Data Visualisation Expert Assignment

The objective of this assessment task is to evaluate your skills and expertise in data visualisation. You will be provided with a dataset and a set of questions that require you to create effective and informative data visualisations. You will be assessed based on the clarity, accuracy, and creativity of your visualisations, as well as your ability to choose appropriate visual encodings and design principles.

Instructions:

Dataset:

You will be provided with a dataset in CSV format. The dataset contains information about customer transactions for an e-commerce company. It includes columns such as customer ID, transaction date, product category, quantity, price, and total amount.

This is a sales transaction data set of UK-based e-commerce (online retail) for one year. This London-based shop has been selling gifts and homewares for adults and children through the website since 2007. Their customers come from all over the world and usually make direct purchases for themselves. There are also small businesses that buy in bulk and sell to other customers through retail outlet channels.

The data set contains 500K rows and 9 columns. The following is the description of each column.

TransactionNo (categorical): a six-digit unique number that defines each transaction. The letter "C" in the code indicates a cancellation.

Date (numeric): the date when each transaction was generated.

ProductNo (categorical): a five or six-digit unique character used to identify a specific product.

Product (categorical): product/item name.

Price (numeric): the price of each product per unit in pound sterling (£).

Quantity (numeric): the quantity of each product per transaction. Negative values related to cancelled transactions.

CustomerNo (categorical): a five-digit unique number that defines each customer.

Country (categorical): name of the country where the customer resides.

Category (categorical): category of each product/item.

There is a small percentage of order cancellation in the data set. Most of these cancellations

were due to out-of-stock conditions on some products. Under this situation, customers tend to cancel an order as they want all products delivered all at once.

Questions:
Your task is to create data visualisations that effectively answer the following questions:
a. What is the distribution of transaction amounts?
b. How does the total transaction amount vary over time?
c. Which product categories generate the most revenue?
d. Can you identify any trends or patterns in customer purchasing behaviour?

Requirements:
a. You are free to use any data visualisation tool or programming language of your choice.
b. Your visualisations should be clear, concise, and visually appealing.
c. You may create multiple visualisations to address each question if necessary.
d. Include appropriate labels, titles, and legends to enhance understanding.
e. Provide a brief explanation or interpretation of each visualisation.

Submission:
Provide the following for your submission:
a. Source code or a script file to reproduce the visualisations.
b. Image or PDF files of the visualisations.
c. A document (PDF or Word) containing your explanations and interpretations.

Assessment Criteria:
You will be assessed based on the following criteria:
- Clarity and effectiveness of the visualisations
- Accuracy and relevance in addressing the questions
- Creativity in visual design and choice of encodings
- Consistency and coherence in the explanations and interpretations
- Overall presentation and organisation of the submission

Note:
Please ensure that your code and visualisations are self-contained and can be executed independently. If you use any external libraries or tools, provide instructions on how to install and run them.