



COFFEE SHOP TRANSACTION

Aliriza Hamonangan
Matondang

Outline

01

Data understanding & manipulating

Understanding the dataset with simple query and manipulate the data for preparation

02

Datamart Preparation

Preparing, summarising, and exporting the dataset with postgresql

03

Data Visualization

Created data visualization on Tableau & Exploratory Data Analysis

04

Creating Dashboard & Story

Created Dashboard and Data Story on Tableau & Exploratory Data Analysis

05

Conclusion & Recommendation

The Conclusion & Recommendation of the Coffee Transaction Analysis

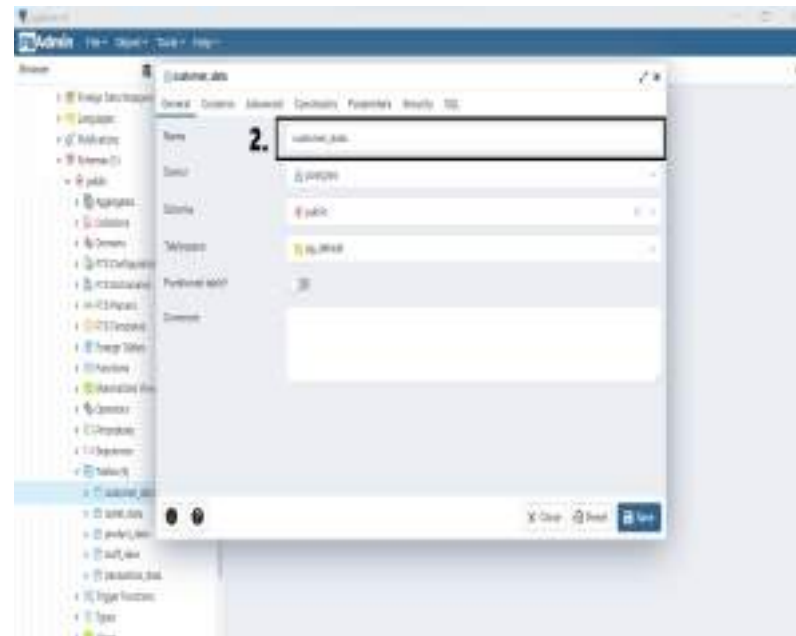
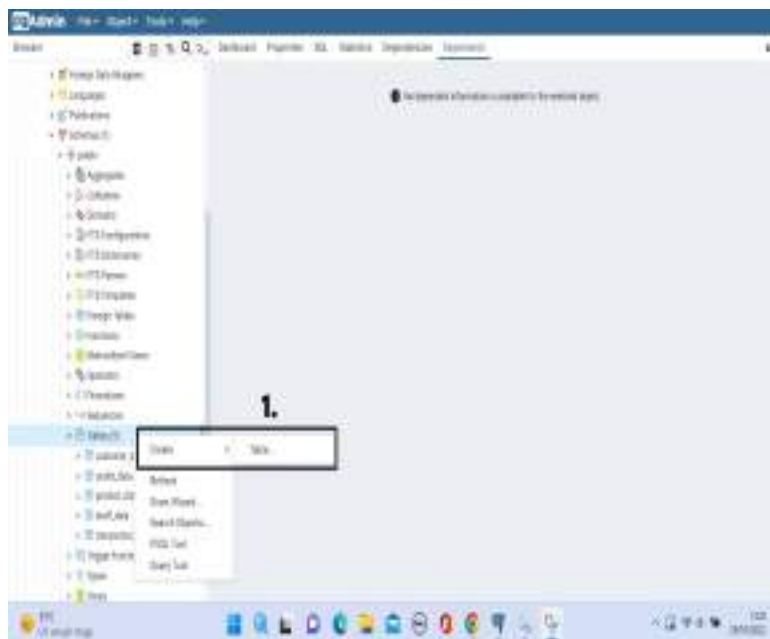
To find the Data Story Dashboard, you can click [here](#)

If you want to see only the analysis, you can click [here](#)

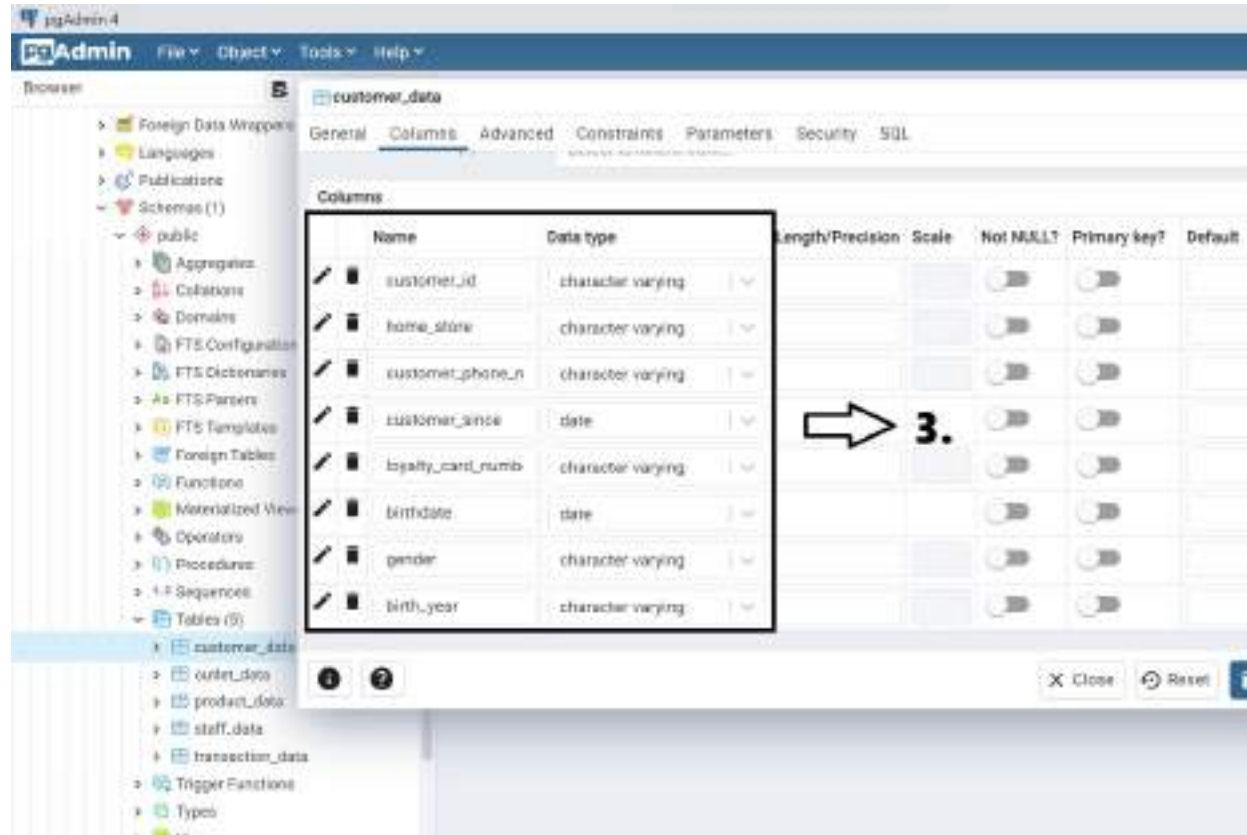
Import the dataset



Import the dataset

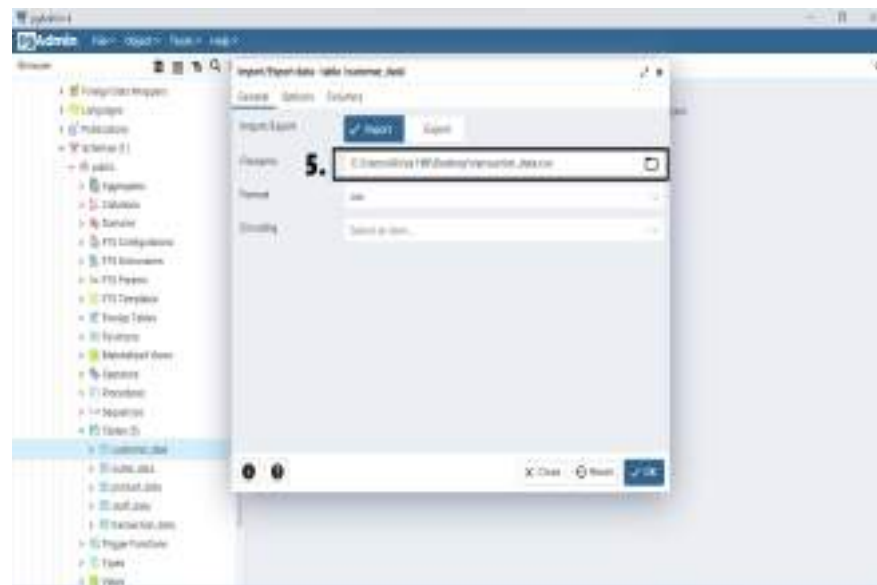
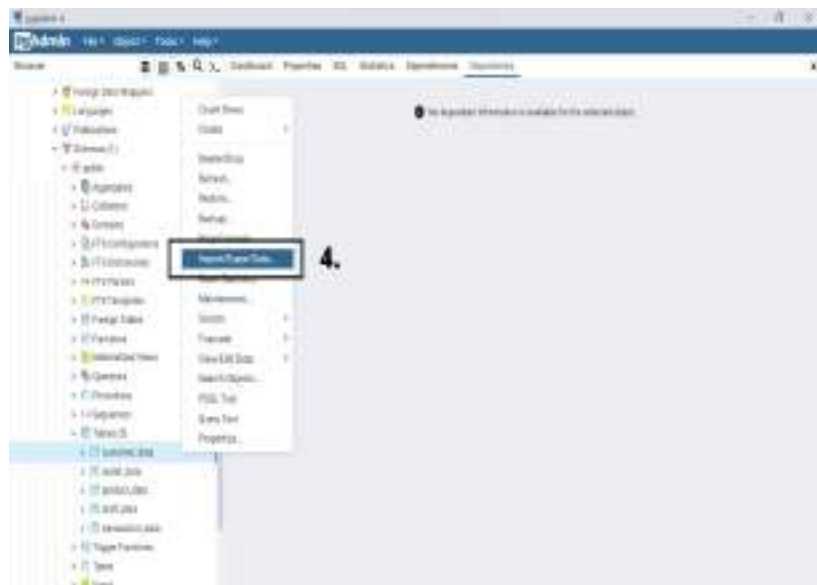


Import the dataset



Insert the columns and Data types from the csv file. Ensure the name of the columns and Data types are same

Import the dataset



Data manipulation



Data manipulation

```

with
transaction as (
select d."customer_id",
      EXTRACT(YEAR FROM age(cast(birthdate as date))) as age,
      d."gender",
      age(d."customer_since") as customer_since,
      b."staff_id",
      e."first_name",
      e."position",
      c."store_city",
      b."order",
      b."transaction_date",
      b."transaction_time",
      a."product",
      a."product_category",
      a."new_product_yn",
      b."sales_outlet_id" as outlet,
      b."quantity",
      a."current_retail_price" as unit_price
from product_data a
left join transaction_data b
on a."product_id" = b."product_id"
left join outlet_data c
on c.sales_outlet_id = b.sales_outlet_id
left join customer_data d
on d.customer_id = b.customer_id
left join staff_data e
on e.staff_id = b.staff_id
)

select * from transaction
  
```

Did join the table with relevant columns for analyze the transaction using ats for make easier to do simple query

customer_id	age	gender	customer_since	staff_id	first_name	position	store_city	order	transaction_date	transaction_time	product	product_category	new_product_yn	outlet	quantity	unit_price
1	100	M	1	1	John	Staff Manager	Atlanta	1	2018-01-01	10:00:00	Product A	Food & Beverage	N	1	10	10.00
2	100	M	1	2	Andrew	Staff Manager	Atlanta	2	2018-01-01	10:00:00	Product B	Food & Beverage	N	2	20	20.00
3	100	M	1	3	Andrew	Staff Manager	Atlanta	3	2018-01-01	10:00:00	Product C	Food & Beverage	N	3	30	30.00
4	100	M	1	4	Andrew	Staff Manager	Atlanta	4	2018-01-01	10:00:00	Product D	Food & Beverage	N	4	40	40.00
5	100	M	1	5	Andrew	Staff Manager	Atlanta	5	2018-01-01	10:00:00	Product E	Food & Beverage	N	5	50	50.00
6	100	M	1	6	Andrew	Staff Manager	Atlanta	6	2018-01-01	10:00:00	Product F	Food & Beverage	N	6	60	60.00
7	100	M	1	7	Andrew	Staff Manager	Atlanta	7	2018-01-01	10:00:00	Product G	Food & Beverage	N	7	70	70.00
8	100	M	1	8	Andrew	Staff Manager	Atlanta	8	2018-01-01	10:00:00	Product H	Food & Beverage	N	8	80	80.00
9	100	M	1	9	Andrew	Staff Manager	Atlanta	9	2018-01-01	10:00:00	Product I	Food & Beverage	N	9	90	90.00
10	100	M	1	10	Andrew	Staff Manager	Atlanta	10	2018-01-01	10:00:00	Product J	Food & Beverage	N	10	100	100.00
11	100	M	1	11	Andrew	Staff Manager	Atlanta	11	2018-01-01	10:00:00	Product K	Food & Beverage	N	11	110	110.00
12	100	M	1	12	Andrew	Staff Manager	Atlanta	12	2018-01-01	10:00:00	Product L	Food & Beverage	N	12	120	120.00
13	100	M	1	13	Andrew	Staff Manager	Atlanta	13	2018-01-01	10:00:00	Product M	Food & Beverage	N	13	130	130.00
14	100	M	1	14	Andrew	Staff Manager	Atlanta	14	2018-01-01	10:00:00	Product N	Food & Beverage	N	14	140	140.00
15	100	M	1	15	Andrew	Staff Manager	Atlanta	15	2018-01-01	10:00:00	Product O	Food & Beverage	N	15	150	150.00
16	100	M	1	16	Andrew	Staff Manager	Atlanta	16	2018-01-01	10:00:00	Product P	Food & Beverage	N	16	160	160.00
17	100	M	1	17	Andrew	Staff Manager	Atlanta	17	2018-01-01	10:00:00	Product Q	Food & Beverage	N	17	170	170.00
18	100	M	1	18	Andrew	Staff Manager	Atlanta	18	2018-01-01	10:00:00	Product R	Food & Beverage	N	18	180	180.00
19	100	M	1	19	Andrew	Staff Manager	Atlanta	19	2018-01-01	10:00:00	Product S	Food & Beverage	N	19	190	190.00
20	100	M	1	20	Andrew	Staff Manager	Atlanta	20	2018-01-01	10:00:00	Product T	Food & Beverage	N	20	200	200.00
21	100	M	1	21	Andrew	Staff Manager	Atlanta	21	2018-01-01	10:00:00	Product U	Food & Beverage	N	21	210	210.00
22	100	M	1	22	Andrew	Staff Manager	Atlanta	22	2018-01-01	10:00:00	Product V	Food & Beverage	N	22	220	220.00
23	100	M	1	23	Andrew	Staff Manager	Atlanta	23	2018-01-01	10:00:00	Product W	Food & Beverage	N	23	230	230.00
24	100	M	1	24	Andrew	Staff Manager	Atlanta	24	2018-01-01	10:00:00	Product X	Food & Beverage	N	24	240	240.00
25	100	M	1	25	Andrew	Staff Manager	Atlanta	25	2018-01-01	10:00:00	Product Y	Food & Beverage	N	25	250	250.00
26	100	M	1	26	Andrew	Staff Manager	Atlanta	26	2018-01-01	10:00:00	Product Z	Food & Beverage	N	26	260	260.00
27	100	M	1	27	Andrew	Staff Manager	Atlanta	27	2018-01-01	10:00:00	Product A	Food & Beverage	N	27	270	270.00
28	100	M	1	28	Andrew	Staff Manager	Atlanta	28	2018-01-01	10:00:00	Product B	Food & Beverage	N	28	280	280.00
29	100	M	1	29	Andrew	Staff Manager	Atlanta	29	2018-01-01	10:00:00	Product C	Food & Beverage	N	29	290	290.00
30	100	M	1	30	Andrew	Staff Manager	Atlanta	30	2018-01-01	10:00:00	Product D	Food & Beverage	N	30	300	300.00
31	100	M	1	31	Andrew	Staff Manager	Atlanta	31	2018-01-01	10:00:00	Product E	Food & Beverage	N	31	310	310.00
32	100	M	1	32	Andrew	Staff Manager	Atlanta	32	2018-01-01	10:00:00	Product F	Food & Beverage	N	32	320	320.00
33	100	M	1	33	Andrew	Staff Manager	Atlanta	33	2018-01-01	10:00:00	Product G	Food & Beverage	N	33	330	330.00
34	100	M	1	34	Andrew	Staff Manager	Atlanta	34	2018-01-01	10:00:00	Product H	Food & Beverage	N	34	340	340.00
35	100	M	1	35	Andrew	Staff Manager	Atlanta	35	2018-01-01	10:00:00	Product I	Food & Beverage	N	35	350	350.00
36	100	M	1	36	Andrew	Staff Manager	Atlanta	36	2018-01-01	10:00:00	Product J	Food & Beverage	N	36	360	360.00
37	100	M	1	37	Andrew	Staff Manager	Atlanta	37	2018-01-01	10:00:00	Product K	Food & Beverage	N	37	370	370.00
38	100	M	1	38	Andrew	Staff Manager	Atlanta	38	2018-01-01	10:00:00	Product L	Food & Beverage	N	38	380	380.00
39	100	M	1	39	Andrew	Staff Manager	Atlanta	39	2018-01-01	10:00:00	Product M	Food & Beverage	N	39	390	390.00
40	100	M	1	40	Andrew	Staff Manager	Atlanta	40	2018-01-01	10:00:00	Product N	Food & Beverage	N	40	400	400.00
41	100	M	1	41	Andrew	Staff Manager	Atlanta	41	2018-01-01	10:00:00	Product O	Food & Beverage	N	41	410	410.00
42	100	M	1	42	Andrew	Staff Manager	Atlanta	42	2018-01-01	10:00:00	Product P	Food & Beverage	N	42	420	420.00
43	100	M	1	43	Andrew	Staff Manager	Atlanta	43	2018-01-01	10:00:00	Product Q	Food & Beverage	N	43	430	430.00
44	100	M	1	44	Andrew	Staff Manager	Atlanta	44	2018-01-01	10:00:00	Product R	Food & Beverage	N	44	440	440.00
45	100	M	1	45	Andrew	Staff Manager	Atlanta	45	2018-01-01	10:00:00	Product S	Food & Beverage	N	45	450	450.00
46	100	M	1	46	Andrew	Staff Manager	Atlanta	46	2018-01-01	10:00:00	Product T	Food & Beverage	N	46	460	460.00
47	100	M	1	47	Andrew	Staff Manager	Atlanta	47	2018-01-01	10:00:00	Product U	Food & Beverage	N	47	470	470.00
48	100	M	1	48	Andrew	Staff Manager	Atlanta	48	2018-01-01	10:00:00	Product V	Food & Beverage	N	48	480	480.00
49	100	M	1	49	Andrew	Staff Manager	Atlanta	49	2018-01-01	10:00:00	Product W	Food & Beverage	N	49	490	490.00
50	100	M	1	50	Andrew	Staff Manager	Atlanta	50	2018-01-01	10:00:00	Product X	Food & Beverage	N	50	500	500.00
51	100	M	1	51	Andrew	Staff Manager	Atlanta	51	2018-01-01	10:00:00	Product Y	Food & Beverage	N	51	510	510.00
52	100	M	1	52	Andrew	Staff Manager	Atlanta	52	2018-01-01	10:00:00	Product Z	Food & Beverage	N	52	520	520.00

Data understanding



Data understanding

```
1 with
2 transaction as (
3   select s."customer_id",
4         EXTRACT(YEAR FROM age(date(s.transaction_date))) as age,
5         s.gender,
6         age(s."customer_id") as customer_age,
7         s."staff_id",
8         s."first_name",
9         s."last_name",
10        s."work_email",
11        s."phone_number",
12        s."email_address",
13        s."product_id",
14        s."product_name",
15        s."new_product_id",
16        s."sales_order_id" as order_id,
17        s."quantity",
18        s."current_stock_id" as unit_stock
19  from product_data s
20  left join transaction_data t
21    on s."product_id" = t."product_id"
22  left join sales_order s2
23    on s2.sales_order_id = t.sales_order_id
24  left join customer_data c
25    on c.customer_id = t.customer_id
26  left join staff_data s3
27    on s3.staff_id = t.staff_id
28 )
29
30 select count(*) as total_transaction,
31        count(distinct customer_id) as total_customer,
32        extract(year from transaction_date) as year
33  from transaction
34  where transaction_date is not null
35  group by year
```

Data output Messages Notifications

	total_transaction bigint	total_customer bigint	year numeric
1	49902	2245	2019

There are
49.902
transactions
from 2.245
Customers at
The Coffee
Shop in 2019

Data understanding

```
Query Query History
1 with
2 transaction as (
3     select a."customer_id",
4           EXTRACT(YEAR FROM extract(epoch(timestamp) as text)) as year,
5           a."gender",
6           age(a."customer_name") as customer_age,
7           a."self_id",
8           a."first_name",
9           a."location",
10          a."store_id",
11          b."order_id",
12          a."transaction_date",
13          a."transaction_time",
14          b."promo_item_id",
15          a."product",
16          a."product_category",
17          a."new_product_id",
18          b."sales_outlet_id" as outlet,
19          b."quantity",
20          a."current_item_price" as unit_price
21 from product_data a
22 left join transaction_data b
23 on a."product_id" = b."product_id"
24 left join outlet_data c
25 on c.sales_outlet_id = b.sales_outlet_id
26 left join customer_data d
27 on d.customer_id = b.customer_id
28 left join staff_data e
29 on e.staff_id = b.staff_id
30 )
31
32 select product,
33        sum(order_id) as total_order,
34        sum(quantity) as total_quantity,
35        promo_item_id,
36        unit_price
37 from transaction
38 where (order_id IS NOT null) and (quantity IS NOT null)
39 GROUP BY 1,4,5
40 ORDER BY total_order DESC
41 Limit 10
```

Date output Messages Notifications

product	total_order	total_quantity	promo_item_id	unit_price
chocolate croissant	1182	1008	M	48000
Brazilian Rg	1016	5558	M	36000
Latte	1015	5551	M	45000
Dark chocolate Lg	1004	5548	M	34000
Sustainable Scone Organic Lg	995	5498	M	57000
Columbian Medium Roast Cg	891	5562	M	38000
Morning Bunni Chai Rg	890	5510	M	38000
Latte Rg	887	5488	M	51000
Traditional Blend Chai Rg	885	5487	M	38000
Cappuccino	880	5485	M	48000

The top 10 products order of the Coffee Shop

1. Chocolate Croissant has the highest order with price Rp45.000,-
2. Brazilian Rg has the lowest order from the top ten order with price Rp36.00,-
3. The top 10 products order of the Coffee Shop are from 80 products
4. If we add promo item, there are 3 products that have promo, specifically :
 - Chocolate Croissant 2 orders,
 - Ginger Scone 232 Orders,
 - Ouro Brasileiro shot 256 Orders

Data understanding

```
Query Query History
1 with
2 transaction as (
3 select c."customer_id",
4        EXTRACT(YEAR FROM age(cast(birthdate as date))) as age,
5        c."gender",
6        age(d, (customer_since)) as customer_since,
7        b."staff_id",
8        a."first_name",
9        a."position",
10       c."store_city",
11       b."order",
12       b."transaction_date",
13       b."transaction_time",
14       b."promo_item_yr",
15       a."product",
16       a."product_category",
17       a."new_product_yr",
18       b."sales_outlet_id" as outlet,
19       b."quantity",
20       a."current_retail_price" as unit_price
21 from product_data a
22 left join transaction_data b
23 on a."product_id" = b."product_id"
24 left join outlet_data c
25 on c."sales_outlet_id" = b."sales_outlet_id"
26 left join customer_data d
27 on d."customer_id" = b."customer_id"
28 left join staff_data e
29 on e."staff_id" = b."staff_id"
30 )
31
32 select product_category,
33        sum("order") as total_order,
34        sum("quantity") as total_quantity
35 from transaction
36 WHERE ("order" IS NOT NULL) and ("quantity" IS NOT NULL)
37 GROUP BY 1
38 ORDER BY total_order DESC
```

Data output Messages Notifications			
	product_category character varying	total_order bigint	total_quantity bigint
1	Coffee	19802	30058
2	Tea	15164	23199
3	Bakery	7948	7764
4	Drinking Chocolate	3844	5848
5	Coffee beans	3842	616
6	Loose Tea	2551	401
7	Flavours	2257	3466
8	Branded	2173	255
9	Packaged Chocol...	966	151

The top order by products category of the Coffee Shop

1. Coffee has the highest order with 19.802 orders
2. Packaged Chocolate has the lowest order
3. Based on the highest order the main products has the highest order, which is coffee is the superior same as the name of the shop and still relevant with the name of the store

Data understanding

```
Query Query History
1 with
2 transaction as (
3 select a."customer_id",
4        EXTRACT(YEAR FROM age(cast(birthdate as date))) as age,
5        a."gender",
6        age/d."customer_since" as customer_since,
7        b."staff_id",
8        a."first_name",
9        a."position",
10       a."store_city",
11       b."order",
12       b."transaction_date",
13       b."transaction_time",
14       b."promo_item_yr",
15       a."product",
16       a."product_category",
17       a."new_product_yr",
18       b."sales_outlet_id" as outlet,
19       b."quantity",
20       b."current_retail_price" as unit_price
21 from product_data a
22 left join transaction_data b
23 on a."product_id" = b."product_id"
24 left join outlet_data c
25 on c."sales_outlet_id" = b."sales_outlet_id"
26 left join customer_data d
27 on d."customer_id" = b."customer_id"
28 left join staff_data e
29 on e."staff_id" = b."staff_id"
30 )
31
32
33 select outlet,
34        store_city,
35        sum("order") as total_order,
36        sum("quantity") as total_quantity
37 from transaction
38 WHERE ("order" IS NOT NULL) and ("quantity" IS NOT NULL)
39 GROUP BY 1,2
40 ORDER BY total_order DESC
```

	outlet	store_city	total_order	total_quantity
	character varying	character varying	bigint	bigint
1	8	Jakarta Selatan	20139	24109
2	3	Jakarta Barat	19251	23620
3	5	Jakarta Selatan	19157	24029

1. Store 8 has the highest order in Jakarta Selatan
2. Store 5 has the lowest order in Jakarta Selatan. Based on the order, the different order is not really far, which is good because they got so many order in Jakarta Selatan
3. The quantity of products at store 5 needs an improvement based on the order because the different total quantity and total order are still so far

Data understanding

```
Query Query History
1 with
2 transaction as (
3 select d."customer_id",
4        EXTRACT(YEAR FROM age(cast(birthdate as date))) as age,
5        d."gender",
6        age(d."customer_since") as customer_since,
7        s."staff_id",
8        s."first_name",
9        s."position",
10       s."store_city",
11       s."order",
12       s."transaction_date",
13       s."transaction_time",
14       s."promo_item_id",
15       s."product",
16       s."product_category",
17       s."res_product_yr",
18       s."sales_outlet_id" as outlet,
19       s."quantity",
20       s."current_retail_price" as unit_price
21 from product_data s
22 left join transaction_data b
23 on s."product_id" = b."product_id"
24 left join outlet_data c
25 on c.sales_outlet_id = b.sales_outlet_id
26 left join customer_data d
27 on d.customer_id = b.customer_id
28 left join staff_data e
29 on e.staff_id = b.staff_id
30 )
31
32 select distinct product_category,
33                gender,
34                sum("order") as total_order
35 from transaction
36 WHERE ("order" IS NOT NULL) and ("gender" IS NOT NULL)
37 GROUP BY 1,2
38 ORDER BY product_category
39
```

Data output Messages Notifications			
	product_category character varying	gender character varying	total_order bigint
1	Bakery	L	1081
2	Bakery	P	1280
3	Branded	L	297
4	Branded	P	425
5	Coffee	L	4178
6	Coffee	P	5031
7	Coffee beans	L	525
8	Coffee beans	P	819
9	Drinking Chocolate	L	819
10	Drinking Chocolate	P	1115
11	Flavours	L	297
12	Flavours	P	490
13	Loose Tea	L	391
14	Loose Tea	P	576
15	Packaged Chocolate	L	185
16	Packaged Chocolate	P	238
17	Tea	L	3280
18	Tea	P	1622

1. The women have the highest order in every category products
2. Needs improvement of the products that attract the men to increase the order
3. Create the event or design interior that make the men feel comfort

Data understanding

Query QueryHistory

```
1 with
2 transaction as (
3   select c."customer_id",
4         EXTRACT(YEAR FROM age(cast(birthdate as date))) as age,
5         c."gender",
6         age(c."customer_since") as customer_age,
7         b."staff_id",
8         a."first_order",
9         a."product",
10        a."store_city",
11        b."order",
12        b."transaction_date",
13        b."transaction_time",
14        b."gross_total",
15        a."product",
16        a."product_category",
17        a."new_product_id",
18        b."sales_outlet_id" as outlet,
19        b."quantity",
20        a."current_retail_price" as unit_price
21 from product_data a
22 left join transaction_data b
23 on a."product_id" = b."product_id"
24 left join outlet_data c
25 on c."sales_outlet_id" = b."sales_outlet_id"
26 left join customer_data d
27 on d."customer_id" = b."customer_id"
28 left join staff_data e
29 on e."staff_id" = b."staff_id"
30 )
31
32 select distinct gender,
33        avg("order") as average_order,
34        sum("order") as total_order,
35        outlet,
36        store_city
37 from transaction
38 where ("order" IS NOT NULL) and ("gender" IS NOT NULL) and ("outlet" IS NOT NULL) and ("store_city" IS NOT NULL)
39 group by 1,4,5
40 order by total_order desc
```

Data output Messages Notifications					
	gender	average_order	total_order	outlet	store_city
	character varying	numeric	bigint	character varying	character varying
1	F	1.1307564268548	6114	5	Jakarta Selatan
2	F	1.1683166621761	5613	8	Jakarta Selatan
3	F	1.0996275603214	4724	3	Jakarta Barat
4	L	1.0400882010769	4697	3	Jakarta Barat
5	L	1.1576251115280	3832	5	Jakarta Selatan
6	L	1.1353696482501	3285	8	Jakarta Selatan

1. The women have dominant order at store 5
2. The average order is 1
3. The men have lower order than women

Data understanding

```
Query Queryhistory
1 with
2 transaction as (
3 select d."customer_id",
4        EXTRACT(YEAR FROM age((cast(birthdate as date))) as age,
5        e."gender",
6        age(d."customer_since") as customer_since,
7        b."staff_id",
8        a."first_name",
9        e."postal_code",
10       c."store_city",
11       b."order_id",
12       b."transaction_date",
13       b."transaction_time",
14       b."promo_name_yr",
15       a."product",
16       a."product_category",
17       a."new_product_yr",
18       b."sales_outlet_id" as outlet,
19       b."quantity",
20       a."current_retail_price" as unit_price
21 from product_data a
22 left join transaction_data b
23 on a."product_id" = b."product_id"
24 left join outlet_data c
25 on c.sales_outlet_id = b.sales_outlet_id
26 left join customer_data d
27 on d.customer_id = b.customer_id
28 left join staff_data e
29 on e.staff_id = b.staff_id
30 )
31
32 select sum("order") as total_order,
33        product_category,
34        outlet,
35        extract(YEAR from transaction_time) as transaction_time
36 from transaction
37 WHERE ("order" IS NOT NULL) AND ("transaction_time" IS NOT NULL)
38 GROUP BY 2,3,4
39 ORDER BY total_order DESC
40 limit 10
```

Data output					Messages	Notifications
	total_order	product_category	outlet	transaction_time		
1	805	Coffee	8	8		
2	803	Coffee	8	10		
3	859	Coffee	8	9		
4	771	Coffee	8	10		
5	726	Coffee	8	8		
6	725	Coffee	8	9		
7	723	Coffee	8	7		
8	580	Tea	8	8		
9	071	Tea	8	9		
10	667	Coffee	8	9		

1. The highest order is in the morning until afternoon
2. Everyone needs coffee in the morning in every store

Data understanding

Query Query History

```
1 with
2 transaction as (
3 select d."customer_id",
4        EXTRACT(YEAR FROM age(cast(birthdate as date)) -> age)
5        d."gender",
6        Age(d."customer_since") as customer_since,
7        b."staff_id",
8        a."first_name",
9        a."position",
10       c."store_city",
11       b."order",
12       b."transaction_date",
13       b."transaction_time",
14       b."promo_item_nm",
15       a."product",
16       a."product_category",
17       a."new_product_yn",
18       b."sales_outlet_id" as outlet,
19       b."quantity",
20       a."current_retail_price" as unit_price
21 from product_data a
22 left join transaction_data b
23 on a."product_id" = b."product_id"
24 left join outlet_data d
25 on d."sales_outlet_id" = b."sales_outlet_id"
26 left join customer_data c
27 on c."customer_id" = b."customer_id"
28 left join staff_data e
29 on e."staff_id" = b."staff_id"
30 )
31
32 select AVG(age) as average_age,
33        AVG(customer_since) as average_customer_since,
34        gender,
35        count(gender) as total_gender
36 from transaction
37 where ("age" IS NOT NULL) AND ("customer_since" IS NOT NULL) AND ("gender" IS NOT NULL)
38 group by 3
```

1. The average age of the customers between 42 - 46 years old
2. The customers are loyal with the average age of has been becoming customers 4 years

Data output Messages Notifications

	average_age interval	average_customer_since interval	gender character varying	total_gender bigint
1	46 years 4 mo...	4 years 8 mons 19 day...	L	10324
2	42 years 4 mo...	4 years 7 mons 27 day...	P	14528

Data understanding

Query Query History

```
1 with
2 transaction as (
3 select d."customer_id",
4        EXTRACT(YEAR FROM age(cast(birthdate as date))) as age,
5        d."gender",
6        age(d."customer_since") as customer_since,
7        b."staff_id",
8        a."first_name",
9        c."position",
10       c."store_city",
11       b."order",
12       b."transaction_date",
13       b."transaction_time",
14       b."promo_item_yr",
15       a."product",
16       a."product_category",
17       a."new_product_yr",
18       b."sales_outlet_id" as outlet,
19       b."quantity",
20       e."current_retail_price" as unit_price
21 from product_data a
22 left join transaction_data b
23 on a."product_id" = b."product_id"
24 left join outlet_data c
25 on c."sales_outlet_id" = b."sales_outlet_id"
26 left join customer_data d
27 on d."customer_id" = b."customer_id"
28 left join staff_data e
29 on e."staff_id" = b."staff_id"
30 )
31
32 select product_category,
33        gender,
34        SUM("unit_price") as total_price
35 from transaction
36 WHERE ("gender" IS NOT NULL)
37 GROUP BY 1,2
38 ORDER BY product_category
```

Data output Messages Notifications

	product_category character varying	gender character varying	total_price double precision
1	Bakery	L	6798000
2	Bakery	P	9291000
3	Branded	P	10708000
4	Branded	L	8600000
5	Coffee	L	151060800
6	Coffee	P	210322200
7	Coffee beans	P	21257000
8	Coffee beans	L	22608200
9	Drinking Chocolate	P	57078000
10	Drinking Chocolate	L	40299000
11	Flavours	L	2811800
12	Flavours	P	4665600
13	Loose Tea	L	7092000
14	Loose Tea	P	10072800
15	Packaged Chocola...	L	8010920
16	Packaged Chocola...	P	3969000
17	Tea	L	109829800
18	Tea	P	156215200

1. The different total price between women and is men so high
2. Even for the men product, which is coffee, the total price has spent by the men is still lower than women

Datamart Preparation



Datamart Preparation

Query Query/History

```
1 with
2 transaction as (
3   select d."customer_id",
4         EXTRACT(YEAR FROM age(cast(birthdate as date))) as age,
5         d."gender",
6         age(d."customer_since") as customer_since,
7         b."staff_id",
8         a."first_name",
9         e."position",
10        c."store_city",
11        b."order",
12        b."transaction_date",
13        b."transaction_time",
14        a."product",
15        a."product_category",
16        a."new_product_y_n",
17        b."sales_outlet_id" as outlet,
18        b."quantity",
19        a."current_retail_price" as unit_price
20  from product_data a
21  left join transaction_data b
22  on a."product_id" = b."product_id"
23  left join outlet_data c
24  on c.sales_outlet_id = b.sales_outlet_id
25  left join customer_data d
26  on d.customer_id = b.customer_id
27  left join staff_data e
28  on e.staff_id = b.staff_id
29 );
```

```
rfm as (
  select DISTINCT customer_id,
         extract(day from max(transaction_date)) as last_order_day,
         count("order") as frequency,
         SUM(product_price * quantity) as total_price
  from transaction
  WHERE ("customer_id" IS NOT NULL)
  AND (product_price != 0)
  GROUP BY 1
  order by total_price DESC
);
```

```
rfm_final as (
  select customer_id,
         last_order_day,
         frequency,
         total_price,
         NTILE(4) OVER (ORDER BY last_order_day) AS rfm_recency,
         NTILE(4) OVER (ORDER BY frequency) AS rfm_frequency,
         NTILE(4) OVER (ORDER BY total_price) AS rfm_monetary
  from rfm
);
```

1. Joined to the relevant columns from all tables with ats
2. Created the second ats and new columns are last_order_day, frequency, and total price to do RFM Analysis for Customer Segmentation
3. Created the third as final dataset to do RFM Analysis for Customer Segmentation
4. Divided the new columns into 4 groups for each dimensions (R, F and M)

Datamart Preparation

```
rfm_join as (select customer_id,  
    last_order_day,  
    frequency,  
    total_price,  
    rfm_recency,  
    rfm_frequency,  
    rfm_monetary,  
    rfm_recency*100 + rfm_frequency*10 + rfm_monetary AS rfm_combined  
from rfm_final  
)
```

1. Created new column which is rfm_combined
2. Created the fourth ats for joining to others columns



Datamart Preparation

1. Selected relevant columns to analyze the dataset
2. Downloaded the dataset as csv

Note :

1. We didn't create Segmentation or loyalty member here, because we need to understand the best group for this dataset
2. We would do create Segmentation or loyalty member on tableau because it is easier to see the median, or mean values with data visualization

[illegible]

Export The dataset as CSV

[illegible]

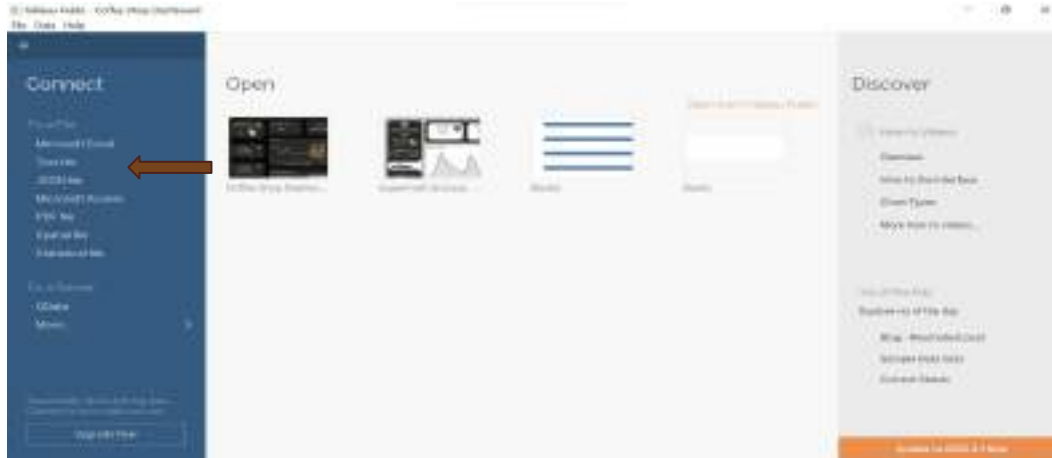
1. After the dataset was ready, clicked download icon
2. Then, saved the dataset



Data Visualization



Data Visualization



1. Added the dataset that have prepared from Postgresql by clicking Microsoft Excel
2. Then, we would go to Data Source for editing the dataset on Data Source such as Split, Join, Change the data type, etc

Note : this visualization for create Dashboard so, the visualization fit with the template of the Dashboard



Data Visualization



1. Total Customers of dataset are 2.245
2. By dragging the feature Customer id to Text and set to Count Distinct



Data Visualization



1. The total GMV of the dataset is Rp1.376B (GMV = Sales Price x Quantity)
2. By dragging the feature GMV to Text



Data Visualization



1. Total Orders of dataset are 28.060
2. By dragging the feature order to Text



Data Visualization

Outlet 3

Rp464M

Outlet 5

Rp430M

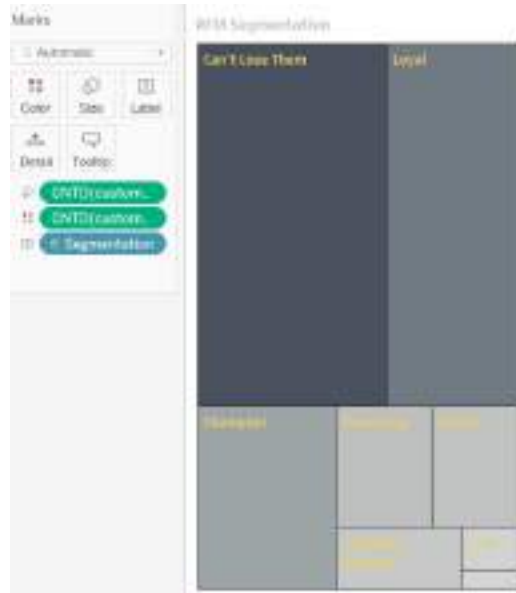
Outlet 8

Rp430M

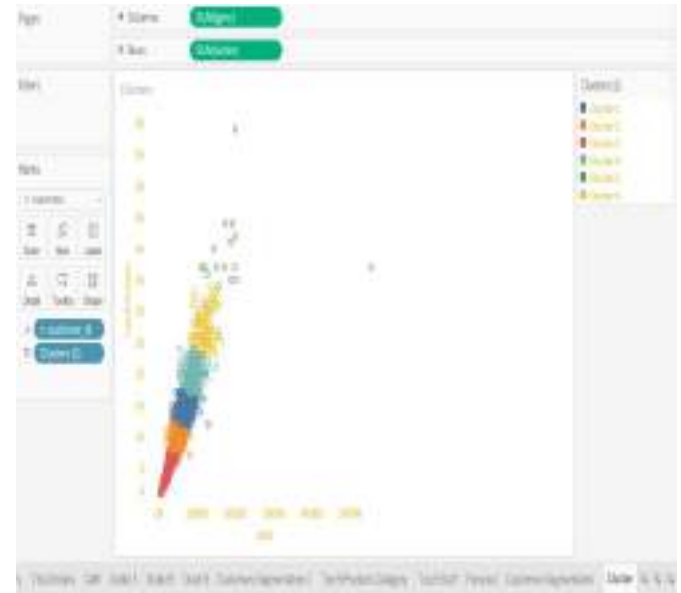


1. The different of total GMV on every outlet is not really different, these indicated that every outlet has the same SOP
2. By dragging the feature GMV to Text and dragging the feature GMV to Filters

Data Visualization



VS



1. RFM Segmentation vs Cluster
2. The Cluster divided into 6 Clusters and The RFM Segmentation divided into 8 Segments
3. The result of cluster based on K-Means is so good, but I prefer to choose RFM Segmentation, because the groups are more segmented based on rank of the GMV, Frequency, and Recency than Cluster.

Data Visualization



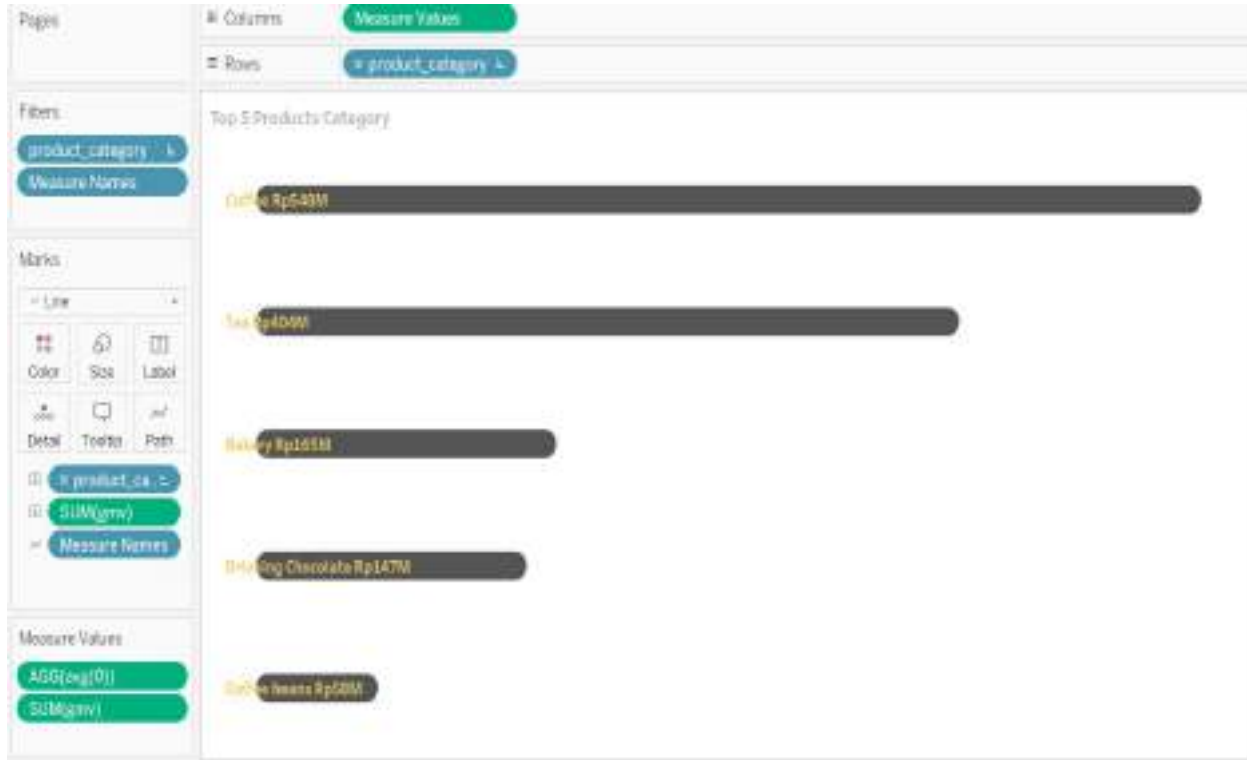
The details of RFM Segmentation

Data Visualization



The details of Cluster

Data Visualization



1. The Top 5 Products Category by GMV
2. Coffee & Tea are the best Products Category
3. To create the visualization, we created the zero axis to separate the GMV with zero axis, so for each bar is started from zero
4. Then we have changed the bar to line in bigger form

Data Visualization



1. The Top 5 Staffs by GMV
2. Adam is the best employee on this dataset, because he works at all outlet, which are Outlet 3,5 and 8
3. To create the visualization, we created the zero axis to separate the GMV with zero axis, so the first number of bar are zero
4. Then we have changed the bar to line in bigger form

Data Visualization



1. The forecast of the GMV for next 13 days
2. Based on the forecast GMV, The Coffee Shop needs to interrupt the drop of GMV
3. The Coffee Shop could use the Segmentation of the Customers to create marketing products to increase the GMV

Data Visualization

1. The Details of the forecast
2. The forecast is quite good based on the model evaluations and summary
3. The MAPE value of 26% means that the difference between the average forecast value and the actual value is 26%

Describe Forecast

Summary Models

Options Used to Create Forecasts

Time series: Day of transaction_date

Measures: Sum of gmv

Forecast forward: 13 days (29 April 2019 – 11 May 2019)

Forecast based on: 1 April 2019 – 28 April 2019

Ignore last: 1 day (29 April 2019)

Seasonal pattern: 7 day cycle

Sum of gmv

Initial	Change From Initial	Seasonal Effect	Contribution	Quality
29 April 2019	29 April 2019 – 11 May 2019	High	Trend Season	
0.000.100 a 0.000.251	1.340.351	0 May 2019 - 1 0 May 2019 - 1	0.0% - 100.0%	C6

☐ Show values as percentages

Copy to Clipboard [Learn more about the forecast summary](#)

Close

Describe Forecast

Summary Models

All forecasts were computed using exponential smoothing.

Sum of gmv

Model			Quality Metrics					Smoothing Coefficients		
Level	Trend	Season	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Multiplicative	None	Multiplicative	4,838.569	3,723.289	0,47	26.0%	883	0.500	0.000	0.481

Copy to Clipboard [Learn more about the forecast models](#)

Close

Forecast Options

Forecast Length

☒ Automatic: Next 13 days

☐ Exactly 1

Years

☐ Until 1

Years

Source Data

Aggregate by: Automatic (Days)

Ignore last: 1 Days

☐ Fill in missing values with zeroes

Forecast Model

Custom

Trend: None Season: Multiplicative

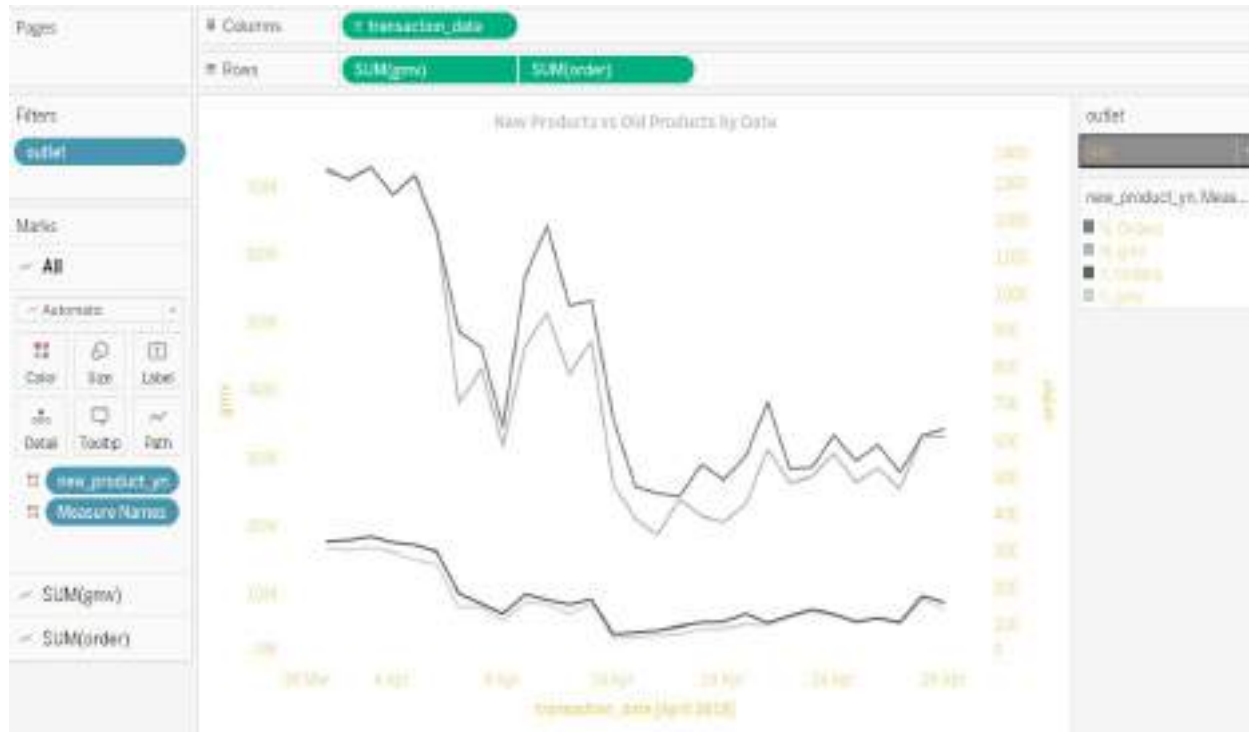
☒ Show prediction intervals 95%

Currently using source data from 1 April 2019 to 28 April 2019 to create a forecast through 11 May 2019. Looking for potential seasonal patterns every 7 Days.

[Learn more about forecast options](#)

OK

Data Visualization



1. The old products have downward trend, and the new products seem steady, and tend to downward trend.
2. Slightly different between order and GMV means, the new products indicate their fair price
3. The old products price is not quite fair so they got so many orders, but received low GMV
4. Overall of the new products performance is quite good for a month based on the trend

Data Visualization



1. The top and bottom 7 new and old products
2. Based on the visualization, 11 new products can compete with old products
3. This is indicated that we can remove the bottom products and change with the top 11 new products to streamline operational costs
4. However, we need deeply analyze because the data provided is only a month.

Creating Dashboard

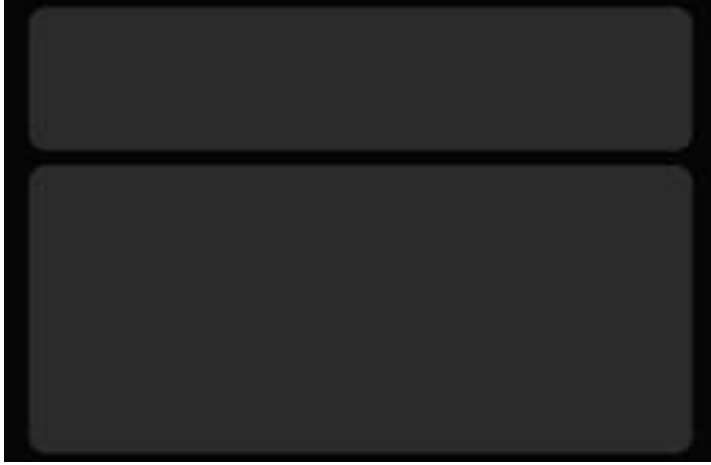
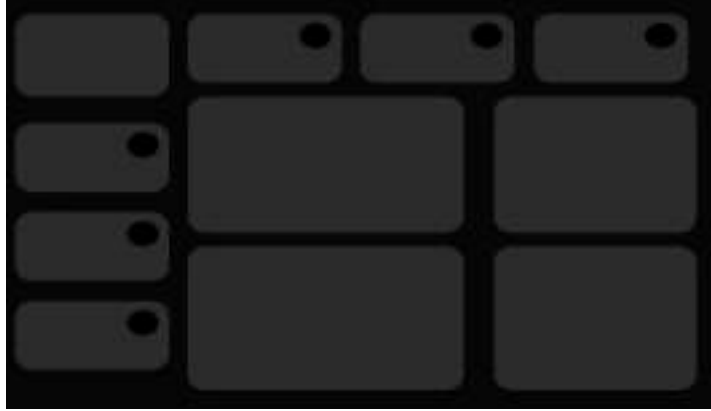




Note

These Dashboard use Filter on Outlet and Icon on the Title. You can click the Icon and the notification “Go to Outlet Details”, on Outlet.

Creating Dashboard



1. First, create template on Figma
2. Second, upload the template to Tableau
3. Make sure the size of the template fit with the Dashboard
4. We created 3 Dashboards on this dataset

The Dashboard



1. The overall performance of Outlet 3 is good, because there is slightly different among GMV in every outlet
2. The Outlet 3 has so many Customers & Orders
3. The Outlet 3, based on the Forecast needs to interrupt the drop or stable on low of GMV
4. The Outlet 3 needs to do marketing strategic based on the segmentation
5. The outlet 3 can focus on the customers on the Segmentation Can't Lose Them, At Risk, and Customer Needing Attention to increase the GMV

The Dashboard



1. The overall performance of Outlet 5 is good, because there is slightly different among GMV in every outlet
2. The Outlet 5 is the best Outlet on the Coffee Shop
3. The Outlet 5 can focus on Promising Customers to increase the GMV and find the new customers

The Dashboard



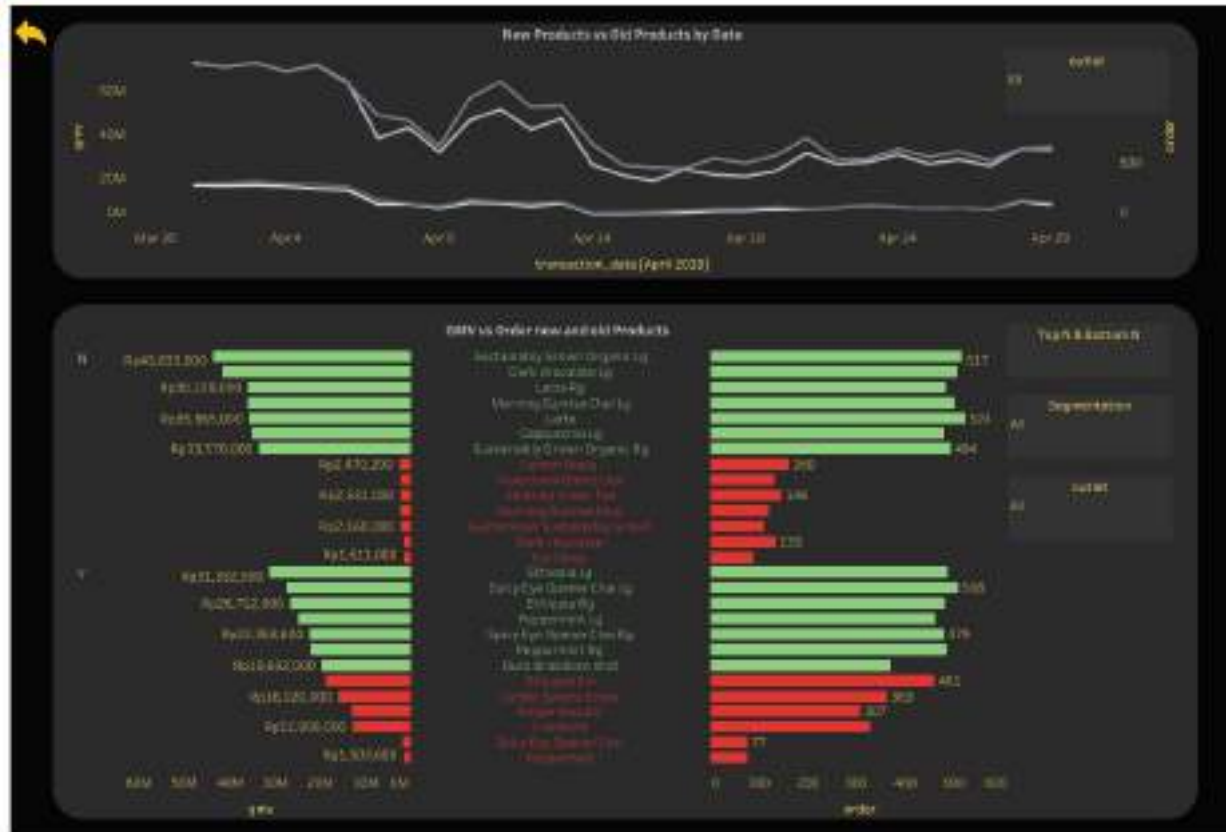
1. The overall performance of Outlet 3 is good, because there is slightly different among GMV in every outlet
2. The Outlet 8 is the worst Outlet on the Coffee Shop based on the GMV, Total Customers, and Total Orders
3. The Outlet 8 has so many Loyal and Champion Customers, which means the customers feel comfort with Outlet 8
4. The outlet 8 can focus on Can't Lose Them Segmentation and find new customers to increase the GMV and Orders

The Dashboard



1. The details of RFM Segmentation
2. I prefer to choose RFM Segmentation because the groups are more segmented based on rank of the GMV, Frequency, and Recency than Cluster
3. Female spent more than Male
4. More female than male Customers
5. The avg of the age is > 40 years old
6. If the Coffee Shop wants to get the new customers, it is better to get the young generation customers, because the Coffee Shop dominated by old generation customers

The Dashboard



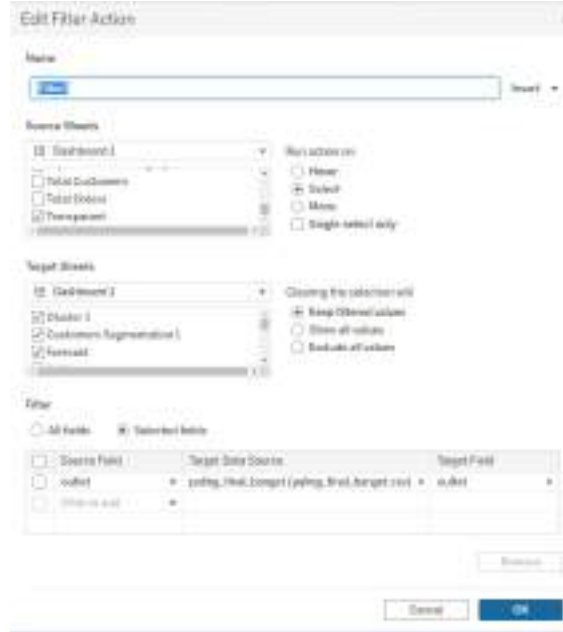
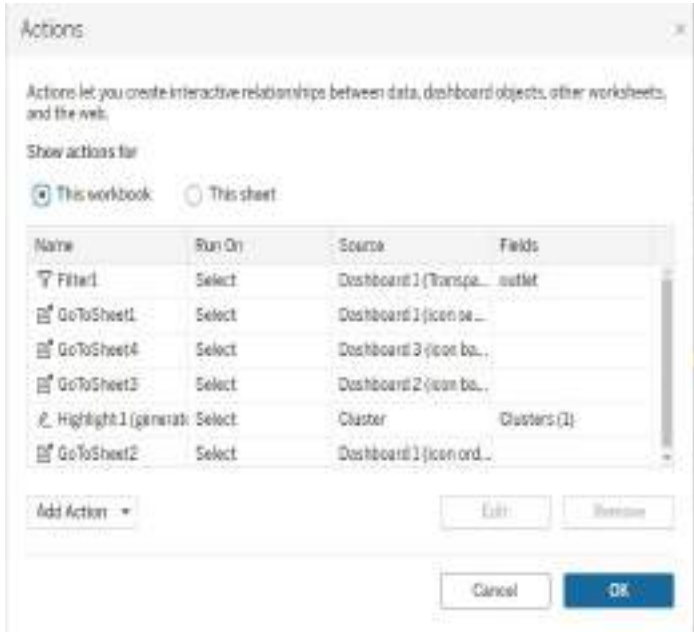
1. The details of New and Old Products
2. As we said before, this visualization indicates that we have to replace the bottom products with the top new products for make more efficient of operational costs, however we need deeply analyze because the data only provided a month
3. You can filter the dashboard with the outlet and the segmentation to make sure the best products for every outlet and segmentation

Data Visualization

Order Details										
customer_id	age	gender	Segment	Outlet	Store	Product	Product	Unit Price	Orders	Quantity
1	72	L	Recent Customer	3	Jakarta Barat	Bakery	Oatmeal	36.000	1	1
						Coffee	Columb.	48.000	2	4
							Espres.	36.000	1	1
							Ethiopi.	36.000	1	1
							Our Ol.	30.000	1	1
						Drinkin.	Dark ch.	42.000	1	1
						Tea	Spicy E.	37.200	1	2
2	72	L	Can't Lose Them	3	Jakarta Barat	Bakery	Cranbe.	39.000	1	1
							Croissa.	39.000	1	1
							Hazeln.	42.000	1	1
						Branded	I Need ..	144.000	9	1
						Coffee	Brazili.	42.000	1	1
							Brazili.	36.000	1	1
							Cappuc.	51.000	1	1
							Columb.	48.000	2	2
							Latte Rg	51.000	1	2
							Our Ol.	24.000	1	2
						Drinkin.	Dark ch.	42.000	1	2
						Tea	Earl Gr.	36.000	1	2
							English.	108.000	3	4
3	72	L	Can't Lose Them	3	Jakarta Barat	Bakery	Chocol.	45.000	1	1
							Cranbe.	39.000	1	1
							Croissa.	78.000	2	2
							Ginger	78.000	2	2

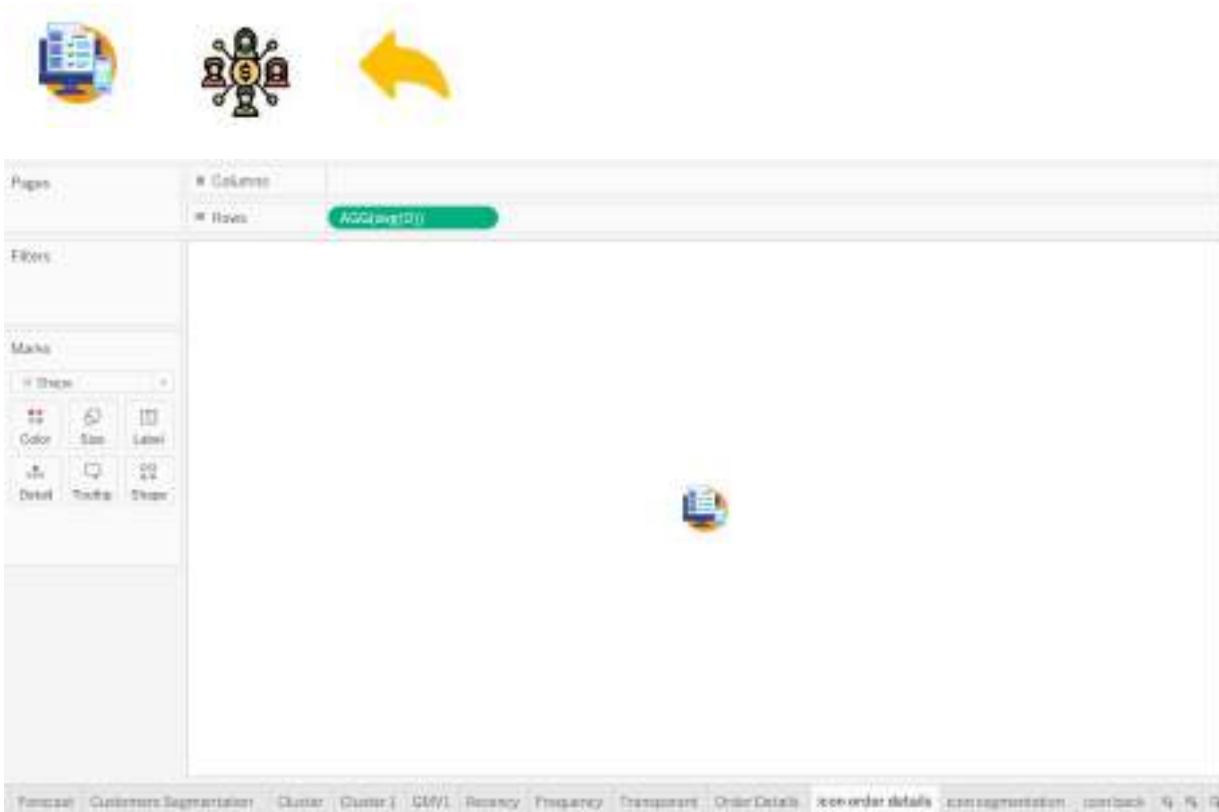
The details of the orders

Creating Dashboard



1. We used Actions as filter and shortcut on these Dashboard to filter by outlet and go to others dashboard
2. So you can click the description on the outlet and you can click icon on title to go to others dashboard

Creating Dashboard



1. To create an icon you can add the icon on Shape
2. Create the zero axis
3. Fill the tooltip

Creating Dashboard



1. The function of the transparent is the device to click on the Outlet filter
2. Drag the outlet to rows, remove everything's and make it white, then you can fill the tooltip

Data Story



Main Dashboard



1. The Main Dashboard is the analysis result of the Coffee Shop Transaction
2. You can click the outlet to filter the Dashboard

RFM Segmentation

Coffee Shop Transaction Dashboard



1. The RFM Segmentation divided into 8 Segments
2. I prefer to choose RFM Segmentation, because the groups are more segmented based on rank of the GMV, Frequency, and Recency than Cluster
3. The Coffee Shop needs to improve the recent customer and Can't Lose Them Segment depend on the Characteristic of the Segment

Cluster

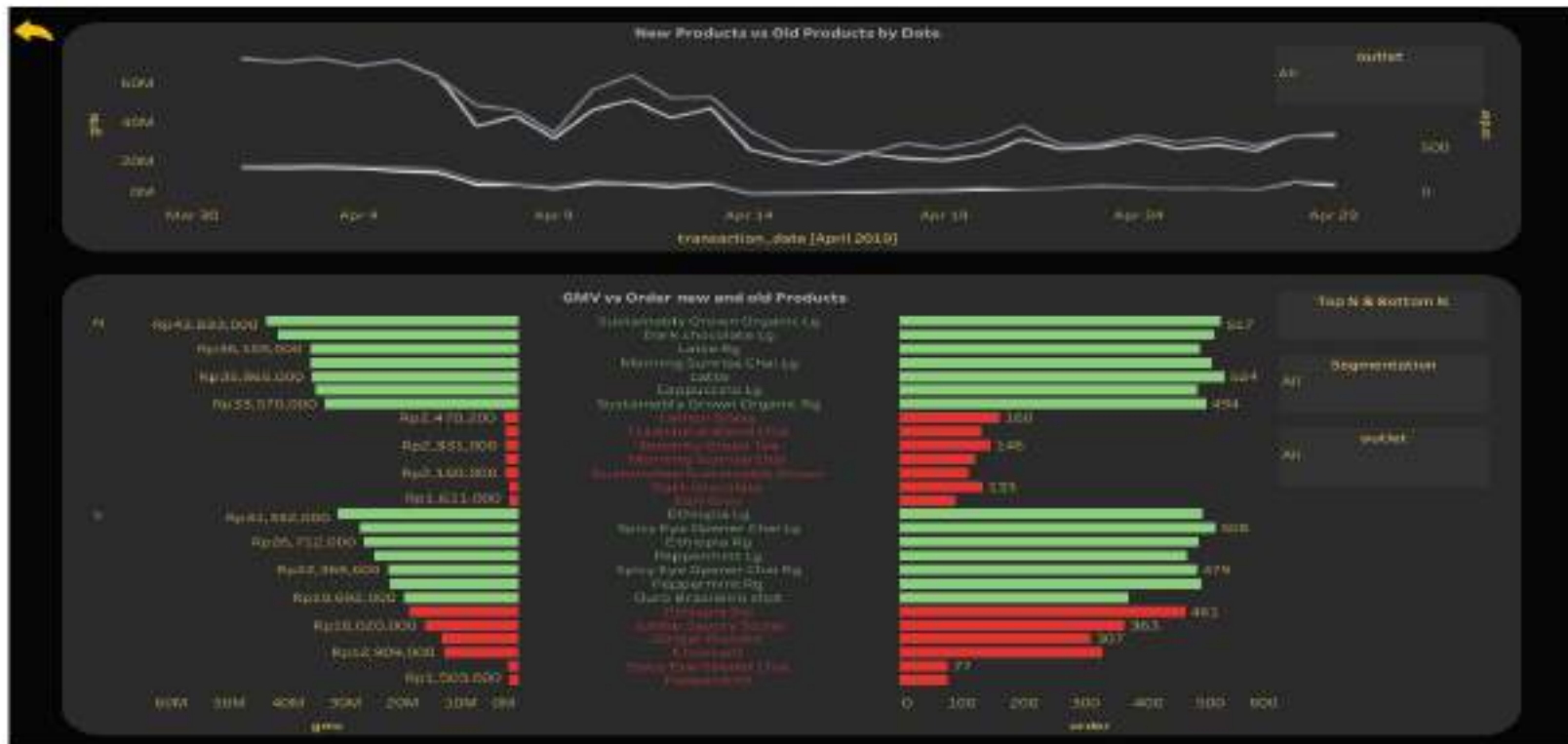
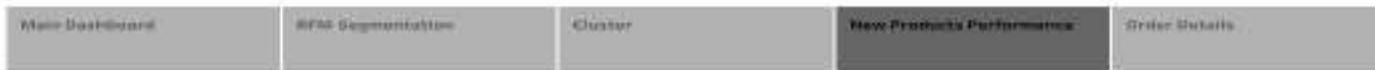
Coffee Shop Transaction Dashboard



1. The Cluster divided into 6 Clusters dan The RFM Segmentation divided into 8 Segments
2. The cluster based on K-Means, the result of the cluster is so good
3. Based on the Cluster Need Attention & Can't Lose Them Cluster are really dominant
4. The result o the Cluster is quite similar with RFM Segmentation

New Products Performance

Coffee Shop Transaction Dashboard



Order Details

Coffee Shop Transaction Dashboard

Main Dashboard	RFM Segmentation	Cluster	New Products Performance	Order Details
----------------	------------------	---------	--------------------------	---------------

Order Details													
customer_id	age	gender	Segment	cluster	store_city	new_products	product_cat	product	unit_price	Orders	quantity	gross	
1	32	L	Recent Customers	3	Jakarta Barat	N	Bakery	Chocolate Brownie	36.000	1	1	36,00	
								Coffee	Starbucks Medium	48.000	2	4	36,00
							Coffee	Espresso Shot	36.000	1	1	30,00	
								Old Old Time Blend	80.000	1	1	80,00	
								Orchard Ch.	Dark Chocolate Rp	42.000	1	1	42,00
								Coffee	Ethiopia Rp	36.000	1	1	36,00
2	32	L	Cart Lovers	3	Jakarta Barat	N	Bakery	Spicy Eye Specie C.	47.200	1	2	71,40	
								Cookies Brownie	38.000	1	1	38,00	
							Coffee	Espresso Shot	40.000	1	1	40,00	
								Orchard Ch.	Dark Chocolate Rp	144.000	3	1	144,00
								Coffee	Weather Rp	32.000	1	1	32,00
								Espresso Shot	36.000	1	1	36,00	
							Coffee	Espresso Shot	51.000	1	1	51,00	
								Columbian Medium	48.000	2	2	48,00	
								Coffee Rp	31.000	1	2	102,00	
								Old Old Time Blend	38.000	1	2	80,00	
							Coffee	Orchard Ch.	Dark Chocolate Rp	42.000	1	2	84,00
								Tee	Red Deep Rp	38.000	1	2	76,00
3	32	L	Cart Lovers	3	Jakarta Barat	N	Bakery	English Breakfast	109.000	2	4	144,00	
								Serenity Green Tea	36.000	1	2	72,00	
							Coffee	Serenity Green Tea	60.000	2	2	60,00	
								Espresso	38.000	1	1	38,00	
							Bakery	Spicy Eye Specie C.	61.200	2	2	61,20	
								Espresso	45.000	1	1	45,00	
								Cookies Brownie	38.000	1	1	38,00	
								Old Old Time Blend	78.000	2	2	78,00	
							Coffee	Starbucks Medium	36.000	1	1	36,00	
								Starbucks Green Rp	54.000	1	1	54,00	
								Orchard Ch.	Dark Chocolate Rp	288.000	3	2	288,00
								Coffee	Starbucks Small	26.400	1	1	26,40
							Coffee	Columbian Medium	36.000	1	2	72,00	
								Espresso Shot	72.000	2	2	144,00	
								Old Old Time Blend	48.000	2	2	96,00	
								Coffee Starbucks Medium Organic	216.000	2	1	216,00	

Conclusion & Recommendation



Conclusion :

1. Overall of The Coffee Shop Transaction is so good, the GMV is so high for a month
2. The different GMV between every Outlet not really high which means, the SOP of each Outlet are same
3. The Coffee Shop needs do improvement for all Outlet to giant the GMV
4. It's better to choose RFM Segmentation because the groups are more segmented based on rank of the GMV, Frequency, and Recency than Cluster
5. The Outlet 3 needs to focus on At Risk & Can't Lose Them Segment
6. The Outlet 5 needs to focus on Promising Segment and Service
7. The Outlet 8 needs to increase the new customer's, and focus on Recent Customers and Can't Lose Them Segment
8. The Coffee Shop needs to compline the data longer for analyze the products performance, to streamline operational costs

Conclusion & Recommendation



Recommendation :

1. Outlet 3 :

- At Risk is the customers who rarely shop and come.
- Can't Lose Them is the customers who rarely come but once they come, they do big order and spent much money.
- We need to create good service for them, so they feel comfortable to come (provided wifi, music, the menu which they really love, etc).
- Create discount promo for who haven't had transaction for a long time for calling them to come
- Create the event that relate to At Risk & Can't Lose Them Segment.
- The management should focus on outlet 3 because the forecasting result of the GMV is so low and stagnant
- The management should do the intervention of the forecast result.

2. Outlet 5 :

- The Outlet 5 is the best outlet of the Coffee Shop.
- Promising Segment is the customers who frequently come and the frequent of order and spent much money.
- So the best way for this Segment is the prime service.
- We need to make them feel comfortable (provided wifi, music, the menu which they really love for gift, booking place or menu, and membership).

3. Outlet 8:

- The Outlet 8 is the lowest GMV and Orders.
- The Outlet 8 needs to find more new customers.
- Recent Customers is the new customers.
- For the Can't Lose Them segment, they can use the same recommendation with Outlet 3.
- For the new customers, Outlet 8 can create discount promo referral code for the new customers by utilizing the Loyal and Champion Customers because the Outlet 8 has so many Loyal and Champion Customers.
- The Outlet 8 should be careful to create decision because the MAPE value is quite high but still fine.



THANK YOU