Project Requirements

Find out female or male who does more shopping Item purchased vs purchased amount(USD) Catogory vs purchased amount(USD)

Find out the top5 locations which has more sales Find the season which has more sales

Find out review rating of Item purchased

In [2]:

**import** pandas **as** pd

**import** seaborn **as** sns df**=**pd**.**read\_csv("shopping\_trends.csv") df**.**head()

Out[2]:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | **(USD)** |  |
|  | **0** | 1 | 55 | Male | Blouse | Clothing | 53 | Kentucky |
|  | **1** | 2 | 19 | Male | Sweater | Clothing | 64 | Maine |
|  | **2** | 3 | 50 | Male | Jeans | Clothing | 73 | Massachusetts |
|  | **3** | 4 | 21 | Male | Sandals | Footwear | 90 | Rhode Island |
|  | **4** | 5 | 45 | Male | Blouse | Clothing | 49 | Oregon |
| In [3]: | df**.**shape |  |  |  |  |  |  |  |
| Out[3]: | (3900, 19) |  |  |  |  |  |  |  |
| In [4]: | df**.**info() |  |  |  |  |  |  |  |
| <class 'pandas.core.frame.DataFrame'> RangeIndex: 3900 entries, 0 to 3899 Data columns (total 19 columns):  # Column Non-Null Count Dtype | | | | | | | | |

# Customer

**ID**

# Age Gender Item Purchased

**Purchase Category Amount**

# Location Siz

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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

In [5]:

pd**.**isnull(df)**.**sum()

Out[5]:

In [6]:

Out[6]:

Customer ID 0

Age 0

Gender 0

Item Purchased 0

Category 0

Purchase Amount (USD) 0

Location 0

Size 0

Color 0

Season 0

Review Rating 0

Subscription Status 0

Payment Method 0

Shipping Type 0

Discount Applied 0

Promo Code Used 0

Previous Purchases 0

Preferred Payment Method 0

Frequency of Purchases 0

dtype: int64

df**.**columns

Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category', 'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season', 'Review Rating', 'Subscription Status', 'Payment Method', 'Shipping Type', 'Discount Applied', 'Promo Code Used', 'Previous Purchases', 'Preferred Payment Method',

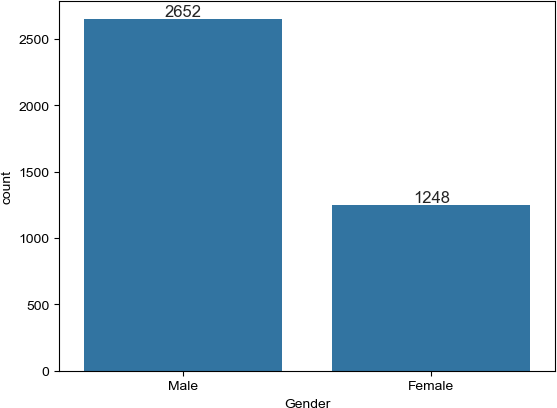
'Frequency of Purchases'], dtype='object')

EDA

In [8]:

ax**=**sns**.**countplot(data**=**df,x**=**'Gender') sns**.**set(rc**=**{'figure.figsize':(15,5)}) **for** bars **in** ax**.**containers:

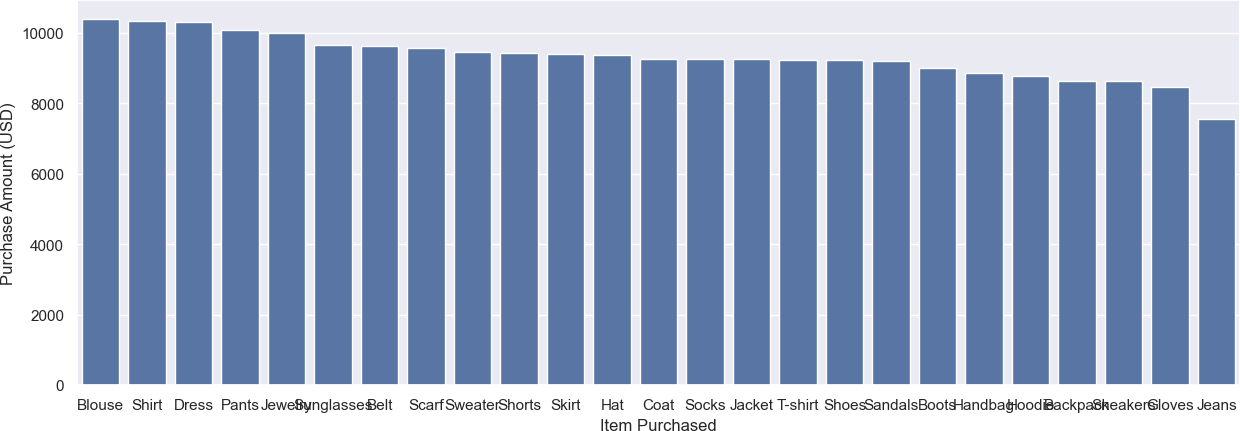
ax**.**bar\_label(bars)



In [9]:

sns\_data**=**df**.**groupby(['Item Purchased'],as\_index**=False**)['Purchase Amount (USD)']**.**sum sns**.**set(rc**=**{'figure.figsize':(15,5)})

sns**.**barplot(x**=**'Item Purchased',y**=**'Purchase Amount (USD)',data**=**sns\_data)

Out[9]:

In [10]:

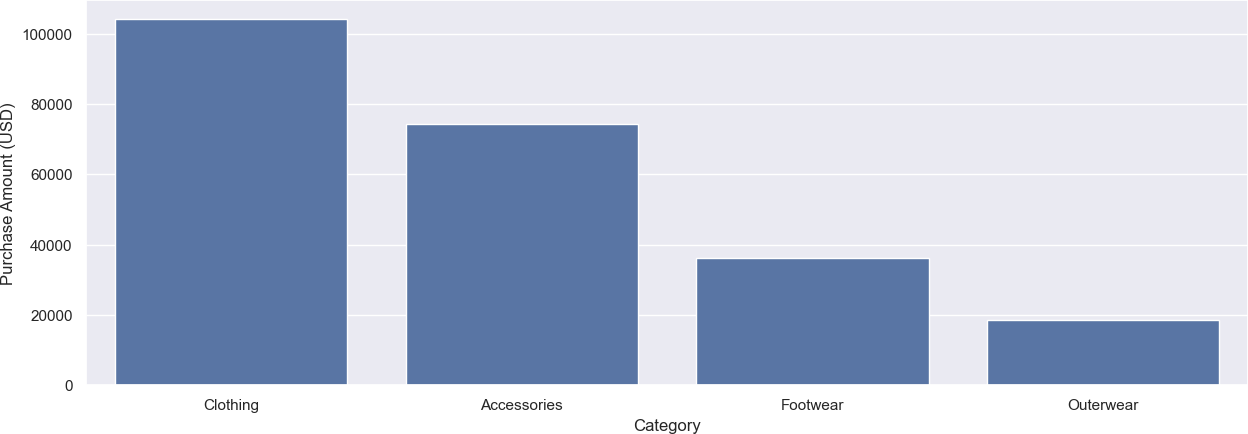
Out[10]:

<Axes: xlabel='Item Purchased', ylabel='Purchase Amount (USD)'>

sns\_data**=**df**.**groupby(['Category'],as\_index**=False**)['Purchase Amount (USD)']**.**sum()**.**sor sns**.**set(rc**=**{'figure.figsize':(15,5)})

sns**.**barplot(x**=**'Category',y**=**'Purchase Amount (USD)',data**=**sns\_data)

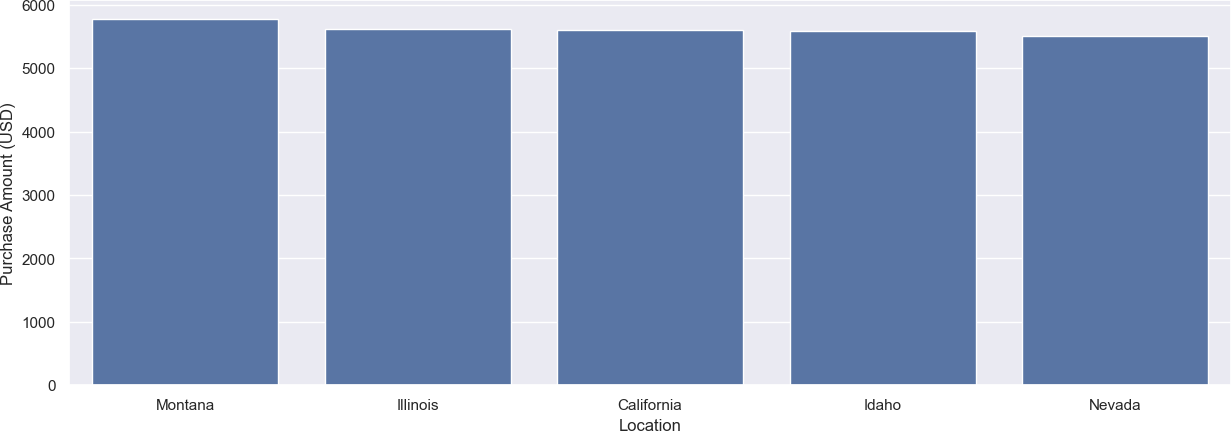
<Axes: xlabel='Category', ylabel='Purchase Amount (USD)'>



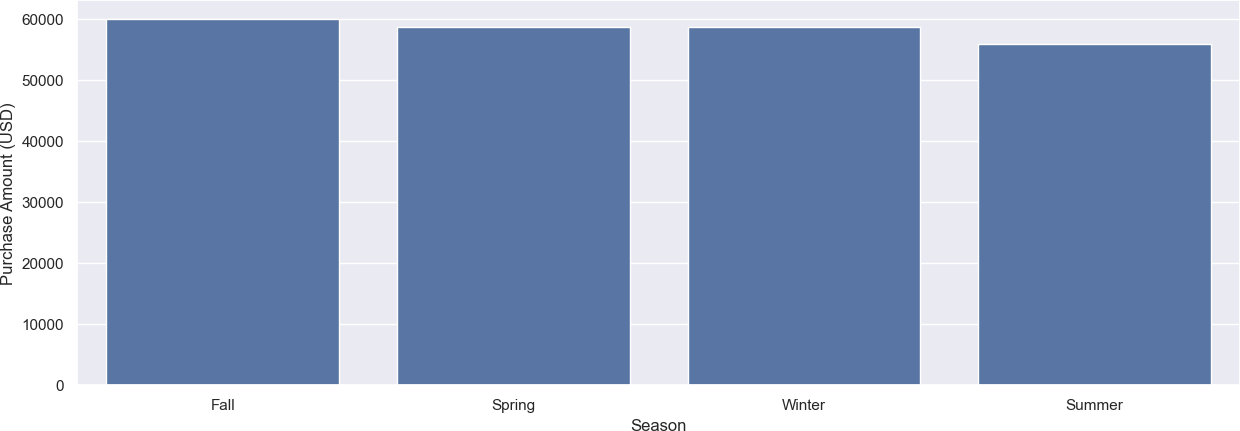
In [11]:

sns\_data**=**df**.**groupby(['Location'],as\_index**=False**)['Purchase Amount (USD)']**.**sum()**.**sor sns**.**set(rc**=**{'figure.figsize':(15,5)})

sns**.**barplot(x**=**'Location',y**=**'Purchase Amount (USD)',data**=**sns\_data)

Out[11]:

In [12]:

Out[12]:

In [13]:

sns\_data**=**df**.**groupby(['Item Purchased'],as\_index**=False**)['Review Rating']**.**sum()**.**sort\_ sns**.**set(rc**=**{'figure.figsize':(15,5)})

sns**.**barplot(x**=**'Item Purchased',y**=**'Review Rating',data**=**sns\_data)

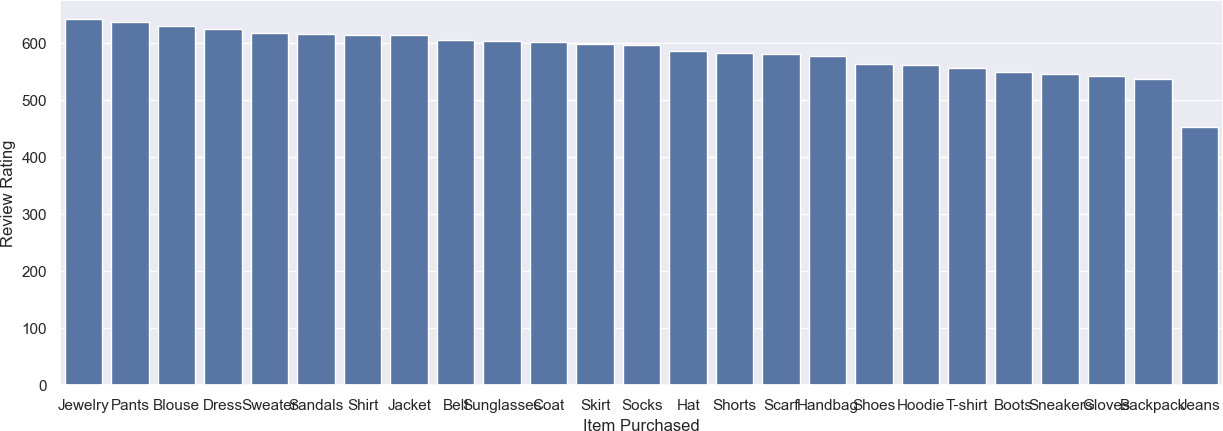
<Axes: xlabel='Location', ylabel='Purchase Amount (USD)'>

sns\_data**=**df**.**groupby(['Season'],as\_index**=False**)['Purchase Amount (USD)']**.**sum()**.**sort\_ sns**.**set(rc**=**{'figure.figsize':(15,5)})

sns**.**barplot(x**=**'Season',y**=**'Purchase Amount (USD)',data**=**sns\_data)

<Axes: xlabel='Season', ylabel='Purchase Amount (USD)'>

Out[13]: <Axes: xlabel='Item Purchased', ylabel='Review Rating'>



Insights

Males do more shopping than females Sales are more for blouse,shirt,pants,dress Clothing category has more sales

Montana,illinois,California,Idaho and Nevada are top 5 locations where sales are high

In fall,winter seasons the sales are high Customer review rating for jewelry is high

In [ ]: