```python
import pandas as pd
import seaborn as sns
import plotly.express as px
from datetime import datetime as dt
```

```python
df = pd.read_csv('https://raw.githubusercontent.com/plotly/datasets/master/global_super_store_orders.tsv',sep = '\t')
```

▾ Column list

Order ID: Each order receives its own Order ID

Ship Date:The definition of ship date is the date that the order is shipped from the seller or warehouse to the customer.

Ship Mode : The shipping mode is a way of shipping goods. More specifically, a shipping mode is the combination of a shipping carrier and the shipping service that is offered by that carrier.That column have a three types of mode **(Standard Class,Second Class,Same Day, First Class)**

Customer ID: Customer ID means the unique means of identification allocated to or selected by the Customer in relation to any one or more of the Services whether in the form of a password, PIN or other form of personal identification, or any combination of any of them.

Customer Name: Name of the Perion who purchase the product

Segment :In marketing, market segmentation is the process of dividing a broad consumer into sub-groups of consumers based on some type of shared characteristics

City:The definition of a city is a town of significant size or an urban area with self-government

State:State is defined as a territory with its own government and borders within a larger country

Country:A country is defined as a nation, the people of the nation or land in a rural area

Postal Code:A postal code is a series of letters or digits or both, sometimes including spaces or punctuation, included in a postal address for the purpose of sorting mail

Market:A marker identifies a location on a map

Region : A region is a large area of land that is different from other areas of land

Product ID : Product identifiers are a series of numerical or alphanumerical digits that are used to identify a specific product

Category :The definition of a category is any sort of division or class

Sub-Category: A category that is a subdivision of a larger category

Product Name: A product name identifies a specific product

Sales : Sales are activities related to selling or the number of goods sold in a given targeted time period

Quantity : an indefinite amount or number

Discount :A discount is the reduction of either the monetary amount or a percentage of the normal selling price of a product or service

Profit : Profit is the money a business pulls in after accounting for all expenses

Shipping Cost: Just weigh the package and use a shipping cost calculator to get a shipping price

Order Priority:Order of Priority means, in respect of any payment to be made by the Issuer before the Enforcement Date

```python
df.head()
```

1 to 5 of 5 entries   Filter

index	Unnamed: 0	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country	Postal Code	Market	Region	Product ID	Category	Sub-Category	Pro
0	35180	29165	IN-2014-24531	8/15/14	8/19/14	Standard Class	AS-10285	Alejandro Savely	Corporate	Guangzhou	Guangdong	China	NaN	APAC	North Asia	OFF-SU-10001343	Office Supplies	Supplies	Elite High
1	39850	39615	CA-2014-130904	4/12/14	4/17/14	Standard Class	HM-14980	Henry MacAllister	Consumer	Burlington	North Carolina	United States	27217.0	US	South	OFF-AR-10000127	Office Supplies	Art	Nev
2	41575	39158	CA-2012-142993	10/12/12	10/17/12	Standard Class	KA-16525	Kelly Andreada	Consumer	Seattle	Washington	United States	98103.0	US	West	TEC-AC-10003038	Technology	Accessories	King Digi Dat 16G
3	12393	2193	MX-2013-169502	6/13/13	6/15/13	Second Class	AJ-10960	Astrea Jones	Consumer	Santo Domingo	Santo Domingo	Dominican Republic	NaN	LATAM	Caribbean	TEC-AC-10001341	Technology	Accessories	Ene Mer USE
4	45530	23217	IN-2014-50627	9/14/14	9/18/14	Standard Class	LW-16825	Laurel Workman	Corporate	Jinhua	Zhejiang	China	NaN	APAC	North Asia	OFF-EN-10001977	Office Supplies	Envelopes	Can Bus Env Rec

Show 25 per page
Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.
```
Warning: Total number of columns (25) exceeds max_columns (20) limiting to first (20) columns.
```

```python
df.tail()
```

1 to 5 of 5 entries   Filter

index	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country	Postal Code	Market	Region	Product ID	Category	Sub-Category	Product Name	Sales
995	36853	CA-2012-125066	2012-12-14 00:00:00	2012-12-14 00:00:00	Standard Class	KD-16495	Keith Dawkins	Corporate	Jackson	Mississippi	United States	39212.0	US	South	FUR-FU-10003829	Furniture	Furnishings	Stackable Trays	616

DATA PREPROSSING

Removed unwanted columns

```
df.drop(['Unnamed: 0'],axis=1,inplace=True)
```

Data type conversion

```
df['Sales'] = df['Sales'].str.replace(',','')
df.astype({'Sales':int})
```

```
df['Profit'] = df['Profit'].str.replace(',','')
df.astype({'Profit':int})
```

```
df.astype({'Row ID':str})
```

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

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

```
df['Discount'] = df['Discount'].str.replace(',','')
df.astype({'Discount':int})
```

```
df['Shipping Cost'] = df['Shipping Cost'].str.replace(',','')
df.astype({'Shipping Cost':int})
```

```
new_df=df.astype({'Postal Code':str})
```

Describe of Data Frame

```
new_df.describe()
```

	Row ID	Quantity
count	1000.000000	1000.000000
mean	25571.221000	3.388000
std	14570.032549	2.168908
min	19.000000	1.000000
25%	13678.250000	2.000000
50%	25229.000000	3.000000
75%	38116.750000	4.000000
max	51285.000000	14.000000

```
ship_mode = df.groupby('Ship Mode', as_index=False, sort=False)['Ship Mode'].agg({'Ship Mode':'count'})
ship_mode['Ship_Mode_Name']=new_df['Ship Mode'].unique()
px.pie(ship_mode,values='Ship Mode',names='Ship_Mode_Name')
```

Information of Data frame

```
new_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 24 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Row ID          1000 non-null   int64
 1   Order ID        1000 non-null   object
 2   Order Date      1000 non-null   datetime64[ns]
 3   Ship Date       1000 non-null   datetime64[ns]
 4   Ship Mode       1000 non-null   object
 5   Customer ID     1000 non-null   object
 6   Customer Name   1000 non-null   object
 7   Segment         1000 non-null   object
 8   City            1000 non-null   object
 9   State           1000 non-null   object
 10  Country         1000 non-null   object
 11  Postal Code     1000 non-null   object
 12  Market          1000 non-null   object
 13  Region          1000 non-null   object
 14  Product ID      1000 non-null   object
 15  Category        1000 non-null   object
 16  Sub-Category    1000 non-null   object
 17  Product Name    1000 non-null   object
 18  Sales           1000 non-null   object
 19  Quantity        1000 non-null   int64
 20  Discount        1000 non-null   object
 21  Profit          1000 non-null   object
 22  Shipping Cost   1000 non-null   object
 23  Order Priority  1000 non-null   object
dtypes: datetime64[ns](2), int64(2), object(20)
memory usage: 187.6+ KB
```

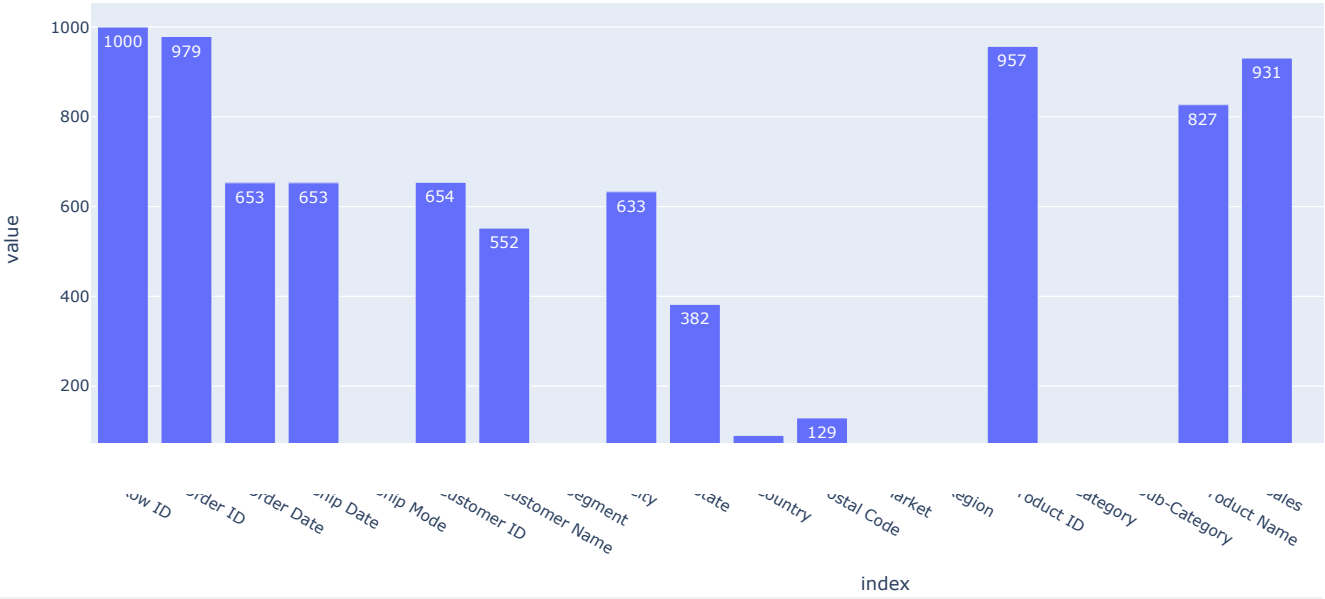### Define the Roll of Data Frame

```
new_df.shape[0]
```

```
1000
```

Define the Unique value of Data Frame

```
unique=pd.DataFrame(new_df.nunique())
unique
```

	0
Row ID	1000
Order ID	979
Order Date	653
Ship Date	653
Ship Mode	4
Customer ID	654
Customer Name	552
Segment	3
City	633
State	382
Country	90
Postal Code	129
Market	7
Region	13
Product ID	957
Category	3
Sub-Category	17
Product Name	827
Sales	931
Quantity	14
Discount	21
Profit	903
Shipping Cost	848
Order Priority	4

```
px.bar(unique,text_auto=True)
```

✓ 0s   completed at 5:58 PM   ● ✕