Practical 1: Data Loading, Storage, and File Formats

1. **Q:** What is the main goal of this practical?
   A: To load, clean, and analyze sales data from CSV, Excel, and JSON files.
2. **Q:** Which library is used to handle different file formats in Python?
   A: pandas.
3. **Q:** Which function is used to read a CSV file?
   A: pd.read_csv().
4. **Q:** Which function is used to read an Excel file?
   A: pd.read_excel().
5. **Q:** Which function is used to read a JSON file?
   A: pd.read_json().
6. **Q:** What is a DataFrame?
   A: A 2D table-like data structure in pandas.
7. **Q:** How can you check for missing values?
   A: Using df.isnull().sum().
8. **Q:** How can you remove duplicate records?
   A: Using df.drop_duplicates().
9. **Q:** Why do we convert data into one unified format?
   A: For easier analysis and consistency.
10. **Q:** Which function is used to merge datasets?
    A: pd.merge() or pd.concat().
11. **Q:** How can you get basic statistics of a dataset?
    A: Using df.describe().
12. **Q:** Which libraries are used for data visualization?
    A: matplotlib and seaborn.

Practical 2: Interacting with Web APIs

1. **Q:** What is an API?
   A: It's an interface that allows two systems to communicate and exchange data.
2. **Q:** What is an API key?
   A: A unique code used to authenticate and access an API service.
3. **Q:** Which Python library is commonly used to interact with APIs?
   A: requests.
4. **Q:** How do you send a GET request in Python?
   A: Using requests.get(url).
5. **Q:** What is the format of data received from most APIs?
   A: JSON format.
6. **Q:** How do you convert API response to a Python dictionary?
   A: Using .json() method.
7. **Q:** What are common weather attributes retrieved from the API?
   A: Temperature, humidity, wind speed, and precipitation.
8. **Q:** How can you handle missing or inconsistent data?
   A: By cleaning or replacing missing values using pandas.
9. **Q:** Which plots are suitable for showing temperature trends?
   A: Line charts or bar plots.
10. **Q:** What is data aggregation?
    A: Summarizing data by specific time periods like daily or monthly.
11. **Q:** Which libraries are used for visualizing weather data?
    A: matplotlib and seaborn.

Practical 3: Data Cleaning and Preparation

1. **Q:** Which library is commonly used for data cleaning in Python?
   A: pandas.
2. **Q:** What does "churn" mean in this dataset?
   A: It means customers who have stopped using the company's services.
3. **Q:** How can you check for missing values in the dataset?
   A: Using df.isnull().sum().
4. **Q:** How can you handle missing values?
   A: By filling them with mean, median, mode, or removing the rows.
5. **Q:** How can you remove duplicate records?
   A: Using df.drop_duplicates().
6. **Q:** What does inconsistent data mean?
   A: Data with irregular formats or spelling variations that need standardization.
7. **Q:** How do you convert a column's data type?
   A: Using df['column'] = df['column'].astype(type).
8. **Q:** What are outliers?
   A: Extreme values that differ significantly from other observations.
9. **Q:** What is feature engineering?
   A: Creating new useful features from existing data to improve analysis.
10. **Q:** Why do we normalize or scale data?
    A: To bring all features to a similar range for better model performance.
11. **Q:** Why do we split the dataset into training and testing sets?
    A: To evaluate model performance on unseen data.
12. **Q:** How can you export the cleaned dataset?
    A: Using df.to_csv('cleaned_data.csv', index=False).

Practical 4: Data Wrangling

1. **Q:** What is the main goal of this practical?
   A: To clean, organize, and prepare real estate data for analysis or modeling.

2. **Q:** Which library is commonly used for data wrangling in Python?
   A: pandas.

3. **Q:** What is data wrangling?
   A: The process of cleaning, transforming, and organizing raw data into a usable format.

4. **Q:** Why do we clean column names?
   A: To remove spaces or special characters and make column names easier to use.

5. **Q:** How can missing values be handled?
   A: By filling them using mean, median, or mode, or by removing the missing rows/columns.

6. **Q:** What is the purpose of merging datasets?
   A: To combine information from multiple sources for a more complete analysis.

7. **Q:** How can you filter or subset data in pandas?
   A: Using conditional statements like df[df['column'] == 'value'].

8. **Q:** What is encoding in data wrangling?
   A: Converting categorical data into numerical format using label or one-hot encoding.

9. **Q:** What is aggregation?
   A: Summarizing data using functions like mean, sum, or count, often grouped by a specific column.

10. **Q:** Which pandas function is used for aggregation?
    A: df.groupby() combined with aggregation functions like .mean() or .sum().

11. **Q:** What are outliers?
    A: Values that are much higher or lower than the majority of the data.

12. **Q:** Why is it important to handle outliers?
    A: Because they can distort analysis and affect model accuracy.

Practical 5: Data Visualization using Matplotlib

1. **Q:** Which library is used for data visualization in Python?
   A: matplotlib.

2. **Q:** What is a line plot used for?
   A: To show changes or trends in data over time.

3. **Q:** What type of chart is suitable for comparing AQI values across dates?
   A: Bar plot or stacked bar plot.

4. **Q:** What is a scatter plot used for?
   A: To show relationships between two variables.

5. **Q:** What is the purpose of a box plot?
   A: To show the distribution, median, and outliers of data.

6. **Q:** How do you add a title to a matplotlib plot?
   A: Using plt.title("Title").

7. **Q:** How can you label the x-axis and y-axis in matplotlib?
   A: Using plt.xlabel() and plt.ylabel().

8. **Q:** How do you display a legend on a plot?
   A: Using plt.legend().

9. **Q:** How can you show the plot after creating it?
   A: Using plt.show().

10. **Q:** Which matplotlib function is used to create a line graph?
    A: plt.plot().

11. **Q:** Why is color customization important in plots?
    A: To make the visualization clearer and easier to interpret.

Practical 6: Data Aggregation

1. **Q:** What is the main goal of this practical?
   A: To analyze and compare sales performance by region using data aggregation.

2. **Q:** Which Python library is used for data aggregation?
   A: pandas.

3. **Q:** What does data aggregation mean?
   A: Summarizing data by grouping and applying functions like sum, mean, or count.

4. **Q:** Which pandas function is used to group data?
   A: df.groupby().

5. **Q:** How can you calculate total sales by region?
   A: Using df.groupby('Region')['Sales'].sum().

6. **Q:** Which plot can be used to show sales distribution by region?
   A: Bar plot or pie chart.

7. **Q:** How can you identify the top-performing region?
   A: By finding the region with the highest total sales.

8. **Q:** What does grouping by multiple columns help in?
   A: It helps analyze combined effects, like region and product category together.

9. **Q:** Which visualization is suitable for comparing sales by region and category?
   A: Stacked bar plot or grouped bar plot.

10. **Q:** Why is data aggregation important in sales analysis?
    A: It helps summarize large data to identify trends and patterns.

11. **Q:** How can you sort the aggregated results in descending order?
    A: Using .sort_values(ascending=False).

12. **Q:** Which function is used to calculate average sales instead of total sales?
    A: mean().