Business Analytics for an UK Online Retailer

Data Set Information:

This is a transnational data set which contains all the transactions occurring between 2010 and 2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts

Attribute Information:

- InvoiceNo: Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. **If this code starts with 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.
- Quantity: The quantities of each product (item) per transaction (Numeric)
- Invoice_time: Invice Date and time - the day and time when each transaction was generated.
- Price: Unit price. Numeric, Product price per unit in sterling.
- CustomerID: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.
- Country: Country name. Nominal, the name of the country where each customer resides.
- Purchase_dt: the date when purchase happened

Data sources:

- Raw csv file: *onlinepurchase.csv*
- MySQL database on AWS: *schema – ecommerce; table - onlinetrans*

Part One – BI Reporting

Use Tableau to connect the AWS SQL server: *schema – ecommerce; table - onlinetrans*

Quickly create views for the following questions and think about the insights

I. Overall business performance:
   1. Number of customers and number of order quantities by month from 2010 to 2011
   2. Year over year performance by quarter for total quantity, total sales and average purchase amount per customer
      Think about: Does the business perform well overall?

II. Customer activity by daily hour
    Think about: When does the majority of purchases happen? What can you do from marketing perspective?

III. Sales by Country
     Think about: Do you have any recommendation?

Part Two – Recommendation

I. Use *python_Recommendation_student.ipynb* to extract the data into Pandas Dataframe for analysis

Part Three – Calculate the CLV

Load dataset 'online_clean.csv' into Python pandas and develop a query to get RFM and calculate CLV for each customer.

- **Frequency:** represents the number of repeated purchases the customer has made.
- **Recency**: represents the age of the customer when they made their most recent purchases. This is equal **to the duration between current date and their latest purchase**.
- **Monetary:** represents the average value of a given customer's purchases. This is equal to the sum of all a customer's purchases divided by the total number of purchases.