Assignment

Week 9

Step 3: Data Preprocessing & Visualization (Week 9-11)

Week 9: Data Cleaning - Missing Values, Outliers, Encoding, Scaling

Q1. Count how many missing values are in a DataFrame.

Q2. Show which values are missing in a DataFrame (True/False).

Q3. Remove all rows with missing values.

Q4. Fill missing values in a column with 0.

Q5. Fill missing values in a column with "Not Available".

Q6. Print the maximum and minimum value of a column.

Q7. Find values in a column that are greater than 100.

Q8. Replace all values greater than 100 with 100.

Q9. Sort a column to check for extreme values.

Q10. Remove the highest value from a numeric column.

Q11. Convert column Gender with Male/Female into 0/1.

Q12. Create separate columns for each City from a City column.

Q13. Count how many times each category appears in a column.

Q14. Change all values in a column to lowercase.

Q15. Divide all values in a numeric column by 100.

Q16. Subtract the minimum value from a numeric column.

Q17. Convert all column values to range 0 to 1 manually.

Q18. Show the mean and standard deviation of a numeric column.

Q19. Subtract column mean from each value (basic standardization).

Q20. Round all values in a numeric column to 2 decimals.

Assignment

Week 10

Week 10: pandas + numpy Deep Dive - DataFrames, Aggregations, Joins

Q1. Create a DataFrame of 5 students with columns: Name, Age, Marks.

Q2. Display the first 3 rows and last 2 rows of the DataFrame.

Q3. Show the shape of the DataFrame (rows × columns).

Q4. Access only the Name column from the DataFrame.

Q5. Filter all rows where Marks are greater than 50.

Q6. Add a new column Result with values: "Pass" or "Fail".

Q7. Update the Marks of the first student to 95.

Q8. Delete the Result column from the DataFrame.

Q9. Sort the DataFrame by Marks in descending order.

Q10. Count the number of unique names in the Name column.

Q11. Create a DataFrame of Sales with columns: Region, Product, Sales.

Q12. Calculate the total sales.

Q13. Calculate the average sales per product.

Q14. Count how many times each product appears in the dataset.

Q15. Find maximum and minimum sales per region using groupby and agg.

Q16. Create a pivot table showing total sales per Region vs Product.

Q17. Create a pivot table showing average sales per Region.

Q18. Create a NumPy array of numbers from 1 to 10.

Q19. Find the sum, mean, and standard deviation of the array.

Q20. Reshape an array of numbers from 1 to 12 into a 3×4 matrix.

Q21. Create a 2D NumPy array of random integers (3x3).

Q22. Find the sum of each row and sum of each column.

Q23. Multiply two

3×3 matrices using NumPy.

Q24. Find the maximum and minimum value in the array and their indices.

Q25. Slice a NumPy array to get only even numbers.

Week 11: Data Visualization - matplotlib, seaborn, EDA Techniques

Q1. Create a dataset with the following columns-> Products, Regions, Sales, Profit and 10 rows.

Q2. Plot a line chart of Sales for all 10 rows (index as X-axis).

Q3. Plot a bar chart showing total sales for each product.

Q4. Plot a scatter plot of Sales vs Profit.

Q5. Create a pie chart showing total sales by region.

Q6. Plot a histogram of the Sales column.

Q7. Create a boxplot of Profit to detect outliers.

Q8. Create a countplot showing how many products are sold in each region.

Q9. Create a heatmap to show correlation between Sales and Profit.

Q10. Show total and average sales per region using pandas.