Very often we come across retail/ecom/FMCG job openings – though the business varies in each sector/company, some concepts remain almost similar. This project is designed to give you that common perspective. This project is based on retail sale data, please find below the problem statement and try to come up with an answer to the same.

Don't just complete a question and write the answer but it would be far important to understand the concept behind the problem statement – technical and functional both. When I say technical means related to Python concepts, when I say functional means business aspects. In real world, you will have to make a good balance of these two things to come up with the solution to the problem statement.

Across any problem statement, whatever visual you are choosing, make sure why you are choosing that and why not others?

Part A - EDA

1. Understand the columns in the dataset and their datatypes. Wherever required, change the datatype to suitable one.
   For instance, quantity column should be numbers, Unit price should be float, Invoice Data should be date, likewise check the available datatype and change wherever required.
2. Find out the count and distinct count of each column.
3. Find out if any of the column has null values or not, if yes, how to deal with them?
4. Calculate Revenue per month and show in a data frame and a visual of your choice Write down the inference out of this.
5. Calculate cumulative revenue for across all months and show in a visual.
6. Calculate monthly percent change in growth rate and show in a visual. Note down the inference.
7. Extract and Plot the following:
   a. Revenue by country
   b. Total active customer(unique count of customer id) by country and month
   c. Total orders by country and month
   d. Total orderlines by country and month
   e. Total SKU (distinct count of Stock code) by country and month
   f. Monthly revenue (avg) per order
8. Find whether a customer is the new customer or not. A new customer would be figured out based on their first date of purchase. Figure out new customer on monthly basis.
9. Find total revenue per month for new and existing customer per month. Show it in a visual
10. Calculate total revenue per month by day for new and existing customer – in same graph. (You can use line graph)
11. Calculate monthly retention rate (using crosstab() function of pandas) and find out total retained user on a monthly basis.
    Note: Good research and understanding would be required over here.

Part B – Customer Segmentation

Before you start working on this part, let's first make clear what do you mean by customer segmentation, why do we need it and what are different ways to do it?

Go through the below links to understand the theoretical cum practical part.

https://www.customer.com/blog/retail-marketing/customer-segmentation-techniques/

https://www.optimove.com/blog/customer-segmentation-for-more-effective-marketing

https://www.tutorialspoint.com/retail_management/retail_market_segmentation_strategies.htm

In our project, we are going to do RFM clustering. Understand the concepts via these two articles – watch videos if required.

https://medium.com/swlh/exploring-customers-segmentation-with-rfm-analysis-and-k-means-clustering-93aa4c79f7a7

https://medium.com/clustering-segmentation/introduction-7321e39c595e

As a part of customer segmentation, we would be using unsupervised machine learning to deal with this problem. Why should we use unsupervised machine learning? – Find answer before moving forward.

1. Calculate recency by finding inactive days for each customer. And then apply K – means clustering to figure out recency score against each customer
2. Calculate frequency by finding total number of orders for each customer and then apply K – means clustering to figure out the frequency cluster for each customer.
3. Do the same for monetary / revenue
4. Add RFM to find out total score.
5. Put an overall score and then categorise customer among best, medium and worst customer.
6. You can deploy this model on Heroku too. Refer to the session that was taken by Board Infinity to deploy this model on Heroku. Before deploying make sure your model is running and working.

All the best!