**Hackathon Subject 2: Option 2: Data Exploration and Analysis from Public APIs**

**The Monthly station data**

The essence of this subject is to encourage you to dive into the world of data using public APIs.

The code has been taken from the Meteostat API (API key from RapidAPI) using Python

"https://meteostat.p.rapidapi.com/point/monthly"

 API **Key**: The 'x-rapidapi-key': "Sign Up for Key" line suggests that you need to sign

 Response **Handling**: The response data is being read and printed out. It is likely in JSON format, so you may want to parse the response as JSON for easier handling.

To extract data from the Meteostat API and save it as a CSV file.

1. **Make the API Request**
2. **Parse the Response**: The response will likely be in JSON format
3. **Convert to CSV**: Use Python's csv module or pandas to convert the parsed JSON data into a CSV file.

To clean and preprocess your data, follow these steps. These steps will help ensure your data is in good shape for analysis:

1. Load the Data: Read the CSV file into a Pandas DataFrame.
2. Inspect the Data: Check for missing values, data types, and other basic characteristics.
3. Handle Missing Data: Decide how to deal with missing data (e.g., drop, fill with mean/median, etc.).
4. Convert Data Types: Ensure that data types are appropriate for analysis (e.g., convert dates to datetime).
5. Remove Duplicates: Identify and remove duplicate rows, if any.
6. Normalize/Standardize Data: Depending on your analysis, normalize or standardize numeric columns.
7. Basic Transformation: Create new features if needed or transform existing ones (e.g., create month/year columns from dates).

To unearth valuable insights from a dataset, we can apply various data analysis techniques. Here’s a structured approach I have taken:

**1. Understand the Dataset**

* **Overview**: Check the size, structure, and types of data. Understand the columns and their significance.
* **Summary Statistics**: Compute mean, median, standard deviation, and other statistical measures for numeric columns.
* **Missing Values**: Identify and address any missing or null values.

**2. Exploratory Data Analysis (EDA)**

* **Distribution Analysis**: Plot histograms or density plots for numerical columns to understand their distributions.
* **Categorical Analysis**: Use bar charts or pie charts to visualize the frequency distribution of categorical variables.
* **Correlation Analysis**: Use heatmaps or pair plots to explore relationships between numeric variables.

**3. Visualizations**

* **Trends Over Time**: For time-series data, plot trends using line charts.
* **Relationships Between Variables**: Use scatter plots to identify relationships between two numeric variables.
* **Comparisons**: Use box plots or violin plots to compare distributions across different groups or categories.

**4. Statistical Testing**

* **Hypothesis Testing**: Perform tests (e.g., t-tests, chi-square tests) to validate any assumptions or claims about the data.
* **Regression Analysis**: Apply linear or logistic regression to understand the impact of one variable on another.

**5. Advanced Techniques**

* **Clustering**: Use algorithms like K-means or hierarchical clustering to identify groups within the data.
* **Dimensionality Reduction**: Apply PCA (Principal Component Analysis) to reduce the number of variables while preserving important information.
* **Anomaly Detection**: Identify any outliers or unusual data points that may be significant.

**6. Insights and Reporting**

* **Summarize Findings**: Create a summary of key insights and findings from your analysis.
* **Visual Storytelling**: Use visualizations to clearly communicate the insights.
* **Actionable Recommendations**: Provide recommendations based on your findings that can inform decision-making.

**Example Using Your E-commerce Dataset**

1. **Distribution of Total Spend**: Plot a histogram to see how customer spending is distributed.
2. **Average Rating by Membership Type**: Use a box plot to compare the average ratings across different membership types.
3. **Correlation Between Age and Total Spend**: Create a scatter plot to explore if there's a relationship between age and spending.
4. **Satisfaction Level Analysis**: Analyze how satisfaction levels vary across different cities or membership types.