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

In [2]:
```
df=pd.read_excel(r"C:\Users\VIIT\OneDrive - klit.ac.in\desktop\data analystics\python\Superstore_USA.xlsx")
df.head()
pd.set_option('display.max_columns', None)
df.head()
```

Out[2]:

	Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	Ship Mode	Customer Segment	Product Category	Product Sub-Category	Product Container	Product Name	Product Base Margin	Region	Ship Date	P
0	18606	Not Specified	0.01	2.88	0.50	2	Janice Fletcher	Regular Air	Corporate	Office Supplies	Labels	Small Box	Avery 49	0.36	Central		
1	20647	High	0.01	2.84	0.93	3	Bonnie Potter	Express Air	Corporate	Office Supplies	Pens & Art Supplies	Wrap Bag	SANFORD Liquid Accent™ Tank-Style Highlighters	0.54	West	Was	
2	23086	Not Specified	0.03	6.68	6.15	3	Bonnie Potter	Express Air	Corporate	Office Supplies	Paper	Small Box	Xerox 1968	0.37	West	Was	
3	23087	Not Specified	0.01	6.68	3.60	3	Bonnie Potter	Regular Air	Corporate	Office Supplies	Scissors, Rulers and Trimmers	Small Pack	Acme® Preferred Stainless Steel Scissors	0.56	West	Was	
4	23088	Not Specified	0.00	205.99	2.50	3	Bonnie Potter	Express Air	Corporate	Technology	Telephones and Communication	Small Box	V70	0.59	West	Was	

In [3]:
```
df.isnull().sum()
```

Out[3]:
```
Row ID                  0
Order Priority          0
Discount                0
Unit Price              0
Shipping Cost           0
Customer ID             0
Customer Name           0
Ship Mode               0
Customer Segment        0
Product Category        0
Product Sub-Category    0
Product Container       0
Product Name            0
Product Base Margin    72
Region                  0
State or Province       0
City                    0
Postal Code             0
Order Date              0
Ship Date               0
Profit                  0
Quantity ordered new    0
Sales                   0
Order ID                0
dtype: int64
```

In [4]:
```
df['Product Base Margin'].fillna(df['Product Base Margin'].mean(),inplace=True)
```

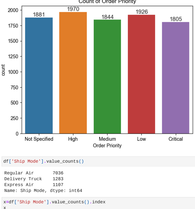
In [5]:
```
df['Order Priority'].value_counts()
```

Out[5]:
```
High              1970
Low               1926
Not Specified     1881
Medium            1844
Critical          1805
Critical             1
Name: Order Priority, dtype: int64
```

In [6]:
```
df['Order Priority']=df['Order Priority'].replace('Critical ','Critical')
```

In [7]:
```
sns.set(rc={'figure.figsize':(15,8)})
ax=sns.countplot(x='Order Priority',data=df,hue='Order Priority')

plt.title('Count of Order Priority')
for bars in ax.containers:
    ax.bar_label(bars)
```



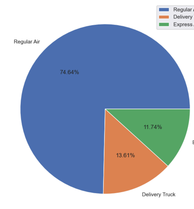
In [8]:
```
df['Ship Mode'].value_counts()
```

Out[8]:
```
Regular Air    7036
Delivery Truck    1283
Express Air    1187
Name: Ship Mode, dtype: int64
```

In [9]:
```
x=df['Ship Mode'].value_counts().index
x
y=df['Ship Mode'].value_counts().values
y
```

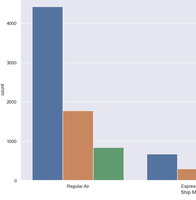
Out[9]:
```
array([7036, 1283, 1187], dtype=int64)
```

In [10]:
```
plt.pie(y,labels=x,autopct='%.2f%%')
plt.legend()
plt.show()
```



In [11]:
```
sns.countplot(x='Ship Mode',data=df,hue='Product Category')
```

Out[11]:
```
<Axes: xlabel='Ship Mode', ylabel='count'>
```



In [12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9426 entries, 0 to 9425
Data columns (total 24 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Row ID                 9426 non-null   int64
 1   Order Priority         9426 non-null   object
 2   Discount               9426 non-null   float64
 3   Unit Price             9426 non-null   float64
 4   Shipping Cost          9426 non-null   float64
 5   Customer ID            9426 non-null   int64
 6   Customer Name          9426 non-null   object
 7   Ship Mode              9426 non-null   object
 8   Customer Segment       9426 non-null   object
 9   Product Category       9426 non-null   object
 10  Product Sub-Category   9426 non-null   object
 11  Product Container      9426 non-null   object
 12  Product Name           9426 non-null   object
 13  Product Base Margin    9426 non-null   float64
 14  Region                 9426 non-null   object
 15  State or Province      9426 non-null   object
 16  City                   9426 non-null   object
 17  Postal Code            9426 non-null   int64
 18  Order Date             9426 non-null   datetime64[ns]
 19  Ship Date              9426 non-null   datetime64[ns]
 20  Profit                 9426 non-null   float64
 21  Quantity ordered new   9426 non-null   int64
 22  Sales                  9426 non-null   float64
 23  Order ID               9426 non-null   int64
dtypes: datetime64[ns](2), float64(6), int64(6), object(11)
memory usage: 1.7+ MB
```

In [13]:
```
df['Order Year']=df['Order Date'].dt.year
```

In [14]:
```
df.info()
```

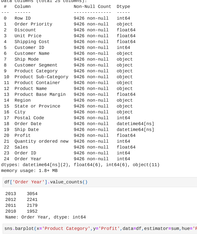
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9426 entries, 0 to 9425
Data columns (total 25 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Row ID                 9426 non-null   int64
 1   Order Priority         9426 non-null   object
 2   Discount               9426 non-null   float64
 3   Unit Price             9426 non-null   float64
 4   Shipping Cost          9426 non-null   float64
 5   Customer ID            9426 non-null   int64
 6   Customer Name          9426 non-null   object
 7   Ship Mode              9426 non-null   object
 8   Customer Segment       9426 non-null   object
 9   Product Category       9426 non-null   object
 10  Product Sub-Category   9426 non-null   object
 11  Product Container      9426 non-null   object
 12  Product Name           9426 non-null   object
 13  Product Base Margin    9426 non-null   float64
 14  Region                 9426 non-null   object
 15  State or Province      9426 non-null   object
 16  City                   9426 non-null   object
 17  Postal Code            9426 non-null   int64
 18  Order Date             9426 non-null   datetime64[ns]
 19  Ship Date              9426 non-null   datetime64[ns]
 20  Profit                 9426 non-null   float64
 21  Quantity ordered new   9426 non-null   int64
 22  Sales                  9426 non-null   float64
 23  Order ID               9426 non-null   int64
 24  Order Year             9426 non-null   int64
dtypes: datetime64[ns](2), float64(6), int64(6), object(11)
memory usage: 1.8+ MB
```

In [15]:
```
df['Order Year'].value_counts()
```

Out[15]:
```
2013    3054
2012    2241
2011    2178
2014    1953
Name: Order Year, dtype: int64
```

In [16]:
```
sns.barplot(x='Product Category',y='Profit',data=df,estimator=sum,hue='Product Category')
```

Out[16]:
```
<Axes: xlabel='Product Category', ylabel='Profit'>
```

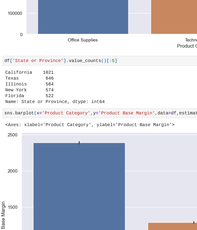


In [17]:
```
df['State or Province'].value_counts()[:5]
```

Out[17]:
```
California    1821
Texas          646
Illinois       583
New York       574
Florida        522
Name: State or Province, dtype: int64
```

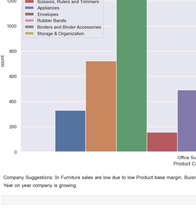
In [18]:
```
sns.barplot(x='Product Category',y='Product Base Margin',data=df,estimator=sum,hue='Product Category')
```

Out[18]:
```
<Axes: xlabel='Product Category', ylabel='Product Base Margin'>
```



In [19]:
```
sns.countplot(x='Product Category',data = df[df['Product Category']=='Office Supplies'],hue='Product Sub-Category')
```

Out[19]:
```
<Axes: xlabel='Product Category', ylabel='count'>
```



Company Suggestions: In Furniture sales are low due to low Product base margin. Business in california is great so introduce new offers there to increase sales.
Year on year company is growing

In [ ]:
```
```