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**
Preparating, 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**
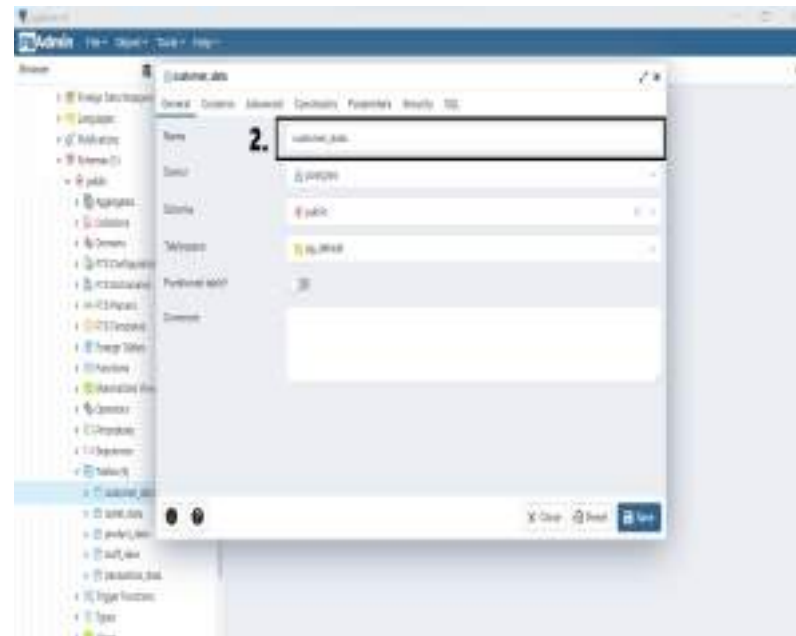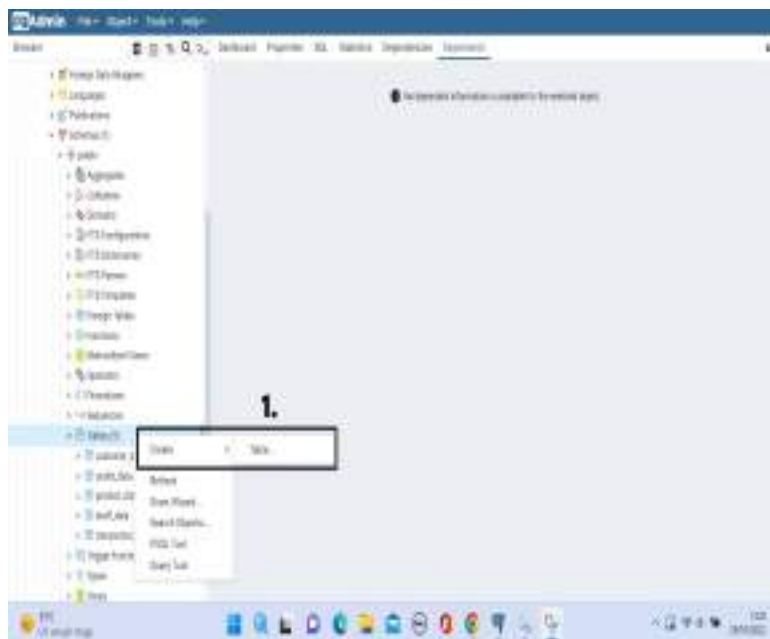The Conclusion & Recommendation of the Coffee Transaction Analysis

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

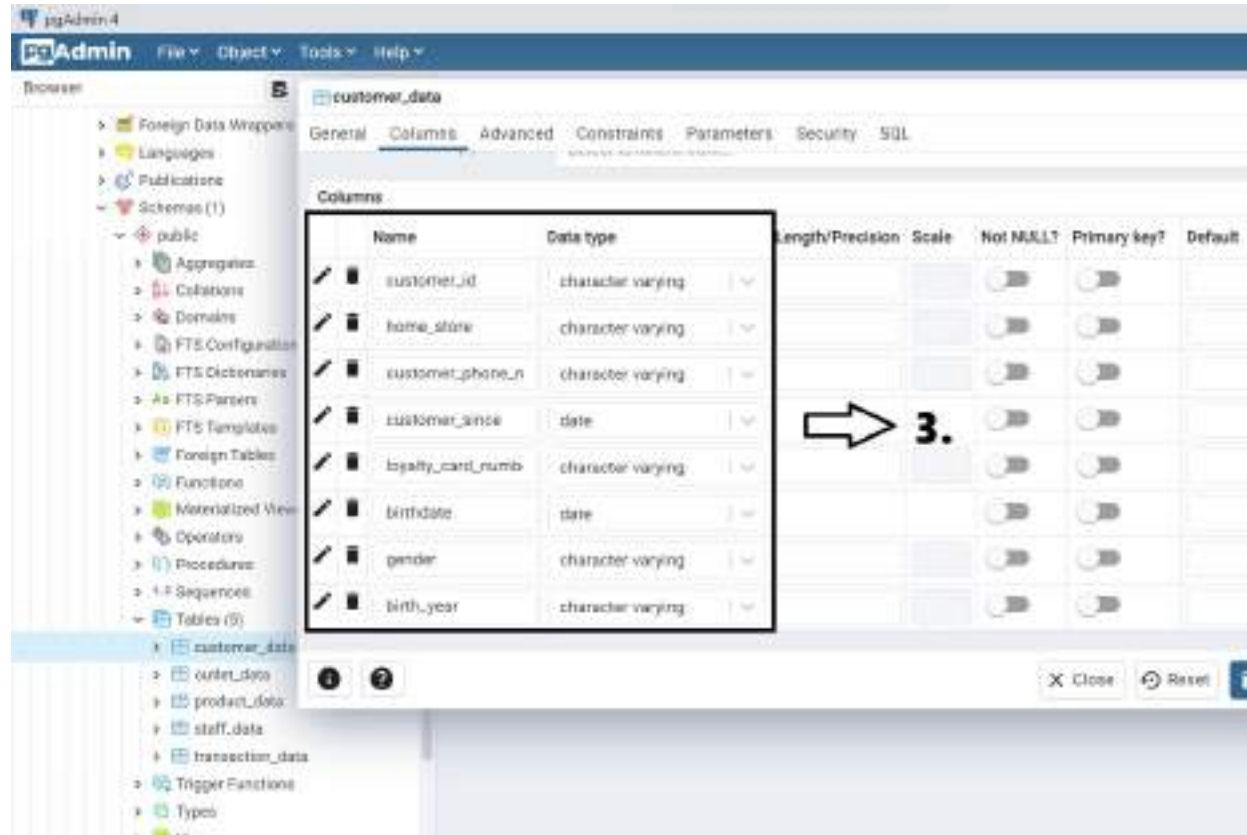If you want to see only the analysis, you can click [here](#)
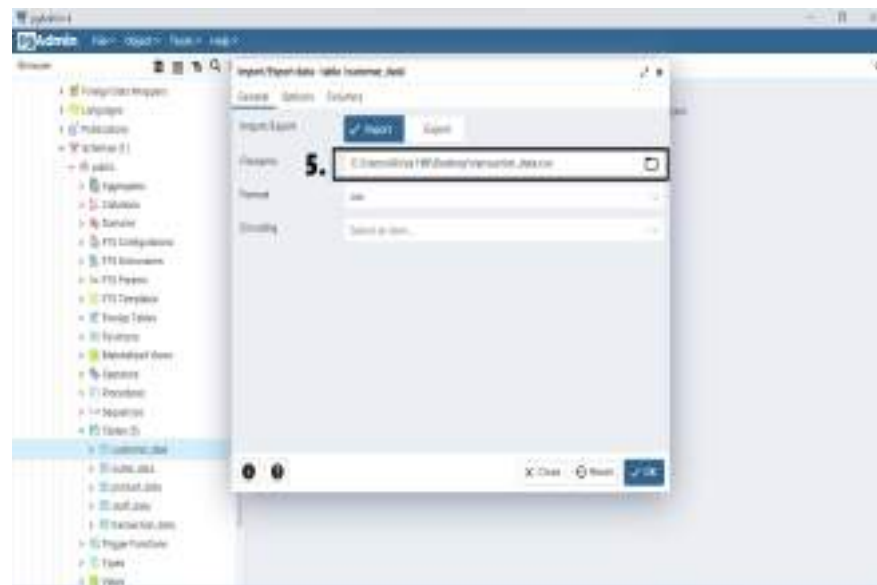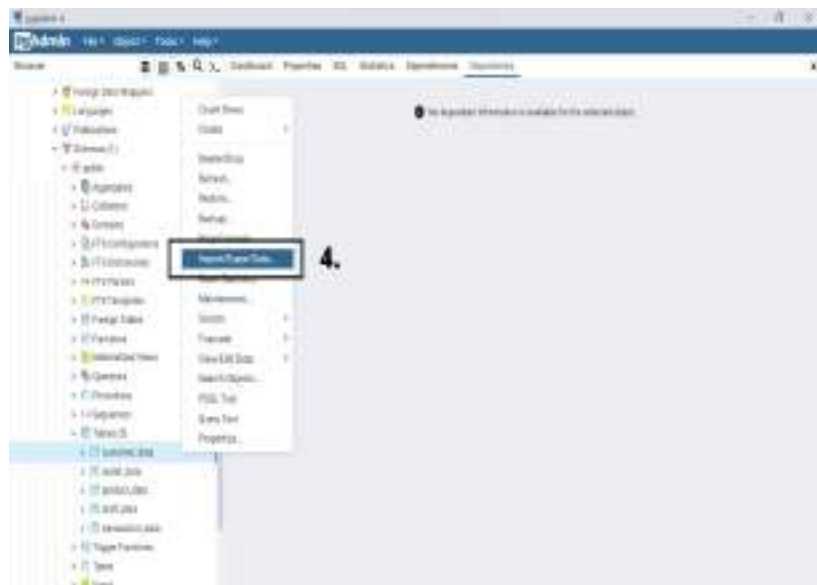
Import the dataset

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."ffirst_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

Data understanding

Data understanding

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

Data understanding

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 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      e."first_name",
9      e."position",
10     c."store_city",
11     b."order",
12     b."transaction_date",
13     b."transaction_time",
14     b."promo_item_yn",
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 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

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 sore 5 needs an improvement based on the order because the different total quantity and total order are still so far

Data understanding

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

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

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))) as age,
5      d."gender",
6      age(d."customer_since") as customer_since,
7      b."staff_id",
8      e."first_name",
9      e."position",
10     c."store_city",
11     b."order",
12     b."transaction_date",
13     b."transaction_time",
14     b."promo_item_yn",
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 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 AVG(age) as avarange_age,
33      AVG(customer_since) as avarange_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

	averange_age interval	averange_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

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

```
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       e."first_name",
9       e."position",
10      e."store_city",
11      b."order",
12      b."transaction_date",
13      b."transaction_time",
14      a."product",
15      a."product_category",
16      a."new_product_yn",
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

Export The dataset as CSV

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

Data Visualization

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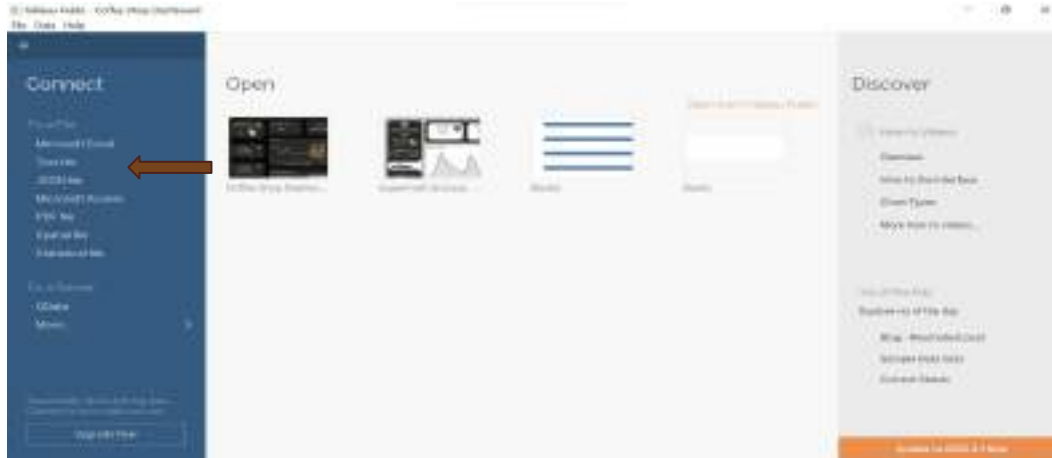
1. Added the dataset that have prepared from Postgresql by clicking Microsoft Excel
2. Then, we would go to Data Source for editing the dataset on Data Source such as Split, Join, Change the data type, etc

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

GMV

Rp1,376B

(Tableau worksheet interface showing Pages, Columns, Rows, Filters, and Marks panels. The Marks panel shows Automatic, Color, Size, Text, Detail, Tooltip, and SUM(gmv).)

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   Outlet 5   Outlet 8

Rp464M   **Rp430M**   **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
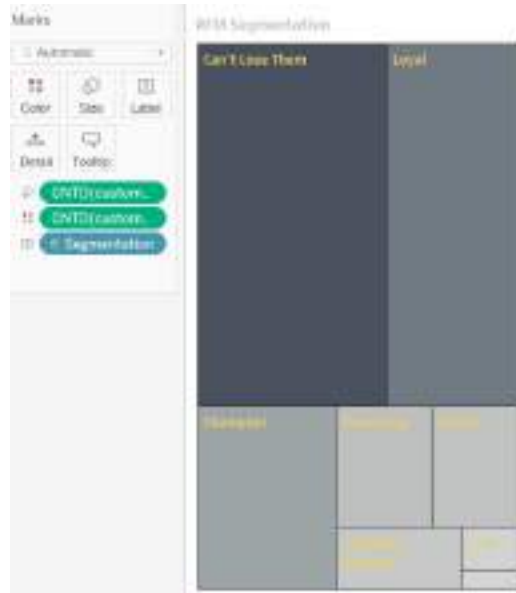
VS
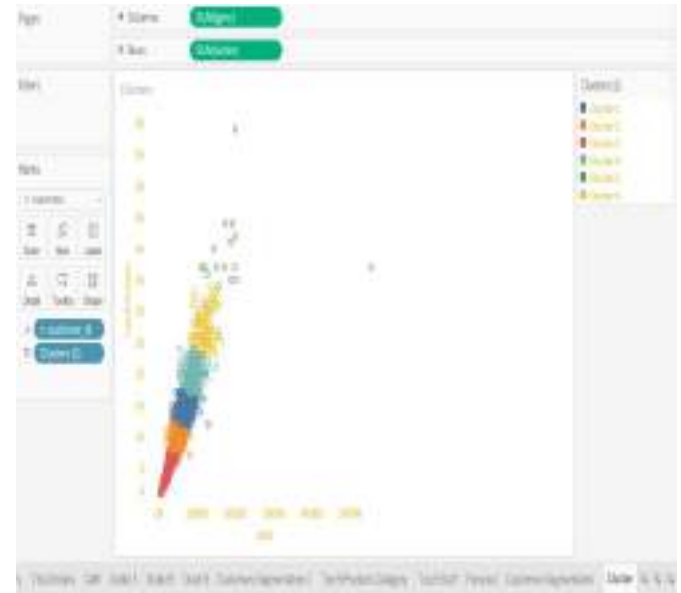
1. RFM Segmentation vs Cluster
2. The Cluster divided into 6 Clusters and The RFM Segmentation divided into 8 Segments
3. The result of cluster based on K-Means is so good, but I prefer to choose RFM Segmentation, because the groups are more segmented based on rank of the GMV, Frequency, and Recency than Cluster.

Data Visualization

The details of RFM Segmentation

Data Visualization

The details of Cluster

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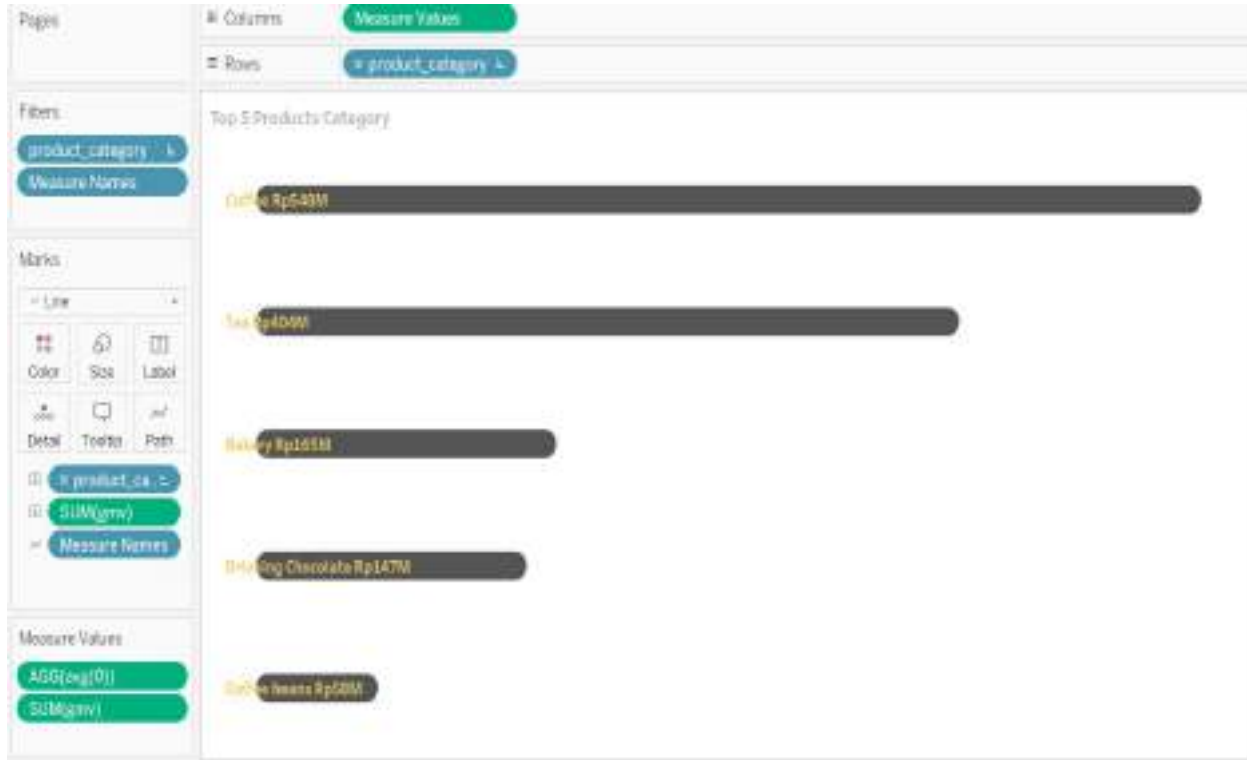
1. The Top 5 Products Category by GMV
2. Coffee & Tea are the best Products Category
3. To create the visualization, we created the zero axis to separate the GMV with zero axis, so for each bar is started from zero
4. Then we have changed the bar to line in bigger form

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%

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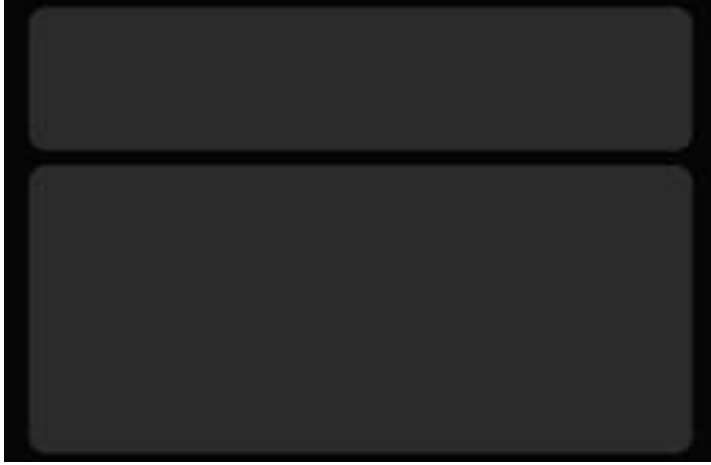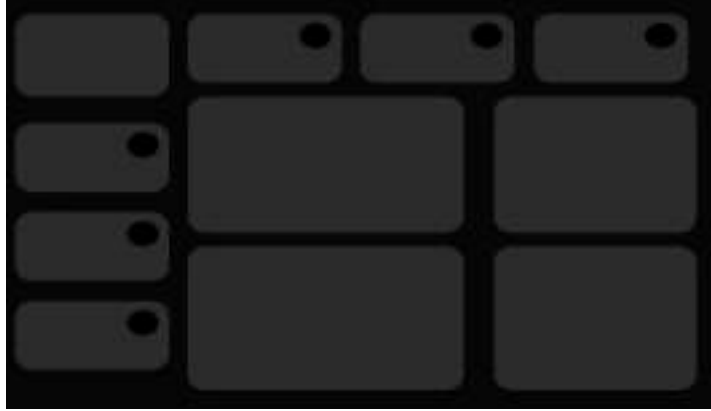
1. First, create template on Figma
2. Second, upload the template to Tableau
3. Make sure the size of the template fit with the Dashboard
4. We created 3 Dashboards on this dataset

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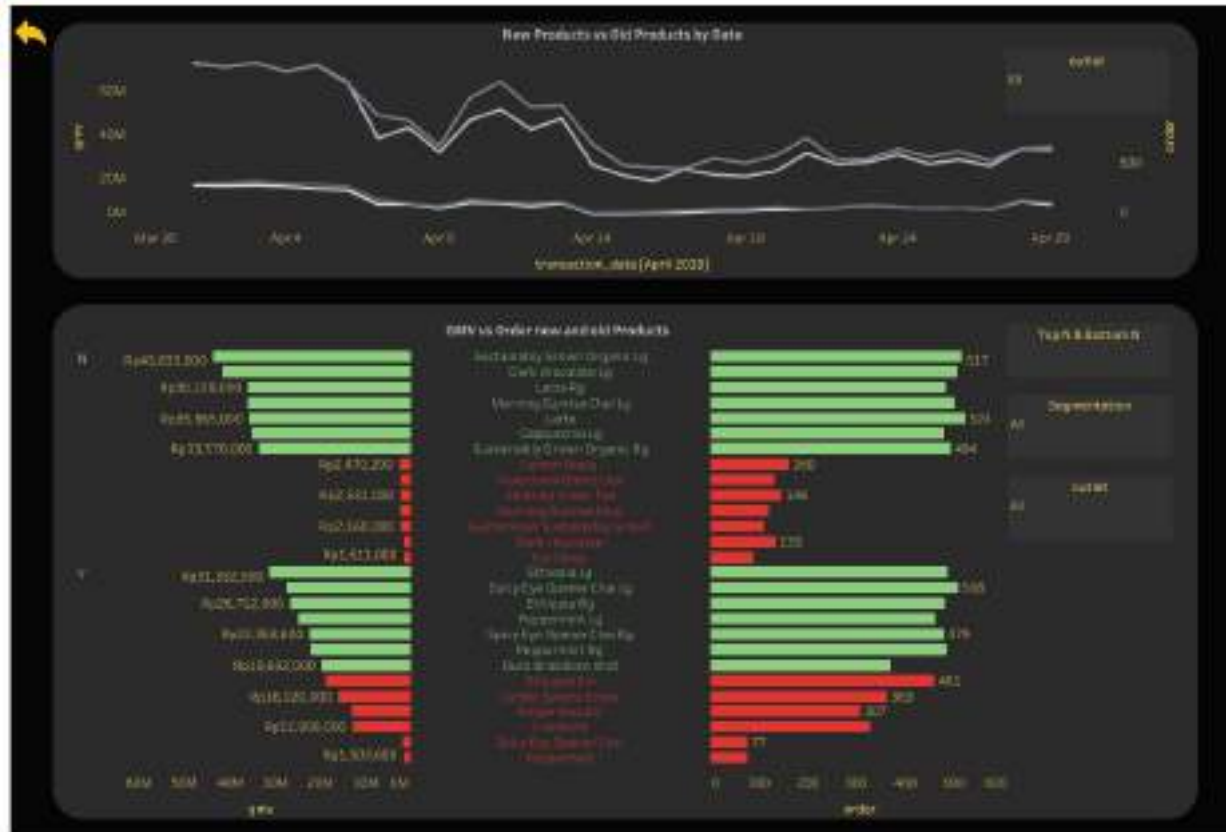
1. The details of New and Old Products
2. As we said before, this visualization indicates that we have to replace the bottom products with the top new products for make more efficient of operational costs, however we need deeply analyze because the data only provided a month
3. You can filter the dashboard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

Data Visualization

The details of the orders

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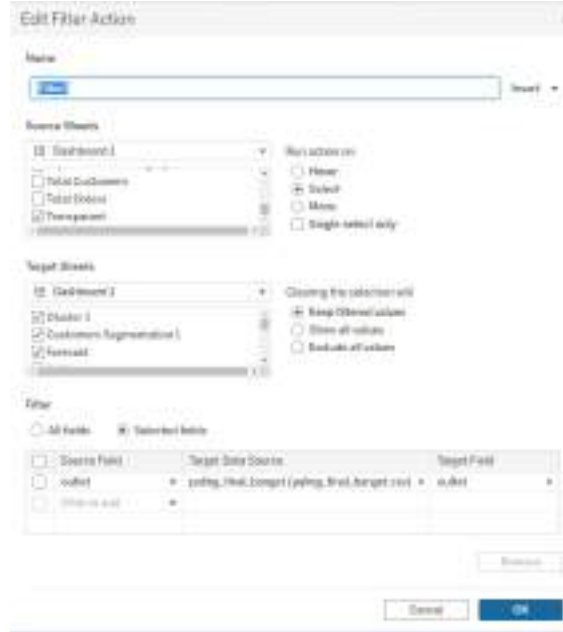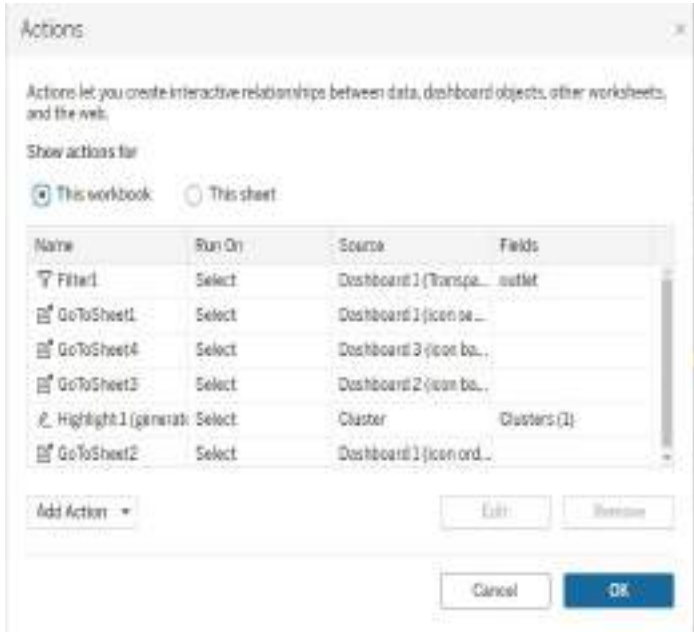
1. We used Actions as filter and shortcut on these Dashboard to filter by outlet and go to others dashboard
2. So you can click the description on the outlet and you can click icon on title to go to others dashboard

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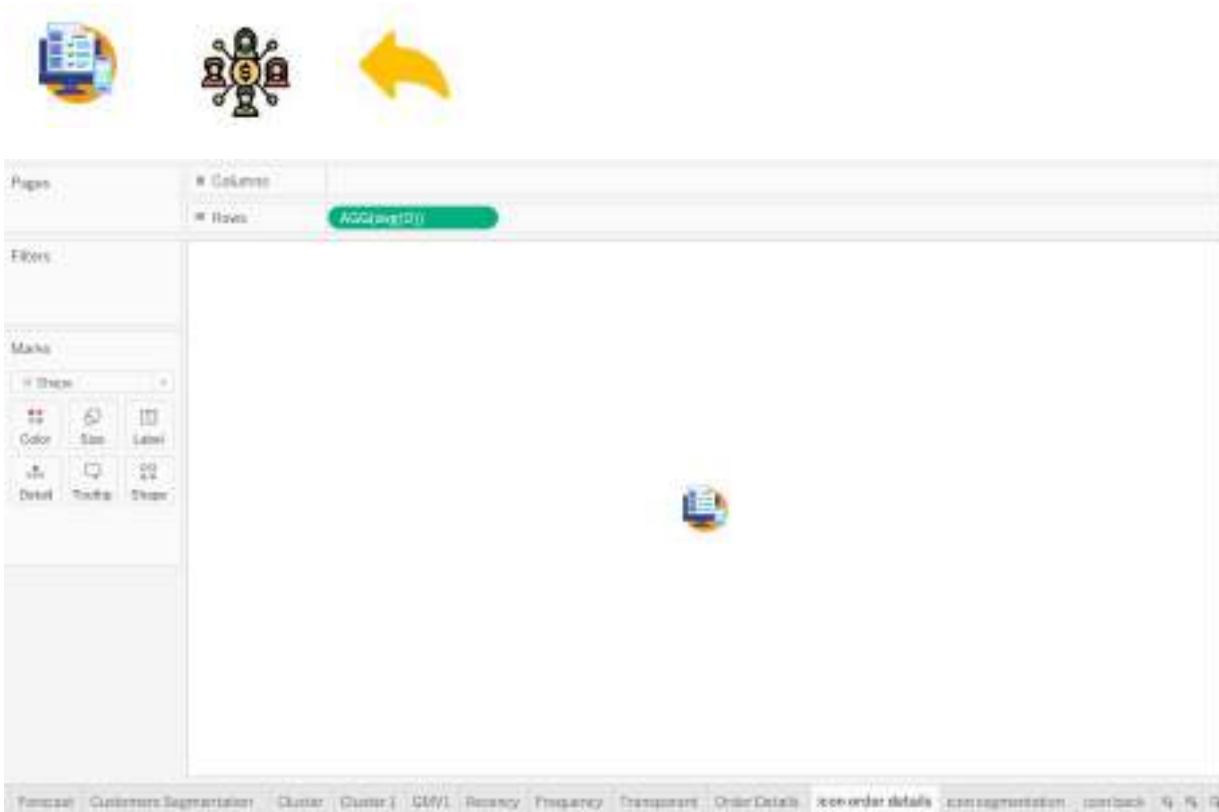
1. To create an icon you can add the icon on Shape
2. Create the zero axis
3. Fill the tooltip

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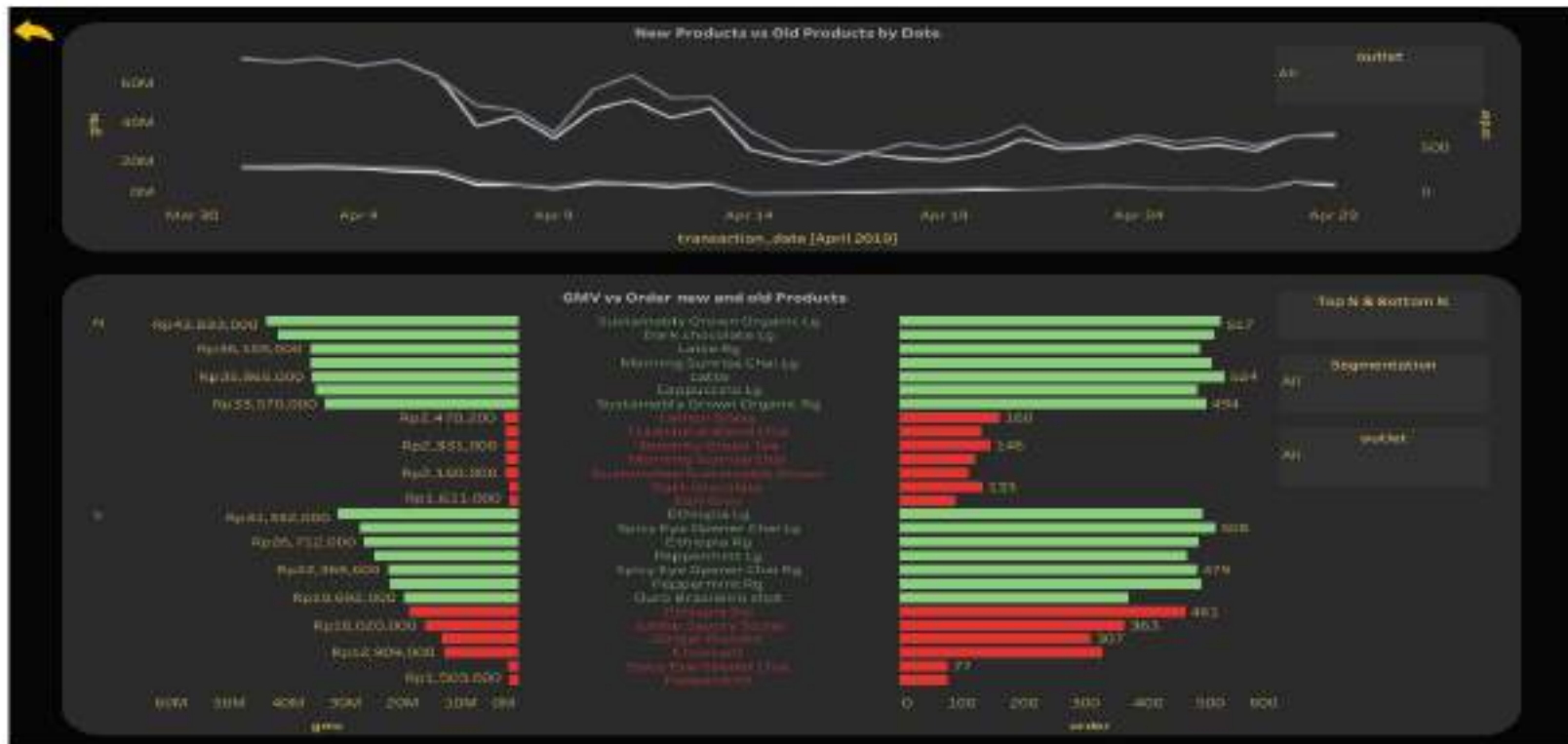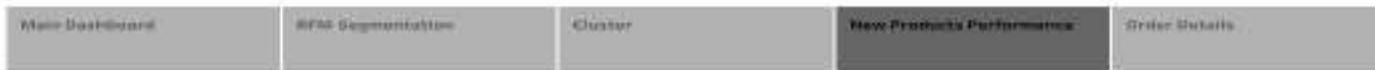
1. The Cluster divided into 6 Clusters dan The RFM Segmentation divided into 8 Segments
2. The cluster based on K-Means, the result of the cluster is so good
3. Based on the Cluster Need Attention & Can't Lose Them Cluster are really dominant
4. The result o the Cluster is quite similar with RFM Segmentation

New Products Performance

Order Details

Coffee Shop Transaction Dashboard

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