

```

In [19]: import pandas as pd
import sqlite3
import requests
from bs4 import BeautifulSoup
import matplotlib.pyplot as plt
import seaborn as sns

### Milestone 2 ###

def load_csv_data(file_path):
    """Load and clean CSV dataset."""
    data = pd.read_csv(file_path)
    data.columns = data.columns.str.replace(' ', '_')
    data_cleaned = data.dropna().drop_duplicates()
    return data_cleaned

def display_data_info(data, title):
    """Display first few rows and info of the dataframe."""
    if not data.empty:
        print(f"\n{title}")
        print(data.head())
        print(data.info())
        print(data.describe())
    else:
        print(f"\n{title} is empty.")

#### Milestone 3 ####

def parse_html_file(html_file_path):
    """Parse HTML file and return cleaned data as a list of dictionaries."""
    with open(html_file_path, 'r', encoding='utf-8') as file:
        soup = BeautifulSoup(file.read(), 'html.parser')

    table = soup.find('table')
    if not table:
        print("Table not found in the HTML.")
        return []

    headers = [header.text.strip() for header in table.find('thead').find_all('th')]
    rows = table.find_all('tr')[1:]
    data = []

```

```

for row in rows:
    cells = row.find_all('td')
    row_data = [cell.text.strip() for cell in cells]
    row_dict = dict(zip(headers, row_data))
    data.append(row_dict)

cleaned_data = [
    {
        **item,
        'numberOfOffers': int(item['numberOfOffers']) if item.get('numberOfOffers') else None,
        'price_value': float(item['price/value'].replace('$', '').replace(',', '')) if item.get('price/value') else
    }
    for item in data if item.get('price/value')
]

unique_data = [dict(t) for t in {tuple(d.items()) for d in cleaned_data}]
return unique_data

#### Milestone 4 ####

def fetch_weather_data(api_key, city):
    """Fetch weather data from OpenWeatherMap API."""
    url = f'https://api.openweathermap.org/data/2.5/weather?q={city}&appid={api_key}'
    response = requests.get(url)
    if response.status_code == 200:
        return response.json()
    else:
        print(f"Failed to fetch weather data for {city} from the API. Status code:", response.status_code)
        return None

def format_weather_data(weather_data):
    """Format raw weather data into a dictionary."""
    if not weather_data:
        return None
    formatted_data = {
        'city': weather_data['name'],
        'country': weather_data['sys']['country'],
        'temperature_celsius': round(weather_data['main']['temp'] - 273.15, 2),
        'description': weather_data['weather'][0]['description']
    }
    formatted_data['temperature_fahrenheit'] = round((formatted_data['temperature_celsius'] * 9/5) + 32, 2)
    formatted_data['description'] = formatted_data['description'].replace('sky', '').strip()
    return formatted_data

```

```
def create_database(db_name):
    """Create SQLite database and return connection and cursor."""
    conn = sqlite3.connect(db_name)
    cursor = conn.cursor()
    return conn, cursor

def create_tables(cursor):
    """Create necessary tables in the database."""
    create_table_website = """
    CREATE TABLE IF NOT EXISTS website_data (
        asin TEXT,
        categoryFullName TEXT,
        categoryName TEXT,
        categoryUrl TEXT,
        name TEXT,
        numberOfOffers INTEGER,
        position INTEGER,
        price_currency TEXT,
        price_value REAL,
        reviewsCount INTEGER,
        stars REAL,
        thumbnailUrl TEXT,
        url TEXT
    );
    """
    cursor.execute(create_table_website)
    print("Table website_data created.")

    create_table_weather = """
    CREATE TABLE IF NOT EXISTS weather_data (
        city TEXT,
        country TEXT,
        temperature_celsius REAL,
        temperature_fahrenheit REAL,
        description TEXT
    );
    """
    cursor.execute(create_table_weather)
    print("Table weather_data created.")

    create_table_csv = """
    CREATE TABLE IF NOT EXISTS csv_data (
        product_name TEXT,
```

```

        category TEXT,
        categoryRank TEXT,
        noRatings TEXT,
        cost TEXT,
        REVIEWLIST TEXT,
        product_url TEXT,
        customer_city TEXT,
        ASIN TEXT
    );
    """

    cursor.execute(create_table_csv)
    print("Table csv_data created.")

def insert_data_to_db(df, table_name, conn):
    """Insert data from DataFrame into the specified table."""
    df.to_sql(table_name, conn, if_exists='replace', index=False)
    print(f"Data inserted into {table_name}.")

def merge_data(cursor):
    """Merge data from website_data, csv_data, and weather_data tables."""
    cursor.execute("""
    SELECT w.asin, w.categoryFullName, w.categoryName, w.categoryUrl, w.name, w.numberOfOffers, w.position, w.reviewsCount,
           c.product_name, c.category, c.categoryRank, c.noRatings, c.cost, c.REVIEWLIST, c.product_url, wd.city, wd.temperature
    FROM website_data w
    JOIN csv_data c ON w.asin = c.ASIN
    JOIN weather_data wd ON c.customer_city = wd.city;
    """)
    columns = [
        'asin', 'categoryFullName', 'categoryName', 'categoryUrl', 'name', 'numberOfOffers', 'position', 'reviewsCount',
        'product_name', 'category', 'categoryRank', 'noRatings', 'cost', 'REVIEWLIST', 'product_url', 'city', 'temperature'
    ]
    return cursor.fetchall(), columns

def normalize_keys(df, key):
    """Normalize join keys: lowercase and strip whitespace."""
    df[key + '_normalized'] = df[key].str.lower().str.strip()
    return df

def plot_visualizations(merged_df):
    """Create and display various visualizations."""
    if not merged_df.empty:
        plt.figure(figsize=(12, 8))
        sns.histplot(data=merged_df, x='cost', bins=30, kde=True)
        plt.title('Distribution of Product Costs')
        plt.xlabel('Cost ($)')

```

```
plt.ylabel('Frequency')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

plt.figure(figsize=(12, 8))
sns.scatterplot(data=merged_df, x='cost', y='stars')
plt.title('Product Costs vs. Ratings')
plt.xlabel('Cost ($)')
plt.ylabel('Ratings (Stars)')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

plt.figure(figsize=(12, 8))
sns.scatterplot(data=merged_df, x='temperature_celsius', y='stars')
plt.title('Temperature in Customer Cities vs. Ratings')
plt.xlabel('Temperature (°C)')
plt.ylabel('Ratings (Stars)')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

plt.figure(figsize=(14, 10))
sns.scatterplot(data=merged_df, x='reviewsCount', y='stars')
plt.title('Product Ratings vs. Number of Reviews')
plt.xlabel('Number of Reviews')
plt.ylabel('Ratings (Stars)')
plt.tight_layout()
plt.show()

plt.figure(figsize=(12, 8))
sns.countplot(data=merged_df, y='category', order=merged_df['category'].value_counts().index)
plt.title('Distribution of Product Categories')
plt.xlabel('Count')
plt.ylabel('Category')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

else:
    print("No data available to plot.")

def print_final_dataset(merged_df):
```

```
"""Print the final dataset after all transformations."""
print("\nHuman Readable Dataset After All Transformations:")
with pd.option_context('display.max_rows', None, 'display.max_columns', None):
    print(merged_df)

def main():
    csv_file_path = 'C://Users//deekc//OneDrive//Documents//Deeksha_BV//DSC540//Project//abs080922 - clear.csv'
    html_file_path = 'C://Users//deekc//OneDrive//Documents//Deeksha_BV//DSC540//Project//dataset_amazon-bestsellers_sc
    api_key = 'ae897397968e3c35aff065764c4b72b4'

    print("Loading CSV data...")
    csv_data = load_csv_data(csv_file_path)
    display_data_info(csv_data, "CSV Data")

    print("\nParsing HTML data...")
    website_data = parse_html_file(html_file_path)
    website_df = pd.DataFrame(website_data)
    display_data_info(website_df, "Website Data")

    print("\nCreating database and tables...")
    conn, cursor = create_database('merged_data.db')
    create_tables(cursor)

    print("\nInserting data into database...")
    insert_data_to_db(csv_data, 'csv_data', conn)
    insert_data_to_db(website_df, 'website_data', conn)

    print("\nIdentifying common ASINs and fetching weather data...")
    common_asins = set(csv_data['asin']).intersection(set(website_df['asin']))
    common_cities = csv_data[csv_data['asin'].isin(common_asins)]['customer_city'].unique()

    weather_data_list = []
    for city in common_cities:
        weather_data = fetch_weather_data(api_key, city)
        cleaned_weather_data = format_weather_data(weather_data)
        if cleaned_weather_data:
            weather_data_list.append(cleaned_weather_data)

    weather_df = pd.DataFrame(weather_data_list)
    display_data_info(weather_df, "Weather Data")

    insert_data_to_db(weather_df, 'weather_data', conn)

    print("\nMerging data...")
    merged_data, columns = merge_data(cursor)
```

```

merged_df = pd.DataFrame(merged_data, columns=columns)
display_data_info(merged_df, "Merged Data")

conn.close()

print("\nCreating visualizations...")
plot_visualizations(merged_df)

# Print the final dataset
print_final_dataset(merged_df)

# Summary of the project
summary = """
### Project Summary

In this project, we integrated product data from various sources to analyze and visualize product trends and correlations. We loaded, cleaned, and merged CSV data with scraped HTML data and enriched it with weather data from an API. We used SQL to store and manage the data, and created visualizations to uncover insights.

#### Key Learnings

- Data Integration: Combining datasets from different sources requires meticulous data cleaning and normalization.
- Visualization: Visualizations are crucial for understanding data distributions and relationships.
- APIs: Leveraging APIs can enrich datasets and provide additional context.

#### Ethical Implications

- Data Cleansing: Removing duplicates and null values can bias the dataset.
- Privacy: Using city names and weather data should be handled with caution to avoid potential privacy issues.
- Accuracy: Data inaccuracies during transformations can lead to misleading insights.

#### Mitigation Strategies

- Transparency: Document all data cleansing steps and transformations.
- Compliance: Ensure data usage complies with relevant privacy laws and regulations.
- Validation: Validate data sources and transformations to maintain accuracy and integrity.

This project highlighted the importance of ethical considerations in data handling and the value of thorough data integration.
"""
print(summary)

if __name__ == '__main__':
    main()

```

Loading CSV data...

CSV Data

	asin	product_name
0	B7920378403	Fire TV Stick with Alexa Voice Remote (include...
1	B6533933898	All-new Ring Video Doorbell (2nd Gen) - 1080p ...
2	B0BTYCRJSS	Echo Dot (3rd Gen, 2018 release) - Smart speak...
3	B9526586604	Fire TV Stick 4K Max streaming device, Wi-Fi 6...
4	B4426290690	Amazon Smart Plug

	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

	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...}

	product_url	customer_city
0	https://www.amazon.com/fire-tv-stick-with-3rd-...	New York
1	https://www.amazon.com/All-new-Ring-Video-Door...	Los Angeles
2	https://www.amazon.com/Echo-Dot/dp/B07FZ8S74R/...	Chicago
3	https://www.amazon.com/fire-tv-stick-4k-max-wi...	Houston
4	https://www.amazon.com/Amazon-Smart-Plug/dp/B0...	Phoenix

<class 'pandas.core.frame.DataFrame'>

Index: 1314 entries, 0 to 1315

Data columns (total 9 columns):

#	Column	Non-Null Count	Dtype
0	asin	1314 non-null	object
1	product_name	1314 non-null	object
2	category	1314 non-null	object
3	categoryRank	1314 non-null	object
4	noRatings	1314 non-null	object
5	cost	1314 non-null	object
6	REVIEWLIST	1314 non-null	object
7	product_url	1314 non-null	object
8	customer_city	1314 non-null	object

dtypes: object(9)

memory usage: 102.7+ KB

None

	asin	product_name \
count	1314	1314
unique	1313	1195
top	B2255878648	2017 Topps WWE Road to Wrestlemania HUGE Facto...
freq	2	7

	category	categoryRank	noRatings	cost	REVIEWLIST \
count	1314	1314	1314	1314	1314
unique	40	50	1079	853	1117
top	Amazon Devices & Accessories	#1	1	[]	[]
freq	50	40	14	33	57

	product_url	customer_city
count	1314	1314
unique	1216	415
top	https://www.amazon.com/Wrestlemania-Factory-EX...	Springfield
freq	7	9

Parsing HTML data...

Website Data

	asin	categoryFullName	categoryName \
0	B09PDLBFKY	Best Sellers in Electronics	Electronics
1	B0B3DC27Q3	Best Sellers in Electronics	Electronics
2	B09BVXT8TJ	Best Sellers in Electronics	Electronics
3	B08MTBCXWX	Best Sellers in Electronics	Electronics
4	B09BKHHL5H	Best Sellers in Electronics	Electronics

	categoryUrl \
0	https://www.amazon.com/Best-Sellers-Electronic...
1	https://www.amazon.com/Best-Sellers-Electronic...
2	https://www.amazon.com/Best-Sellers-Electronic...
3	https://www.amazon.com/Best-Sellers-Electronic...
4	https://www.amazon.com/Best-Sellers-Electronic...

	name	numberOfOffers	position \
0	6 Ft Surge Protector Power Strip - 8 Widely Ou...	None	23
1	Cruise Essentials,6 Ft Power Strip with 3 Outl...	None	93
2	SHOKZ OpenRun Pro - Open-Ear Bluetooth Bone Co...	None	96
3	6Ft Power Strip Surge Protector - Yintar Exten...	None	62
4	5-Outlet Surge Protector Wall Charger with 4 U...	None	19

price	price/currency	price/value	reviewsCount	stars \
-------	----------------	-------------	--------------	---------

0	\$	12.98	20456	4.7
1	\$	13.98	5335	4.7
2	\$	179.95	18399	4.6
3	\$	12.99	21036	4.7
4	\$	8.99	23095	4.7

	thumbnailUrl \
0	https://images-na.ssl-images-amazon.com/images...
1	https://images-na.ssl-images-amazon.com/images...
2	https://images-na.ssl-images-amazon.com/images...
3	https://images-na.ssl-images-amazon.com/images...
4	https://images-na.ssl-images-amazon.com/images...

	url	price_value
0	https://www.amazon.com/dp/B09PDLBFKY	12.98
1	https://www.amazon.com/dp/B0B3DC27Q3	13.98
2	https://www.amazon.com/dp/B09BVXT8TJ	179.95
3	https://www.amazon.com/dp/B08MTBCXWX	12.99
4	https://www.amazon.com/dp/B09BKHHL5H	8.99

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 33 entries, 0 to 32
```

```
Data columns (total 15 columns):
```

#	Column	Non-Null Count	Dtype
0	asin	33 non-null	object
1	categoryFullName	33 non-null	object
2	categoryName	33 non-null	object
3	categoryUrl	33 non-null	object
4	name	33 non-null	object
5	numberOfOffers	0 non-null	object
6	position	33 non-null	object
7	price	33 non-null	object
8	price/currency	33 non-null	object
9	price/value	33 non-null	object
10	reviewsCount	33 non-null	object
11	stars	33 non-null	object
12	thumbnailUrl	33 non-null	object
13	url	33 non-null	object
14	price_value	33 non-null	float64

```
dtypes: float64(1), object(14)
```

```
memory usage: 4.0+ KB
```

```
None
```

	price_value
count	33.000000
mean	44.395152

```

std      46.244532
min      8.990000
25%     18.990000
50%     29.000000
75%     41.990000
max     179.950000

```

Creating database and tables...

Table website_data created.

Table weather_data created.

Table csv_data created.

Inserting data into database...

Data inserted into csv_data.

Data inserted into website_data.

Identifying common ASINs and fetching weather data...

Weather Data

	city	country	temperature_celsius	description \
0	Indianapolis	US	15.35	clear
1	Raleigh	US	19.97	few clouds
2	Madison	US	19.49	clear
3	Chesapeake	US	19.84	clear
4	San Bernardino	US	28.48	overcast clouds

	temperature_fahrenheit
0	59.63
1	67.95
2	67.08
3	67.71
4	83.26

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 32 entries, 0 to 31

Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype
0	city	32 non-null	object
1	country	32 non-null	object
2	temperature_celsius	32 non-null	float64
3	description	32 non-null	object
4	temperature_fahrenheit	32 non-null	float64

dtypes: float64(2), object(3)

memory usage: 1.4+ KB

None

	temperature_celsius	temperature_fahrenheit
count	32.000000	32.000000
mean	20.071250	68.127813
std	5.638739	10.148617
min	12.240000	54.030000
25%	14.932500	58.882500
50%	19.665000	67.395000
75%	24.190000	75.540000
max	32.640000	90.750000

Data inserted into weather_data.

Merging data...

Merged Data

	asin	categoryFullName	categoryName	\
0	B09PDLBFKY	Best Sellers in Electronics	Electronics	
1	B0B3DC27Q3	Best Sellers in Electronics	Electronics	
2	B09BVXT8TJ	Best Sellers in Electronics	Electronics	
3	B08MTBCXWX	Best Sellers in Electronics	Electronics	
4	B09BKHHL5H	Best Sellers in Electronics	Electronics	

	categoryUrl	\
0	https://www.amazon.com/Best-Sellers-Electronic...	
1	https://www.amazon.com/Best-Sellers-Electronic...	
2	https://www.amazon.com/Best-Sellers-Electronic...	
3	https://www.amazon.com/Best-Sellers-Electronic...	
4	https://www.amazon.com/Best-Sellers-Electronic...	

	name	numberOfOffers	position	\
0	6 Ft Surge Protector Power Strip - 8 Widely Ou...	None	23	
1	Cruise Essentials,6 Ft Power Strip with 3 Outl...	None	93	
2	SHOKZ OpenRun Pro - Open-Ear Bluetooth Bone Co...	None	96	
3	6Ft Power Strip Surge Protector - Yintar Exten...	None	62	
4	5-Outlet Surge Protector Wall Charger with 4 U...	None	19	

	reviewsCount	stars	thumbnailUrl	...	\
0	20456	4.7	https://images-na.ssl-images-amazon.com/images...	...	
1	5335	4.7	https://images-na.ssl-images-amazon.com/images...	...	
2	18399	4.6	https://images-na.ssl-images-amazon.com/images...	...	
3	21036	4.7	https://images-na.ssl-images-amazon.com/images...	...	
4	23095	4.7	https://images-na.ssl-images-amazon.com/images...	...	

	category	categoryRank	noRatings	cost	\
0	Amazon Launchpad	#13	16,891	\$17.50	
1	Sports Collectibles	#19	2	\$126.00	

2	Handmade Products	#8	2,024	\$9.99
3	Sports & Outdoors	#2	7,126	\$35.71 - \$80.00
4	Cell Phones & Accessories	#5	35,096	\$7.63

```

                                REVIEWLIST \
0  [{'review_meta': 'Reviewed in the United State...
1                                     []
2  [{'review_meta': 'Reviewed in the United State...
3                                     []
4  [{'review_meta': 'Reviewed in the United State...

```

```

                                product_url          city \
0  https://www.amazon.com/COSRX-Repairing-Hydrati...  Chesapeake
1  https://www.amazon.com/Panini-Prizm-Football-M...  Davenport
2  https://www.amazon.com/Amazing-Mirror-Bathroom...  Centennial
3  https://www.amazon.com/LifeStraw-Filter-2-Stag...  Sandy Springs
4  https://www.amazon.com/Ailun-Protector-Compati...  Palm Desert

```

	temperature_celsius	temperature_fahrenheit	description
0	19.84	67.71	clear
1	14.29	57.72	clear
2	15.02	59.04	scattered clouds
3	18.88	65.98	clear
4	28.10	82.58	clear

[5 rows x 22 columns]

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 33 entries, 0 to 32

Data columns (total 22 columns):

#	Column	Non-Null Count	Dtype
0	asin	33 non-null	object
1	categoryFullName	33 non-null	object
2	categoryName	33 non-null	object
3	categoryUrl	33 non-null	object
4	name	33 non-null	object
5	numberOfOffers	0 non-null	object
6	position	33 non-null	object
7	reviewsCount	33 non-null	object
8	stars	33 non-null	object
9	thumbnailUrl	33 non-null	object
10	url	33 non-null	object
11	product_name	33 non-null	object
12	category	33 non-null	object
13	categoryRank	33 non-null	object

```
14 noRatings      33 non-null  object
15 cost           33 non-null  object
16 REVIEWLIST     33 non-null  object
17 product_url    33 non-null  object
18 city           33 non-null  object
19 temperature_celsius  33 non-null  float64
20 temperature_fahrenheit 33 non-null  float64
21 description    33 non-null  object
```

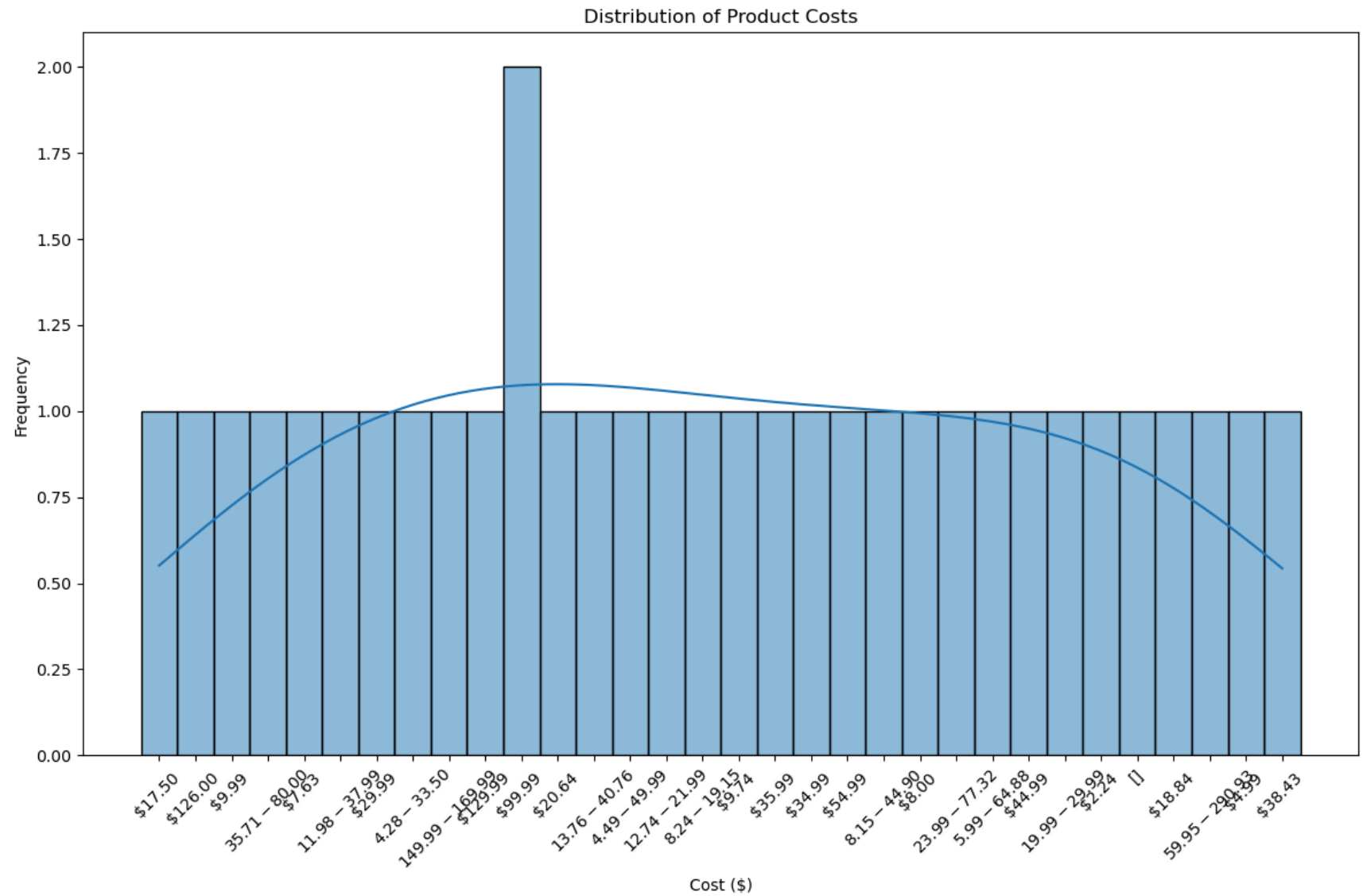
```
dtypes: float64(2), object(20)
```

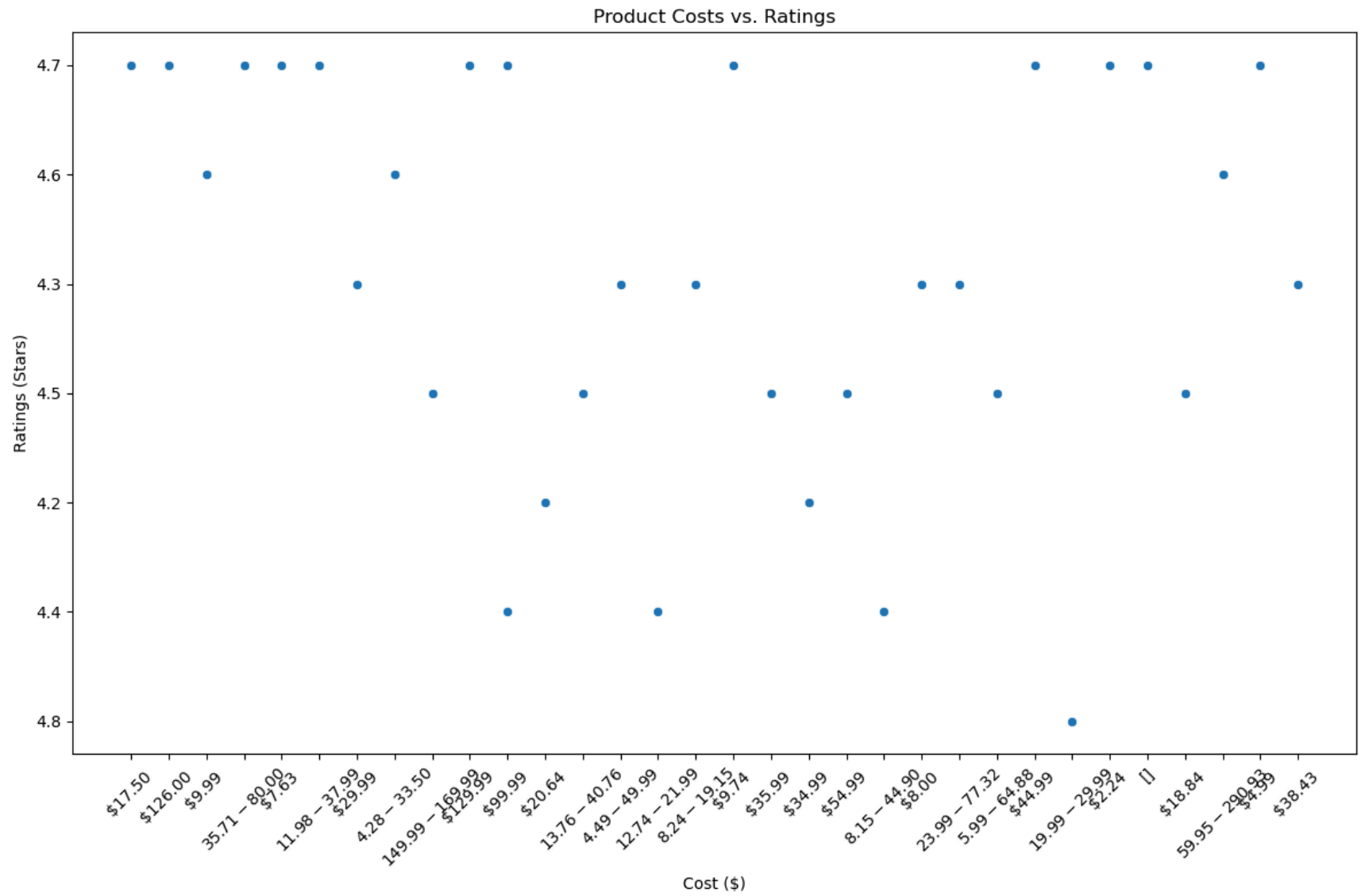
```
memory usage: 5.8+ KB
```

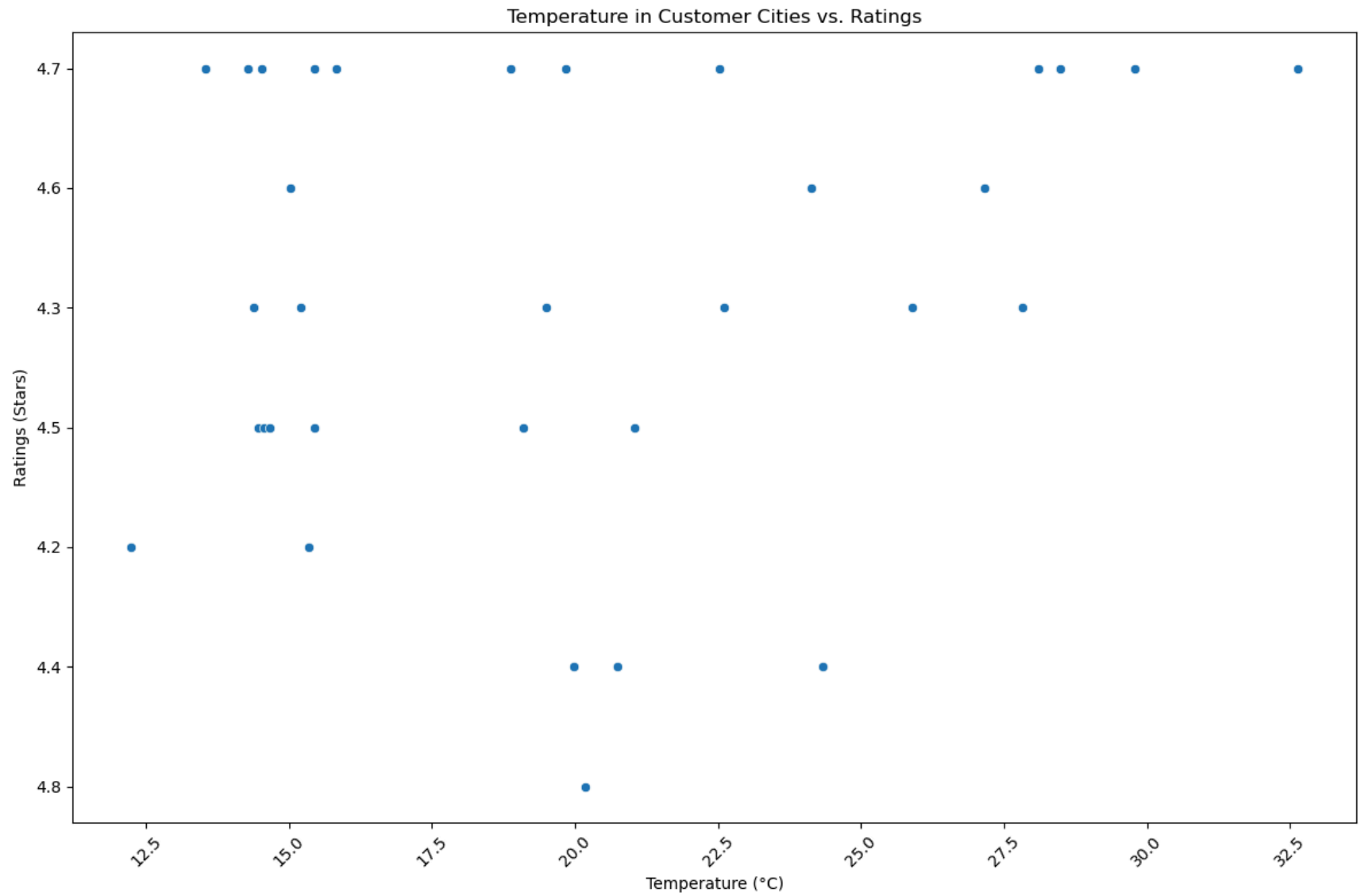
```
None
```

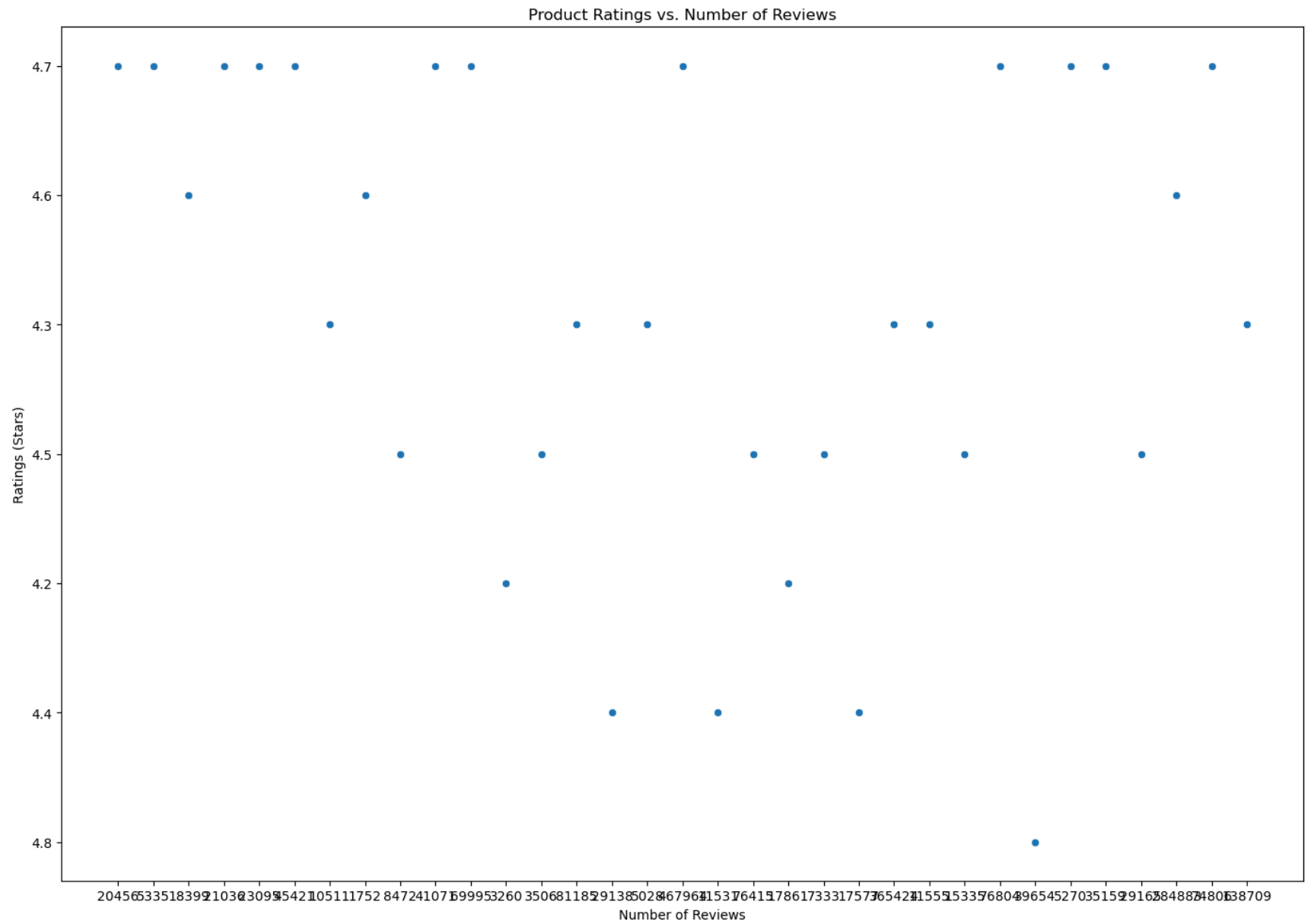
	temperature_celsius	temperature_fahrenheit
count	33.000000	33.000000
mean	19.931212	67.875758
std	5.607934	10.093186
min	12.240000	54.030000
25%	15.020000	59.040000
50%	19.490000	67.080000
75%	24.140000	75.450000
max	32.640000	90.750000

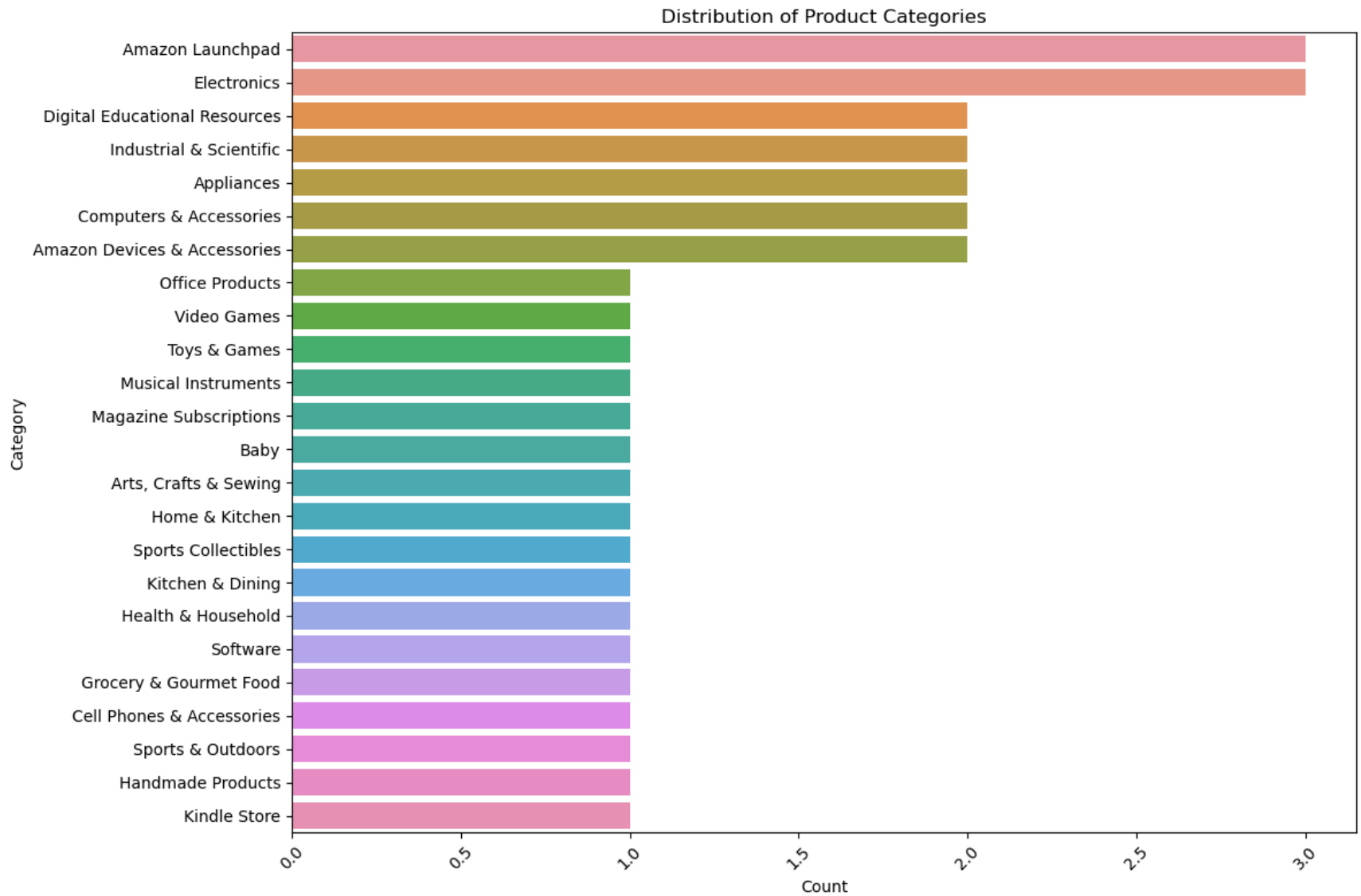
```
Creating visualizations...
```











Human Readable Dataset After All Transformations:

	asin	categoryFullName	categoryName \
0	B09PDLBFKY	Best Sellers in Electronics	Electronics
1	B0B3DC27Q3	Best Sellers in Electronics	Electronics
2	B09BVXT8TJ	Best Sellers in Electronics	Electronics
3	B08MTBCXWX	Best Sellers in Electronics	Electronics
4	B09BKHH15H	Best Sellers in Electronics	Electronics
5	B07CXF452P	Best Sellers in Electronics	Electronics
6	B092Q1TRJC	Best Sellers in Electronics	Electronics
7	B0CLF3VPMV	Best Sellers in Electronics	Electronics
8	B0BQS9Y69G	Best Sellers in Electronics	Electronics
9	B01KBEOL5E	Best Sellers in Electronics	Electronics
10	B08R6S1M1K	Best Sellers in Electronics	Electronics
11	B0BQPZ7DMV	Best Sellers in Electronics	Electronics
12	B0B1HJ666G	Best Sellers in Electronics	Electronics
13	B09FT58QQP	Best Sellers in Electronics	Electronics
14	B0756BXDTX	Best Sellers in Electronics	Electronics
15	B0BTCZ2RR9	Best Sellers in Electronics	Electronics
16	B08C1W5N87	Best Sellers in Electronics	Electronics
17	B0B5F9SZW7	Best Sellers in Electronics	Electronics
18	B07NM3RSRQ	Best Sellers in Electronics	Electronics
19	B07HH1QSLB	Best Sellers in Electronics	Electronics
20	B088D8JG3L	Best Sellers in Electronics	Electronics
21	B0BTYCJXBK	Best Sellers in Electronics	Electronics
22	B07J2Z5DBM	Best Sellers in Electronics	Electronics
23	B0B9TWY11Q	Best Sellers in Electronics	Electronics
24	B0C3W4MNN1	Best Sellers in Electronics	Electronics
25	B0916TKFF2	Best Sellers in Electronics	Electronics
26	B08P5LRY37	Best Sellers in Electronics	Electronics
27	B000HZDYAG	Best Sellers in Electronics	Electronics
28	B0BCH5H2R3	Best Sellers in Electronics	Electronics
29	B09Z6Q2MLC	Best Sellers in Electronics	Electronics
30	B06X16Z7DZ	Best Sellers in Electronics	Electronics
31	B09BKCDXZC	Best Sellers in Electronics	Electronics
32	B08SG2MS3V	Best Sellers in Electronics	Electronics

	categoryUrl \
0	https://www.amazon.com/Best-Sellers-Electronic...
1	https://www.amazon.com/Best-Sellers-Electronic...
2	https://www.amazon.com/Best-Sellers-Electronic...
3	https://www.amazon.com/Best-Sellers-Electronic...
4	https://www.amazon.com/Best-Sellers-Electronic...
5	https://www.amazon.com/Best-Sellers-Electronic...
6	https://www.amazon.com/Best-Sellers-Electronic...
7	https://www.amazon.com/Best-Sellers-Electronic...

8 <https://www.amazon.com/Best-Sellers-Electronic...>
 9 <https://www.amazon.com/Best-Sellers-Electronic...>
 10 <https://www.amazon.com/Best-Sellers-Electronic...>
 11 <https://www.amazon.com/Best-Sellers-Electronic...>
 12 <https://www.amazon.com/Best-Sellers-Electronic...>
 13 <https://www.amazon.com/Best-Sellers-Electronic...>
 14 <https://www.amazon.com/Best-Sellers-Electronic...>
 15 <https://www.amazon.com/Best-Sellers-Electronic...>
 16 <https://www.amazon.com/Best-Sellers-Electronic...>
 17 <https://www.amazon.com/Best-Sellers-Electronic...>
 18 <https://www.amazon.com/Best-Sellers-Electronic...>
 19 <https://www.amazon.com/Best-Sellers-Electronic...>
 20 <https://www.amazon.com/Best-Sellers-Electronic...>
 21 <https://www.amazon.com/Best-Sellers-Electronic...>
 22 <https://www.amazon.com/Best-Sellers-Electronic...>
 23 <https://www.amazon.com/Best-Sellers-Electronic...>
 24 <https://www.amazon.com/Best-Sellers-Electronic...>
 25 <https://www.amazon.com/Best-Sellers-Electronic...>
 26 <https://www.amazon.com/Best-Sellers-Electronic...>
 27 <https://www.amazon.com/Best-Sellers-Electronic...>
 28 <https://www.amazon.com/Best-Sellers-Electronic...>
 29 <https://www.amazon.com/Best-Sellers-Electronic...>
 30 <https://www.amazon.com/Best-Sellers-Electronic...>
 31 <https://www.amazon.com/Best-Sellers-Electronic...>
 32 <https://www.amazon.com/Best-Sellers-Electronic...>

	name	numberOfOffers	position	\
0	6 Ft Surge Protector Power Strip - 8 Widely Ou...	None	23	
1	Cruise Essentials,6 Ft Power Strip with 3 Outl...	None	93	
2	SHOKZ OpenRun Pro - Open-Ear Bluetooth Bone Co...	None	96	
3	6Ft Power Strip Surge Protector - Yintar Exten...	None	62	
4	5-Outlet Surge Protector Wall Charger with 4 U...	None	19	
5	USX MOUNT Full Motion TV Wall Mount for Most 4...	None	86	
6	VIZIO 40-inch D-Series Full HD 1080p Smart TV ...	None	88	
7	SAMSUNG Galaxy Tab A9+ Tablet 11" 64GB Android...	None	81	
8	Roku Voice Remote TV Remote Control with Voi...	None	65	
9	Amazon Basics Full Motion Articulating TV Moni...	None	55	
10	Wall Charger, Surge Protector, QINLIANF 5 Outl...	None	15	
11	JBL Vibe Buds - True Wireless Earbuds, Smart A...	None	69	
12	Mac Book Pro Charger - 118W USB C Charger Fast...	None	87	
13	TOZO A1 Mini Wireless Earbuds Bluetooth 5.3 in...	None	11	
14	Occer 12x25 Compact Binoculars for Adults and ...	None	90	
15	Digital Camera, FHD 1080P Camera, Digital Poin...	None	82	
16	Amazon Fire TV Stick, HD, sharp picture qualit...	None	4	
17	Fitbit Inspire 3 Health &-Fitness-Tracker with...	None	84	

18	Soundcore Anker Life Q20 Hybrid Active Noise C...	None	59
19	Amazon Basics In Ear Wired Headphones, Earbuds...	None	98
20	KYY Portable Monitor 15.6inch 1080P FHD USB-C,...	None	99
21	Soundcore by Anker P20i True Wireless Earbuds,...	None	89
22	TOZO T10 Bluetooth 5.3 Wireless Earbuds with W...	None	30
23	WYZE Cam Pan v3 Indoor/Outdoor IP65-Rated 1080...	None	64
24	Bluetooth Headphones Wireless Earbuds 80hrs Pl...	None	24
25	Roku Express 4K+ Roku Streaming Device 4K/HD...	None	10
26	Power Strip, ALESTOR Surge Protector with 12 O...	None	75
27	Kodak Funsaver One Time Use Film Camera (2-pack)	None	95
28	Roku Express HD Roku Streaming Device with S...	None	17
29	(Pack of 2) Replacement Remote Control Only fo...	None	26
30	Apple EarPods Headphones with 3.5mm Plug, Wire...	None	67
31	Roku Streaming Stick 4K Portable Roku Stream...	None	13
32	Blink Video Doorbell Two-way audio, HD video...	None	12

	reviewsCount	stars	thumbnailUrl \
0	20456	4.7	https://images-na.ssl-images-amazon.com/images...
1	5335	4.7	https://images-na.ssl-images-amazon.com/images...
2	18399	4.6	https://images-na.ssl-images-amazon.com/images...
3	21036	4.7	https://images-na.ssl-images-amazon.com/images...
4	23095	4.7	https://images-na.ssl-images-amazon.com/images...
5	45421	4.7	https://images-na.ssl-images-amazon.com/images...
6	10511	4.3	https://images-na.ssl-images-amazon.com/images...
7	1752	4.6	https://images-na.ssl-images-amazon.com/images...
8	8472	4.5	https://images-na.ssl-images-amazon.com/images...
9	41071	4.7	https://images-na.ssl-images-amazon.com/images...
10	69995	4.7	https://images-na.ssl-images-amazon.com/images...
11	3260	4.2	https://images-na.ssl-images-amazon.com/images...
12	3506	4.5	https://images-na.ssl-images-amazon.com/images...
13	81185	4.3	https://images-na.ssl-images-amazon.com/images...
14	29138	4.4	https://images-na.ssl-images-amazon.com/images...
15	5028	4.3	https://images-na.ssl-images-amazon.com/images...
16	467964	4.7	https://images-na.ssl-images-amazon.com/images...
17	11531	4.4	https://images-na.ssl-images-amazon.com/images...
18	76415	4.5	https://images-na.ssl-images-amazon.com/images...
19	17861	4.2	https://images-na.ssl-images-amazon.com/images...
20	7333	4.5	https://images-na.ssl-images-amazon.com/images...
21	17577	4.4	https://images-na.ssl-images-amazon.com/images...
22	365424	4.3	https://images-na.ssl-images-amazon.com/images...
23	11555	4.3	https://images-na.ssl-images-amazon.com/images...
24	15335	4.5	https://images-na.ssl-images-amazon.com/images...
25	76804	4.7	https://images-na.ssl-images-amazon.com/images...
26	39654	4.8	https://images-na.ssl-images-amazon.com/images...
27	5270	4.7	https://images-na.ssl-images-amazon.com/images...

28	35159	4.7	https://images-na.ssl-images-amazon.com/images...
29	29165	4.5	https://images-na.ssl-images-amazon.com/images...
30	284883	4.6	https://images-na.ssl-images-amazon.com/images...
31	74806	4.7	https://images-na.ssl-images-amazon.com/images...
32	138709	4.3	https://images-na.ssl-images-amazon.com/images...

url \

0	https://www.amazon.com/dp/B09PDLBFBKY
1	https://www.amazon.com/dp/B0B3DC27Q3
2	https://www.amazon.com/dp/B09BVXT8TJ
3	https://www.amazon.com/dp/B08MTBCXWX
4	https://www.amazon.com/dp/B09BKHHL5H
5	https://www.amazon.com/dp/B07CXF452P
6	https://www.amazon.com/dp/B092Q1TRJC
7	https://www.amazon.com/dp/B0CLF3VPMV
8	https://www.amazon.com/dp/B0BQS9Y69G
9	https://www.amazon.com/dp/B01KBEOL5E
10	https://www.amazon.com/dp/B08R6S1M1K
11	https://www.amazon.com/dp/B0BQPZ7DMV
12	https://www.amazon.com/dp/B0B1HJ666G
13	https://www.amazon.com/dp/B09FT58QQP
14	https://www.amazon.com/dp/B0756BXDTX
15	https://www.amazon.com/dp/B0BTCZ2RR9
16	https://www.amazon.com/dp/B08C1W5N87
17	https://www.amazon.com/dp/B0B5F9SZW7
18	https://www.amazon.com/dp/B07NM3RSRQ
19	https://www.amazon.com/dp/B07HH1QSLB
20	https://www.amazon.com/dp/B088D8JG3L
21	https://www.amazon.com/dp/B0BTYCJXBK
22	https://www.amazon.com/dp/B07J2Z5DBM
23	https://www.amazon.com/dp/B0B9TWY11Q
24	https://www.amazon.com/dp/B0C3W4MNN1
25	https://www.amazon.com/dp/B0916TKFF2
26	https://www.amazon.com/dp/B08P5LRY37
27	https://www.amazon.com/dp/B000HZDYAG
28	https://www.amazon.com/dp/B0BCH5H2R3
29	https://www.amazon.com/dp/B09Z6Q2MLC
30	https://www.amazon.com/dp/B06X16Z7DZ
31	https://www.amazon.com/dp/B09BKCDXZC
32	https://www.amazon.com/dp/B08SG2MS3V

product_name \

0	COSRX Snail Mucin 96% Power Repairing Essence ...
1	2021 Panini Prizm NFL Football MEGA box (40 ca...
2	You Look Amazing Mirror Decal Vinyl Decal Bath...

3 LifeStraw Go Water Filter Bottle with 2-Stage ...
 4 Ailun 2 Pack Screen Protector Compatible for i...
 5 OLLY Women's Multivitamin Gummy
 6 Fire TV Stick Lite with latest Alexa Voice Rem...
 7 (10 Pack) EXTRA Gum Sugar Free Chewing Gum Bul...
 8 Ice Maker Countertop, Fast ice Making in 6 Min...
 9 TP-Link Deco Mesh WiFi System (Deco S4) - Up t...
 10 Windows 11 Pro Upgrade, from Windows 11 Home (...
 11 Bounty Quick Size Paper Towels, White, 4 Packs...
 12 THERMOS FUNTAINER 10 Ounce Food Jar
 13 Cricut Everyday Iron On
 14 Simple Modern Kids Water Bottle with Straw Lid...
 15 Dude Wipes Flushable Wipes Dispenser,
 16 Carson Dellosa | The 100+ Series Chemistry Wor...
 17 Insignia 32-inch Class F20 Series Smart HD 720...
 18 Mac Book Pro Charger - 118W USB C Charger Powe...
 19 Ring Rechargeable Battery Pack
 20 Fire TV Stick 4K Max streaming device, Wi-Fi 6...
 21 Amazon Brand - Mama Bear Gentle Touch Diapers
 22 HGTV Magazine
 23 Vandoren Iarinet Traditional Reeds Strength
 24 Crayola Colored Pencils, Assorted Colors
 25 Roblox Digital Gift Card - 4,500 Robux [Includ...
 26 Keangs Disposable Face Masks, 3 Layer Masks 50...
 27 Carson Dellosa | Skill Builders Reading Compre...
 28 Pencils, Wood-Cased, Pre-Sharpened, Graphite #...
 29 Command General Purpose Variety Kit, Various S...
 30 Broan-NuTone 24-inch Under-Cabinet Range Hood ...
 31 The Last Lie Told (Finley O'Sullivan Book 1)
 32 Fujifilm Instax Mini Instant Film, 10 Sheetsx5...

	category	categoryRank	noRatings	cost \
0	Amazon Launchpad	#13	16,891	\$17.50
1	Sports Collectibles	#19	2	\$126.00
2	Handmade Products	#8	2,024	\$9.99
3	Sports & Outdoors	#2	7,126	\$35.71 - \$80.00
4	Cell Phones & Accessories	#5	35,096	\$7.63
5	Amazon Launchpad	#19	31,032	\$11.98 - \$37.99
6	Electronics	#19	15,667	\$29.99
7	Grocery & Gourmet Food	#21	46,303	\$4.28 - \$33.50
8	Appliances	#25	910	\$149.99 - \$169.99
9	Computers & Accessories	#12	22,955	\$129.99
10	Software	#24	39	\$99.99
11	Health & Household	#4	90,107	\$20.64
12	Kitchen & Dining	#3	17,106	\$13.76 - \$40.76

13	Arts, Crafts & Sewing	#30	10,922	\$4.49 - \$49.99
14	Home & Kitchen	#23	16,714	\$12.74 - \$21.99
15	Amazon Launchpad	#3	81,953	\$8.24 - \$19.15
16	Digital Educational Resources	#26	510	\$9.74
17	Amazon Devices & Accessories	#42	79,842	\$99.99
18	Computers & Accessories	#30	1,500	\$35.99
19	Amazon Devices & Accessories	#14	98,318	\$34.99
20	Electronics	#8	69,684	\$54.99
21	Baby	#10	41,817	\$8.15 - \$44.90
22	Magazine Subscriptions	#9	10,286	\$8.00
23	Musical Instruments	#21	11,257	\$23.99 - \$77.32
24	Toys & Games	#1	58,204	\$5.99 - \$64.88
25	Video Games	#26	112,204	\$44.99
26	Industrial & Scientific	#50	17,043	\$19.99 - \$29.99
27	Digital Educational Resources	#29	3,496	\$2.24
28	Office Products	#1	56,049	[]
29	Industrial & Scientific	#21	4,740	\$18.84
30	Appliances	#24	4,918	\$59.95 - \$290.93
31	Kindle Store	#17	11,733	\$4.99
32	Electronics	#21	7,618	\$38.43

REVIEWLIST \

```

0  [{'review_meta': 'Reviewed in the United State...
1  []
2  [{'review_meta': 'Reviewed in the United State...
3  []
4  [{'review_meta': 'Reviewed in the United State...
5  [{'review_meta': 'Reviewed in the United State...
6  [{'review_meta': 'Reviewed in the United State...
7  [{'review_meta': 'Reviewed in the United State...
8  [{'review_meta': 'Reviewed in the United State...
9  [{'review_meta': 'Reviewed in the United State...
10 [{'review_meta': 'Reviewed in the United State...
11 [{'review_meta': 'Reviewed in the United State...
12 [{'review_meta': 'Reviewed in the United State...
13 [{'review_meta': 'Reviewed in the United State...
14 [{'review_meta': 'Reviewed in the United State...
15 [{'review_meta': 'Reviewed in the United State...
16 [{'review_meta': 'Reviewed in the United State...
17 [{'review_meta': 'Reviewed in the United State...
18 [{'review_meta': 'Reviewed in the United State...
19 [{'review_meta': 'Reviewed in the United State...
20 [{'review_meta': 'Reviewed in the United State...
21 [{'review_meta': 'Reviewed in the United State...
22 [{'review_meta': 'Reviewed in the United State...

```

```

23 [{'review_meta': 'Reviewed in the United State...
24 [{'review_meta': 'Reviewed in the United State...
25 [{'review_meta': 'Reviewed in the United State...
26 [{'review_meta': 'Reviewed in the United State...
27 [{'review_meta': 'Reviewed in the United State...
28                                     []
29 [{'review_meta': 'Reviewed in the United State...
30 [{'review_meta': 'Reviewed in the United State...
31 [{'review_meta': 'Reviewed in the United State...
32 [{'review_meta': 'Reviewed in the United State...

```

	product_url	city \
0	https://www.amazon.com/COSRX-Repairing-Hydrati...	Chesapeake
1	https://www.amazon.com/Panini-Prizm-Football-M...	Davenport
2	https://www.amazon.com/Amazing-Mirror-Bathroom...	Centennial
3	https://www.amazon.com/LifeStraw-Filter-2-Stag...	Sandy Springs
4	https://www.amazon.com/Ailun-Protector-Compati...	Palm Desert
5	https://www.amazon.com/OLLY-Womens-Multivitami...	San Bernardino
6	https://www.amazon.com/fire-tv-stick-lite-late...	Pasadena
7	https://www.amazon.com/Pack-EXTRA-Sugar-Chewin...	Round Rock
8	https://www.amazon.com/Countertop-26-5lbs-Port...	Garden Grove
9	https://www.amazon.com/TP-Link-Deco-Replacemen...	Scottsdale
10	https://www.amazon.com/Windows-Upgrade-Home-Di...	Lewisville
11	https://www.amazon.com/Bounty-Quick-Size-Towel...	Kenosha
12	https://www.amazon.com/THERMOS-FUNTAINER-Ounce...	Columbus
13	https://www.amazon.com/Cricut-Everyday-Iron-On...	Visalia
14	https://www.amazon.com/Simple-Modern-Insulated...	Palm Coast
15	https://www.amazon.com/Dude-Wipes-Flushable-Di...	Madison
16	https://www.amazon.com/Chemistry-Grades-9-12/d...	Garden Grove
17	https://www.amazon.com/Insignia-NS-32DF310NA19...	Raleigh
18	https://www.amazon.com/Mac-Book-Pro-Charger-Co...	Fayetteville
19	https://www.amazon.com/Ring-Rechargeable-Batte...	Indianapolis
20	https://www.amazon.com/fire-tv-stick-4k-max-wi...	Peoria
21	https://www.amazon.com/Mama-Bear-Gentle-Touch-...	Wichita Falls
22	https://www.amazon.com/Hearst-Magazines-HGTV-M...	Cleveland
23	https://www.amazon.com/Vandoren-larinet-Tradit...	Cape Coral
24	https://www.amazon.com/Crayola-Colored-Pencils...	Redlands
25	https://www.amazon.com/Robux-Roblox-Online-Gam...	Royal Oak
26	https://www.amazon.com/Keangs-Disposable-Masks...	Bowie
27	https://www.amazon.com/Reading-Comprehension-G...	Santa Rosa
28	https://www.amazon.com/Pencils-Wood-Cased-Pre-...	Hollywood
29	https://www.amazon.com/Command-General-Purpose...	Cleveland Heights
30	https://www.amazon.com/Broan-NuTone-24-inch-Un...	Fort Lauderdale
31	https://www.amazon.com/Last-Told-Finley-OSulli...	Redondo Beach
32	https://www.amazon.com/Fujifilm-Instax-Instant...	Hayward

	temperature_celsius	temperature_fahrenheit	description
0	19.84	67.71	clear
1	14.29	57.72	clear
2	15.02	59.04	scattered clouds
3	18.88	65.98	clear
4	28.10	82.58	clear
5	28.48	83.26	overcast clouds
6	22.61	72.70	broken clouds
7	24.14	75.45	overcast clouds
8	15.45	59.81	clear
9	32.64	90.75	clear
10	22.53	72.55	clear
11	12.24	54.03	light rain
12	14.47	58.05	scattered clouds
13	25.89	78.60	clear
14	24.34	75.81	broken clouds
15	19.49	67.08	clear
16	15.45	59.81	clear
17	19.97	67.95	few clouds
18	21.04	69.87	clear
19	15.35	59.63	clear
20	14.57	58.23	clear
21	20.75	69.35	clear
22	15.21	59.38	broken clouds
23	27.83	82.09	clear
24	19.09	66.36	clear
25	13.54	56.37	broken clouds
26	20.18	68.32	scattered clouds
27	29.79	85.62	overcast clouds
28	15.82	60.48	clear
29	14.67	58.41	overcast clouds
30	27.16	80.89	thunderstorm
31	14.52	58.14	overcast clouds
32	14.38	57.88	clear

Project Summary

In this project, we integrated product data from various sources to analyze and visualize product trends and correlations.

We loaded, cleaned, and merged CSV data with scraped HTML data and enriched it with weather data from an API. We used

SQL to store and manage the data, and created visualizations to uncover insights.

Key Learnings

- **Data Integration**: Combining datasets from different sources requires meticulous data cleaning and normalization.
- **Visualization**: Visualizations are crucial for understanding data distributions and relationships.
- **APIs**: Leveraging APIs can enrich datasets and provide additional context.

Ethical Implications

- **Data Cleansing**: Removing duplicates and null values can bias the dataset.
- **Privacy**: Using city names and weather data should be handled with caution to avoid potential privacy issues.
- **Accuracy**: Data inaccuracies during transformations can lead to misleading insights.

Mitigation Strategies

- **Transparency**: Document all data cleansing steps and transformations.
- **Compliance**: Ensure data usage complies with relevant privacy laws and regulations.
- **Validation**: Validate data sources and transformations to maintain accuracy and integrity.

This project highlighted the importance of ethical considerations in data handling and the value of thorough data integration and analysis.

In []: