```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("/content/sample_-_superstore.csv")

df
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Produ
0	1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-B 1000175
1	2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-C 1000045
2	3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	90036	West	OFF-L 100002
3	4	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	FUR-T 100005
4	5	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	OFF-S 100007
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9989	9990	CA-2014-110422	1/21/2014	1/23/2014	Second Class	TB-21400	Tom Boeckenhauer	Consumer	United States	Miami	...	33180	South	FUR-F 100018
9990	9991	CA-2017-121258	2/26/2017	3/3/2017	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	92627	West	FUR-F 100007
9991	9992	CA-2017-121258	2/26/2017	3/3/2017	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	92627	West	TEC-P 100036
9992	9993	CA-2017-121258	2/26/2017	3/3/2017	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	92627	West	OFF-P 100040
9993	9994	CA-2017-119914	5/4/2017	5/9/2017	Second Class	CC-12220	Chris Cortes	Consumer	United States	Westminster	...	92683	West	OFF-A 100026

9994 rows × 21 columns

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
```

```
 2   Order Date     9994 non-null   object
 3   Ship Date      9994 non-null   object
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Sub-Category   9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
dtypes: float64(3), int64(3), object(15)
memory usage: 1.6+ MB
```

```
df['Profit'] = df['Profit'].astype('int')
```

```
df['Profit'].dtypes
```

```
dtype('int64')
```

```
df.shape
```

```
(9994, 21)
```

```
df.describe()
```

	Row ID	Postal Code	Sales	Quantity	Discount	Profit
count	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	4997.500000	55190.379428	229.858001	3.789574	0.156203	28.656896
std	2885.163629	32063.693350	623.245101	2.225110	0.206452	234.260108
min	1.000000	1040.000000	0.444000	1.000000	0.000000	-6599.978000
25%	2499.250000	23223.000000	17.280000	2.000000	0.000000	1.728750
50%	4997.500000	56430.500000	54.490000	3.000000	0.200000	8.666500
75%	7495.750000	90008.000000	209.940000	5.000000	0.200000	29.364000
max	9994.000000	99301.000000	22638.480000	14.000000	0.800000	8399.976000

```
df.describe(include="object")
```

	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Region	Product ID	Category	Sub-Category
count	9994	9994	9994	9994	9994	9994	9994	9994	9994	9994	9994	9994	9994	9994
unique	5009	1237	1334	4	793	793	3	1	531	49	4	1862	3	17
top	CA-2017-100111	9/5/2016	12/16/2015	Standard Class	WB-21850	William Brown	Consumer	United States	New York City	California	West	OFF-PA-10001970	Office Supplies	Binders
freq	14	38	35	5968	37	37	5191	9994	915	2001	3203	19	6026	1523

```
df.isnull().sum()
```

```
Row ID          0
Order ID        0
Order Date      0
Ship Date       0
Ship Mode       0
Customer ID     0
Customer Name   0
Segment         0
Country         0
```

```
City             0
State            0
Postal Code      0
Region           0
Product ID       0
Category         0
Sub-Category     0
Product Name     0
Sales            0
Quantity         0
Discount         0
Profit           0
dtype: int64
```

```
df['Ship Date'] = pd.to_datetime(df['Ship Date'])
```

```
df['Ship Year'] =df['Ship Date'].dt.year
```

```
df['Ship Month'] =df['Ship Date'].dt.month
df['Ship Day'] =df['Ship Date'].dt.day
```

```
df['Order Date'] = pd.to_datetime(df['Order Date'])
```

```
df['Order Year'] =df['Order Date'].dt.year
df['Order Month'] =df['Order Date'].dt.month
df['Order Day'] =df['Order Date'].dt.day
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 27 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   datetime64[ns]
 3   Ship Date      9994 non-null   datetime64[ns]
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Sub-Category   9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
 21  Ship Year      9994 non-null   int32
 22  Ship Month     9994 non-null   int32
 23  Ship Day       9994 non-null   int32
 24  Order Year     9994 non-null   int32
 25  Order Month    9994 non-null   int32
 26  Order Day      9994 non-null   int32
dtypes: datetime64[ns](2), float64(3), int32(6), int64(3), object(13)
memory usage: 1.8+ MB
```

```
df
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Sales	Quantity	Discount	Pr
0	1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	261.9600	2	0.00	41.
1	2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	731.9400	3	0.00	219.
2	3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	14.6200	2	0.00	6.
3	4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	957.5775	5	0.45	-383.
4	5	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	22.3680	2	0.20	2.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
9989	9990	CA-2014-110422	2014-01-21	2014-01-23	Second Class	TB-21400	Tom Boeckenhauer	Consumer	United States	Miami	...	25.2480	3	0.20	4.
9990	9991	CA-2017-121258	2017-02-26	2017-03-03	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	91.9600	2	0.00	15.
9991	9992	CA-2017-121258	2017-02-26	2017-03-03	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	258.5760	2	0.20	19.
9992	9993	CA-2017-121258	2017-02-26	2017-03-03	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	29.6000	4	0.00	13.
9993	9994	CA-2017-119914	2017-05-04	2017-05-09	Second Class	CC-12220	Chris Cortes	Consumer	United States	Westminster	...	243.1600	2	0.00	72.

9994 rows × 27 columns

```
df.head(100)
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Sales	Quantity	Discount	Profit	S Y
0	1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	261.9600	2	0.00	41.9136	2(
1	2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	731.9400	3	0.00	219.5820	2(
2	3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	14.6200	2	0.00	6.8714	2(
3	4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	957.5775	5	0.45	-383.0310	2(
4	5	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	22.3680	2	0.20	2.5164	2(
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
95	96	US-2017-109484	2017-11-06	2017-11-12	Standard Class	RB-19705	Roger Barcio	Home Office	United States	Portland	...	5.6820	1	0.70	-3.7880	2(
96	97	CA-2017-161018	2017-11-09	2017-11-11	Second Class	PN-18775	Parhena Norris	Home Office	United States	New York City	...	96.5300	7	0.00	40.5426	2(
97	98	CA-2017-157833	2017-06-17	2017-06-20	First Class	KD-16345	Katherine Ducich	Consumer	United States	San Francisco	...	51.3120	3	0.20	17.9592	2(
98	99	CA-2016-149223	2016-09-06	2016-09-11	Standard Class	ER-13855	Elpida Rittenbach	Corporate	United States	Saint Paul	...	77.8800	6	0.00	22.5852	2(
99	100	CA-2016-158568	2016-08-29	2016-09-02	Standard Class	RB-19465	Rick Bensley	Home Office	United States	Chicago	...	64.6240	7	0.20	22.6184	2(

100 rows × 27 columns

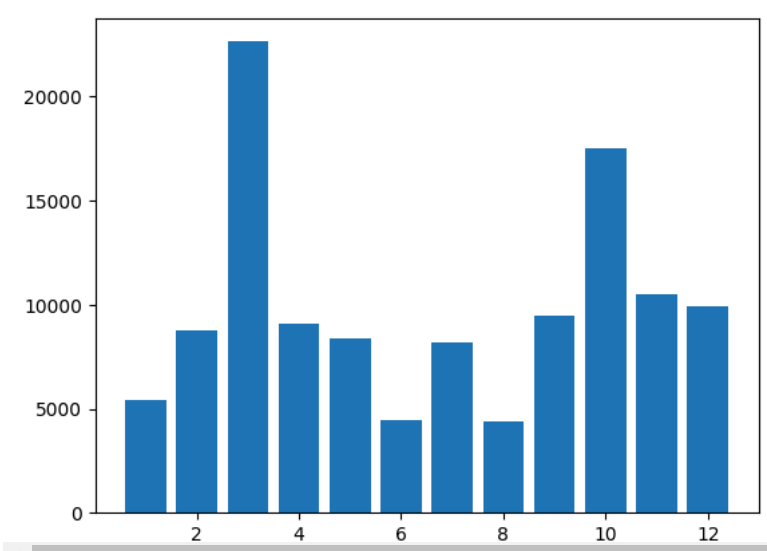
```
df.tail(5)
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Sales	Quantity	Discount	Profi
9989	9990	CA-2014-110422	2014-01-21	2014-01-23	Second Class	TB-21400	Tom Boeckenhauer	Consumer	United States	Miami	...	25.248	3	0.2	4.102
9990	9991	CA-2017-121258	2017-02-26	2017-03-03	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	91.960	2	0.0	15.633
9991	9992	CA-2017-121258	2017-02-26	2017-03-03	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	258.576	2	0.2	19.393
9992	9993	CA-2017-121258	2017-02-26	2017-03-03	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	29.600	4	0.0	13.320
9993	9994	CA-2017-119914	2017-05-04	2017-05-09	Second Class	CC-12220	Chris Cortes	Consumer	United States	Westminster	...	243.160	2	0.0	72.948

5 rows × 27 columns

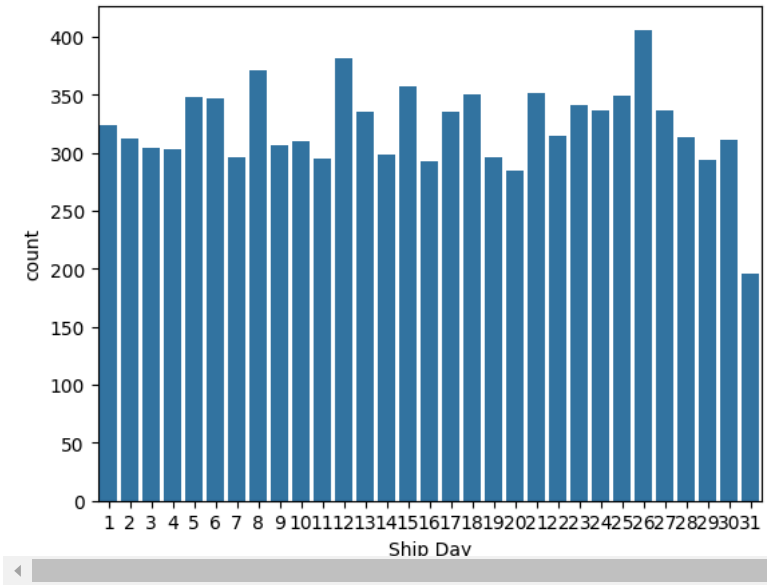
```
x = df['Order Month']
y = df['Sales']


plt.bar(x,y)
plt.show()
```




```
sns.countplot(x = df['Ship Day'])
```

```
<Axes: xlabel='Ship Day', ylabel='count'>
```




```
df_new = df.groupby(['Sales','City'])['Order Year'].value_counts().unstack()
```

```
df['Country'].value_counts()
```

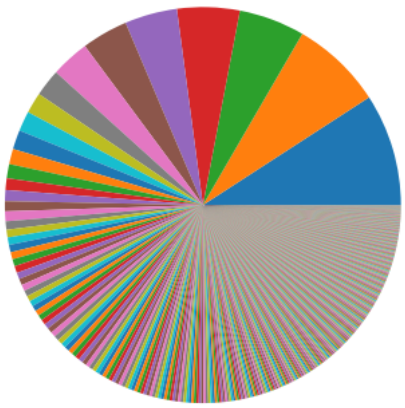
```
Country
United States    9994
Name: count, dtype: int64
```

```
df['City'].value_counts()
```

```
City
New York City      915
Los Angeles        747
Philadelphia       537
San Francisco      510
Seattle            428
                  ...
Glenview             1
Missouri City        1
Rochester Hills      1
Palatine             1
Manhattan            1
Name: count, Length: 531, dtype: int64
```

```
plt.pie(df['City'].value_counts())
plt.show()
```

```
plt.show()
```



```
df['Ship Mode'].value_counts()
```

```
Ship Mode
Standard Class    5968
Second Class      1945
First Class       1538
Same Day           543
Name: count, dtype: int64
```

```
mylabel = ['Standard Class','Second Class','First Class','Same Day']
plt.pie(df['Ship Mode'].value_counts(),startangle=90,counterclock=False,autopct="%0.2f%%",explode=(0.05,0,0,0),labels=mylabel,wedgeprops={"
plt.legend(loc="upper right")
plt.show()
```