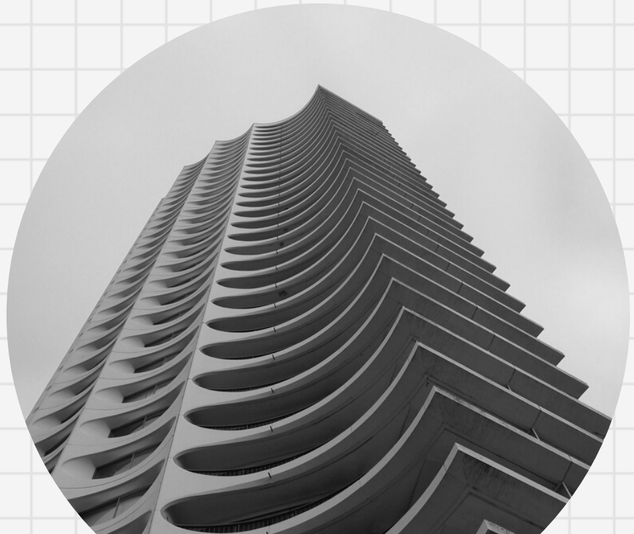
MINI PORTFOLIO SQL

Portfolio by Muhammad Farras Rizki



Walkthrough

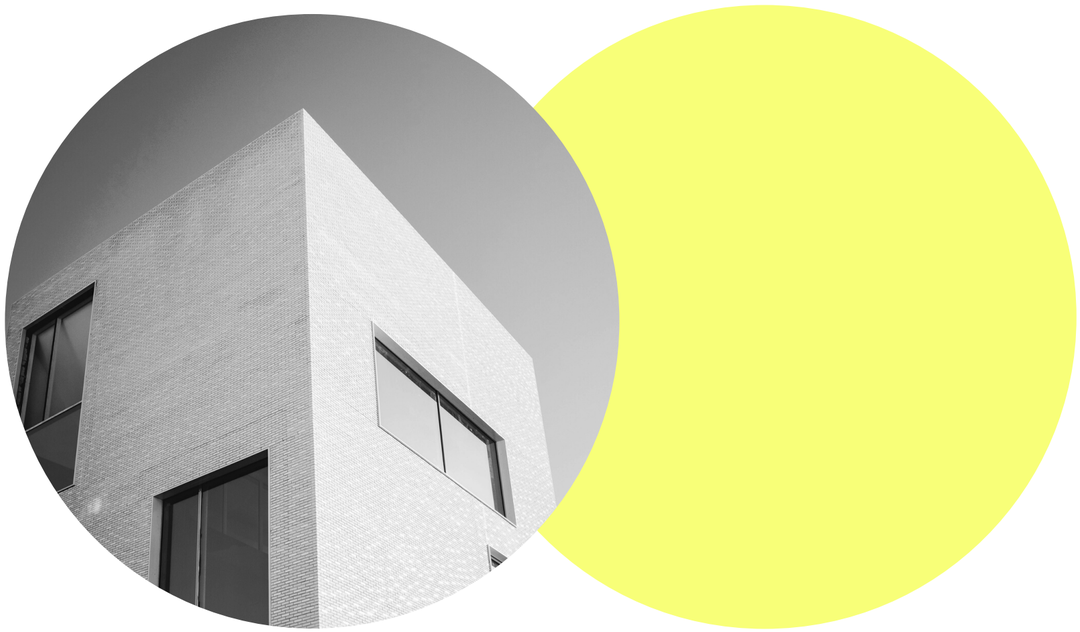
1. Introduction
2. Create Cloud Database
3. Get To Know The Data
4. Actionable Knowledge

1

Introduction

1. Background Data
2. Tools

Background Data



Bee Cycle adalah sebuah perusahaan manufaktur sepeda, menjual beberapa produk yang terbagi menjadi 4 kategori, antara lain :

- Bikes (Touring, Mountain, etc)
- Accessories (Helm, Lights, etc)
- Clothing (Sock, Jersey, etc)
- Component (Wheels, Pedal, etc)

Secara arsitektur, Bee Cycle menggunakan database potsgresql dan memiliki 5 tabel utama, antara lain :

1. **dim_geography** : Informasi geografi customer
2. **dim_customer** : Informasi personal customer
3. **dim_product** : Informasi product
4. **dim_territory** : Informasi toko cabang
5. **fact_sales** : Informasi detail transaksi

Link : github.com/FarrasRizki/BeeCycle_SQL

Tools

Jupyter Notebook

Running SQL syntax

DBeaver

Database management tools and see the relation between databases

2

Create Cloud Database

1. Create Framework
2. Install Database

dim_geography

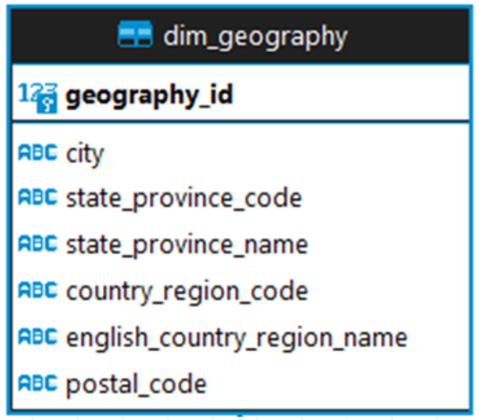
Create Framework

Run Query :

```
%%sql

CREATE TABLE public.dim_geography (
    geography_id int4 primary key,
    city varchar(50),
    state_province_code varchar(5),
    state_province_name varchar(50),
    country_region_code varchar(5),
    english_country_region_name varchar(50),
    postal_code varchar(10)
);
```

Check on DBeaver :



Column Name	#	Data type	Identity	Collation	Not Null
geography_id	1	int4			[v]
city	2	varchar(50)		default	[ ]
state_province_code	3	varchar(5)		default	[ ]
state_province_name	4	varchar(50)		default	[ ]
country_region_code	5	varchar(5)		default	[ ]
english_country_region_name	6	varchar(50)		default	[ ]
postal_code	7	varchar(10)		default	[ ]

dim_customer

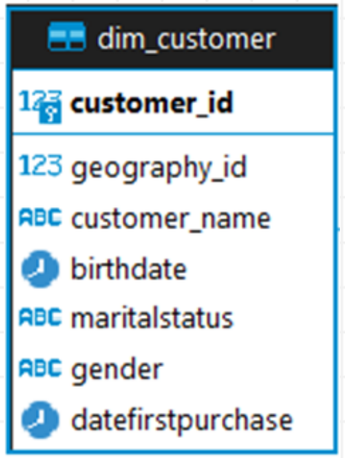
Create Framework

Run Query :

```
%%sql

CREATE TABLE public.dim_customer (
    customer_id int4 primary KEY,
    geography_id int4 ,
    customer_name varchar(100),
    birthdate date,
    maritalstatus varchar(5),
    gender varchar(5),
    datefirstpurchase date,
    foreign key(geography_id) references dim_geography(geography_id)
);
```

Check on DBeaver :



Column Name	#	Data type	Identity	Collation	Not Null
customer_id	1	int4			[v]
geography_id	2	int4			[ ]
customer_name	3	varchar(100)		default	[ ]
birthdate	4	date			[ ]
maritalstatus	5	varchar(5)		default	[ ]
gender	6	varchar(5)		default	[ ]
datefirstpurchase	7	date			[ ]

dim_product

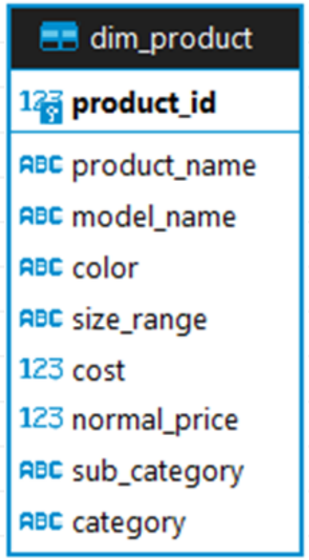
Create Framework

Run Query :

```
%%sql

CREATE TABLE public.dim_product (
    product_id int4 primary key,
    product_name varchar(100),
    model_name varchar(100),
    color varchar(30),
    size_range varchar(30),
    "cost" numeric(10),
    normal_price numeric(10),
    sub_category varchar(100),
    category varchar(100)
);
```

Check on DBeaver :



Column Name	#	Data type	Identity	Collation	Not Null
product_id	1	int4			[v]
product_name	2	varchar(100)		default	[ ]
model_name	3	varchar(100)		default	[ ]
color	4	varchar(30)		default	[ ]
size_range	5	varchar(30)		default	[ ]
cost	6	numeric(10)			[ ]
normal_price	7	numeric(10)			[ ]
sub_category	8	varchar(100)		default	[ ]
category	9	varchar(100)		default	[ ]

dim_territory

Create Framework

Run Query :

```
%%sql

CREATE TABLE public.dim_territory (
    territory_id int4 primary key,
    region varchar(30),
    country varchar(30),
    "groups" varchar(30)
);
```

Check on DBeaver :

Column Name	#	Data type	Identity	Collation	Not Null
territory_id	1	int4			[v]
region	2	varchar(30)		default	[ ]
country	3	varchar(30)		default	[ ]
groups	4	varchar(30)		default	[ ]

fact_sales

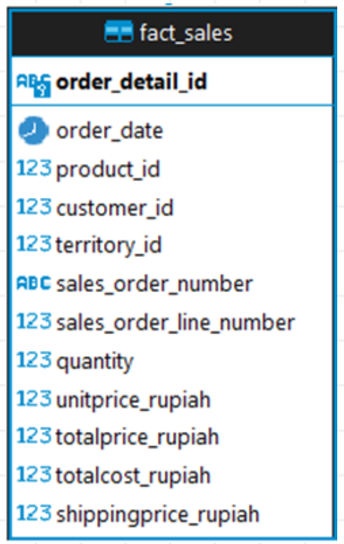
Create Framework

Run Query :

```
%%sql

CREATE TABLE public.fact_sales (
    order_detail_id varchar(30) primary key,
    order_date date ,
    product_id int4 ,
    customer_id int4 ,
    territory_id int4 ,
    sales_order_number varchar(30),
    sales_order_line_number int4,
    quantity int4,
    unitprice_rupiah numeric(12),
    totalprice_rupiah numeric(12),
    totalcost_rupiah numeric(12),
    shippingprice_rupiah numeric(12),
    foreign key(product_id) references dim_product(product_id),
    foreign key(customer_id) references dim_customer(customer_id),
    foreign key(territory_id) references dim_territory(territory_id)
);
```

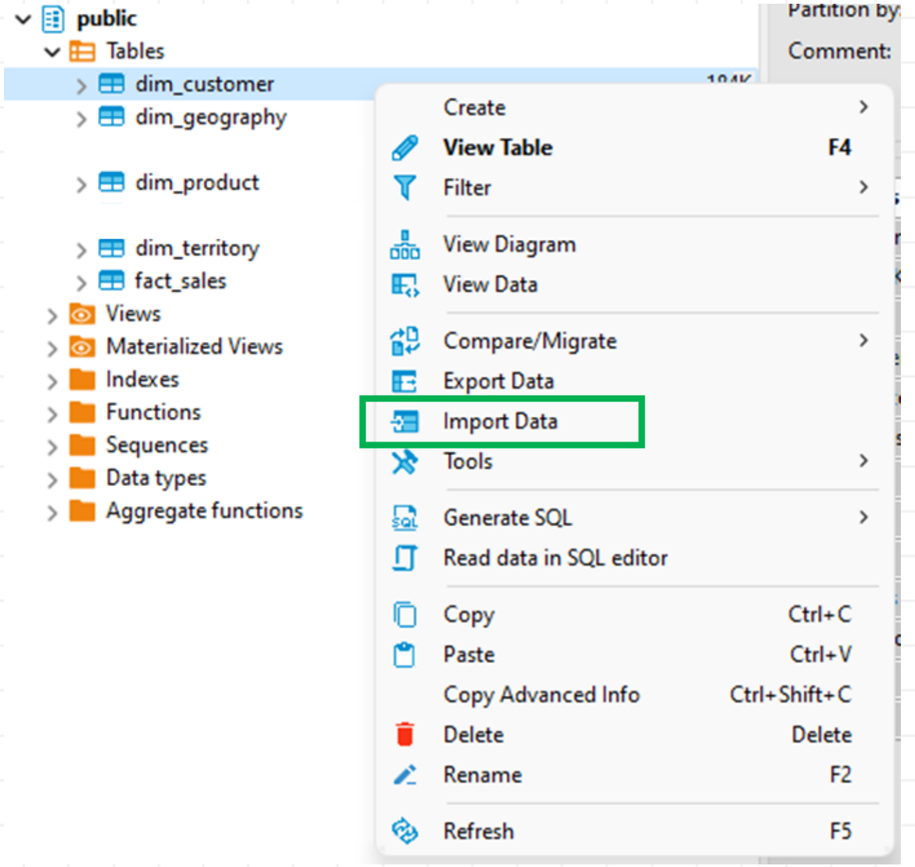
Check on DBeaver :



Column Name	#	Data type	Identity	Collation	Not Null
order_detail_id	1	varchar(30)		default	[v]
order_date	2	date			[ ]
product_id	3	int4			[ ]
customer_id	4	int4			[ ]
territory_id	5	int4			[ ]
sales_order_number	6	varchar(30)		default	[ ]
sales_order_line_number	7	int4			[ ]
quantity	8	int4			[ ]
unitprice_rupiah	9	numeric(12)			[ ]
totalprice_rupiah	10	numeric(12)			[ ]
totalcost_rupiah	11	numeric(12)			[ ]
shippingprice_rupiah	12	numeric(12)			[ ]

Install Database

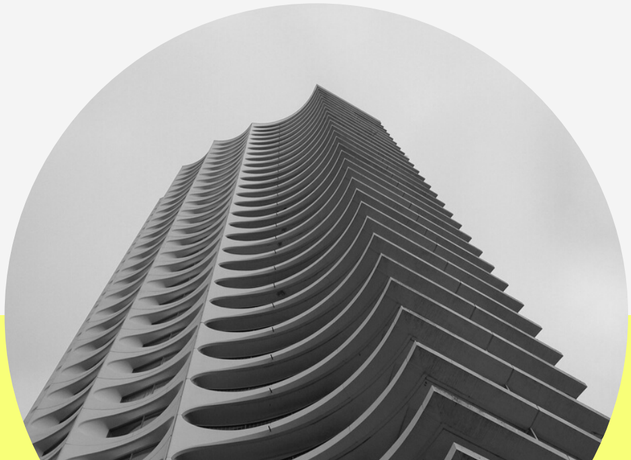
1. Buka Dbeaver
2. Klik kanan pada kerangka yang akan diinstal data, kemudian pilih 'Import Data'
3. Pilih file CSV yang akan diimport
4. Klik Proceed
5. Ulangi Langkah-langkahnya untuk setiap kerangka yang akan diinstall data



3

Get To Know The Data

1. Check Table Name
2. Check ER Diagram
3. Check Data



Check Table Name

Run Query :

```
%%sql

select table_name
from information_schema.tables
where table_schema = 'public'
```

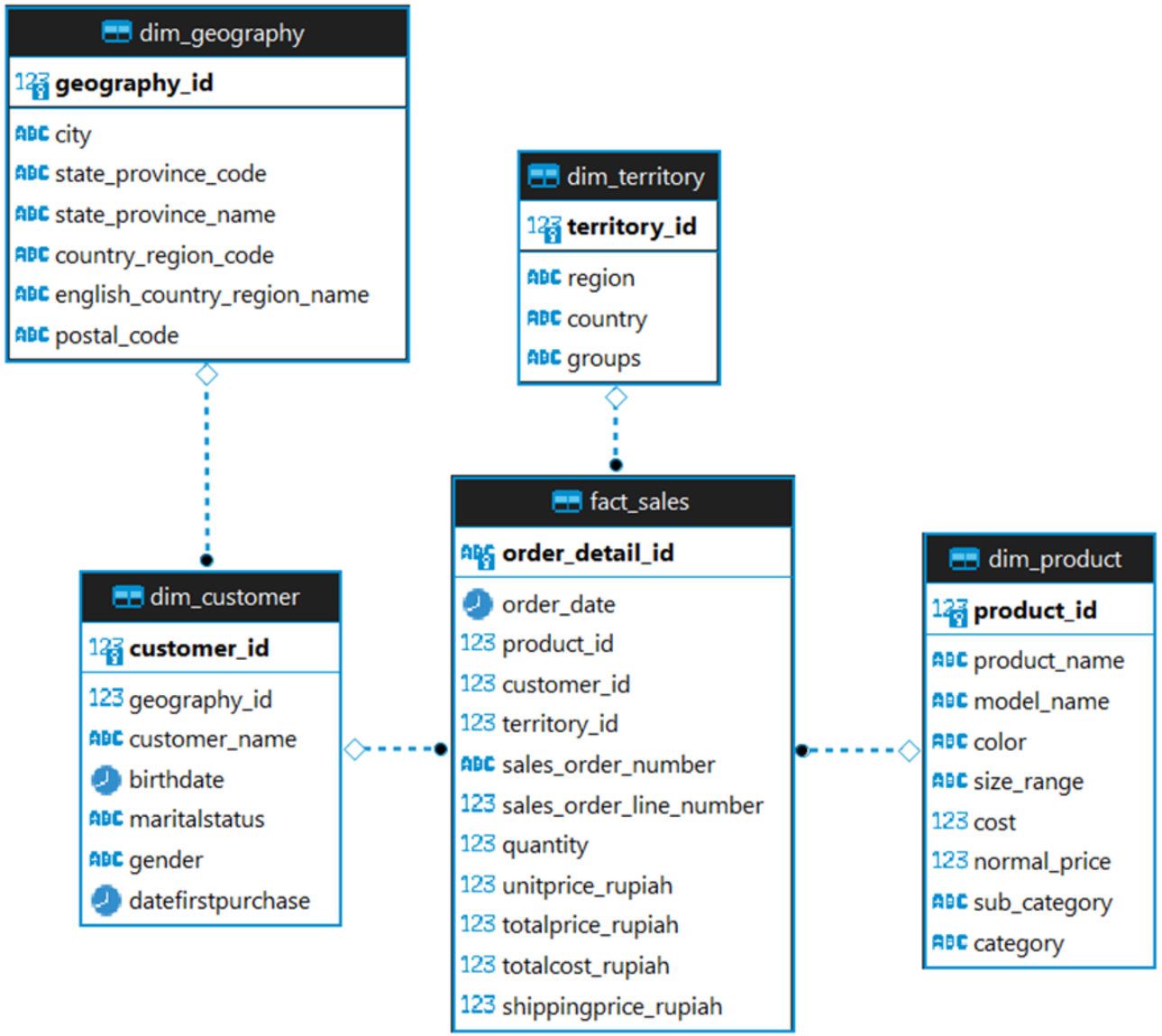
Output :

table_name
dim_geography
dim_customer
dim_product
fact_sales
dim_territory

Conclution :

5 Table utama telah berhasil diinstall. Selanjutnya akan dilakukan pengecekan ER Diagram dan Data Table

Check ER Diagram



Check Data

dim_geography

Run Query :

```
%%sql

select *
from dim_geography
limit 5
```

Output :

geography_id	city	state_province_code	state_province_name	country_region_code	english_country_region_name	postal_code
2	Coffs Harbour	NSW	New South Wales	AU	Australia	2450
3	Darlinghurst	NSW	New South Wales	AU	Australia	2010
4	Goulburn	NSW	New South Wales	AU	Australia	2580
5	Lane Cove	NSW	New South Wales	AU	Australia	1597
6	Lavender Bay	NSW	New South Wales	AU	Australia	2060

Column :

- **geography_id** → identifier [Primary Key]
- **city** → kota
- **state_province_code** → kode provinsi
- **state_province_name** → nama provinsi
- **country_region_code** → kode negara
- **english_country_region_name** → nama negara
- **postal_code** → kode pos

Check Data

dim_customer

Run Query :

```
%%sql

select *
from dim_customer
limit 5
```

Column :

- **customer_id** → identifier [Primary Key]
- **customer_name** → nama customer
- **birthdate** → tanggal lahir
- **maritalstatus** → status pernikahan
- **gender** → jenis kelamin
- **datefirstpurchase** → tanggal pertama kali transaksi

Output :

customer_id	geography_id	customer_name	birthdate	maritalstatus	gender	datefirstpurchase
11000	26	Jon Yang	1986-04-08	M	M	2016-07-22
11001	37	Eugene Huang	1985-05-14	S	M	2016-07-18
11002	31	Ruben Torres	1985-08-12	M	M	2016-07-10
11004	19	Elizabeth Johnson	1988-08-08	S	F	2016-07-26
11005	22	Julio Ruiz	1985-08-05	S	M	2016-07-02

Check Data

dim_product

Run Query :

```
%%sql

select *
from dim_product
limit 5
```

Output :

product_id	product_name	model_name	color	size_range	cost	normal_price	sub_category	category
210	HL Road Frame - Black, 58	HL Road Frame	Black	54-58 CM	11000	11000	Road Frames	Components
211	HL Road Frame - Red, 58	HL Road Frame	Red	54-58 CM	11000	11000	Road Frames	Components
480	Patch Kit/8 Patches	Patch kit	NA	NA	11991	32060	Tires and Tubes	Accessories
529	Road Tire Tube	Road Tire Tube	NA	NA	20892	55860	Tires and Tubes	Accessories
477	Water Bottle - 30 oz.	Water Bottle	NA	NA	26128	69860	Bottles and Cages	Accessories

Column :

- **product_id** → identifier [Primary Key]
- **product_name** → nama produk
- **model_name** → nama model produk
- **color** → warna produk
- **size_range** → ukuran produk
- **cost** → biaya produksi
- **normal_price** → harga jual
- **sub_category** → sub kategori product
- **category** → kategori produk

Check Data

dim_territory

Run Query :

```
%%sql

select *
from dim_territory
limit 5
```

Output :

territory_id	region	country	groups
1	Northwest	United States	North America
2	Northeast	United States	North America
3	Central	United States	North America
4	Southwest	United States	North America
5	Southeast	United States	North America

Column :

- **territory_id** → identifier [Primary Key]
- **region** → nama wilayah
- **country** → nama negara
- **groups** → group wilayah

Check Data

fact_sales

Run Query :

```
%%sql

select *
from fact_sales
limit 5
```

Output :

order_detail_id	order_date	product_id	customer_id	territory_id	sales_order_number	sales_order_line_number	quantity	unitprice_rupiah	totalprice_rupiah	totalcost_rupiah	shippingprice_rupiah
SO43698-1	2016-07-01	346	28389	7	SO43698	1	1	47599860	47599860	26770162	1189997
SO43704-1	2016-07-02	351	11005	9	SO43704	1	1	47249860	47249860	26573322	1181247
SO43705-1	2016-07-02	344	11011	9	SO43705	1	1	47599860	47599860	26770162	1189997
SO43713-1	2016-07-05	310	27601	4	SO43713	1	1	50095780	50095780	30398119	1252395
SO43714-1	2016-07-05	311	13591	10	SO43714	1	1	50095780	50095780	30398119	1252395

Column :

- **order_detail_id** → identifier [Primary Key]
- **order_date** → tanggal transaksi
- **product_id** → product identifier
- **customer_id** → customer identifier
- **territory_id** → territory identifier
- **sales_order_number** → nomor sales order
- **sales_order_line_number** → jumlah line product pada 1 sales order
- **quantity** → jumlah quantity 1 product
- **unitprice_rupiah** → harga per 1 product
- **totalprice_rupiah** → harga total per product (harga product * quantity)
- **totalcost_rupiah** → biaya total per product
- **shippingprice_rupiah** → biaya shipping per product

4

Actionable Knowledge

1. Tampilkan data toko cabang berdasarkan jumlah customer dan total transaksinya

1

Objective : Perusahaan ingin mengevaluasi performa dari masing-masing toko cabang dan mencari toko cabang yang memiliki potensi untuk dikembangkan berdasarkan transaksi yang tertinggi

Query :

```
%%sql

select fs.territory_id,
       sum(totalprice_rupiah) total_transaksi,
       count(distinct customer_id) jumlah_cust,
       region,
       country
from fact_sales fs left join dim_territory dt on fs.territory_id = dt.territory_id
group by 1,4,5
order by 2 desc
```

Output :

territory_id	total_transaksi	jumlah_cust	region	country
9	27054890692	357	Australia	Australia
4	10311461708	243	Southwest	United States
10	10178191455	135	United Kingdom	United Kingdom
8	9158798700	121	Germany	Germany
7	9083327871	140	France	France
1	6780941263	175	Northwest	United States
6	4624895625	184	Canada	Canada
5	33610640	2	Southeast	United States

Conclution:

Toko cabang di Australia memiliki pontensi yang tinggi untuk dikembangkan karena memiliki total transaksi dan jumlah customer yang tertinggi. Sedangkan untuk toko cabang di Southeast US merupakan toko dengan performa terendah, sehingga perlu dilakukan evaluasi.

2. Top 10 sub kategori dan kategori produk yang memberikan profit terbesar pada perusahaan

2

Objective : Untuk membentuk strategi pemasaran selanjutnya, perushaan Ingin mengetahui kategori dan sub kategori produk apa saja yang memiliki kontribusi terbesar pada profit perushaan

Query :

```
%%sql

with
sp as
(select category,
       sub_category,
       sum(totalprice_rupiah-totalcost_rupiah) as profit
from fact_sales fs left join dim_product dp on fs.product_id = dp.product_id
group by 1,2),

cp as
(select category, sum(profit) as category_profit
from sp
group by 1)

select sp.category, sp.sub_category, profit, category_profit
from sp left join cp on sp.category = cp.category
order by 3 desc
```

Output :

category	sub_category	sub_profit	category_profit
Bikes	Mountain Bikes	17100270907	31772096742
Bikes	Road Bikes	9813019228	31772096742
Bikes	Touring Bikes	4858806607	31772096742
Accessories	Helmets	183071244	540413900
Accessories	Tires and Tubes	161341388	540413900
Accessories	Bottles and Cages	50885080	540413900
Accessories	Hydration Packs	44819676	540413900
Clothing	Jerseys	41678216	107401322
Accessories	Fenders	36215004	540413900
Accessories	Bike Racks	30498720	540413900

Conclution:

Jika dilihat dari profit yang dihasilkan, kategori produk Bikes dan Accessories memiliki sub kategori produk yang paling banyak memberikan profit pada perusahaan. Namun dari kedua kategori produk tersebut, Bikes menghasilkan profit yang lebih tinggi daripada Accessories

3. Top 5 kota tempat tinggal customer yang paling banyak membeli kategori produk Bikes

3

Objective : Karena produk dengan kategori Bikes adalah produk yang paling banyak memberikan profit pada perusahaan, perusahaan ingin mencari lokasi customer yang sering membeli produk Bikes berdasarkan kota tempat tinggal

Query :

```
%%sql

select category,
        city,
        count(city) as jumlah
from fact_sales fs
left join dim_customer dc on fs.customer_id = dc.customer_id
left join dim_geography dg on dg.geography_id = dc.geography_id
left join dim_product dp on fs.product_id = dp.product_id
where category = 'Bikes'
group by 1,2
order by 1,3 desc
limit 5
```

Output :

category	city	jumlah
Bikes	London	73
Bikes	Paris	44
Bikes	Warrnambool	34
Bikes	Geelong	33
Bikes	Newcastle	32

Conclution:

5 kota tempat tinggal customer yang paling banyak membeli kategori produk Bikes adalah London, Paris, Warrnambool, Geelong dan Newcastle

4. Top 3 sub kategori produk yang paling diminati masing-masing gender

4

Objective : Perusahaan Ingin memaksimalkan penggunaan voucher discount pada masing-masing customer berdasarkan gender dan minat sesuai sub kategori produk

Query :

```
%%sql

with
ranking as
(select count(order_detail_id),
       gender,
       sub_category,
       row_number() over (partition by gender order by count(order_detail_id) desc) as rank
from fact_sales fs
left join dim_customer dc on fs.customer_id = dc.customer_id
left join dim_product dp on fs.product_id = dp.product_id
where gender != 'None'
group by 2,3)

select gender, rank, sub_category
from ranking
where rank <= 3
order by 1,2
```

Output :

gender	rank	sub_category
F	1	Tires and Tubes
F	2	Mountain Bikes
F	3	Bottles and Cages
M	1	Tires and Tubes
M	2	Mountain Bikes
M	3	Bottles and Cages

Conclution:

Ternyata, 3 sub kategori produk yang paling diminati oleh customer laki-laki maupun perempuan sama. Maka, strategi pembagian voucher berdasarkan sub kategori produk, pada customer laki-laki maupun perempuan dapat disamakan

5. Siapa 5 customer ID yang memiliki total transaksi tertinggi ? Sebutkan nama dan kota tempat tinggal customer

5

Objective : Perusahaan BeeCycle ingin mengapresiasi customer yang paling loyal dengan perusahaan

Query :

```
%%sql

select sum(totalcost_rupiah) as total_transaksi, customer_name, city
from fact_sales fs
left join dim_customer dc on fs.customer_id = dc.customer_id
left join dim_geography dg on dc.geography_id = dg.geography_id
group by 2,3
order by 1 desc
limit 5
```

Output :

total_transaksi	customer_name	city
112629431	Nichole Nara	Saint-Denis
112475544	Randall Dominguez	Dunkerque
112286440	Kaitlyn Henderson	Tremblay-en-France
112205380	Margaret He	Metz
112076453	Adriana Gonzalez	Colomiers

Conclution:

Dari analisa tersebut, dapat membantu perusahaan dalam menentukan 5 customer yang memiliki total transaksi tertinggi di BeeCycle. Dan rata-rata total transaksi dari kelima customer adalah > Rp 110.000.000

6. Top 3 product apa yang paling populer dari tiap toko cabang pada bulan Oktober?

6

Objective : Karena kuota campaign pada bulan Oktober terbatas untuk tiap toko cabang, maka akan diambil top 3 product dari masing-masing toko cabang pada bulan Oktober untuk dijadikan referensi

Query :

```
%%sql

with
ranking as
(select count(order_detail_id) as quantity,
       date_part('month', order_date) as month,
       territory_id,
       product_name,
       row_number() over (partition by territory_id order by count(order_detail_id) desc) as rank
from fact_sales fs left join dim_product dp
      on fs.product_id = dp.product_id
where date_part('month', order_date) = 10.0
group by 2,3,4)

select territory_id, rank, product_name
from ranking
where rank <= 3
order by 1,2
```

Output :

territory_id	rank	product_name
1	1	Road Tire Tube
1	2	Patch Kit/8 Patches
1	3	HL Mountain Tire
4	1	Road-150 Red, 48
4	2	Road Tire Tube
4	3	Road-150 Red, 52
6	1	Touring Tire Tube
6	2	Patch Kit/8 Patches
6	3	Road Tire Tube
7	1	Water Bottle - 30 oz.
7	2	Road Bottle Cage
7	3	Mountain-200 Black, 42
8	1	AWC Logo Cap
8	2	Water Bottle - 30 oz.
8	3	Road Bottle Cage
9	1	Mountain-200 Black, 42
9	2	Mountain-200 Silver, 42
9	3	Mountain-200 Silver, 38
10	1	Mountain Tire Tube
10	2	Sport-100 Helmet, Blue
10	3	ML Mountain Tire

Conclution:

Dari hasil analisa tersebut, maka dapat ditentukan produk-produk apa saja yang akan diprioritaskan untuk campaign di bulan Oktober pada masing-masing toko cabang (territory_id)

7. Grouping age dan gender apa yang memiliki transaksi paling tinggi di perusahaan?

7

Objective : Untuk mengetahui target customer perusahaan, akan dilihat distribusi customer berdasarkan gender dan umur

Query :

```
%%sql

with
query as
(select sum(totalprice_rupiah) transaksi, gender, date_part('year', current_date) - date_part('year' , birthdate) as age
from fact_sales fs left join dim_customer dc
      on fs.customer_id = dc.customer_id
where gender != 'None'
group by 2,3)


select transaksi,
       gender,
       SUM(CASE WHEN age <= 20 THEN 1 ELSE 0 END) AS "Group <=20",
       SUM(CASE WHEN age BETWEEN 21 AND 40 THEN 1 ELSE 0 END) AS "Group 21 - 40",
       SUM(CASE WHEN age BETWEEN 41 AND 60 THEN 1 ELSE 0 END) AS "Group 41 - 60",
       SUM(CASE WHEN age > 60 THEN 1 ELSE 0 END) AS "Group >60"
from query
group by 1,2
order by transaksi desc
limit 1
```

Output:

transaksi	gender	Group <=20	Group 21 - 40	Group 41 - 60	Group >60
2281102135	F	0	1	0	0

Conclution:

Kategori customer dengan transaksi paling tinggi di perusahaan adalah customer Perempuan dengan rentang umur 21-40 tahun, dengan total transaksi sebesar Rp 2.281.102.135
Maka, perusahaan dapat memfokuskan strategi marketing pada customer dengan kategori tersebut

8. Warna apa di tiap tahun yg menjadi warna paling populer dibeli oleh customer?

8

Objective : Karena akan diadakan campaign baru, maka perlu menghilight warna product tertentu yang menjadi kegemaran customer untuk dijadikan referensi

Query :

```
%%sql

with
ranking as
(select count(order_detail_id),
       date_part('year', order_date) as year,
       color,
       row_number() over (partition by date_part('year', order_date) order by count(order_detail_id) desc) as rank
from fact_sales fs join dim_product dp
    on fs.product_id = dp.product_id
where color != 'NA'
group by 2,3)

select *
from ranking
where rank = 1
order by 2
```

Output :

count	year	color	rank
216	2016.0	Red	1
223	2017.0	Red	1
503	2018.0	Black	1
281	2019.0	Black	1

Conclution:

Dari tahun 2016 sampai 2019, warna yang paling populer dibeli oleh customer adalah merah dan hitam. Namun karena dari 2 tahun terakhir warna hitam menjadi warna yang paling populer, maka saya merekomendasikan produk dengan warna hitam yang lebih ditonjolkan pada campaign yang akan datang

Thank you

Accept positive criticism and suggestions
Please contact me

Email

mfarrasrizki@gmail.com

LinkedIn

linkedin.com/in/mfarrasrizki

Github

github.com/FarrasRizki/BeeCycle_SQL