|  |
| --- |
| **Question:1:** |
| Create the following data frame using suitable commands such as np.arange, list etc. |

|  |
| --- |
| **Question 2:** |
| Data for this question can be found from "tweets" sheet in assignment-data.xlsx.  “ABC Company” has collected “tweets” from tweet.com and instructed its junior data scientist “Mr. Jo Jo” to mask sensitive data so that they can use the masked data for testing.  ***Tasks:***  ***Task-1:***  Mr.Jo Jo likes to do his experiment on small amount of data thus decided to play with only 10 rows.  Read only the rows 3-12 from tweets sheet and name it as "df" and display the type of df. The output should be shown as follows:      ***Task-2*:**  Mr.Jo Jo found that “tweetid, created-at and username” columns are sensitive thus decided to mask the values from those columns. He created new columns “new-tweet-id”,”created-at4” to store the masked “tweet-ids” and masked “created-at” values. He decided to use “username” column to store masked usernames.  Now, analyze the following output figure to do essential mask operations. |

|  |
| --- |
| **Question 3:** |
| Data for this question can be found in "online-retail" sheet from assignment-data.xlsx. **Since it is a big data, load first 200 rows and keep it in the data frame called "dataset". This “dataset” is used for all tasks in this question**  Assume that you are a data scientist in Amazon. Since the company is celebrating Silver Jubilee this year, it has decided to reward their customers. Your Manager handed over last 2 years retail data and asked you to do certain tasks. The tasks are as follows:  ***Task1-1:***  When you started working with data, you’ve realized that it needs cleaning to produce better results.  Do essential data cleaning. The final output should be the one as follows      ***Task-2:***  Amazon wants to reward the customer who has done the most number of transactions, by providing a gift hamper worth 1000 pesos.  Find out the customer who has done the most number of transactions [Note: One invoice number corresponds to one transaction]. The output should the one as below:    ***Task-3:***  Amazon wants to find the best customer and likes to reward 50% discount on his next purchase. Find the Customer with the highest amount of total purchase amount. The output should the one as below:    ***Task-4:***  Amazon wants to find out the most and the least bought items. It likes to increase the sale of least bought item by bundling it with the most bought item and ready to give special 10% discount for this bundle. The output should be the one as below.    ***Task-5:***  Company likes to find out overall average purchase-amount among all transactions [ one transaction indicates one invoice number]. The expected output is as follows:    ***Task-6:***  Company also likes to find out the total purchase amount and the number of transactions made by each customer in their lifetime  Create new data frame called "new-dataset", containing the columns ['CustomerID', 'TotalPurchaseAmount', 'NumberofTransactions']. Display the last 10 rows. Expected output is as follows:    ***Task-7:***  Company also likes to know average purchase amount made by every customer to get an idea of how much they’ve spent with their business in average.  Append a column "AveragePurchase" in the new-dataset. Get the average total purchase per CustomerID. Display first 5 results. Expected output is as follows: |