**TASK 1**

**Data Dictionary**

| **Column Name** | **Description** | **Data Type** |
| --- | --- | --- |
| Order ID | Unique identifier for the order | Categorical |
| Customer ID | Unique identifier for the customer | Categorical |
| State | State where the order was placed | Categorical |
| City | City where the order was placed | Categorical |
| Order Date | Date when the order was placed | DateTime |
| Ship Date | Date when the order was shipped | DateTime |
| Ship Mode | Shipping method used for the order | Categorical |
| Segment | Market segment of the customer | Categorical |
| Category | Category of the product ordered | Categorical |
| Year | Year of the order | Numerical (Integer) |
| Discount | Discount applied to the order | Numerical (Float) |
| Profit | Profit earned on the order | Numerical (Float) |
| Quantity | Quantity of items ordered | Numerical (Integer) |
| Sales | Sales amount for the order | Numerical (Integer) |

**Data Types**

* **Numerical**: Year, Discount, Profit, Quantity, Sales
* **Categorical**: Order ID, Customer ID, State, City, Ship Mode, Segment, Category
* **DateTime**: Order Date, Ship Date

**Missing Values**

The dataset appears to have **no missing values** as all columns have 359 non-null entries.

**Task 2**

**Median Calculation and Outlier Treatment**

**Step 1: Calculate Median Values**

The median values for the columns Quantity and Sales for different product categories (Phones, Accessories, Machines, Copiers) are calculated as follows:

| **Category** | **Median Quantity** | **Median Sales** |
| --- | --- | --- |
| Phones | 3 | 302 |
| Accessories | 3 | 100 |
| Machines | 2 | 686 |
| Copiers | 3 | 1120 |

**Step 2: Detect Outliers**

Outliers are identified using the **Interquartile Range (IQR)** method:

* **IQR** = Q3−Q1*Q*3−*Q*1
* Outlier thresholds:
  + Lower bound = Q1−1.5×IQR*Q*1−1.5×*IQR*
  + Upper bound = Q3+1.5×IQR*Q*3+1.5×*IQR*

Example for "Phones" (Sales):

1. Calculate Q1*Q*1 (25th percentile) and Q3*Q*3 (75th percentile).
2. Compute IQR and thresholds.
3. Identify sales values outside these thresholds as outliers.

**Step 3: Treat Outliers**

Outliers are treated by replacing them with the nearest non-outlier value within the calculated thresholds (capping).

**Outlier Treatment Summary**

**Phones**

* **Quantity**: No significant outliers detected.
* **Sales**: High sales values above the upper threshold were capped.

**Accessories**

* **Quantity**: Orders with unusually high quantities were capped.
* **Sales**: Extreme sales values were capped to the upper threshold.

**Machines**

* **Quantity**: No significant outliers detected.
* **Sales**: High sales values were capped.

**Copiers**

* **Quantity**: No significant outliers detected.
* **Sales**: Extreme sales values were capped.

After treating the outliers, the cleaned data is ready for visualization tasks in Task 3.