Junior Data Science Assessment test.

By: Alphaeus mutemi

- **<u>General view of the Data.</u>**

Total number of customers: 5272

Kenya: 4177

Nigeria Customers: 1095

- **<u>Overview of orders.</u>**

Total orders: 4395.

Total Kenya orders: 3451.

Total Nigeria orders: 944.

First Purchase order in Kenya: 2022/01/01   - Last Purchase order: 2022/17/02

First Purchase order in Nigeria:2022/01/01   - Last Purchase order: 2022/17/02

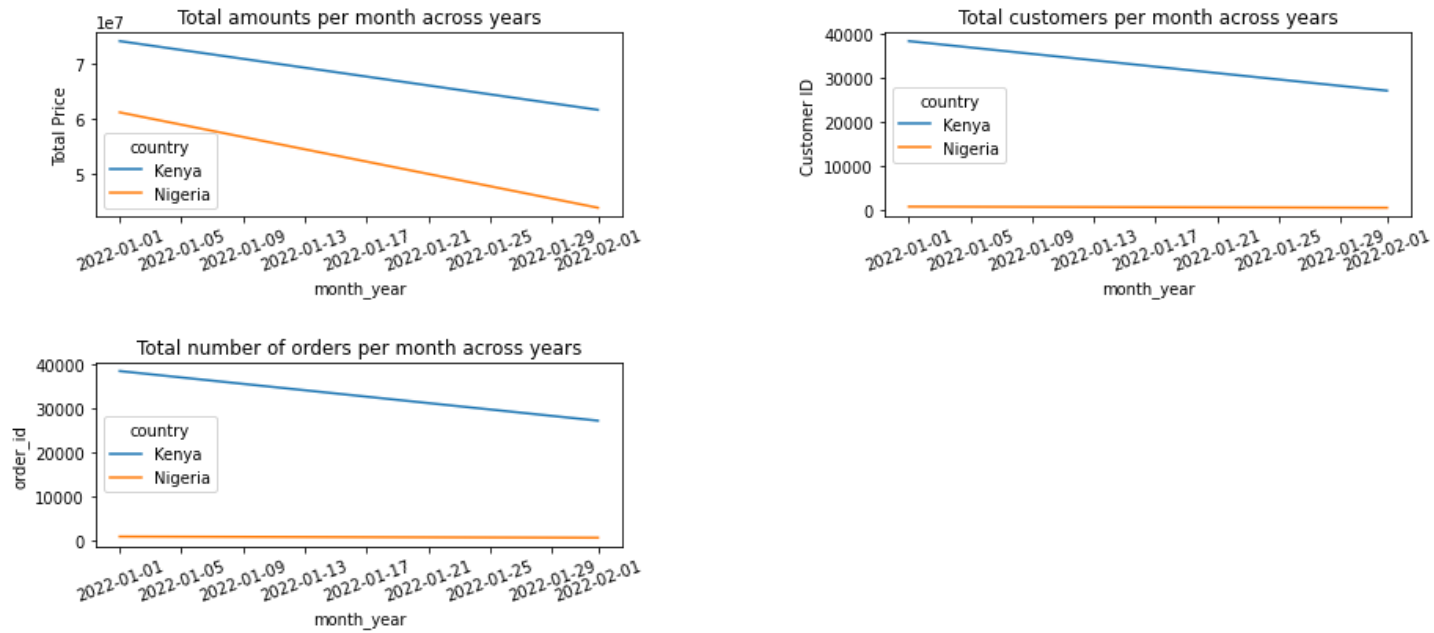Last Delivery Kenya: 31st Oct. 2021

Last Delivery Nigeria: 31st Oct 2021

From the above there is some inconsistency, I used complete time to truck when orders were delivered but didn't much data from Orders Data.

- # **Data exploration.**

  1. Seasonality:

  Due to the small size of the data I could not find any seasonality/ patterns for both countries.
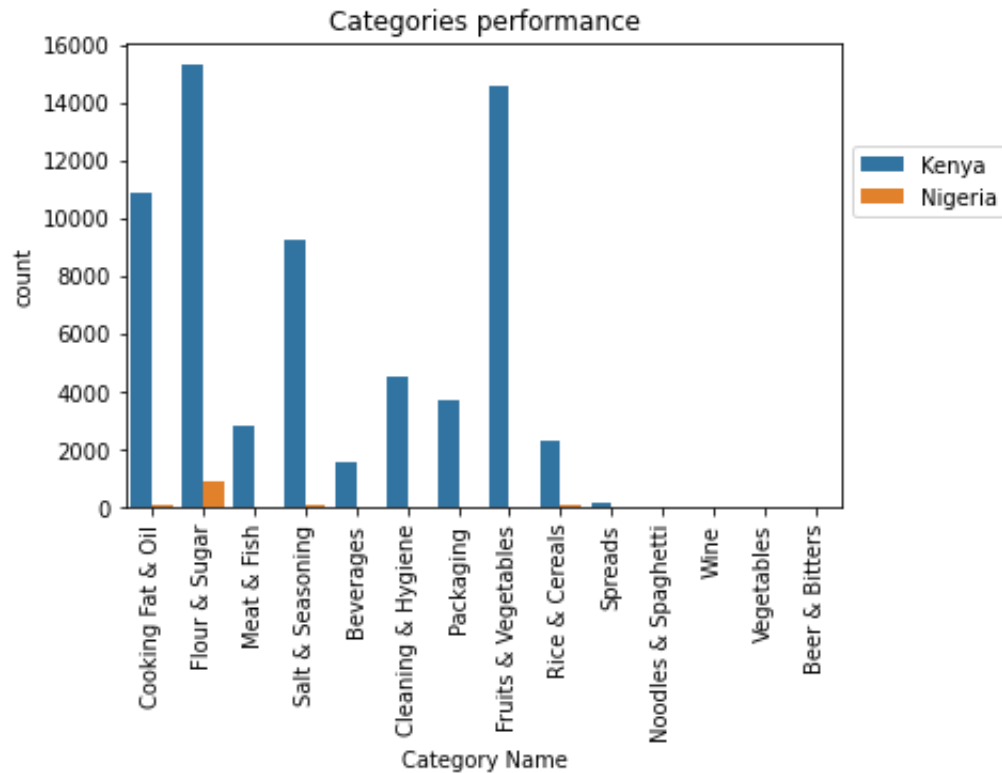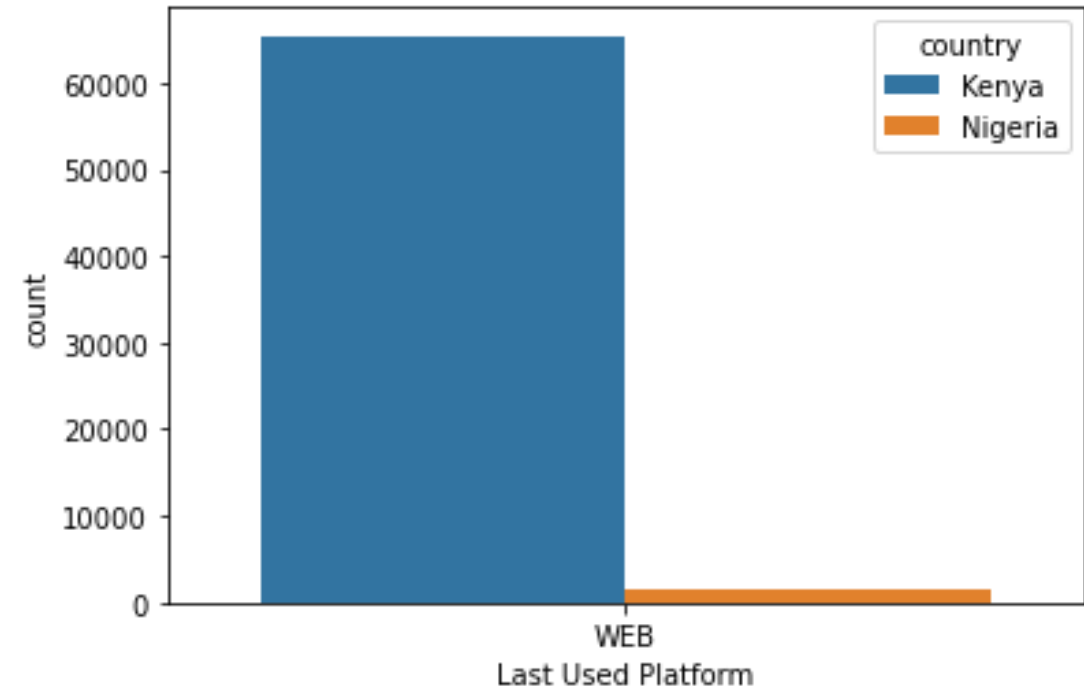

General sales distribution)

Meanwhile you can see a steady decrease of orders and customers placing orders. This could probably be products are not in seasons but can't conclude yet.

Data exploration.

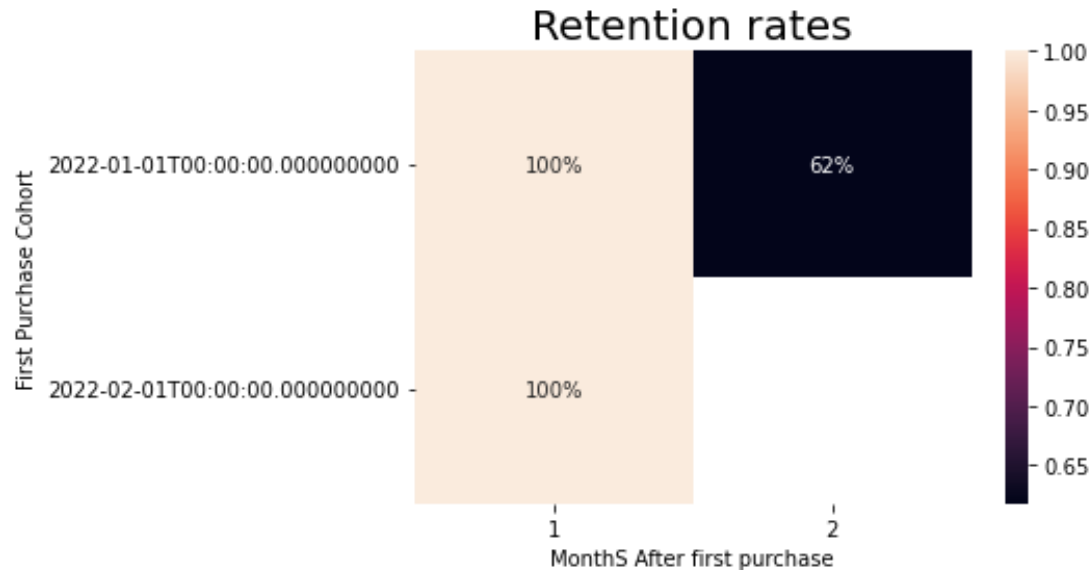1. Products performances per country:
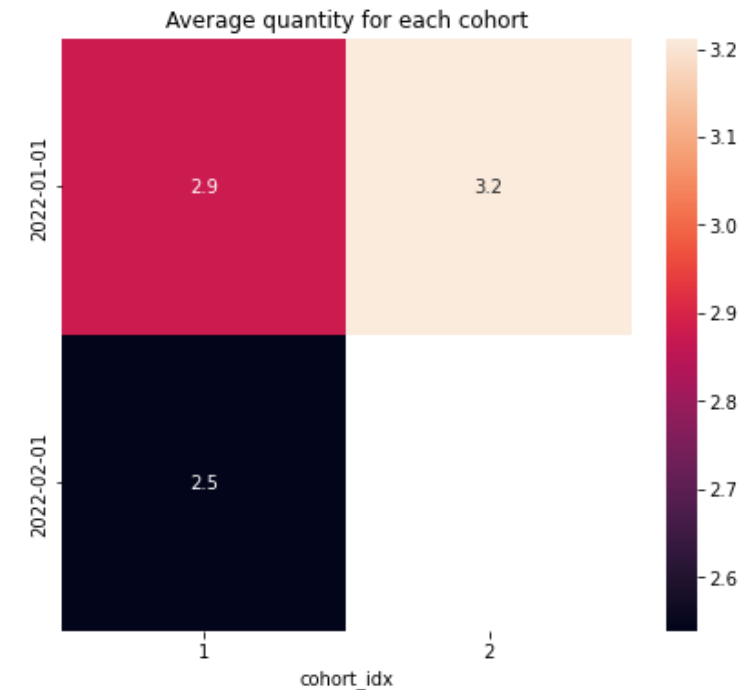
2. Lasts used platforms used :

- Kenya seems to be doing well that Nigeria.

- Flour and sugar seems to be most preferred product in both countries. Probably this the main product for the company.

- Only single platform is captured in the database according to this data.

- **<u>Retention Analysis.</u>**

  1. Customer Retention:

2. Average quantity order per cohort.

- - I combined retention for both countries Kenya and Nigeria just for General overview of the customer retention.

- - From the above we can see only 62% of customers who joined the business in Jan 2022 came back to place order in Feb 2022. Which is above the average.

- - Those who joined on Jan 2022 we can see on average they placed orders of 2.9 which has gone up during Feb 2022.

Conclusions

1. **Customer retention.**

- For now customer retention stands at 62%. Due to nature of my data I could not do the predictions.

2. Classifying customers.

- I used K-Means to classify thee customers. Where I had 4 classes as shown below.

Cluster	Recency mean	Frequency mean	MonetaryValue mean	count
0	6.78	107.80	1.663252e+06	501
1	2.00	9994.00	2.047752e+08	1
2	33.95	14.24	2.740355e+05	200

1. Cluster 0; They ordered recently but not as frequently and they buy high amount. They take around a week
2. Custer 1; They are frequent, buy recent and buy at higher prices. They take less than a week to order.
3. Cluster 2; Do not buy as recent, not frequent, buy at a lower prices. They take more than weeks to place orders.

3. Product recommendation.

- From our visuals we saw Flour and sugar is the most preferred category from both countries. Probably the same technique applied to this product would help boast the rest and see revenue growth.

- I couldn't find any seasonality due to the size of the data. By checking the seasonality would help in determining when customers prefer the different products. This will help in planning on promotions etc. to push the product.

- From these the business can optimize revenue generation.