Pandas Assignment: Data Manipulation, Analysis, and Cleaning

Dataset Link:

Use this dataset: Airbnb Listings Dataset

Easy Level (12 questions)

1. **Load Dataset**:
   o Load the Airbnb Listings dataset using Pandas. How many rows and columns does the dataset have?
2. **Basic Information**:
   o Display the first 5 rows of the dataset.
3. **Column Names**:
   o List all column names from the dataset.
4. **Data Types**:
   o Check the data types of all the columns.
5. **Handling Missing Values**:
   o Count the total number of missing values in the dataset.
6. **Drop Columns**:
   o Drop any irrelevant columns like 'id' or 'host_id' and display the new dataframe.
7. **Basic Summary Statistics**:
   o Show the summary statistics for all numerical columns.
8. **Filtering Data**:
   o Filter the data to show listings where the price is greater than $100.
9. **Sorting Data**:
   o Sort the dataset by 'price' in descending order.
10. **Rename Columns**:

- Rename the 'price' column to 'listing_price'.

11. **Value Counts**:

- Display the count of unique values in the 'room_type' column.

12. **Subset Data**:

- Select and display the 'neighbourhood', 'room_type', and 'price' columns.

Medium Level (12 questions)

1. **Handling Missing Values**:
   o Fill missing values in the 'reviews_per_month' column with the median of the column.
2. **Data Grouping**:

- Group the dataset by 'neighbourhood' and calculate the mean price for each neighborhood.
3. **Aggregating Data**:
   - Aggregate the data to find the minimum, maximum, and average prices for each 'room_type'.
4. **Filtering by Date**:
   - Filter the dataset to show listings that were last reviewed after '2020-01-01'.
5. **New Feature Creation**:
   - Create a new column called 'price_per_accommodates' by dividing 'price' by 'accommodates'.
6. **Convert Data Types**:
   - Convert the 'price' column to a float after removing special characters (like $).
7. **String Operations**:
   - Extract the first word from the 'name' column and create a new column called 'listing_name_first_word'.
8. **Descriptive Statistics by Group**:
   - Calculate the standard deviation of 'price' for each 'room_type'.
9. **Find Outliers**:
   - Identify listings with a price more than 3 standard deviations above the mean price.
10. **Replace Values**:

- Replace all values in the 'neighbourhood' column where 'neighbourhood' is 'Downtown' with 'Downtown Seattle'.

11. **Pivot Table**:

- Create a pivot table showing the average price for each combination of 'room_type' and 'neighbourhood'.

12. **Cumulative Sum**:

- Compute the cumulative sum of the 'number_of_reviews' column.