ON order_items.product_id = products.id
WHERE order_items.status = "Complete"
GROUP BY brand_barang
ORDER BY total_pendapatan DESC, jumlah_transaksi DESC
LIMIT 5
```

Jika dieksekusi, maka query tersebut akan menghasilkan output sebagai berikut:

Row	brand_barang	jumlah_transaksi	total_pendapatan
1	Diesel	391	52711.40996932...
2	Calvin Klein	790	50056.80012321...
3	True Religion	214	44584.22987747...
4	7 For All Mankind	282	44517.54007339...
5	Carhartt	617	40242.22970676...

Dapat terlihat bahwa 5 brand barang dengan jumlah penjualan terbesar adalah Diesel, Calvin Klein, True Religion, 7 For All Mankind, dan Carhartt. Disini cukup menarik di mana jumlah transaksi tidak sebanding dengan total pendapatan. Bisa kita lihat bahwa pendapatan dari brand 'Diesel' lebih besar dari brand "Calvin Klein" meskipun dari jumlah transaksi, brand "Calvin Klein" lebih unggul.

Dari hasil analisis penjualan, kita mengetahui bahwa fakta di lapangan sesuai dengan yang semestinya dimana jumlah pengguna berbanding lurus dengan jumlah penjualan, semakin besar basis pengguna di suatu negara atau kota, maka semakin besar pula penjualannya. Dalam hal ini, China dan sebagai dari kota-kota di dalamnya masih menjadi konsumen dari The Look e-Commerce yang paling menguntungkan untuk perusahaan. Mungkin bisa menjadi saran kepada perusahaan untuk menggencarkan promosi di China sambil terus mengepakkan sayap di negara lain.

Studi Case 3

Selamat, kita dinyatakan lolos dan diterima menjadi seorang data scientist di perusahaan The Look e-Commerce. Namun jangan senang dulu, karena kita langsung mendapat pekerjaan yang cukup membutuhkan effort. Perusahaan The Look e-Commerce ingin memberikan suatu benefit untuk pengguna The Look e-Commerce yang royal. Tugas kita disini adalah menganalisis siapa pengguna The Look yang royal.

Solution

Untuk menganalisis siapa pengguna The Look e-Commerce yang paling royal, dapat kita lihat dari berapa total pendapatan yang didapat dari seorang pengguna. Hal ini dapat dilakukan dengan cara mengeksekusi query berikut ini:

```
SELECT    CONCAT(users.first_name, " ", users.last_name) as nama_pengguna,
count(order_items.sale_price)                          as          jumlah_transaksi,
sum(order_items.sale_price) as total_pendapatan
FROM `bigquery-public-data.thelook_ecommerce.order_items` as order_items
INNER JOIN `bigquery-public-data.thelook_ecommerce.users` as users
    ON order_items.user_id = users.id
GROUP BY nama_pengguna
ORDER BY total_pendapatan DESC, jumlah_transaksi DESC
LIMIT 5
```

Jika dieksekusi, maka query tersebut akan menghasilkan output sebagai berikut:

Row	nama_pengguna	jumlah_transaksi	total_pendapatan
1	Michael Johnson	68	4555.639988660...
2	John Smith	82	4403.380006790...
3	Michael Miller	53	4331.360011100...
4	Michael Brown	76	4207.680001258...
5	Michael Williams	73	4175.629999399...

Dapat terlihat bahwa 5 pengguna The Look e-Commerce yang paling royal (dilihat dari total pendapatan yang didapat dari pengguna tersebut) adalah Michael Johnson, John Smith, Michael Miller, Michael Brown, dan Michael Williams. Jumlah transaksi yang mereka lakukan bervariasi namun rata-rata pendapatan yang didapat dari mereka di atas 4000 dollar.

Keterangan

1. Berikut ini adalah Study Case:

Studi Case 1

Kita akan melamar menjadi seorang data scientist di suatu perusahaan e-commerce yaitu The Look e-Commerce. Sebagai ujian agar kita bisa diterima di perusahaan, HRD meminta kita untuk menganalisis sebaran dari pengguna The Look e-Commerce. HRD memberikan kita suatu data perusahaan yang dapat kita akses melalui laman berikut ini:

<https://console.cloud.google.com/marketplace/product/bigquery-public-data/thelook-ecommerce>

HRD meminta kita menganalisis data tersebut menggunakan Structured Query Language (SQL)!

2. Berikut ini adalah Query:

```
SELECT  CONCAT(users.first_name, " ", users.last_name) as nama_pengguna,
        count(order_items.sale_price) as jumlah_transaksi,
        sum(order_items.sale_price) as total_pendapatan
FROM `bigquery-public-data.thelook_ecommerce.order_items` as order_items
INNER JOIN `bigquery-public-data.thelook_ecommerce.users` as users
    ON order_items.user_id = users.id
GROUP BY nama_pengguna
ORDER BY total_pendapatan DESC, jumlah_transaksi DESC
LIMIT 5
```

3. Berikut ini adalah Insight (ditandai dengan stabilo berwarna biru muda):

Dapat terlihat bahwa 5 brand barang dengan jumlah penjualan terbesar adalah Diesel, Calvin Klein, True Religion, 7 For All Mankind, dan Carhartt. Disini cukup menarik di mana jumlah transaksi tidak sebanding dengan total pendapatan. Bisa kita lihat bahwa pendapatan dari brand 'Diesel' lebih besar dari brand "Calvin Klein" meskipun dari jumlah transaksi, brand "Calvin Klein" lebih unggul.