In [14]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [16]:
```
df=pd.read_csv('Electric Vehicle Sales by State in India.csv')
```

In [18]:
```
df.head()
```

Out[18]:

| | Year | Month_Name | Date | State | Vehicle_Class | Vehicle_Category | Vehicle_Type | EV_Sales_Quantity |
|---|---|---|---|---|---|---|---|---|
| 0 | 2014.0 | jan | 1/1/2014 | Andhra Pradesh | ADAPTED VEHICLE | Others | Others | 0.0 |
| 1 | 2014.0 | jan | 1/1/2014 | Andhra Pradesh | AGRICULTURAL TRACTOR | Others | Others | 0.0 |
| 2 | 2014.0 | jan | 1/1/2014 | Andhra Pradesh | AMBULANCE | Others | Others | 0.0 |
| 3 | 2014.0 | jan | 1/1/2014 | Andhra Pradesh | ARTICULATED VEHICLE | Others | Others | 0.0 |
| 4 | 2014.0 | jan | 1/1/2014 | Andhra Pradesh | BUS | Bus | Bus | 0.0 |

In [22]:
```
df.describe()
```

Out[22]:

| | Year | EV_Sales_Quantity |
|---|---|---|
| count | 96845.000000 | 96845.000000 |
| mean | 2018.622768 | 37.108896 |
| std | 2.895581 | 431.566675 |
| min | 2014.000000 | 0.000000 |
| 25% | 2016.000000 | 0.000000 |
| 50% | 2019.000000 | 0.000000 |
| 75% | 2021.000000 | 0.000000 |
| max | 2024.000000 | 20584.000000 |

In [26]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96845 entries, 0 to 96844
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Year               96845 non-null  float64
 1   Month_Name         96845 non-null  object 
 2   Date               96845 non-null  object 
 3   State              96845 non-null  object 
 4   Vehicle_Class      96845 non-null  object 
 5   Vehicle_Category   96845 non-null  object 
 6   Vehicle_Type       96845 non-null  object 
 7   EV_Sales_Quantity  96845 non-null  float64
dtypes: float64(2), object(6)
memory usage: 5.9+ MB
```

In [28]:
```
print(df.isnull().sum())
```

```
Year                0
Month_Name          0
Date                0
State               0
Vehicle_Class       0
Vehicle_Category    0
Vehicle_Type        0
EV_Sales_Quantity   0
dtype: int64
```

Change the Data Type of "Date"

In [30]:
```
df['Date']=pd.to_datetime(df['Date'])
```

In [34]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96845 entries, 0 to 96844
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype         
---  ------             --------------  -----         
 0   Year               96845 non-null  float64       
 1   Month_Name         96845 non-null  object        
 2   Date               96845 non-null  datetime64[ns]
 3   State              96845 non-null  object        
 4   Vehicle_Class      96845 non-null  object        
 5   Vehicle_Category   96845 non-null  object        
 6   Vehicle_Type       96845 non-null  object        
 7   EV_Sales_Quantity  96845 non-null  float64       
dtypes: datetime64[ns](1), float64(2), object(5)
memory usage: 5.9+ MB
```

Exploratory Data Analysis

In [40]:
```
df.Year.value_counts()
```

Out[40]:
```
Year
2019.0    10315
2023.0    10279
2018.0    10225
2022.0    10021
2017.0     9799
2016.0     9348
2021.0     9249
2015.0     9052
2014.0     9022
2020.0     8675
2024.0      860
Name: count, dtype: int64
```

In [42]:
```
df.State.value_counts()
```

Out[42]:
```
State
Maharashtra                4912
Karnataka                  4830
Uttar Pradesh              4557
Rajasthan                  4552
Gujarat                    4517
West Bengal                4196
Tamil Nadu                 4063
Odisha                     4027
Haryana                    3842
Kerala                     3666
Chhattisgarh               3590
Madhya Pradesh             3587
Andhra Pradesh             3457
Assam                      3114
Uttarakhand                3045
Himachal Pradesh           2980
Punjab                     2950
Jharkhand                  2773
Bihar                      2544
Jammu and Kashmir          2292
Arunachal Pradesh          2285
Goa                        2139
DNH and DD                 1927
Delhi                      1871
Meghalaya                  1867
Puducherry                 1832
Manipur                    1632
Nagaland                   1588
Tripura                    1564
Mizoram                    1557
Chandigarh                 1554
Sikkim                     1246
Andaman & Nicobar Island   1226
Ladakh                     1063
Name: count, dtype: int64
```

In [44]:
```
df.Vehicle_Class.value_counts()
```

Out[44]:
```
Vehicle_Class
MOTOR CAR                          4111
M-CYCLE/SCOOTER                    4101
GOODS CARRIER                      4096
MOTOR CAB                          3985
BUS                                3813
                                   ... 
SEMI-TRAILER (COMMERCIAL)            18
X-RAY VAN                            12
MOTOR CYCLE/SCOOTER-WITH TRAILER      9
MODULAR HYDRAULIC TRAILER             3
MOTOR CARAVAN                         3
Name: count, Length: 73, dtype: int64
```

In [46]:
```
df.Vehicle_Category.value_counts()
```

Out[46]:
```
Vehicle_Category
Others        54423
2-Wheelers    13121
3-Wheelers    11491
Bus            9119
4-Wheelers     8691
Name: count, dtype: int64
```

In [48]:
```
df.Vehicle_Type.value_counts()
```

Out[48]:
```
Vehicle_Type
Others                54423
2W_Personal           11700
Bus                    7026
4W_Shared              4580
4W_Personal            4111
3W_Shared              3786
3W_Goods               3208
Institution Bus        2093
3W_Shared_LowSpeed     1951
```

```
3W_Goods_LowSpeed      1517
2W_Shared              1421
3W_Personal            1029
Name: count, dtype: int64
```