To approach this dataset for analysis, you'll want to follow a structured process that involves several steps:

1. **Data Exploration**:
   * Begin by exploring the dataset to understand its structure, size, and contents.
   * Identify the columns (attributes) and their data types.
   * Look for any missing values, outliers, or inconsistencies in the data.
2. **Data Cleaning**:
   * Handle missing values: Decide whether to impute missing values or remove rows/columns with missing data.
   * Check for duplicates and remove them if necessary.
   * Correct any data inconsistencies or errors.
3. **Data Analysis**:
   * Define the key metrics and KPIs (Key Performance Indicators) you want to analyze. For example, total revenue, total profit, units sold, etc.
   * Aggregate the data to calculate these metrics, considering different dimensions such as region, country, item type, sales channel, etc.
   * Analyze sales trends over time (monthly, yearly, etc.) and across different dimensions.
   * Explore relationships between variables to identify factors that influence sales performance. For example, the impact of order priority or unit price on total revenue.
4. **Visualization**:
   * Create visualizations (e.g., bar charts, line graphs, pie charts) to represent the analyzed data and insights effectively.
   * Use Tableau or Python libraries like Matplotlib or Seaborn for visualization.
5. **Interpretation**:
   * Interpret the findings from your analysis and visualizations. What patterns or trends do you observe? Are there any interesting insights or anomalies?
   * Identify actionable insights that can help optimize sales strategies or improve business performance.
6. **Reporting and Presentation**:
   * Summarize your findings in a clear and concise report or presentation.
   * Use visualizations and storytelling techniques to communicate your analysis effectively.
   * Provide recommendations based on your insights to guide decision-making.

Here's a general Python-based approach:

* Use pandas library to load the dataset and perform data manipulation.
* Utilize matplotlib and seaborn for visualization.
* Write modular functions/classes for different tasks such as data cleaning, analysis, and visualization to ensure code modularity and testability.

Remember to keep your code well-documented, modular, and easy to understand to meet the evaluation criteria. Additionally, ensure that your analysis addresses the objectives and requirements of the project effectively.

Amazona sales analysis

* Import libraries
* Load the file db = pd.read\_csv(“path of the file”)
* Now **Data Cleaning and Preparation**:
  + 1. Convert the 'Order Date' and 'Ship Date' columns to datetime data type using **pd.to\_datetime()** to facilitate date-based analysis.
    2. Check for any inconsistencies or errors in the data, such as misspellings or invalid values, and address them as needed.
       1. **Check for Missing Values**: Use the **isnull()** method to identify any missing values in your DataFrame:
       2. **Check for Duplicates**: Use the **duplicated()** method to identify duplicate rows in your DataFrame:

If you encounter duplicate records in the future, you can use the **drop\_duplicates()** method to remove them from your DataFrame:

* + - 1. **Check for Unique Values**: Use the **unique()** method to check for unique values in categorical columns:
      2. **Check for Consistency in Categorical Data**: Check for consistent spelling and formatting in categorical columns, such as 'Region', 'Country', 'Item Type', etc. You can use value\_counts() to inspect the frequency of values:
      3. **Check for Consistency in Numeric Data**: Check for outliers or invalid values in numeric columns using descriptive statistics:

**Df.describe()**

* + - 1. **Check for Data Integrity**: Check for any logical inconsistencies or constraints in the data. For example, ensure that 'Total Revenue' is equal to the product of 'Units Sold' and 'Unit Price'.
         1. *[ we found the discrepancies when multiple unit sold \* unit price there were not matching with the total revenue ]*
         2. *[now I have corrected total cost, total revenue, profit]*