Q: Where is the data?

A: Please find the dataset [here](https://docs.google.com/spreadsheets/d/1-31wL1vkNU8-yxfc2ItAX6KE4Re7RuUP/edit?usp=sharing&ouid=107093923581893374366&rtpof=true&sd=true) : https://docs.google.com/spreadsheets/d/1-31wL1vkNU8-yxfc2ItAX6KE4Re7RuUP/edit?usp=sharing&ouid=107093923581893374366&rtpof=true&sd=true.

Q: Tell me about the data.

A: Here is the data discovery summary:

1. Shape: (1067371, 8)
2. Column:
   * InvoiceNo - Invoice number. Nominal. A 6-digit integral number uniquely assigned to each transaction. If this code starts with the letter 'c', it indicates a cancellation.
   * StockCode - Product (item) code. Nominal. A 5-digit integral number uniquely assigned to each distinct product.
   * Description - Product (item) name. Nominal.
   * Quantity - The quantities of each product (item) per transaction. Numeric.
   * InvoiceDate - Invoice date and time. Numeric. The day and time when a transaction was generated.
   * UnitPrice - Unit price. Numeric. Product price per unit in sterling (Â£).
   * CustomerID- Customer number. Nominal. A 5-digit integral number uniquely assigned to each customer.
   * Country - Country name. Nominal. The name of the country where a customer resides.

Q: What is the Expectation from the candidate:

A: The candidate is expected derive 5+ data backed strategies, in total, to answer questions from the following vertical within the organization :

1. Marketing :
   1. “How can we best cluster our customers and get a view on Cluster\_Customer x [RFM](https://www.investopedia.com/terms/r/rfm-recency-frequency-monetary-value.asp) (Recency, Frequency & Monetary)”.
      1. Please create a model to cluster the customers into 3-4 segments.
      2. Please justify quantitatively how you choose and why you finalised the above model.
      3. Create a view ‘similar’ to the below image and derive & present marketing strategies out if it.
         1. A screenshot of a graph

            Description automatically generated
2. Operation team:
   1. “How can we forecast sales to make sure that we have enough stock for the next quarter (i.e. the next 3 months)”
      1. Please create a model to predict the daily sales at country level for the next 2 months.
      2. Please justify quantitatively how you choose and why you finalised the above model.
      3. Create a view similar to the below image and derive & present marketing strategies out if it.
         1. A graph with blue lines and orange dots

            Description automatically generated
3. Retail products team:
   1. “How can we arrange the shelf in each store so that we can promote cross sell & up sell for our products in the physical store. Can we start by you suggesting the ‘next best product in line’ for the top 10 selling stock code?”
      1. Hint : ABI index ( Average Basket item index is a number which depicts how many time two items where bought together. So if the ABI index of one of the top 10 selling item , let’s say MILK, was high then for the MILK x Butter matrix it could be said that the probability of butter being bought every time milk was bought is high)
      2. Create a visualization to the best of your ability as a star member of the Consumer Data science team.
      3. Make sure you handle the elements of time series forecasting such as seasonality.
4. Customer relationship management:
   1. “How can we know which customers are about to churn? If we could know which customer has the highest churn rate in a given a month then we could optimize our customer retention initiatives.”
      1. Please create a model to reduce the churn rate for the organization at country level x month level.
      2. Please justify quantitatively how you choose and why you finalised the above model.
      3. Create a view similar to the below image and derive & present marketing strategies out of it. [Each dashboard should have country as a filter]
         1. 