Pandas & Matplotlib

(solve using SALES.CSV)

Basic Level Exercises

1.   Load the dataset and display the first 5 rows.

2.   Calculate the mean, median, and standard deviation of Item_MRP.

3.   Filter and display rows where Item_Weight is greater than 10.

4.   Check for missing values in the dataset.

5.   Count the occurrences of different Item_Fat_Content values.

6.   Rename the Item_Outlet_Sales column to Sales.

7.   Drop the Outlet_Establishment_Year column from the dataset.

8.   Sort the dataset by Sales in descending order.

9.   Display the unique values in the Outlet_Type column.

10.   Subset the dataset to include only the columns Item_Type, Item_MRP, and Sales.

11.   Create a new column Price_per_Weight by dividing Item_MRP by Item_Weight. Then, find the maximum and minimum values in this new column.

12.   Add a new column High_Sales that assigns True if Sales are greater than the average sales, and False otherwise.

13.   Filter the dataset to include only rows where Item_Type contains the word "Snack."

14.   Print the shape (rows and columns) and size (total elements) of the dataset.

15.   Plot a simple line chart showing the first 50 rows of Sales.

Intermediate Level Exercises

1.   Group the dataset by Outlet_Type and calculate the average Sales for each group.

2.   Replace missing values in Item_Weight with the mean of the column.

3.   Create a new column Sales_per_Unit by dividing Sales by Item_Weight.

4. Create a pivot table showing the sum of Sales for each Item_Type across different Outlet_Location_Type.

5. Assume there is a date column and convert it to datetime format

6. Filter the dataset to display only Low Fat items with Sales greater than 1000.

7. Create a multi-index DataFrame using Outlet_Type and Item_Type, and then display the total Sales for each combination.

8. Calculate the cumulative sum of Sales for each Outlet_Type and plot it.

9. Filter the dataset to include only rows where Item_Visibility is above the median and Item_Fat_Content is 'Low Fat'.

10. If the dataset had a Date column, extract the year and month from the date and add them as new columns.

Advanced:

1. Assume the data has a Date column. Plot the Sales over time to identify trends.

2. Create a pivot table that shows both the average and total Sales for each Outlet_Type and Item_Type.

3. Group the dataset by Outlet_Type and Item_Type, and calculate the sum of Sales.

4. Create a stacked bar plot showing the distribution of Sales across Item_Type for each Outlet_Type.

5. Aggregate the dataset to show the average Sales for each combination of Item_Fat_Content and Outlet_Location_Type.

6. Group the data by Outlet_Type and Item_Type, and calculate both the mean and standard deviation of Sales. Then, plot the results as a grouped bar chart.

7. Write a custom function that categorizes Sales into Low, Medium, High based on specific thresholds, and apply it to create a new column.