I. DEMOGRAPHIC dataset

1. Read demographic.csv into a DataFrame name demo

2. The email attribute contains 'abc@gmail.com' values that indicate values are missing, but they aren't represented as NaN. Replace them with NaN value

3. Calculate the total number of missing values for each column

4. Add a new column named 'Total Null'. It should be the total number of nan values of each row.

5. Drop columns that contain more than 80% missing value using the dropna function

6. Drop rows that contain more than two missing values using the dropna function

7. Reread this dataset. This time, we will fill the missing values with other values instead of dropping them.

- Fill the missing value of the age attribute by the mean value of this attribute

- Fill the missing value of the gender attribute by the most Frequent Occurring. In the case of having more than one with the highest frequency occurring, randomly choose one between them to fill the missing value.

II. Revenue dataset

This dataset provides the revenue created by each customer for each month

1.For typing errors, in this dataset, we see that some customers' revenue appears more than once for a specific month. We need to delete one and keep the row with a larger income.

2.In this dataset, we would like to know the latest revenue created by each customer and the time. We would remove rows that contain a CustomerID already listed earlier. Do it using the drop_duplicates method.

III. Read the apps.csv into a DataFrame named apps

1. Create a new column named currency whose values are the first character of the price column

2. Transform the price column into floating-point number by:

- ✓ Deleting the currency characters
- ✓ Replacing commas by dots
- ✓ Casting the column to float

3. Make currency column a categorical

IV. Clean the phone number column by making all phone numbers use the same format.

1. Read the people.csv into a DataFrame named people

2. Format the phone number column so that each row uses the format 0ddddddddd where d is a digit. For ex: 0389440066 or 0914208888

V. Clean datetime column

1. Read the birthdays.csv into a DataFrame named birthdays

2. Use vectorized string methods to extract the year

3. User the Series.astype() method to convert the type to a float

VI. URL dataset

1. Read the urls.csv into a DataFrame named urls

2. Add a secure Boolean column. It should have the value True if the URL starts with HTTPs and False otherwise

3. Add an extension column. It should have the URL extension(all characters after the last . in the URL). In our case: jpg, html