In [2]:

```python
import pandas as pd
import numpy as np
from sklearn import preprocessing
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="white")#white background for seaborn plots
sns.set(style="whitegrid",color_codes=True)
import warnings
warnings.simplefilter(action='ignore')
```

In [3]:

```
df=pd.read_csv(r"C:\Users\DELL E5490\Downloads\used_cars_data.csv")
df
```

Out[3]:

	S.No.	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Ow
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	
...	...	...	...	...	...	...	...	
7248	7248	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual	
7249	7249	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic	
7250	7250	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	Manual	
7251	7251	Volkswagen Polo GT TSI	Pune	2013	52262	Petrol	Automatic	
7252	7252	Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...	Kochi	2014	72443	Diesel	Automatic	

7253 rows × 14 columns

In [4]:

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

Out[4]:

	S.No.	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_T
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Sec

In [5]:

```python
df.shape
```

Out[5]:

```
(7253, 14)
```

In [6]:

```
df.describe
```

Out[6]:

```
<bound method NDFrame.describe of          S.No.
Name      Location
0          0                          Maruti Wagon R LXI CNG        Mumba
i  \
1          1                     Hyundai Creta 1.6 CRDi SX Option         Pun
e
2          2                                   Honda Jazz V         Chenna
i
3          3                                Maruti Ertiga VDI         Chenna
i
4          4              Audi A4 New 2.0 TDI Multitronic    Coimbator
e
...      ...                                            ...        ...
7248    7248           Volkswagen Vento Diesel Trendline    Hyderaba
d
7249    7249                      Volkswagen Polo GT TSI        Mumba
i
7250    7250                      Nissan Micra Diesel XV        Kolkat
a
7251    7251                      Volkswagen Polo GT TSI         Pun
e
7252    7252  Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...       Koch
i

      Year  Kilometers_Driven Fuel_Type Transmission Owner_Type    Milea
ge
0     2010              72000       CNG      Manual      First  26.6 km/
kg
1     2015              41000    Diesel      Manual      First  19.67 km
pl
2     2011              46000    Petrol      Manual      First   18.2 km
pl
3     2012              87000    Diesel      Manual      First  20.77 km
pl
4     2013              40670    Diesel   Automatic     Second   15.2 km
pl
...    ...                ...       ...        ...        ...        ...
7248  2011              89411    Diesel      Manual      First  20.54 km
pl
7249  2015              59000    Petrol   Automatic      First  17.21 km
pl
7250  2012              28000    Diesel      Manual      First  23.08 km
pl
7251  2013              52262    Petrol   Automatic      Third   17.2 km
pl
7252  2014              72443    Diesel   Automatic      First  10.0 km
pl

      Engine      Power  Seats  New_Price  Price
0     998 CC   58.16 bhp    5.0        NaN   1.75
1    1582 CC   126.2 bhp    5.0        NaN  12.50
2    1199 CC    88.7 bhp    5.0  8.61 Lakh   4.50
3    1248 CC   88.76 bhp    7.0        NaN   6.00
4    1968 CC   140.8 bhp    5.0        NaN  17.74
...      ...         ...    ...        ...    ...
7248 1598 CC   103.6 bhp    5.0        NaN    NaN
7249 1197 CC   103.6 bhp    5.0        NaN    NaN
7250 1461 CC    63.1 bhp    5.0        NaN    NaN
7251 1197 CC   103.6 bhp    5.0        NaN    NaN
```

In [7]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7253 entries, 0 to 7252
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   S.No.              7253 non-null   int64
 1   Name               7253 non-null   object
 2   Location           7253 non-null   object
 3   Year               7253 non-null   int64
 4   Kilometers_Driven  7253 non-null   int64
 5   Fuel_Type          7253 non-null   object
 6   Transmission       7253 non-null   object
 7   Owner_Type         7253 non-null   object
 8   Mileage            7251 non-null   object
 9   Engine             7207 non-null   object
 10  Power              7207 non-null   object
 11  Seats              7200 non-null   float64
 12  New_Price          1006 non-null   object
 13  Price              6019 non-null   float64
dtypes: float64(2), int64(3), object(9)
memory usage: 793.4+ KB
```

In [8]:
```python
df.isna().sum()
```

Out[8]:
```
S.No.                 0
Name                  0
Location              0
Year                  0
Kilometers_Driven     0
Fuel_Type             0
Transmission          0
Owner_Type            0
Mileage               2
Engine               46
Power                46
Seats                53
New_Price          6247
Price              1234
dtype: int64
```

```
7252  2148 CC    170 bhp   5.0          NaN     NaN
```
In [9]:
```
[7253 rows x 14 columns]>
ax=df["Seats"].hist(bins=10,density=True,stacked=True,color='blue',alpha=0.6)
df["Seats"].plot(kind='density',color='violet')
ax.set(xlabel='Seats')
plt.xlim(-0,25)
plt.show()
```
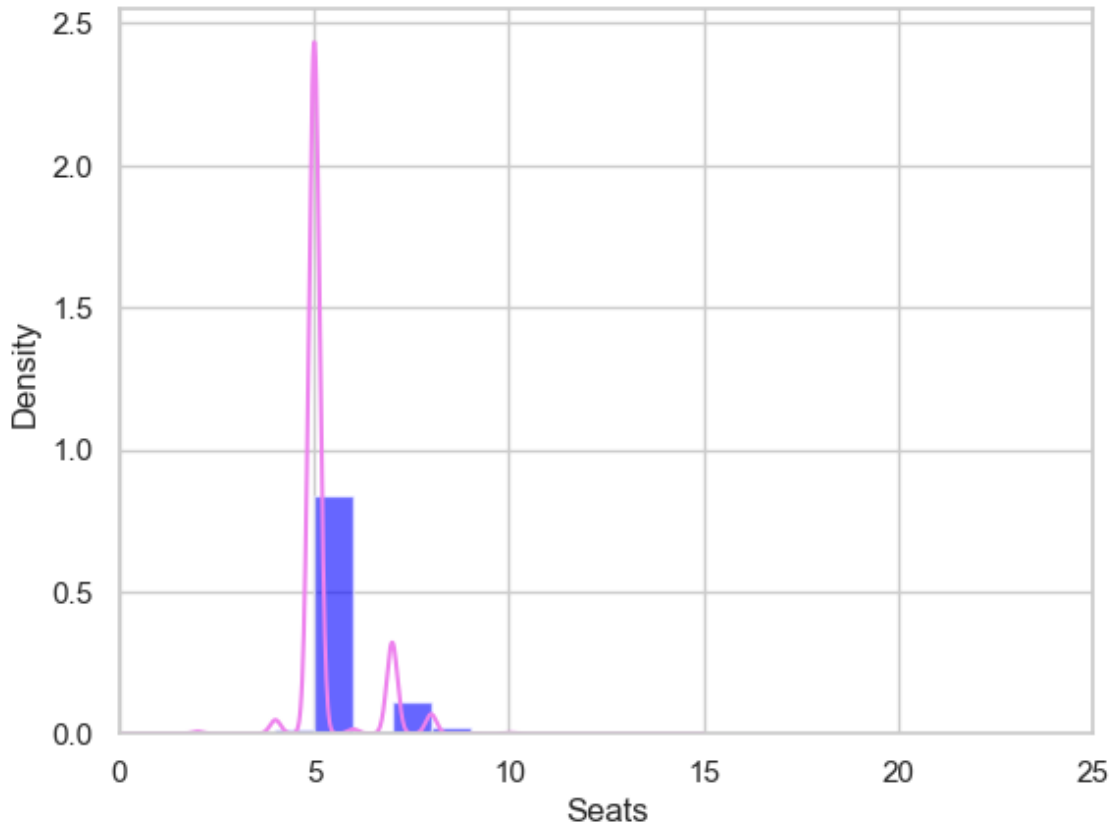

In [10]:
```
 print(df["Seats"].mean(skipna=True))
print(df["Seats"].median(skipna=True))
```

```
5.279722222222222
5.0
```

In [11]:
```
 print(df["New_Price"].isnull().sum()/df.shape[0])
print(df["Price"].isnull().sum()/df.shape[0])
print(df["Mileage"].isnull().sum()/df.shape[0])
print(df["Engine"].isnull().sum()/df.shape[0])
print(df["Power"].isnull().sum()/df.shape[0])
```

```
0.8612987729215497
0.1701364952433476
0.0002757479663587481
0.006342203226251206
0.006342203226251206
```

In [12]:

```python
print(df['Engine'].value_counts())
sns.countplot(x='Engine',data=df,palette='Set3')
plt.xlim(-0,45)
```

```
Engine
1197 CC    732
1248 CC    610
1498 CC    370
998 CC     309
1198 CC    281
            ...
1489 CC      1
1422 CC      1
2706 CC      1
1978 CC      1
1389 CC      1
Name: count, Length: 150, dtype: int64
```

Out[12]:

(0.0, 45.0)
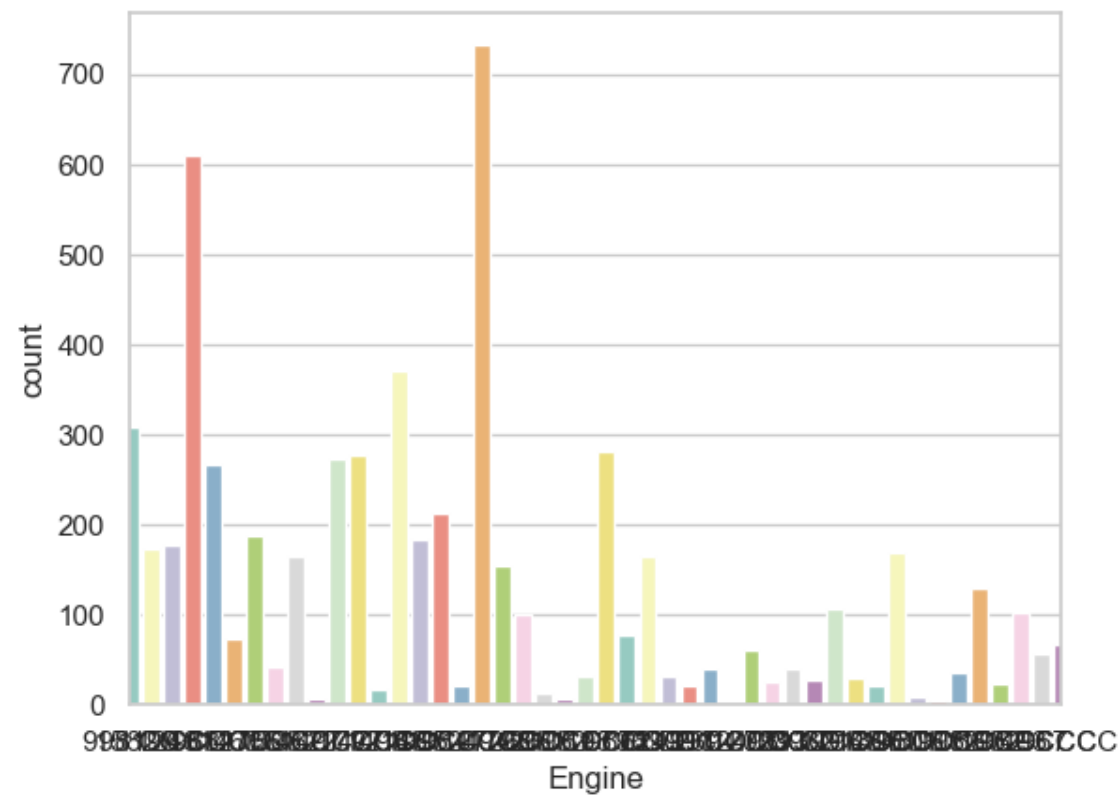
In [13]:

```python
data=df.copy()
data['Seats'].fillna(df['Seats'].median(skipna=True),inplace=True)
data.drop('New_Price',axis=1,inplace=True)
data['Price'].fillna(df['Price'].median(skipna=True),inplace=True)
data['Mileage'].fillna(df['Mileage'].value_counts().idxmax(),inplace=True)
data.drop('Engine',axis=1,inplace=True)
data.drop('Power',axis=1,inplace=True)
```

In [14]:

```python
data.isnull().sum()
```

Out[14]:

```
S.No.               0
Name                0
Location            0
Year                0
Kilometers_Driven   0
Fuel_Type           0
Transmission        0
Owner_Type          0
Mileage             0
Seats               0
Price               0
dtype: int64
```

In [15]:

```python
data.head()
```

Out[15]:

	S.No.	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_T
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Sec

In [16]:

```python
plt.figure(figsize=(15,8))
ax=df["Price"].hist(bins=10,density=True,stacked=True,color='blue',alpha=0.6)
df["Price"].plot(kind='density',color='teal')
ax.legend(['Mileage','Seats'])
ax.set(xlabel='Price')
plt.xlim(-9,80)
plt.show()
```
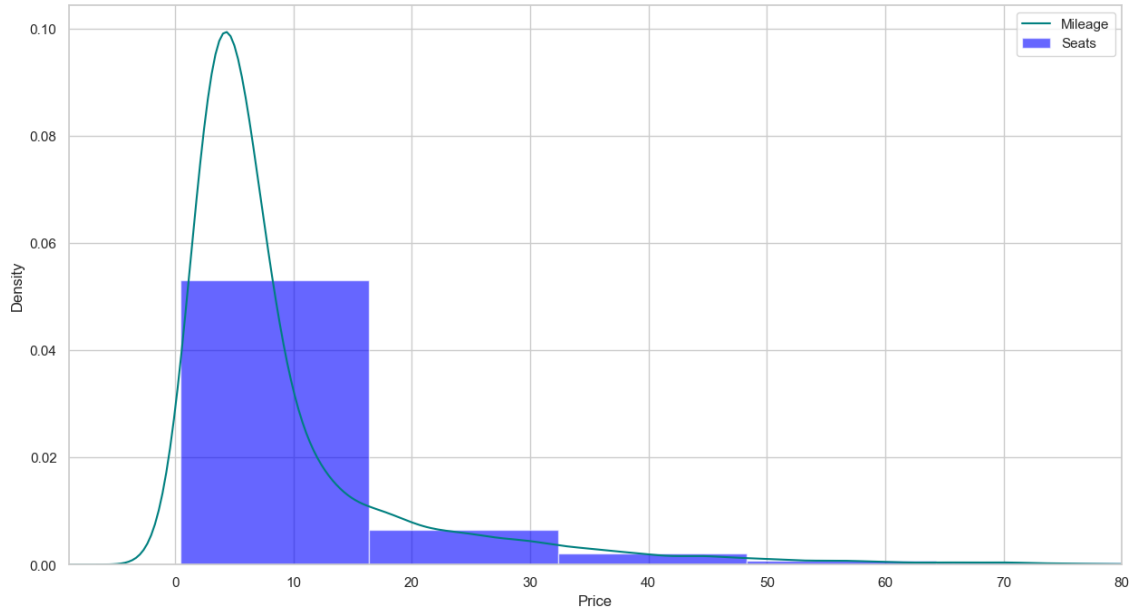

In [17]:

```python
training=pd.get_dummies(data,columns=["S.No."])
final_train=training
final_train.head()
```

Out[17]:

	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	M
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	
1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	
2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	
3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	
4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	

5 rows × 7263 columns
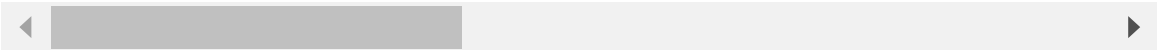
In [18]:

```python
sns.barplot(x='Price',y='Year',data=final_train,color='mediumturquoise')
plt.show()
```
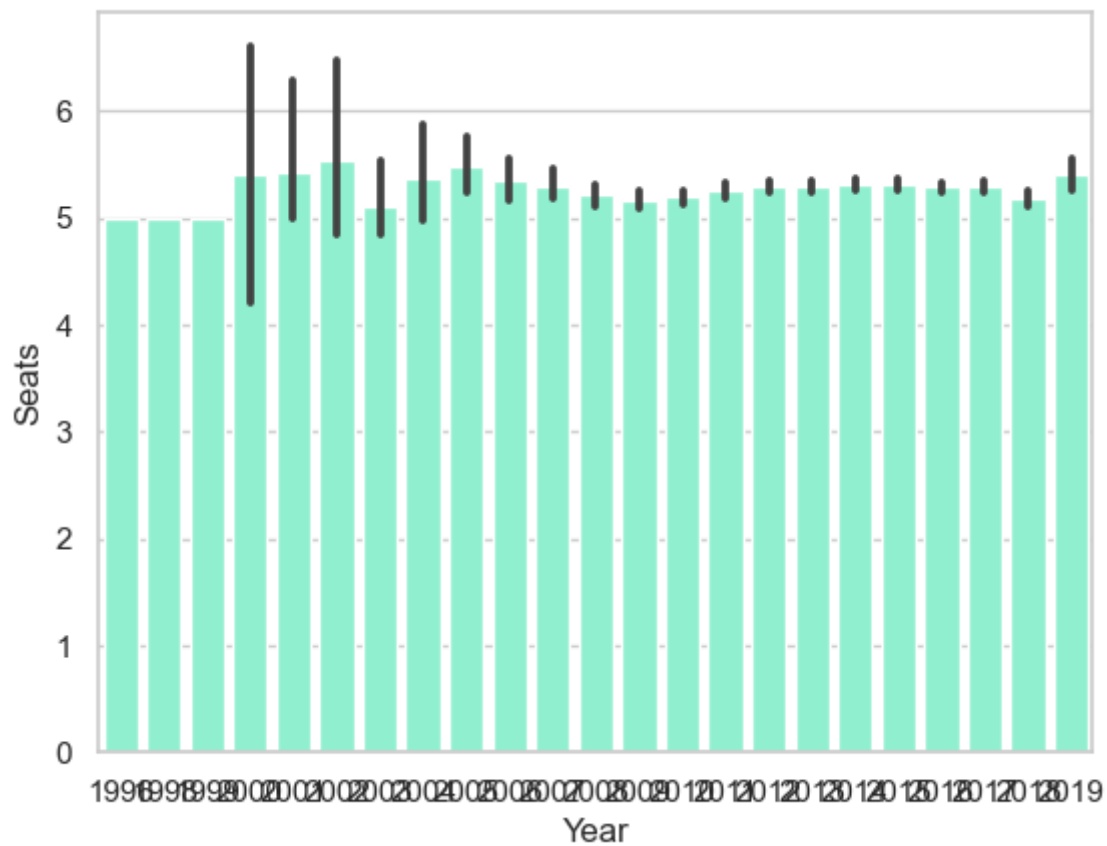
In [19]:

```python
import seaborn as sns
import matplotlib.pyplot as plt
sns.barplot(x='Year',y='Seats',data=df,color='aquamarine')
plt.show()
```


In [ ]: