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

In [ ]:
```
#Load the data set
df=pd.read_csv(r"C:\Users\DELL\Downloads\New folder\DataCoSupplyChainDataset.csv",e
df.head()
```

Out[ ]:

	Type	Days for shipping (real)	Days for shipment (scheduled)	Benefit per order	Sales per customer	Delivery Status	Late_delivery_risk
0	DEBIT	3	4	91.250000	314.640015	Advance shipping	0
1	TRANSFER	5	4	-249.089996	311.359985	Late delivery	1
2	CASH	4	4	-247.779999	309.720001	Shipping on time	0
3	DEBIT	3	4	22.860001	304.809998	Advance shipping	0
4	PAYMENT	2	4	134.210007	298.250000	Advance shipping	0

5 rows × 53 columns

In [ ]:
```
#view columns in data set
df.columns
```

Out[ ]:
```
Index(['Type', 'Days for shipping (real)', 'Days for shipment (scheduled)',
       'Benefit per order', 'Sales per customer', 'Delivery Status',
       'Late_delivery_risk', 'Category Id', 'Category Name', 'Customer City',
       'Customer Country', 'Customer Email', 'Customer Fname', 'Customer Id',
       'Customer Lname', 'Customer Password', 'Customer Segment',
       'Customer State', 'Customer Street', 'Customer Zipcode',
       'Department Id', 'Department Name', 'Latitude', 'Longitude', 'Market',
       'Order City', 'Order Country', 'Order Customer Id',
       'order date (DateOrders)', 'Order Id', 'Order Item Cardprod Id',
       'Order Item Discount', 'Order Item Discount Rate', 'Order Item Id',
       'Order Item Product Price', 'Order Item Profit Ratio',
       'Order Item Quantity', 'Sales', 'Order Item Total',
       'Order Profit Per Order', 'Order Region', 'Order State', 'Order Status',
       'Order Zipcode', 'Product Card Id', 'Product Category Id',
       'Product Description', 'Product Image', 'Product Name', 'Product Price',
       'Product Status', 'shipping date (DateOrders)', 'Shipping Mode'],
      dtype='object')
```

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

Out[ ]: (180519, 53)

In [ ]:
```
#check data types
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180519 entries, 0 to 180518
Data columns (total 53 columns):
 #   Column                         Non-Null Count   Dtype  
---  ------                         --------------   -----  
 0   Type                           180519 non-null  object 
 1   Days for shipping (real)       180519 non-null  int64  
 2   Days for shipment (scheduled)  180519 non-null  int64  
 3   Benefit per order              180519 non-null  float64
 4   Sales per customer             180519 non-null  float64
 5   Delivery Status                180519 non-null  object 
 6   Late_delivery_risk             180519 non-null  int64  
 7   Category Id                    180519 non-null  int64  
 8   Category Name                  180519 non-null  object 
 9   Customer City                  180519 non-null  object 
 10  Customer Country               180519 non-null  object 
 11  Customer Email                 180519 non-null  object 
 12  Customer Fname                 180519 non-null  object 
 13  Customer Id                    180519 non-null  int64  
 14  Customer Lname                 180511 non-null  object 
 15  Customer Password              180519 non-null  object 
 16  Customer Segment               180519 non-null  object 
 17  Customer State                 180519 non-null  object 
 18  Customer Street                180519 non-null  object 
 19  Customer Zipcode               180516 non-null  float64
 20  Department Id                  180519 non-null  int64  
 21  Department Name                180519 non-null  object 
 22  Latitude                       180519 non-null  float64
 23  Longitude                      180519 non-null  float64
 24  Market                         180519 non-null  object 
 25  Order City                     180519 non-null  object 
 26  Order Country                  180519 non-null  object 
 27  Order Customer Id              180519 non-null  int64  
 28  order date (DateOrders)        180519 non-null  object 
 29  Order Id                       180519 non-null  int64  
 30  Order Item Cardprod Id         180519 non-null  int64  
 31  Order Item Discount            180519 non-null  float64
 32  Order Item Discount Rate       180519 non-null  float64
 33  Order Item Id                  180519 non-null  int64  
 34  Order Item Product Price       180519 non-null  float64
 35  Order Item Profit Ratio        180519 non-null  float64
 36  Order Item Quantity            180519 non-null  int64  
 37  Sales                          180519 non-null  float64
 38  Order Item Total               180519 non-null  float64
 39  Order Profit Per Order         180519 non-null  float64
 40  Order Region                   180519 non-null  object 
 41  Order State                    180519 non-null  object 
 42  Order Status                   180519 non-null  object 
 43  Order Zipcode                  24840 non-null   float64
 44  Product Card Id                180519 non-null  int64  
 45  Product Category Id            180519 non-null  int64  
 46  Product Description            0 non-null       float64
 47  Product Image                  180519 non-null  object 
 48  Product Name                   180519 non-null  object 
 49  Product Price                  180519 non-null  float64
 50  Product Status                 180519 non-null  int64  
```

```
 51  shipping date (DateOrders)     180519 non-null  object
 52  Shipping Mode                  180519 non-null  object
dtypes: float64(15), int64(14), object(24)
memory usage: 73.0+ MB
```

In [ ]:
```
#chck null values
df.isnull().sum()
```

Out[ ]:

```
Type                              0
Days for shipping (real)          0
Days for shipment (scheduled)     0
Benefit per order                 0
Sales per customer                0
Delivery Status                   0
Late_delivery_risk                0
Category Id                       0
Category Name                     0
Customer City                     0
Customer Country                  0
Customer Email                    0
Customer Fname                    0
Customer Id                       0
Customer Lname                    8
Customer Password                 0
Customer Segment                  0
Customer State                    0
Customer Street                   0
Customer Zipcode                  3
Department Id                     0
Department Name                   0
Latitude                          0
Longitude                         0
Market                            0
Order City                        0
Order Country                     0
Order Customer Id                 0
order date (DateOrders)           0
Order Id                          0
Order Item Cardprod Id            0
Order Item Discount               0
Order Item Discount Rate          0
Order Item Id                     0
Order Item Product Price          0
Order Item Profit Ratio           0
Order Item Quantity               0
Sales                             0
Order Item Total                  0
Order Profit Per Order            0
Order Region                      0
Order State                       0
Order Status                      0
Order Zipcode                155679
Product Card Id                   0
Product Category Id               0
Product Description          180519
Product Image                     0
Product Name                      0
Product Price                     0
Product Status                    0
shipping date (DateOrders)        0
Shipping Mode                     0
dtype: int64
```

In [ ]:
```
df.describe()
```

Out[ ]:

	Days for shipping (real)	Days for shipment (scheduled)	Benefit per order	Sales per customer	Late_delivery_risk	C
count	180519.000000	180519.000000	180519.000000	180519.000000	180519.000000	180
mean	3.497654	2.931847	21.974989	183.107609	0.548291	
std	1.623722	1.374449	104.433526	120.043670	0.497664	
min	0.000000	0.000000	-4274.979980	7.490000	0.000000	
25%	2.000000	2.000000	7.000000	104.379997	0.000000	
50%	3.000000	4.000000	31.520000	163.990005	1.000000	
75%	5.000000	4.000000	64.800003	247.399994	1.000000	
max	6.000000	4.000000	911.799988	1939.989990	1.000000	

8 rows × 29 columns

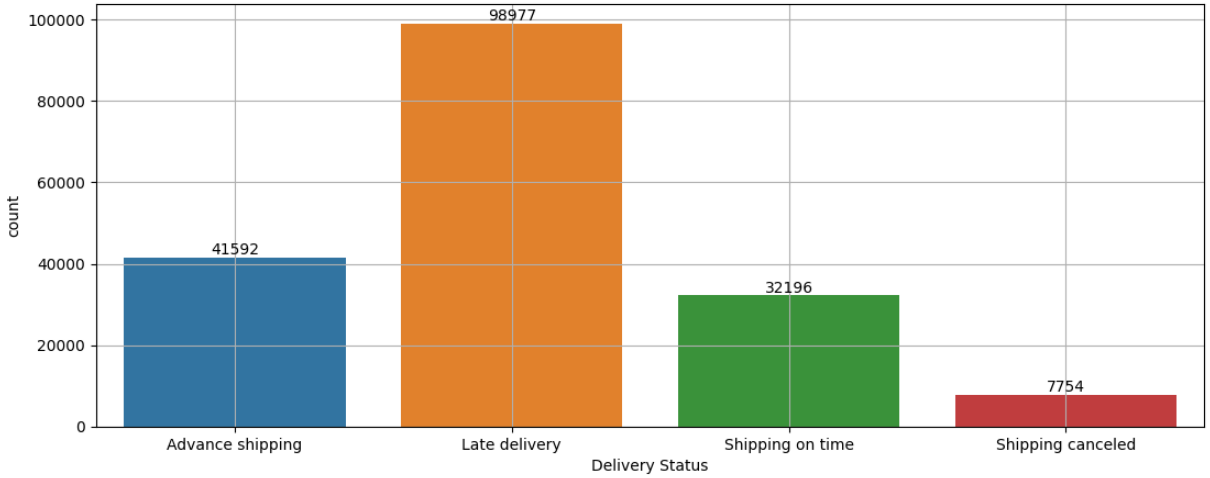
In [ ]:
```
#remove unwanted columns
df.drop(columns=["Days for shipping (real)","Days for shipment (scheduled)","Custom
                 "Customer Password","Customer Street","Customer Zipcode","Department
                 "Order Zipcode","Product Description","Product Image",
                 "Customer City","Customer Fname","Customer Lname"], inplace=True)
```

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

Out[ ]:
```
(180519, 40)
```

Data Visualization

In [ ]:
```
#delivery status
plt.figure(figsize=(13,5))
a=sns.countplot(data=df, x=df["Delivery Status"])
plt.grid()
for i in a.containers:
    a.bar_label(i)
```

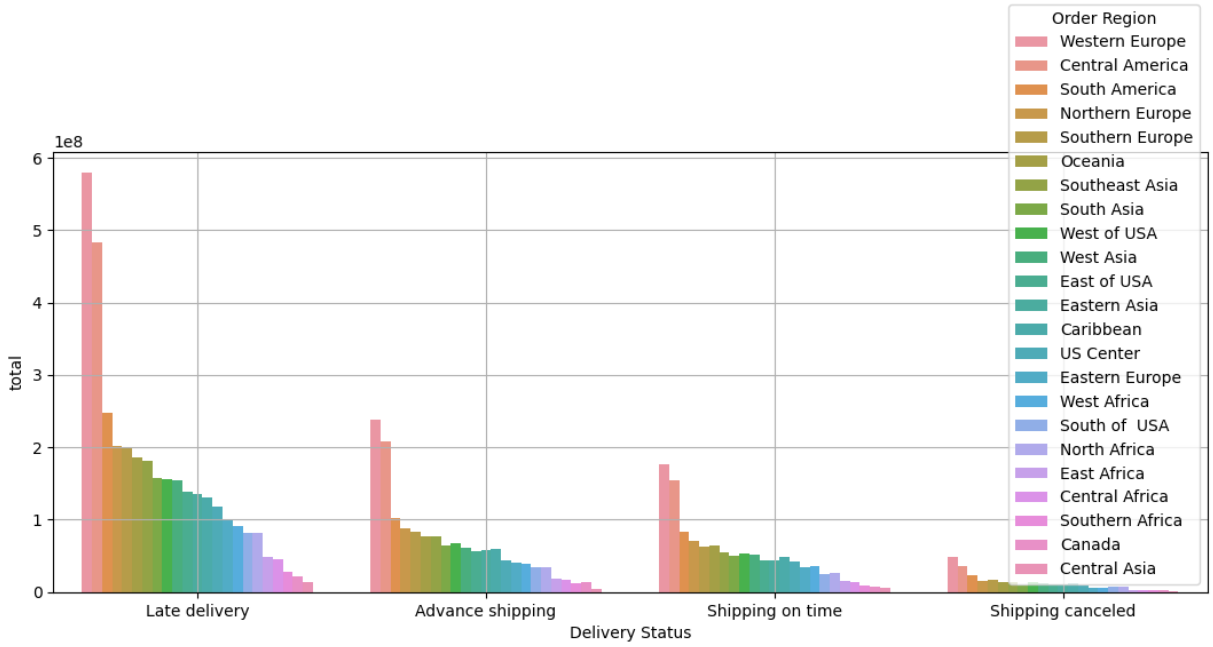


In [ ]:
```
delivery=df.groupby(['Delivery Status', 'Order Region']).agg(
    total=("Order Id", "sum")
).reset_index().sort_values(by="total", ascending=False)
delivery.head()
```

Out[ ]:

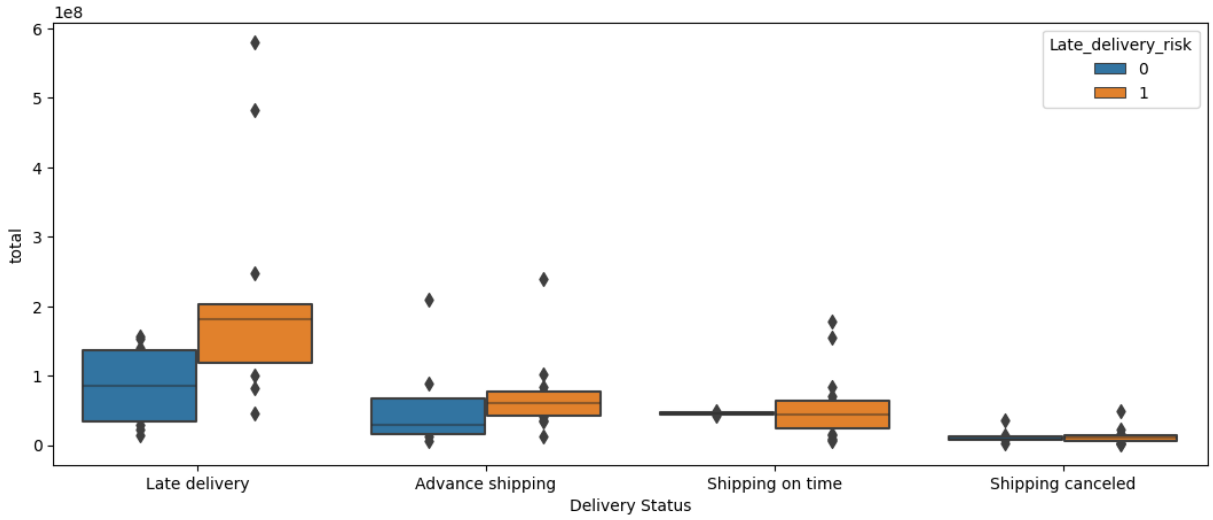
	Delivery Status	Order Region	total
45	Late delivery	Western Europe	579587241
26	Late delivery	Central America	483106830
35	Late delivery	South America	248168906
22	Advance shipping	Western Europe	238874292
3	Advance shipping	Central America	208983770

In [ ]:
```
#order region
plt.figure(figsize=(13,5))
sns.barplot(delivery, x="Delivery Status",y="total", hue="Order Region")
plt.grid()
```



In [ ]:
```
plt.figure(figsize=(13,5))
sns.boxenplot(data=delivery,x="Delivery Status", y="total", hue=df["Late_delivery_r
```

Out[ ]:
```
<Axes: xlabel='Delivery Status', ylabel='total'>
```



Top 20 Customers regarding the quanitity of orders

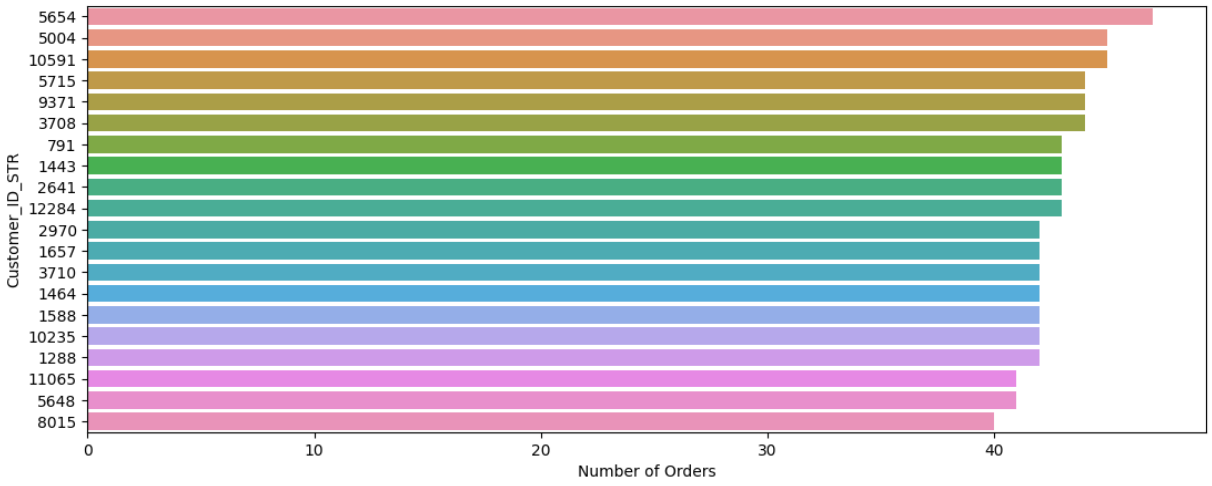
In [ ]:
```
df['Customer_ID_STR']=df['Customer Id'].astype(str)

data_customers=df.groupby(['Customer_ID_STR'])['Order Id'].count().reset_index(name
data_customers.head()
#use-- name = "number of orders" for name changing
```

Out[ ]:

	Customer_ID_STR	Number of Orders
15859	5654	47
15143	5004	45
651	10591	45
15927	5715	44
19958	9371	44

In [ ]:
```
plt.figure(figsize=(13,5))
sns.barplot(data=data_customers.head(20), x="Number of Orders", y="Customer_ID_STR"
```

Out[ ]: <Axes: xlabel='Number of Orders', ylabel='Customer_ID_STR'>



Top 20 Customers regarding profit of all orders

In [ ]:
```
profit=df.groupby(['Customer_ID_STR'])['Order Profit Per Order'].sum().reset_index(
profit
```

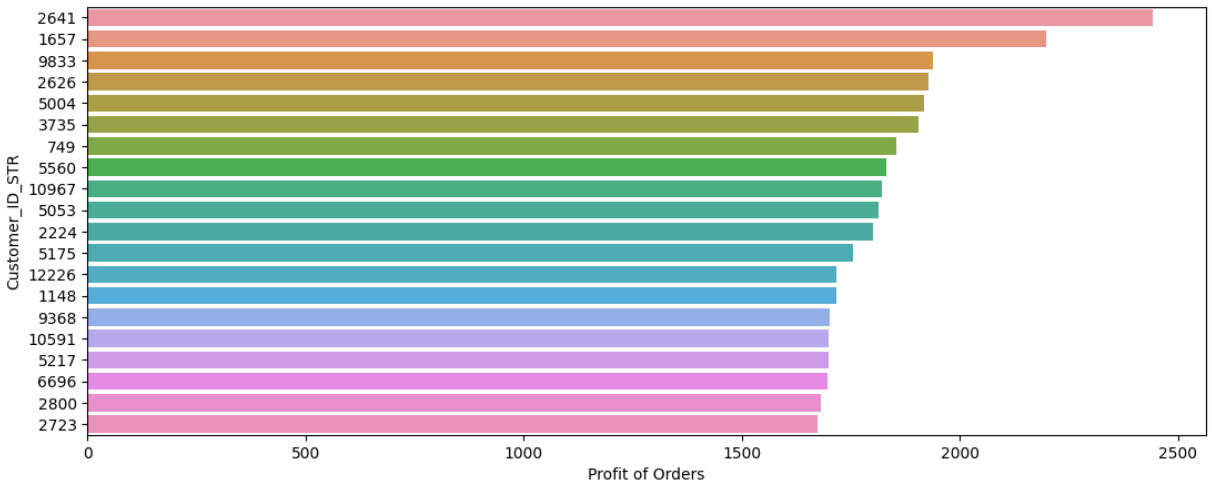
Out[ ]:

	Customer_ID_STR	Profit of Orders
12538	2641	2441.970003
7266	1657	2196.919992
20469	9833	1938.390015
12521	2626	1928.570015
15143	5004	1917.990002
...	...	...
8921	18061	-2592.000000
4763	14313	-3000.000000
8973	18109	-3366.000000
4511	14086	-3442.500000
4725	1428	-3868.559982

20652 rows × 2 columns

In [ ]:
```
plt.figure(figsize=(13,5))
sns.barplot(data=profit.head(20), x="Profit of Orders",y ="Customer_ID_STR")
```

Out[ ]:
```
<Axes: xlabel='Profit of Orders', ylabel='Customer_ID_STR'>
```



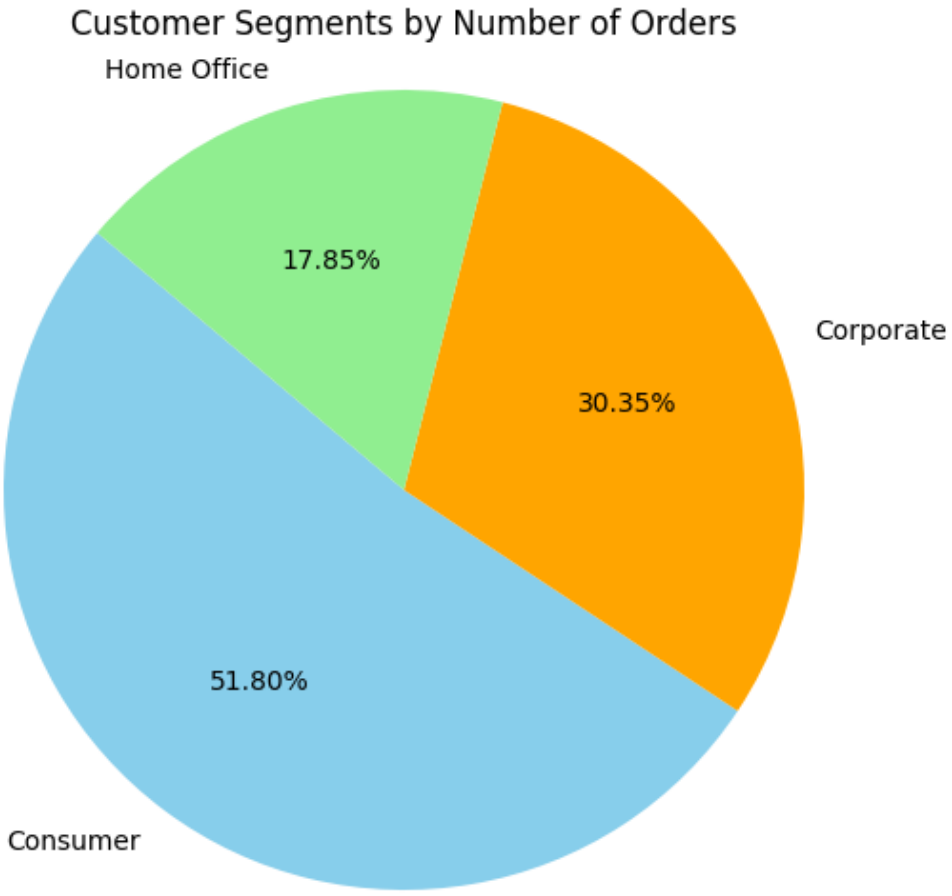
Customer Segment

In [ ]:
```
segment=df.groupby(['Customer Segment'])['Order Id'].count().reset_index(name='Numb
segment
```

Out[ ]:

	Customer Segment	Number of Orders
0	Consumer	93504
1	Corporate	54789
2	Home Office	32226

In [ ]:
```
plt.figure(figsize=(8, 6))
plt.pie(segment["Number of Orders"],labels=segment["Customer Segment"],autopct="%1.
plt.axis('equal')
plt.title('Customer Segments by Number of Orders')
plt.show()
```



Product category

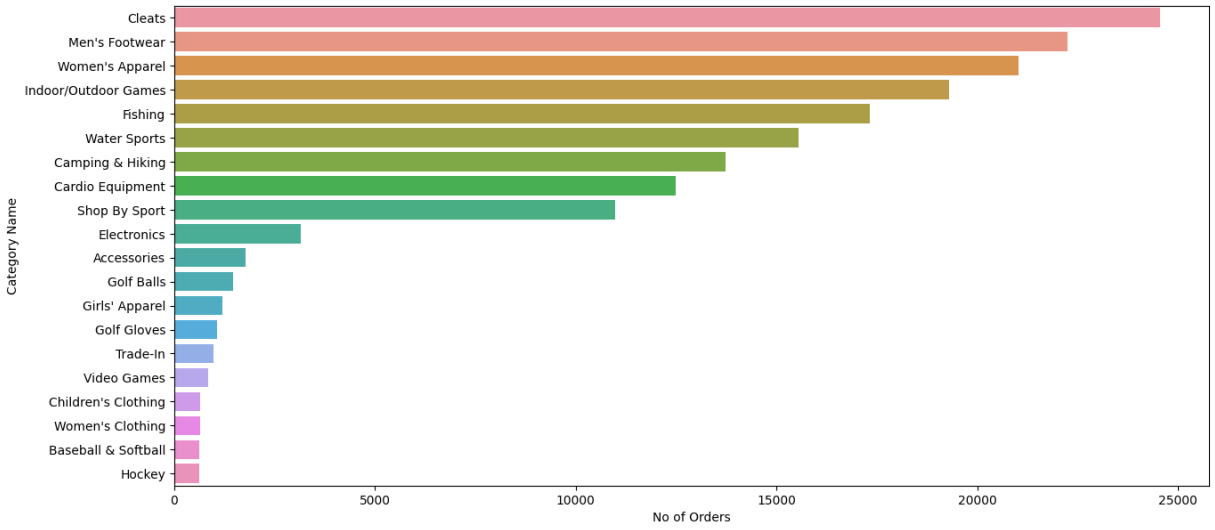
In [ ]:
```
category=df.groupby("Category Name")["Order Id"].count().reset_index(name="No of Or
category.head(10)
```

Out[ ]:

	Category Name	No of Orders
12	Cleats	24551
34	Men's Footwear	22246
47	Women's Apparel	21035
30	Indoor/Outdoor Games	19298
18	Fishing	17325
46	Water Sports	15540
9	Camping & Hiking	13729
10	Cardio Equipment	12487
38	Shop By Sport	10984
17	Electronics	3156

In [ ]:
```
plt.figure(figsize=(15,7))
sns.barplot(data=category.head(20), x='No of Orders',y ='Category Name')
```

Out[ ]: <Axes: xlabel='No of Orders', ylabel='Category Name'>



Sales Analysis

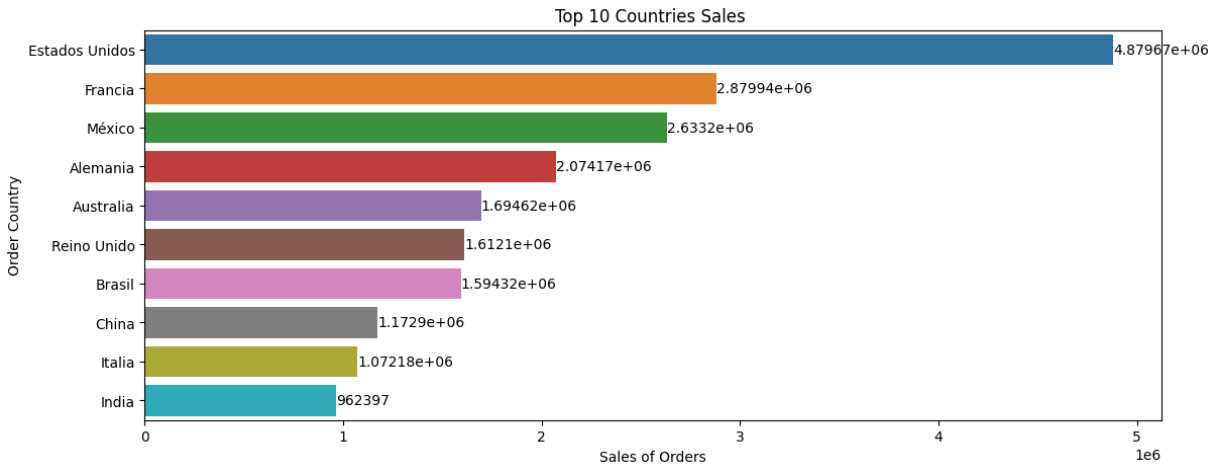
In [ ]:
```
sales_country=df.groupby([ 'Order Country'])['Sales'].sum().round().reset_index(nam
sales_country
```

Out[ ]:

	Order Country	Sales of Orders
48	Estados Unidos	4879668.0
53	Francia	2879942.0
102	México	2633195.0
2	Alemania	2074172.0
8	Australia	1694622.0
...	...	...
82	Kuwait	335.0
62	Guinea Ecuatorial	320.0
44	Eritrea	310.0
23	Burundi	300.0
131	Serbia	120.0

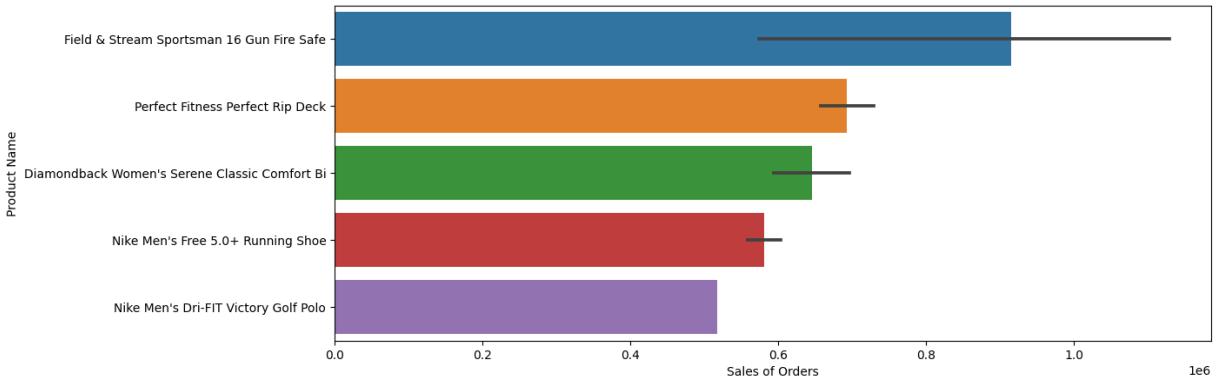
164 rows × 2 columns

In [ ]:
```
plt.figure(figsize=(13,5))
b=sns.barplot(sales_country.head(10), x='Sales of Orders',y = 'Order Country')
plt.title("Top 10 Countries Sales ")
for i in b.containers:
    b.bar_label(i)
```



In [ ]:
```
#Product and order region
plt.figure(figsize=(13,5))
sales_pr=df.groupby([ 'Product Name', 'Order Region'])['Sales'].sum().reset_index(n
sns.barplot(data=sales_pr.head(10), x='Sales of Orders',y = 'Product Name')
```

Out[ ]:
```
<Axes: xlabel='Sales of Orders', ylabel='Product Name'>
```

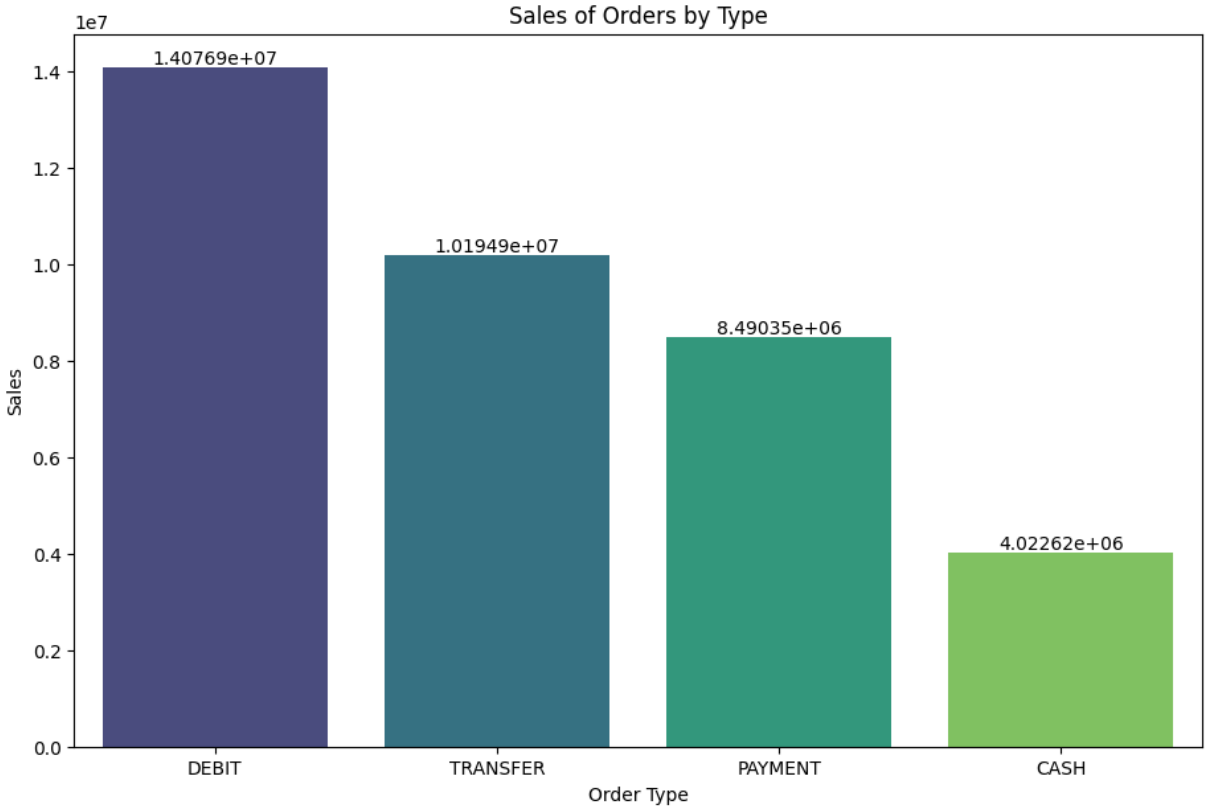


In [ ]:
```
#'Type of payment
sales_pr=df.groupby([ 'Type'])['Sales'].sum().round().reset_index(name='Sales of Or
sales_pr
```

Out[ ]:

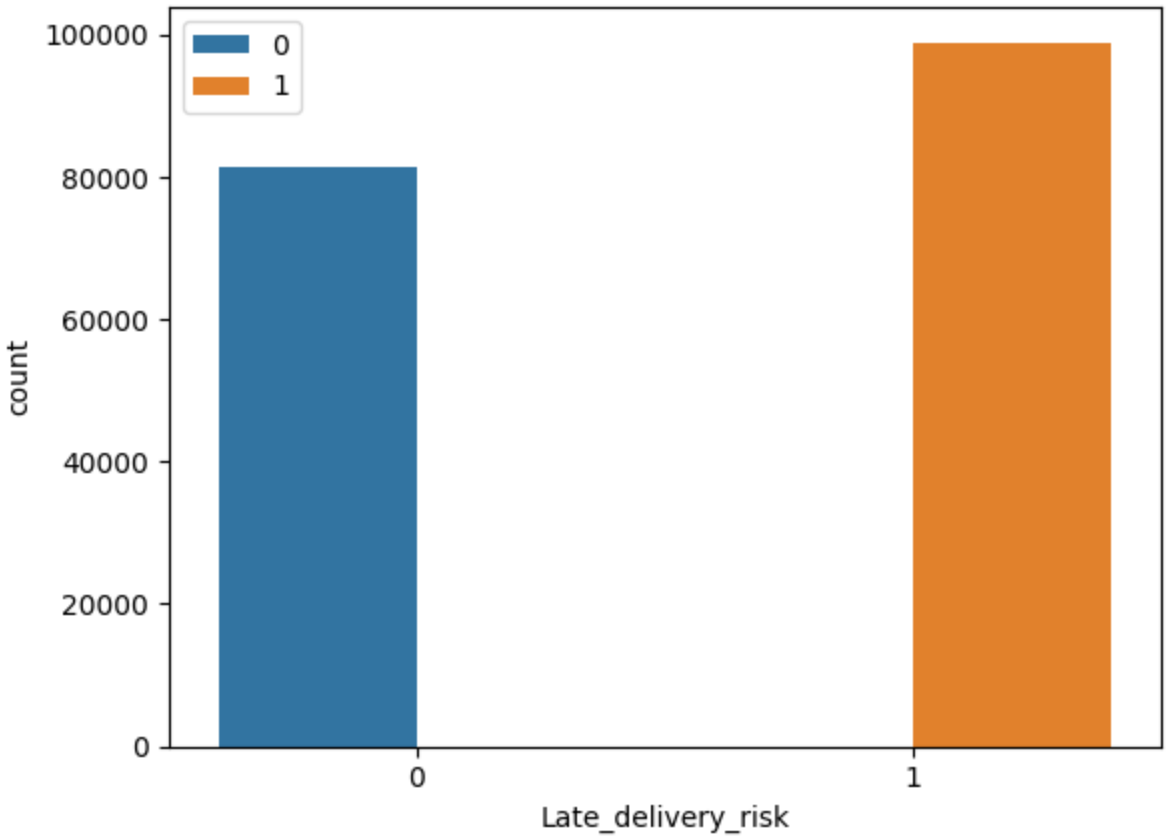
	Type	Sales of Orders
1	DEBIT	14076858.0
3	TRANSFER	10194902.0
2	PAYMENT	8490351.0
0	CASH	4022624.0

In [ ]:
```
plt.figure(figsize=(11, 7))
c=sns.barplot(sales_pr.head(10), y='Sales of Orders',x = 'Type',palette='viridis')
plt.title('Sales of Orders by Type')
plt.xlabel('Order Type')
plt.ylabel('Sales')
for i in c.containers:
    c.bar_label(i)
```



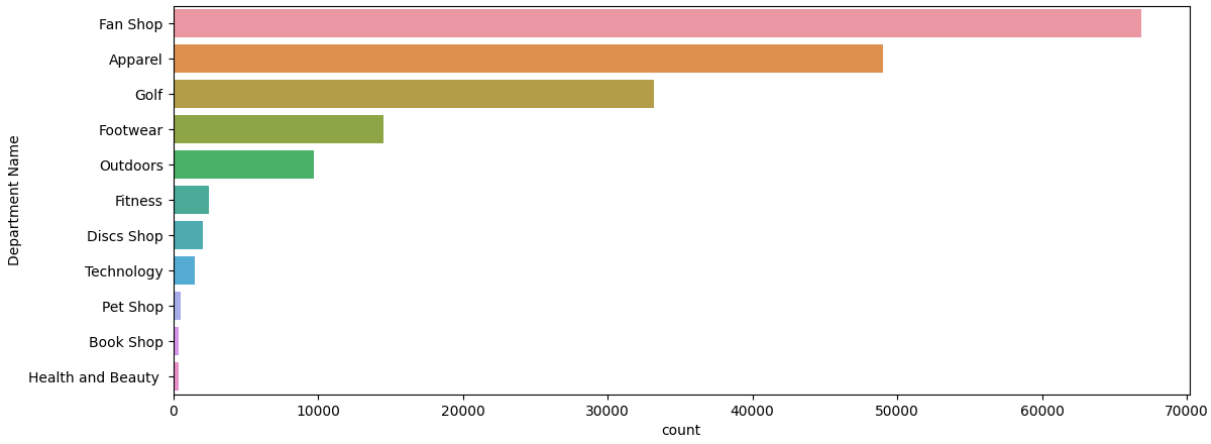
In [ ]:
```
# Delivery Status
sns.countplot(data=df,x="Late_delivery_risk", hue="Late_delivery_risk")
plt.legend()
```

Out[ ]:
```
<matplotlib.legend.Legend at 0x1e35b0ce170>
```



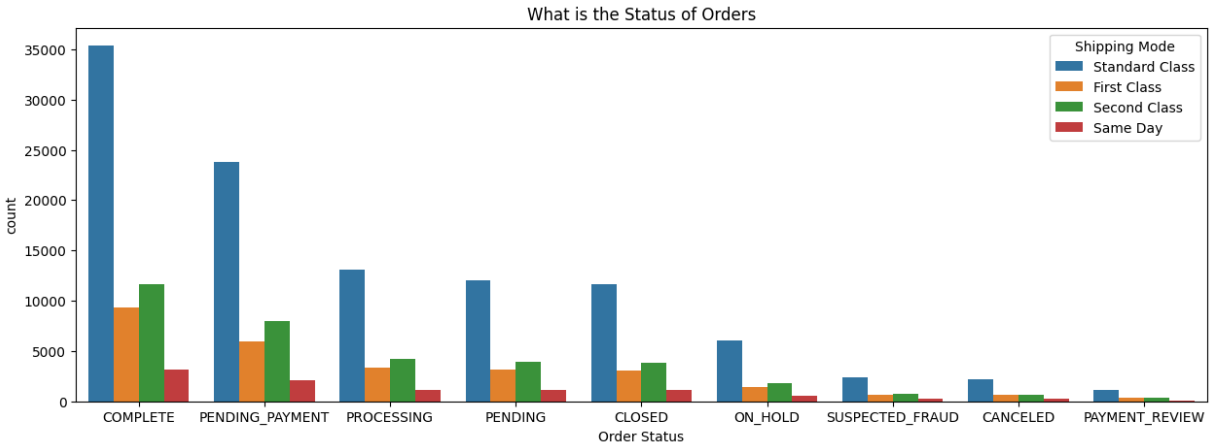
In [ ]:
```
#Top 10 Deparment
plt.figure(figsize=(13,5))
sns.countplot(data=df, y=df["Department Name"],
              order=df["Department Name"].value_counts().index)
```

Out[ ]:
```
<Axes: xlabel='count', ylabel='Department Name'>
```



In [ ]:
```
#Order Status
plt.figure(figsize=(15,5))
sns.countplot(data=df, x=df["Order Status"],
              order=df["Order Status"].value_counts().index, hue="Shipping Mode")
plt.title("What is the Status of Orders")
```

Out[ ]:
```
Text(0.5, 1.0, 'What is the Status of Orders')
```


What is the Status of Orders
Shipping Mode
Standard Class
First Class
Second Class
Same Day
count
35000
30000
25000
20000
15000
10000
5000
0
COMPLETE
PENDING_PAYMENT
PROCESSING
PENDING
CLOSED
ON_HOLD
SUSPECTED_FRAUD
CANCELED
PAYMENT_REVIEW
Order Status