

```
In [17]: import pandas as pd

"""
Step #1: Load the Data
We begin by loading the dataset from the specified path.
"""

# Load dataset
file_path = 'C://Users//deekc//OneDrive//Documents//Deeksha_BV//DSC540//Project//abs080922 - clear.csv'
data = pd.read_csv(file_path)

# Display first few rows of the dataset
data.head()

"""
Step #2: Check and Replace Headers
Inspect the column names and replace any spaces with underscores for consistency.
"""

# Replace spaces in column names with underscores
data.columns = data.columns.str.replace(' ', '_')

# Display updated column names
data.columns

"""
Step #3: Handle Missing Values
Inspect and handle any missing or null values.
"""

# Check for missing values
missing_values = data.isnull().sum()
missing_values[missing_values > 0] # Columns with missing values and their count

# Option 1: Drop rows with missing values
data_cleaned = data.dropna()

# Option 2: Fill missing values with appropriate methods (e.g., mean, median)
# Example: data_cleaned = data.fillna(data.mean())

# Display final cleaned dataset
data_cleaned.head()
```

```
"""
Step #4: Check and Convert Data Types
Ensure appropriate data types for each column.
"""

# Check data types
data_cleaned.dtypes

# Convert specific columns to appropriate data types if needed
# Example: data_cleaned['Date'] = pd.to_datetime(data_cleaned['Date'])

# Display updated data types
data_cleaned.dtypes

"""
Step #5: Remove Duplicates
Check and remove any duplicate rows.
"""

# Check for duplicate rows
duplicate_rows = data_cleaned.duplicated()
duplicate_rows.sum() # Number of duplicate rows

# Remove duplicates
data_cleaned = data_cleaned.drop_duplicates()

# Confirm removal of duplicates
data_cleaned.shape # New shape of the dataframe after removing duplicates

print("Cleaned data after transformation")
print(data_cleaned)

"""
Ethical Implications of Data Wrangling
In the context of analyzing retail sales and customer behavior, ethical considerations are crucial.
Changes made to the data involve handling missing values, ensuring data integrity through data type
conversions, and removing duplicates. Given the potential inclusion of customer-related data
(e.g., reviews, purchases), ethical concerns revolve around data privacy, consent, and confidentiality.
Legal guidelines such as GDPR or consumer protection laws may apply, necessitating responsible data
handling practices. Risks associated with data transformations include unintentional disclosure of sensitive
information or bias propagation. Assumptions in data cleaning might include treating missing values or duplicates
based on standardized procedures. The credibility of the dataset sourced from Kaggle assumes authenticity,
but additional verification and validation could further enhance ethical practices. To mitigate ethical implications,
```

```
anonymization techniques, transparency in data usage, and adherence to legal frameworks are recommended.  
""
```

Cleaned data after transformation

```

      Unnamed: 0      product_name \
0      0      Fire TV Stick with Alexa Voice Remote (include...
1      1      All-new Ring Video Doorbell (2nd Gen) - 1080p ...
2      2      Echo Dot (3rd Gen, 2018 release) - Smart speak...
3      3      Fire TV Stick 4K Max streaming device, Wi-Fi 6...
4      4      Amazon Smart Plug
...      ...      ...
1311      1311      Roblox Digital Gift Card - 4,500 Robux [Includ...
1312      1312      Animal Crossing: New Horizons - Nintendo Switch
1313      1313      Roblox Digital Gift Card - 10,000 Robux [Inclu...
1314      1314      VALORANT $10 Gift Card - PC [Online Game Code]
1315      1315      Xbox Game Pass Ultimate: 3 Month Membership [D...

      category categoryRank noRatings      cost \
0      Amazon Devices & Accessories      #1      256,226      $39.99
1      Amazon Devices & Accessories      #2      112,689      $89.99 - $199.98
2      Amazon Devices & Accessories      #3      936,447      $29.99
3      Amazon Devices & Accessories      #4      69,684      $54.99
4      Amazon Devices & Accessories      #5      508,565      $24.99
...      ...      ...      ...      ...
1311      Video Games      #26      112,204      $44.99
1312      Video Games      #27      67,564      []
1313      Video Games      #28      112,204      $89.99
1314      Video Games      #29      9,153      $10.00
1315      Video Games      #30      27,121      $39.99

      REVIEWLIST \
0      [{'review_meta': 'Reviewed in the United State...
1      [{'review_meta': 'Reviewed in the United State...
2      [{'review_meta': 'Reviewed in the United State...
3      [{'review_meta': 'Reviewed in the United State...
4      [{'review_meta': 'Reviewed in the United State...
...      ...
1311      [{'review_meta': 'Reviewed in the United State...
1312      [{'review_meta': 'Reviewed in the United State...
1313      [{'review_meta': 'Reviewed in the United State...
1314      [{'review_meta': 'Reviewed in the United State...
1315      [{'review_meta': 'Reviewed in the United State...

      product_url
0      https://www.amazon.com/fire-tv-stick-with-3rd-...
1      https://www.amazon.com/All-new-Ring-Video-Door...
2      https://www.amazon.com/Echo-Dot/dp/B07FZ8S74R/...
3      https://www.amazon.com/fire-tv-stick-4k-max-wi...

```

```
4      https://www.amazon.com/Amazon-Smart-Plug/dp/B0...
...
1311 https://www.amazon.com/Robux-Roblox-Online-Gam...
1312 https://www.amazon.com/Animal-Crossing-New-Hor...
1313 https://www.amazon.com/Roblox-Gift-Card-Exclus...
1314 https://www.amazon.com/VALORANT-10-Gift-Card-O...
1315 https://www.amazon.com/Xbox-Game-Pass-Ultimate...
```

```
[1314 rows x 8 columns]
```

```
Out[17]: '\nEthical Implications of Data Wrangling\nIn the context of analyzing retail sales and customer behavior, ethical cons
iderations are crucial.\nChanges made to the data involve handling missing values, ensuring data integrity through data
type\nconversions, and removing duplicates. Given the potential inclusion of customer-related data \n(e.g., reviews, pu
rchases), ethical concerns revolve around data privacy, consent, and confidentiality. \nLegal guidelines such as GDPR o
r consumer protection laws may apply, necessitating responsible data \nhandling practices. Risks associated with data t
ransformations include unintentional disclosure of sensitive\ninformation or bias propagation. Assumptions in data clea
ning might include treating missing values or duplicates\nbased on standardized procedures. The credibility of the data
set sourced from Kaggle assumes authenticity, \nbut additional verification and validation could further enhance ethica
l practices. To mitigate ethical implications,\nanonymization techniques, transparency in data usage, and adherence to
legal frameworks are recommended.\n'
```

```
In [ ]:
```