Pandas Test

Basic Questions (3 marks each)

1. Load the dataset and display the first 5 rows.
2. Check the datatype of each column in the dataset.
3. Find the total number of transactions in the dataset.
4. List all the unique 'Customer Types' present in the dataset.
5. Find the number of transactions where the payment method is 'Cash'.
6. Display the first 10 transactions where the 'Branch' is 'A'.
7. Sort the dataset by 'Total' in descending order and display the top 5 rows.
8. Find the total 'Quantity' sold for the product line 'Health and Beauty'.
9. Check for any missing values in the dataset and display the columns with missing values, if any.
10. Convert the 'Date' column into a datetime format.

Intermediate Questions (5 marks each)

1. Calculate the average 'Total' for each 'Product Line'.
2. Group the data by 'City' and find the sum of 'Gross Income' for each city.
3. Create a new column 'Total Including Tax' by adding 'Tax 5%' to 'Total'.
4. Find the total number of transactions made by 'Members' and 'Normal' customers.
5. Filter the dataset to show transactions where the 'Payment' method is 'Ewallet' and the 'Total' is greater than 500.
6. Find the highest 'Rating' received for each 'Product Line'.
7. Calculate the total 'Quantity' of products sold for each 'Branch'.
8. Display the top 3 cities with the highest total sales ('Total').
9. Replace any missing values in the 'Rating' column with the median of the 'Rating' column.
10. Standardize the format of the 'Date' column and check for any inconsistencies in the date format.

Advanced Questions (10 marks each)

1. Calculate the correlation between 'Quantity' and 'Total'.

2. Group the data by 'Branch' and 'Customer Type' and find the average 'Rating' for each group.
3. Plot a bar chart showing the total sales ('Total') for each 'Product Line'.
4. Determine the top 3 'Product Lines' with the highest 'Gross Income'.
5. Analyze patterns in the use of 'Payment' methods across different 'Customer Types'.
6. Calculate the cumulative sum of 'Gross Income' over time and plot it.
7. Perform a time-series analysis to find the monthly trend in sales by grouping transactions by 'Date'.
8. Using the 'Rating' column, analyze if there is a correlation between customer satisfaction and the product line they purchase.
9. Remove any outliers from the 'Total' column where the value is greater than 1000 and analyze the impact on the overall dataset.