Rapidious Assignment :

Step 1. Importing Libraries:

- First, I import the required libraries:
    - pandas is used to load and work with data.
    - Matplotlib and seaborn are used to create visual charts and graphs to help better understand the data.

Step 2. Loading the Dataset:

- I load a CSV file (epi_r.csv) that contains data.
- This is done using pd.read_csv() which reads the file from specified file path.

Step 3. Checking for Missing Values:

- **Missing Values**: After loading the data, I check if there are any missing values in the dataset. The code counts how many missing or null values each in column.
- **Duplicate Rows**: I also check if there are any duplicate rows in the data. The code counts how many duplicates are present in the dataset.

Step 4. Removing Duplicate Rows:

- The code removes any duplicate rows using the drop_duplicates() function. This ensures that each record in the dataset is unique.
- **Counting Rows**: After removing the duplicates, the code checks how many rows are left in the cleaned dataset. This helps to see how much data was removed.

Step 5. Handling Missing Values:

- The next step I fix missing values in specific columns like calories, fat, protein, and sodium. These columns are important for data, so I fill the missing values with the **median** value of each column.

Step 6. Downloading the Clean Data:

- After cleaning the data (removing duplicates and filling missing values), the code prepares the cleaned dataset for download.
- The file is saved as epi_r.csv, and I download it using files.download().

Step 7. Using PowerBi for Visualization :

- In this section, I load two CSV file in PowerBi to Visualizing the Data ,
    - First I upload the file in transform data, then convert the main file into 2 files name as Dish_Data and Ingredient_Data .
    - After that check all the data is clean and ready to Visualized , I check again the missing values , duplicate values , Null values etc.
    - After that, I visualized the data , In this I used:

      2 Slicer : Filter by Rating and Dish Name

      6 cards : Fat , Protein , Sodium , Calories and Dish Count(Show how many dish in dataset) , Advance Preparation .

      1 line Chart : To show Dishes by Rating

      1 Table : Ingredient Used or Unused

      1 donut chart : Dishes by Rating Status
- After That I Again go to Colab And load the 2 file of Data named as :

  Ingrident_data.xlsx

  Dish_data.xlsx

Step 8. Loading Additional Datasets :

- In this section, I load two Excel files:
    - Ingrident_data.xlsx: This contains information about ingredients.
    - Dish_data.xlsx: This contains information about dishes or recipes.
- I use pd.read_excel() to read these files and load them into the program.
- After loading, the code displays the first few rows of each dataset .

Step 9. Visualizing Data with Seaborn and Matplotlib:

- Now I start creating different charts to understand the data better.

  A. Bar Chart - Average Dish Rating by Preparation Requirement:

  - You want to know if dishes that require advance preparation have higher or lower ratings.
  - The code groups the dishes by whether they require advance preparation and calculates the average rating for each group.
  - **X-axis** : Shows whether advance preparation is required (Yes or No).
      - **Y-axis**: Shows the average rating for each category.

B. Scatter Plot - Caloric Content vs. Protein Content Categorized by Rating :

- This plot compares the calories and protein content of different dishes, while also showing the dish's rating using colours.
- The scatter plot shows :
    - **X-axis**: Calories in each dish.
    - **Y-axis**: Protein content in each dish.
    - **Colour**: Indicates the dish's rating status (showing good vs. poor ratings).
- This helps understand if there's any relationship between calories, protein, and how well a dish is rated.
