Assignment-1: Basics of Programming in Python

Problem Statement:

The accounts section of a company wants to assess their product-wise sales and other factors. Their records are stored in the file: 'sales_data.csv'. Develop a code in Python to clean, analyze, and visualize the assessment from the data.

Implementation:

Use Pandas and Numpy for the implementation and Matplotlib for visualization.

Experiments: [4+4+4+4+4]

(Hints are based on the Google Collab Tutorial covered in class. See here.)
Experiment 1: Load the given dataset into a pandas dataframe. Then, display the first eight rows to get an overview of the data. Hint: Use `df.head()`

Experiment 2 : Tabulate missing values for each group (product, quantity, price, etc). Next, handle them by replacing them with the mean value of the respective group.

Experiment 3: With the results of Experiment 2, create a line plot to visualize the revenue trend over date. Hint: Use `df.mean(), df.fillna()`

Experiment 4: Calculate and print the total number of Orders in the dataset and the total revenue generated from the sales.

Experiment 5: Calculate and print the average price of each product. Create a bar plot to visualize the average price of the product. Then identify and print the top most sold products.

Dataset:

The dataset contains the following groups. Group names and descriptions are:

OrderID: Unique identifier for each order.
Product: Name of the product sold.
Quantity: Number of units of the product sold in each order.
Price: Price per unit of the product.
Date: Date of the transaction.

Submission:

A .zip file containing the python source code and a PDF report file. The final name should follow the template: <Assign-No>_<Your Roll No>.zip. For example, if your roll no is 15CE30021, the filename for Assignment 1 will be: `Assign-1_15ce30021.zip`

1. A single python code (.py) containing the implementations of the models and experiments with comments at function level. The first two lines should contain your name and roll no.

2. A report [PDF] containing **[1+1+1+1+1 points]**
    a. Experiment 1: A table containing the first 8 rows to get an overview of the data.
    b. Experiment 2: A table containing each group name and corresponding number of missing values.
    c. Experiment 3: A line plot of the revenue trend over date. Proper labeling on the plot should be there (X-label, Y-label, Title, and legend).
    d. Experiment 4: Mention total number of orders and total revenue.
    e. Experiment 5: A table containing three individual products and corresponding average price. A bar plot which plots the average price for each product. Mention the most sold product.

Responsible TAs: Please write to the following TAs for any doubt or clarification regarding Assignment 1.

Deepayan Chakraborty: deepayan504@gmail.com

Deadline: The deadline for submission is **15th January (Monday), 11:55 PM, IST**.

Irrespective of the time in your device, once submission in moodle is closed, no request for submission post-deadline will be entertained. No email submission will be considered. So, it is suggested that you start submitting the solution at least one hour before the deadline**.**

Plagiarism policy: Binary marking (two parties)