In [1]:
```
import pandas as pd
import seaborn as sns
import sqlite3
import matplotlib.pyplot as plt
import numpy as np
```

In [2]:
```
df = pd.read_csv("/Users/thomassimmons/c/d/datasets/retail_store_sales.csv")
```

In [4]:
```
df.shape
```

Out[4]:
```
(12575, 11)
```

In [17]:
```
df.isnull().sum()
```

Out[17]:
```
transaction_id      0
customer_id         0
category            0
item                0
price_per_unit      0
quantity            0
total_spent         0
payment_method      0
location            0
transaction_date    0
discount_applied    0
dtype: int64
```

In [12]:
```
df.columns
```

Out[12]:
```
Index(['transaction_id', 'customer_id', 'category', 'item', 'price_per_uni
t',
       'quantity', 'total_spent', 'payment_method', 'location',
       'transaction_date', 'discount_applied'],
      dtype='object')
```

In [11]:
```
df.columns = df.columns.str.strip().str.replace(" ", "_").str.lower()
```

In [16]:
```
df = df.dropna(subset = ['item', 'price_per_unit', 'quantity', 'total_spent'
```

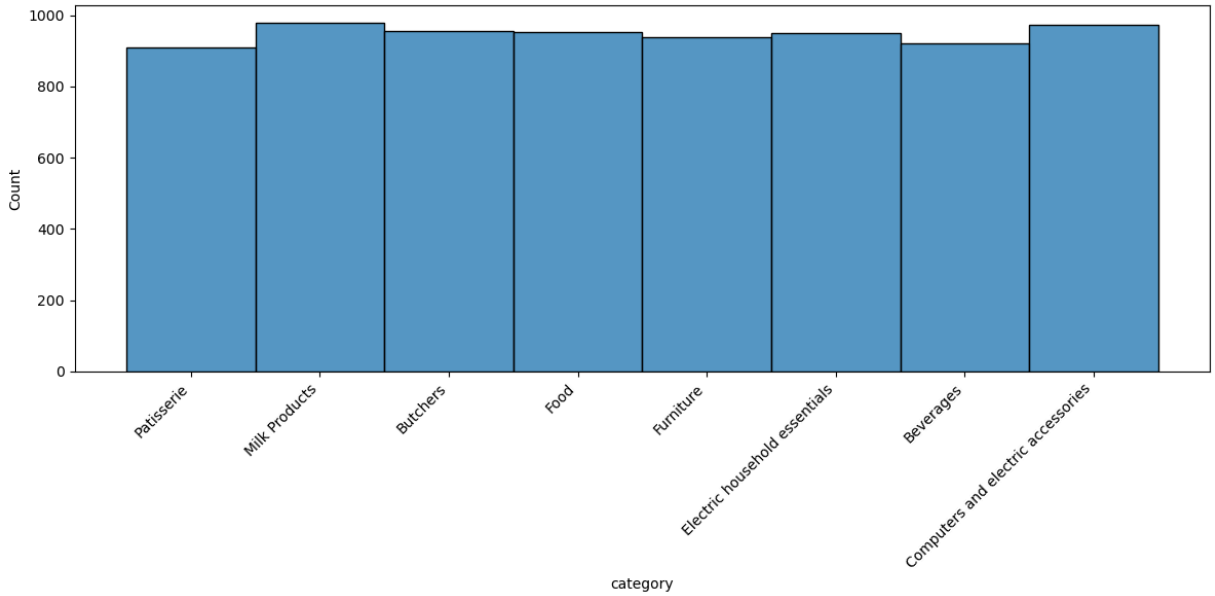
In [18]:
```
df['transaction_date'] = pd.to_datetime(df['transaction_date'], errors='coer
```

```
/var/folders/kf/nt6wlgn90dvc9wsk7ttkqs4w0000gn/T/ipykernel_43368/3405637112.
py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/
stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df['transaction_date'] = pd.to_datetime(df['transaction_date'], errors='co
erce')
```

In [22]:
```
df['month'] = df['transaction_date'].dt.to_period('M')
```

In [26]:
```
plt.figure(figsize=(12,6))
sns.histplot(df['category'])
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



In [32]:
```
print("Dataframe columns:\n", df.columns)
print("\nData types:\n", df.dtypes)
```

```
Dataframe columns:
 Index(['transaction_id', 'customer_id', 'category', 'item', 'price_per_uni
t',
       'quantity', 'total_spent', 'payment_method', 'location',
       'transaction_date', 'discount_applied', 'month'],
      dtype='object')

Data types:
 transaction_id              object
customer_id                 object
category                    object
item                        object
price_per_unit             float64
quantity                   float64
total_spent                float64
payment_method              object
location                    object
transaction_date    datetime64[ns]
discount_applied            object
month                   period[M]
dtype: object
```

In [ ]: