Mini Project SQL

Bee-Cycle

2022

Presented By

ADRISTI SHAFIYA

OUTLINE

1. **About Dataset**
2. **Create Cloud Database**
3. **About The Tables**
4. **Journey to Dataset**
    - Top 10 Most Purchased Product
    - Biggest Transaction Grouping by Gender and Age
    - Top Famous Product Based on Territory
    - Lowest Purchased of The Product
    - Giving a Voucher Discount to Customer
    - Most Purchased Poduct Grouping by Marital Status
    - Product with The Biggest Profit
    - Top 3 Winner Customer in 2018-2019
    - City with the Biggest Transaction of Bikes, Accessories and Clothing
    - Biggest Sales of Bikes Based on City in 2016-2019
    - Biggest Cost of The Product in 2016-2019

ABOUT DATASET

Bee-Cycle

Bee-Cycle is a bicycle manufacturing company. They sold 4 categories of product :

- Bikes (eg Touring, Mountain, etc)
- Accessories (eg Helm, Lights, etc)
- Clothing (eg Sock, Jersey, etc)
- Component (eg wheels, pedal etc)

Architecture

Bee-Cycle is using Database PostgreSQL. It has 5 main table :

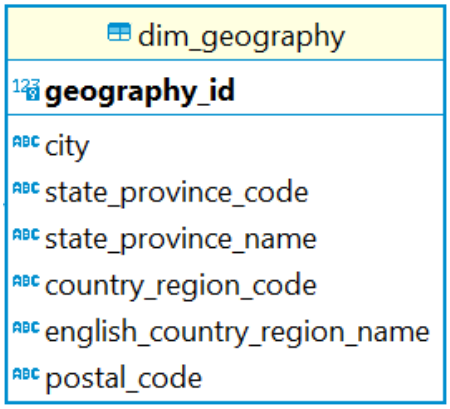
- Dim_Product : Product Information
- Dim_Customer : Customer Information
- Dim_Territory : Branch Store Information
- Dim_Geography : Geography Information, to identify customer adress
- Fact_Sales : Transaction Detail

TOOLS

CREATE CLOUD DATABASE

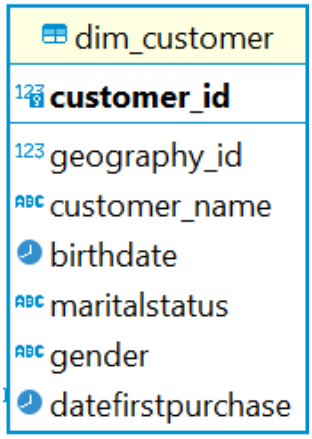
Dim_Geography

```
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))
```



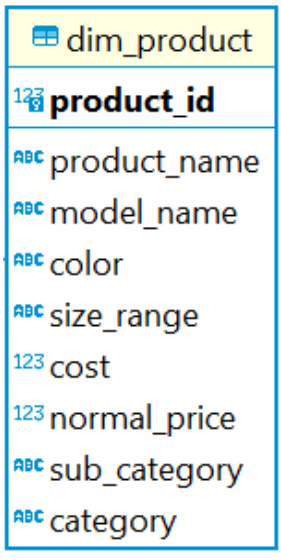
Dim_Customer

```
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))
```



Dim_Product

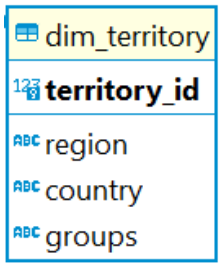
```
CREATE TABLE public.dim_product (
  product_id int4 primary key,
  product_name varchar(100),
  model_name varchar(100),
  color varchar(30),
  size_range varchar(30),
  "cost" numeric(10),
  normal_price numeric(10),
  sub_category varchar(100),
  category varchar(100))
```



CREATE CLOUD DATABASE

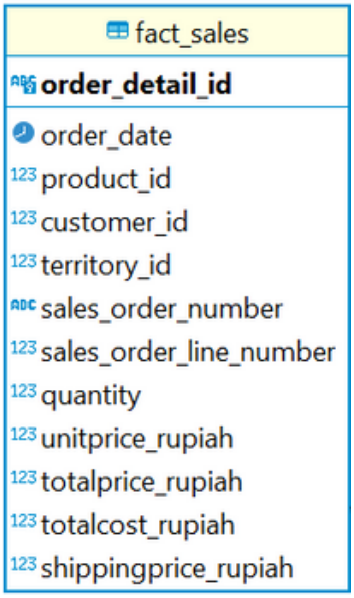
Dim_Territory

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



Fact_Sales

```
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)
```



About The Tables

Dim_Geography

Column:

1. **geography_id** : Identifier (**Primary Key**)
2. **city**
3. **state_province_code** : Province code
4. **state_province_name** : Name of the province
5. **country_region_code** : Country code
6. **postal_code**

QUERY

```
SELECT *
FROM dim_geography
LIMIT 5
```

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

Dim_Customer

Column:

1. **customer_id** : Identifier (**Primary Key**)
2. **customer_name** : Name of the customer
3. **birthdate**
4. **maritalstatus**
5. **gender**
6. **datefirstpurchase** : The date of the first transaction

QUERY

```
SELECT *
FROM dim_customer
LIMIT 5
```

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

Dim_Product

Column:

1. **product_id** : Identifier (**Primary Key**)
2. **product_name** : Name of the product
3. **color** : Color of the product
4. **size_range** : Size of the Product
5. **cost**
6. **normal_price** : Selling price
7. **sub_category** : Product sub-category
8. **category** : Category of the product

QUERY

```
SELECT *
FROM dim_product
LIMIT 5
```

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

About The Tables

Dim_Territory

Column:

1. **territory_id** : Identifier (**Primary Key**)
2. **region**
3. **country** : Name of the country
4. **groups**

QUERY

```
SELECT *
FROM dim_territory
LIMIT 5
```

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

Fact_Sales

Column:

1. **order_detail_id** : Identifier (**Primary Key**)
2. **order_date** : Date of the transaction
3. **product_id**: Product identifier
4. **customer_id** : Customer Identifier
5. **territory_id :** Territory identifier
6. **sales_order_number** : Number of sales order
7. **sales_order_line_number** : Total line product for 1 sales order
8. **quantity** : Number of products purchased
9. **unitprice_rupiah** : Price for each product
10. **totalprice_rupiah** : Total price for each product (**unitprice_rupiah * quantity**)
11. **totalcost_rupiah** : Total cost for each product
12. **shippingprice_rupiah** : Shipping price for each product

QUERY

```
SELECT *
FROM fact_sales
LIMIT 5
```

order_detail_id	order_date	product_id	customer_id	territory_id	sales_order_number
SO43698-1	2016-07-01	346	28389	7	SO43698
SO43704-1	2016-07-02	351	11005	9	SO43704
SO43705-1	2016-07-02	344	11011	9	SO43705
SO43713-1	2016-07-05	310	27601	4	SO43713
SO43714-1	2016-07-05	311	13591	10	SO43714

sales_order_line_number	quantity	unitprice_rupiah	totalprice_rupiah	totalcost_rupiah	shippingprice_rupiah
1	1	47599860	47599860	26770162	1189997
1	1	47249860	47249860	26573322	1181247
1	1	47599860	47599860	26770162	1189997
1	1	50095780	50095780	30398119	1252395
1	1	50095780	50095780	30398119	1252395

1. Top 10 Most Purchased Product

QUERY

```
WITH transaksi AS(
  SELECT product_id, count(distinct order_detail_id) AS count_ts
  FROM fact_sales
  GROUP BY 1
)
SELECT dp.product_id, dp.product_name, dp.model_name, dp.sub_category, dp.category,
ts.count_ts
FROM dim_product dp LEFT JOIN transaksi ts ON dp.product_id = ts.product_id
WHERE count_ts is not null
GROUP BY 1,2,3,4,5,6
ORDER BY count_ts DESC
LIMIT 10
```

RESULT

product_id	product_name	model_name	sub_category	category	count_ts
477	Water Bottle - 30 oz.	Water Bottle	Bottles and Cages	Accessories	363
480	Patch Kit/8 Patches	Patch kit	Tires and Tubes	Accessories	262
478	Mountain Bottle Cage	Mountain Bottle Cage	Bottles and Cages	Accessories	228
222	Sport-100 Helmet, Blue	Sport-100	Helmets	Accessories	221
214	Sport-100 Helmet, Red	Sport-100	Helmets	Accessories	201
528	Mountain Tire Tube	Mountain Tire Tube	Tires and Tubes	Accessories	200
537	HL Mountain Tire	HL Mountain Tire	Tires and Tubes	Accessories	199
479	Road Bottle Cage	Road Bottle Cage	Bottles and Cages	Accessories	191
225	AWC Logo Cap	Cycling Cap	Caps	Clothing	190
485	Fender Set - Mountain	Fender Set - Mountain	Fenders	Accessories	188

INSIGHT

Accessories is the category product that mostly customer buy. Because our store focused on sport, we can communicate to business development team to improve or make inovation especially on the quality and the strenght of product for this category. For the marketing team, they can improve the strategy to promote new product that has been inovated

2. Biggest Transaction Grouping by Gender and Age

QUERY

```
WITH cust_age AS(
  SELECT customer_id, gender,
  CASE WHEN age <= 20 THEN 'Group <= 20'
  WHEN age > 20 AND age <= 40 THEN 'Group 21-40'
  WHEN age > 40 AND age <= 60 THEN 'Group 41-60'
  ELSE 'Group >60'
  END AS grouping_age
  FROM (
    SELECT customer_id, gender, date_part('year', current_date) - date_part('year',
birthdate) AS age
    FROM dim_customer) AS subquery_age
  GROUP BY 1,2,3)
SELECT ca.gender, ca.grouping_age, count(order_detail_id) AS total_transaction
FROM cust_age ca LEFT JOIN fact_sales fs
ON ca.customer_id = fs.customer_id
GROUP BY 1,2
ORDER BY total_transaction DESC
```

RESULT

gender	grouping_age	total_transaction
F	Group 21-40	1551
F	Group 41-60	1454
M	Group 21-40	1401
M	Group 41-60	1304
M	Group >60	161
F	Group >60	82

INSIGHT

Female person with the range of age between 21-60 years old has biggest transaction in our store. This will help the store to adjust the interest item and decide the item to sell based on the categories of the customer above.

3. Top Famous Product Based on Territory

QUERY

```
WITH popular AS(
  SELECT dt.territory_id, dp.product_id, dp.product_name, dp.sub_category,
  count(distinct order_detail_id) jumlah,
  ROW_NUMBER() OVER (partition by dt.territory_id ORDER BY count(distinct
order_detail_id) DESC) AS ranking

  FROM fact_sales fs
  LEFT JOIN dim_territory dt ON fs.territory_id = dt.territory_id
  LEFT JOIN dim_product dp ON fs.product_id = dp.product_id
  GROUP BY 1,2,3,4
)
SELECT *
FROM popular
WHERE ranking <= 3
```

RESULT

territory_id	country	product_id	sub_category	category	jumlah	ranking
1	United States	537	Tires and Tubes	Accessories	39	1
1	United States	480	Tires and Tubes	Accessories	34	2
1	United States	528	Tires and Tubes	Accessories	28	3
4	United States	537	Tires and Tubes	Accessories	48	1
4	United States	480	Tires and Tubes	Accessories	42	2
4	United States	485	Fenders	Accessories	37	3
5	United States	478	Bottles and Cages	Accessories	1	1
5	United States	477	Bottles and Cages	Accessories	1	2
5	United States	353	Mountain Bikes	Bikes	1	3
6	Canada	529	Tires and Tubes	Accessories	83	1
6	Canada	480	Tires and Tubes	Accessories	79	2
6	Canada	528	Tires and Tubes	Accessories	54	3
7	France	477	Bottles and Cages	Accessories	58	1
7	France	479	Bottles and Cages	Accessories	39	2
7	France	222	Helmets	Accessories	36	3
8	Germany	477	Bottles and Cages	Accessories	52	1
8	Germany	478	Bottles and Cages	Accessories	40	2
8	Germany	225	Caps	Clothing	34	3
9	Australia	477	Bottles and Cages	Accessories	124	1
9	Australia	478	Bottles and Cages	Accessories	76	2
9	Australia	214	Helmets	Accessories	68	3
10	United Kingdom	477	Bottles and Cages	Accessories	64	1
10	United Kingdom	222	Helmets	Accessories	42	2
10	United Kingdom	225	Caps	Clothing	41	3

INSIGHT

If we want to make a campaign that is limited for each branch store, we can maximize to make a campaign focusing based on famous product. These will help each branch store to increase the profit

4. Lowest Purchased of The Product

QUERY

```
WITH minus AS(
  SELECT date_part('year', order_date) AS years,
  dp.sub_category, dp.category,
  count(distinct order_detail_id) AS count,
  ROW_NUMBER() OVER (partition by date_part('year', order_date) ORDER BY
count(distinct order_detail_id) ASC) AS rownum

  FROM fact_sales fs LEFT JOIN dim_product dp ON fs.product_id = dp.product_id
  GROUP BY 1,2,3
)
SELECT *
FROM minus
WHERE rownum <=3
```

RESULT

years	sub_category	category	count	rownum
2016.0	Mountain Bikes	Bikes	121	1
2016.0	Road Bikes	Bikes	218	2
2017.0	Mountain Bikes	Bikes	271	1
2017.0	Road Bikes	Bikes	290	2
2018.0	Shorts	Clothing	8	1
2018.0	Bike Stands	Accessories	13	2
2018.0	Vests	Clothing	19	3
2019.0	Bike Racks	Accessories	6	1
2019.0	Bike Stands	Accessories	7	2
2019.0	Shorts	Clothing	8	3

INSIGHT

We see that in 2016 and 2017 the category of the product that has a lowest purchase is **bikes.** In the 2018 and 2019, the category of the product that has a lowest purchase are **accessories and clothing**. In the future we have to evaluate what is the caused of the lowest purchase. To solve it, we can optimize our marketing strategy, doing clearance sale, etc

5. Giving a Voucher Discount to Customer

QUERY

```
SELECT dc.customer_name, count(distinct order_detail_id) AS jumlah_pembelian, dg.city
FROM fact_sales fs
LEFT JOIN dim_customer dc USING (customer_id)
LEFT JOIN dim_geography dg USING (geography_id)
GROUP BY 1,3
ORDER BY 2 DESC
LIMIT 10
```

RESULT

customer_name	jumlah_pembelian	city
Samantha Jenkins	43	Shawnee
Henry Garcia	41	Cliffside
Dalton Perez	38	Haney
Charles Jackson	38	Royal Oak
Mason Roberts	35	Haney
April Shan	35	Cergy
Jason Griffin	34	Victoria
Hailey Patterson	34	Langley
Ryan Thompson	34	Oak Bay
Fernando Barnes	33	N. Vancouver

INSIGHT

The table is consist of customer who has the biggest transaction. We can give some voucher as an appreciation to the customer for entrusting the purchase of items at our store.

6. Most Purchased Poduct Grouping by Marital Status

QUERY

```
WITH ranking AS (
SELECT maritalstatus, count(order_detail_id), sub_category, category,
ROW_NUMBER() OVER (partition by maritalstatus ORDER BY count(order_detail_id) DESC)
AS ranking
FROM fact_sales fs
LEFT JOIN dim_customer dc USING (customer_id)
LEFT JOIN dim_product dp USING (product_id)
GROUP BY 1,3,4)
SELECT *
FROM ranking
WHERE ranking <= 5
AND maritalstatus is not null
```

RESULT

maritalstatus	count	sub_category	category	ranking
M	777	Tires and Tubes	Accessories	1
M	558	Mountain Bikes	Bikes	2
M	448	Bottles and Cages	Accessories	3
M	363	Road Bikes	Bikes	4
M	343	Helmets	Accessories	5
S	554	Mountain Bikes	Bikes	1
S	477	Tires and Tubes	Accessories	2
S	334	Bottles and Cages	Accessories	3
S	307	Road Bikes	Bikes	4
S	254	Helmets	Accessories	5

INSIGHT

We can see that married people has bigger transaction of the product than single people in our store. Married people have interested in category accessories and single people have interested in category bikes. If we try to promote the product , these will help the store to decide the target audience.

7. Product with The Biggest Profit

QUERY

```
SELECT sub_category, category, sum(totalprice_rupiah - totalcost_rupiah) AS profit,
count(order_detail_id) AS jumlah_pembelian
FROM fact_sales fs
LEFT JOIN dim_product dp ON fs.product_id = dp.product_id
GROUP BY 1,2
ORDER BY profit desc
LIMIT 10
```

RESULT

sub_category	category	profit	jumlah_pembelian
Mountain Bikes	Bikes	17100270907	1113
Road Bikes	Bikes	9813019228	670
Touring Bikes	Bikes	4858806607	467
Helmets	Accessories	183071244	597
Tires and Tubes	Accessories	161341388	1254
Bottles and Cages	Accessories	50885080	782
Hydration Packs	Accessories	44819676	93
Jerseys	Clothing	41678216	247
Fenders	Accessories	36215004	188
Bike Racks	Accessories	30498720	29

INSIGHT

Category bikes and accessories is category of the product that has biggest profit. But mountain bikes are the sub of category product that has biggest profit with the big amout of purchase. We can increase the sale of mountain bikes for each branch store to get bigger profit

8. Top 3 Winner Customer in 2018-2019

QUERY

```
WITH winner AS(
  SELECT date_part('year', order_date) AS years,
  dc.customer_id, dc.customer_name,
  count(distinct order_detail_id) AS count,
  ROW_NUMBER() OVER (partition by date_part('year', order_date) ORDER BY
count(distinct order_detail_id) DESC) as rownum
  FROM fact_sales fs
  LEFT JOIN dim_customer dc ON fs.customer_id = dc.customer_id
  GROUP BY 1,2,3)

SELECT *
FROM winner
WHERE rownum <=3
AND years BETWEEN 2018.0 AND 2019.0
```

RESULT

years	customer_id	customer_name	count	rownum
2018.0	11331	Samantha Jenkins	34	1
2018.0	11287	Henry Garcia	32	2
2018.0	11176	Mason Roberts	31	3
2019.0	11300	Fernando Barnes	14	1
2019.0	11185	Ashley Henderson	14	2
2019.0	11223	Hailey Patterson	13	3

INSIGHT

These tables use when we try to find the winner of the customer based on the biggest transaction.

9. City with the Biggest Transaction of Bikes, Accessories and Clothing

QUERY

```
WITH bigger AS (
  SELECT dp.category, dg.city, count(order_detail_id) AS jumlah,
sum(totalprice_rupiah - totalcost_rupiah) AS profit
  ROW_NUMBER() OVER (partition by category ORDER BY count(distinct order_detail_id)
DESC) AS rownum_buyer
  FROM fact_sales fs
  LEFT JOIN dim_customer dc USING (customer_id)
  LEFT JOIN dim_product dp USING (product_id)
  LEFT JOIN dim_geography dg USING (geography_id)
  GROUP BY 1,2)
SELECT *
FROM bigger
WHERE rownum_buyer <= 3
```

RESULT

category	city	jumlah	profit	rownum_buyer
Accessories	Cliffside	123	18630998	1
Accessories	London	79	13720717	2
Accessories	Haney	73	11051905	3
Bikes	London	73	957393695	1
Bikes	Paris	44	580798618	2
Bikes	Warrnambool	34	499496045	3
Clothing	Port Hammond	25	5014704	1
Clothing	Cliffside	23	3912978	2
Clothing	Haney	22	3806792	3

INSIGHT

Bikes is category product that has biggest profit. If we want to increase the sales of bikes, we can promote it in Londin, Paris, and Warrnambol. If we want to increase the sales of chloting, we can promote it in Port Hammond, Cliffside, and Haney

10. Biggest Sales of Bikes Based on City in 2016-2019

QUERY

```
WITH bigger AS(
  SELECT date_part('year', order_date) AS years, dp.category, dg.city, count(distinct
order_detail_id) AS jumlah,
  ROW_NUMBER() OVER (partition by date_part('year', fs.order_date) ORDER BY
count(distinct order_detail_id) DESC) AS rownum_buyer
  FROM fact_sales fs
  LEFT JOIN dim_customer dc USING (customer_id)
  LEFT JOIN dim_product dp USING (product_id)
  LEFT JOIN dim_geography dg USING (geography_id)
  WHERE category = 'Bikes'
  GROUP BY 1,2,3)
SELECT *
FROM bigger
WHERE rownum_buyer <= 3
```

RESULT

years	category	city	jumlah	rownum_buyer
2016.0	Bikes	London	9	1
2016.0	Bikes	Burbank	9	2
2016.0	Bikes	Warrnambool	7	3
2017.0	Bikes	London	15	1
2017.0	Bikes	Newton	13	2
2017.0	Bikes	Oak Bay	11	3
2018.0	Bikes	London	26	1
2018.0	Bikes	Paris	26	2
2018.0	Bikes	Les Ulis	15	3
2019.0	Bikes	London	23	1
2019.0	Bikes	Hawthorne	12	2
2019.0	Bikes	Berlin	11	3

INSIGHT

From 2016 - 2019, London is the city that has biggest sales of the bikes. The trend of bike sales is always increase every year. If we want to increase the sales of bikes in the next year, we can focusing to promote it in London and some of another city

11. Biggest Cost of The Product in 2016-2019

QUERY

```
WITH bigger AS(
  SELECT date_part('year', order_date) AS years, dp.category, dg.city,
totalcost_rupiah,
  ROW_NUMBER() OVER (partition by date_part('year', fs.order_date) ORDER BY
totalcost_rupiah DESC) AS rownum
  FROM fact_sales fs
  LEFT JOIN dim_product dp USING (product_id)
  GROUP BY 1,2,3)
SELECT *
FROM bigger
WHERE rownum <= 1
```

RESULT

years	category	totalcost_rupiah	rownum
2016.0	Bikes	30398119	1
2017.0	Bikes	30398119	1
2018.0	Bikes	21769271	1
2019.0	Bikes	26573322	1

INSIGHT

Bikes is category of the product that has biggest cost for each years. But as we can se the trend of the cost always decrease every year.

Thank You

adristi.shafiya@gmail.com

linkedin.com/in/adristi-shafiya

github.com/adristi99

TAKE A FURTHER LOOK

SCAN ME !