```
import pandas as pd

# Data Loading
df = pd.read_csv('/content/shopping_trends.csv')


print("Data Exploration:")
print(df.head())  # Display the first few rows
print(df.info())  # Display data types and missing values
print(df.describe())  # Summary statistics for numeric columns
```

```
  Payment Method  Shipping Type Discount Applied Promo Code Used  \
0    Credit Card        Express              Yes             Yes
1  Bank Transfer        Express              Yes             Yes
2           Cash  Free Shipping              Yes             Yes
3         PayPal   Next Day Air              Yes             Yes
4           Cash  Free Shipping              Yes             Yes

   Previous Purchases Preferred Payment Method Frequency of Purchases
0                  14                   Venmo            Fortnightly
1                   2                    Cash            Fortnightly
2                  23             Credit Card                 Weekly
3                  49                  PayPal                 Weekly
4                  31                  PayPal               Annually
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 19 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Customer ID               3900 non-null   int64
 1   Age                       3900 non-null   int64
 2   Gender                    3900 non-null   object
 3   Item Purchased            3900 non-null   object
 4   Category                  3900 non-null   object
 5   Purchase Amount (USD)     3900 non-null   int64
 6   Location                  3900 non-null   object
 7   Size                      3900 non-null   object
 8   Color                     3900 non-null   object
 9   Season                    3900 non-null   object
 10  Review Rating             3900 non-null   float64
 11  Subscription Status       3900 non-null   object
 12  Payment Method            3900 non-null   object
 13  Shipping Type             3900 non-null   object
 14  Discount Applied          3900 non-null   object
 15  Promo Code Used           3900 non-null   object
 16  Previous Purchases        3900 non-null   int64
 17  Preferred Payment Method  3900 non-null   object
 18  Frequency of Purchases    3900 non-null   object
dtypes: float64(1), int64(4), object(14)
memory usage: 579.0+ KB
None
       Customer ID          Age  Purchase Amount (USD)  Review Rating  \
count  3900.000000  3900.000000            3900.000000    3900.000000
mean   1950.500000    44.068462              59.764359       3.749949
std    1125.977353    15.207589              23.685392       0.716223
min       1.000000    18.000000              20.000000       2.500000
25%     975.750000    31.000000              39.000000       3.100000
50%    1950.500000    44.000000              60.000000       3.700000
75%    2925.250000    57.000000              81.000000       4.400000
max    3900.000000    70.000000             100.000000       5.000000

       Previous Purchases
count         3900.000000
mean            25.351538
std             14.447125
min              1.000000
25%             13.000000
50%             25.000000
75%             38.000000
max             50.000000
```

```
# Data Cleaning
# Example: Handling missing values by dropping rows with NaN values
df_cleaned = df.dropna()


# Save the Preprocessed Data to a New CSV File
df_cleaned.to_csv('preprocessed_data.csv', index=False)


print("Preprocessed Data Saved to 'preprocessed_data.csv'")
```

```
Preprocessed Data Saved to 'preprocessed_data.csv'
```