**Data Analyst Assessment Task**

Please use Python, R or SQL to answer the below questions. In Part 1 Question 5 you may use any software for visualisation.

**Part 1 – Data Cleaning & Visualisation (Python, R or SQL)**

In the dataset Data – Part 1, there is transactional data from January 2020 to October 2020. It has recorded transactions with the following detail:

* Name of the country to deliver to
* Year and Month of delivery
* Customer ID
* Product ID
* Revenue from transaction ($)
* Cost of revenue (i.e. how much did it cost NCR to make and deliver the product) from transaction ($)
* The Industry the customer belongs to (Banking/Retail/Hospitality)
* Sub-Industry the customer belongs to

1. How much revenue did we get from customers in the Unites States in each quarter?
2. Where did we earn the most?
3. Where did we spend the most?
4. Where did we gain the most overall profit?
5. Visualise the Revenue and cost monthly trend for:
   * 1. the 10 products we made the least profit from **(You may use any software)**
     2. the 10 Customers we made the most profit from **(You may use any software)**

**Part 2 – Data Analysis (Python, R, SQL)**

You have been given 3 datasets, each from a different team at NCR:

1. Customer Information:
   * Customer ID
   * Industry the customer belongs to (Banking/Retail/Hospitality)
   * Customer Account Type
   * Customer Tiering
   * Total number of Employees working for the customer
2. A Country-Region mapping table:
   * Country Name
   * Region Name
3. Invoice Data:
   * Transaction ID
   * Customer ID
   * Invoice Date – date that the invoice was created
   * Payment Date – When the invoice was paid by the customer
   * Invoice Amount ($)
   * Country Name
   * Payment Term – Number of days that the customer has to pay from the invoice date

Questions:

1. How much cash was collected in each region in this data?
2. What proportion of transactions were:
   1. Paid later than the payment term
   2. Not paid at all
3. Identify:
   1. Customer IDs of the customers who contributed the most to the cashflow
   2. Which country had the most cashflow. Was this consistent throughout each month in the data?
   3. Which industry contributes to the most cashflow
4. Is there a relationship between the total amount a customer spends and how many employees it has?
5. Is there a trend in the amount of cash NCR collects?
6. What further insights can you deduce from the data?