

# Ex03-Univariate-Analysis

## › Aim

To read the given data and perform the univariate analysis with different types of plots.

## › Explanation

Univariate analysis is basically the simplest form to analyze data. Uni means one and this means that the data has only one kind of variable. The major reason for univariate analysis is to use the data to describe. The analysis will take data, summarise it, and then find some pattern in the data.

## › Algorithm

### › Step1

Read the given data.

### › Step2

Get the information about the data.

### › Step3

Remove the null values from the data.

### › Step4

Mention the datatypes from the data.

### › Step5

Count the values from the data.

### › Step6

Do plots like boxplots,countplot,distribution plot,histogram plot.

## › Program

Developed by : KERSON P  
Registration Number : 212221230050

```
import pandas as pd
import numpy as np
import seaborn as sns

df=pd.read_csv('superstore.csv')
df
```

```
df.head()
df.info()
df.describe()
df.isnull().sum()

df.dtypes

df['Postal Code'].value_counts()

sns.boxplot(x='Postal Code', data=df)
sns.countplot(x='Postal Code',data=df)
sns.distplot(df["Postal Code"])
sns.histplot(x='Postal Code',data=df)
```

Output

DATA

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category
0	1	CA-2017-152156	06-11-2017	11-11-2017	Second Class	CG-12520	Claire Gulte	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-BO-10001798	Furniture	Bookcases
1	2	CA-2017-152156	06-11-2017	11-11-2017	Second Class	CG-12520	Claire Gulte	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-CH-10000454	Furniture	Chairs
2	3	CA-2017-138688	12-06-2017	16-06-2017	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036.0	West	OFF-LA-10000240	Office Supplies	Labels
3	4	US-2016-108966	11-10-2016	18-10-2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South	FUR-TA-10000577	Furniture	Tables
4	5	US-2016-108966	11-10-2016	18-10-2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South	OFF-ST-10000760	Office Supplies	Storage
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5755	9796	CA-2017-125920	21-05-2017	26-05-2017	Standard Class	SH-19975	Sally Hughsby	Corporate	United States	Chicago	Illinois	60610.0	Central	OFF-BI-10003429	Office Supplies	Binders
5756	9797	CA-2016-128608	12-01-2016	17-01-2016	Standard Class	CS-12490	Cindy Schnelling	Corporate	United States	Toledo	Ohio	43615.0	East	OFF-AR-10001374	Office Supplies	Art
5757	9798	CA-2016-128608	12-01-2016	17-01-2016	Standard Class	CS-12490	Cindy Schnelling	Corporate	United States	Toledo	Ohio	43615.0	East	TEC-PH-10004977	Technology	Phones
5758	9799	CA-2016-128608	12-01-2016	17-01-2016	Standard Class	CS-12490	Cindy Schnelling	Corporate	United States	Toledo	Ohio	43615.0	East	TEC-PH-10000912	Technology	Phones
5759	9800	CA-2016-128608	12-01-2016	17-01-2016	Standard Class	CS-12490	Cindy Schnelling	Corporate	United States	Toledo	Ohio	43615.0	East	TEC-AC-10000487	Technology	Accessories
800 rows x 18 columns																

DATA HEAD

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category
0	1	CA-2017-152156	08-11-2017	11-11-2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-BO-10001798	Furniture	Bookcases
1	2	CA-2017-152156	08-11-2017	11-11-2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-CH-10000454	Furniture	Chairs
2	3	CA-2017-138688	12-06-2017	16-06-2017	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036.0	West	OFF-LA-10000240	Office Supplies	Labels
3	4	US-2016-108966	11-10-2016	18-10-2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South	FUR-TA-10000577	Furniture	Tables
4	5	US-2016-108966	11-10-2016	18-10-2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South	OFF-ST-10000760	Office Supplies	Storage

DATA INFORMATION

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9800 entries, 0 to 9799
Data columns (total 18 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Row ID              9800 non-null   int64
1   Order ID            9800 non-null   object
2   Order Date          9800 non-null   object
3   Ship Date           9800 non-null   object
4   Ship Mode           9800 non-null   object
5   Customer ID         9800 non-null   object
6   Customer Name       9800 non-null   object
7   Segment             9800 non-null   object
8   Country             9800 non-null   object
9   City                9800 non-null   object
10  State               9800 non-null   object
11  Postal Code         9789 non-null   float64
12  Region             9800 non-null   object
13  Product ID          9800 non-null   object
14  Category            9800 non-null   object
15  Sub-Category        9800 non-null   object
16  Product Name        9800 non-null   object
17  Sales               9800 non-null   float64
dtypes: float64(2), int64(1), object(15)
memory usage: 1.3+ MB
```

DATA DESCRIBE

	Row ID	Postal Code	Sales
count	9800.000000	9789.000000	9800.000000
mean	4900.500000	55273.322403	230.769059
std	2829.160653	32041.223413	626.651875
min	1.000000	1040.000000	0.444000
25%	2450.750000	23223.000000	17.248000
50%	4900.500000	58103.000000	54.490000
75%	7350.250000	90008.000000	210.605000
max	9800.000000	99301.000000	22638.480000

DATA NULL VALUES

```
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 11
Region      0
Product ID  0
Category    0
Sub-Category 0
Product Name 0
Sales       0
dtype: int64
```

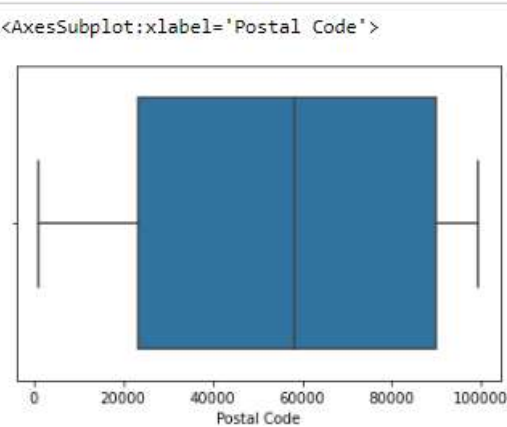
DATA'S DATATYPES

```
Row ID      int64
Order ID    object
Order Date  object
Ship Date   object
Ship Mode   object
Customer ID  object
Customer Name object
Segment     object
Country     object
City        object
State       object
Postal Code  float64
Region      object
Product ID  object
Category    object
Sub-Category object
Product Name object
Sales       float64
dtype: object
```

DATA'S VALUECOUNT

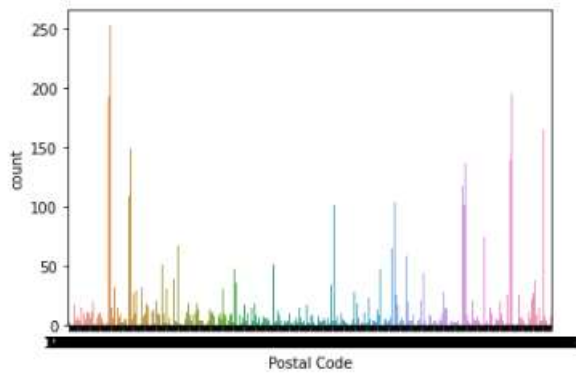
```
10035.0    253
10024.0    225
10009.0    220
94122.0    195
10011.0    193
...
32503.0     1
34741.0     1
61761.0     1
91761.0     1
27514.0     1
Name: Postal Code, Length: 626, dtype: int64
```

BOXPLOT



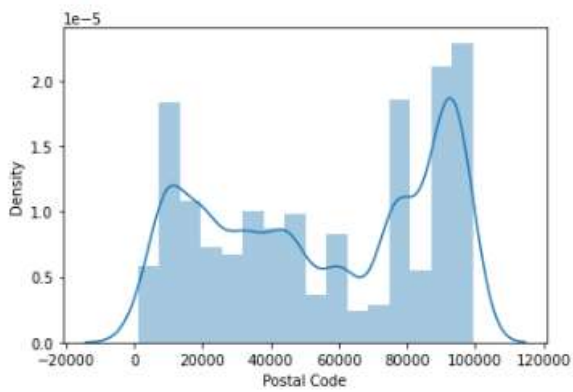
COUNTPLOT

```
<AxesSubplot:xlabel='Postal Code', ylabel='count'>
```



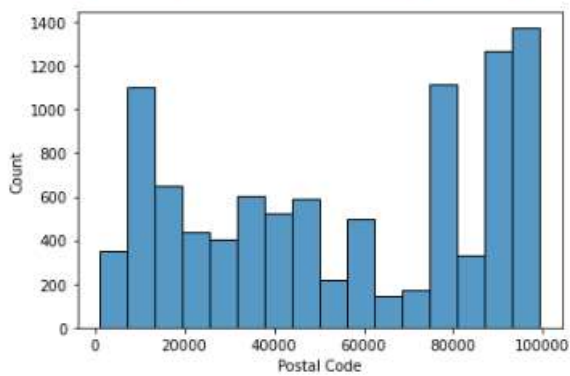
## DISTRIBUTION PLOT

```
<AxesSubplot:xlabel='Postal Code', ylabel='Density'>
```



## HISTOGRAM PLOT

```
<AxesSubplot:xlabel='Postal Code', ylabel='Count'>
```



# Result

Thus we have read the given data and performed the univariate analysis with different types of plots.