Part 1: Identify All Inconsistencies in the Dataset

To identify inconsistencies in the dataset, we will analyze each column for issues such as missing values, incorrect data formats, and logical inconsistencies. Here are the potential inconsistencies:

Missing Values:

- Age: Impute the missing values for custID 3 and 8.
- Price: Impute the missing value for custID 5.
- RatingOfProduct: Impute the missing values for custID 8 and 14.
- AdvertisingAgency: Impute the missing values for custID 4 and 9.
- custName: Impute the missing value for custID 13.
- Product: Impute the missing value for custID 15.
- DatePurchased: Impute the missing value for custID 7.

Inconsistent Data:

Age: Ensure that age values are logical (e.g., no negative or extremely high values).

DatePurchased: Dates should be in a consistent format (YYYY-MM-DD). Ensure no future dates.

Price: Ensure prices are positive and not extremely low or high.

Part 2: Five Techniques/Methods to Solve Inconsistencies

Here are five techniques that can be employed to solve inconsistencies:

Imputation for Missing Values:

Mean/Median/Mode Imputation: For numerical columns like Age and Price, use the mean, median, or mode to fill missing values.

Categorical Imputation: For categorical columns like Product and AdvertisingAgency, use the most frequent value or create a separate category for missing data.

Validation Checks:

Ensure data integrity by applying validation checks during data collection. For example, use age ranges to avoid unrealistic values (e.g., ages below 0 or above 120).

Standardization:

Apply consistent formatting for date fields (e.g., ensuring all dates follow the same format such as YYYY-MM-DD).

Ensure consistency in categorical fields by eliminating typos (e.g., ensure "Social Media" is always spelled the same way).

Removing Outliers:

Identify and remove any outliers in numeric fields like Price and RatingOfProduct to avoid skewing the analysis.

Data Validation Rules:

Set rules during data entry that enforce mandatory fields (e.g., no record should be accepted without a Product or DatePurchased).

Ensuring Correct Data Capture:

Implement real-time validation checks during data collection to catch errors immediately.

Use dropdowns or prefilled lists for categorical fields (e.g., AdvertisingAgency) to reduce the chances of typos.

Implement mandatory fields for critical data (e.g., Product, DatePurchased).

Part 4: Discuss Challenges and Opportunities

Challenges:

Incomplete Data: Missing values can hinder analysis, requiring sophisticated imputation methods.

Data Quality Issues: Inaccuracies and inconsistencies can lead to incorrect conclusions.

Outliers: Extreme values can skew analysis and decision-making.

High Dimensionality: With multiple attributes, maintaining consistent quality across all fields can be challenging.

Data Entry Errors: Manual data entry may introduce errors, leading to the need for extensive cleaning.

Opportunities:

Customer Insights: The dataset can reveal patterns in purchasing behavior, allowing for better-targeted marketing strategies.

Product Optimization: Analyzing customer preferences can help optimize product offerings and pricing.

Advertising Strategy: Insights into the effectiveness of different advertising agencies can help optimize marketing spend.

Customer Segmentation: The dataset allows for segmentation, enabling personalized marketing approaches.

Revenue Growth: By understanding customer preferences and behavior, the company can improve customer satisfaction and drive revenue growth.