LOGISTIC REGRESSION USED CARS

In [1]:

```
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")
df=pd.read_csv(r"C:\Users\prajapath Arjun\Downloads\used_cars_data.csv")
print(df)
```

```
      S.No.                                          Name    Location  \
0         0                       Maruti Wagon R LXI CNG      Mumbai
1         1                 Hyundai Creta 1.6 CRDi SX Option        Pune
2         2                               Honda Jazz V     Chennai
3         3                           Maruti Ertiga VDI     Chennai
4         4                 Audi A4 New 2.0 TDI Multitronic  Coimbatore
...     ...                                          ...         ...
7248   7248              Volkswagen Vento Diesel Trendline   Hyderabad
7249   7249                       Volkswagen Polo GT TSI      Mumbai
7250   7250                       Nissan Micra Diesel XV     Kolkata
7251   7251                       Volkswagen Polo GT TSI        Pune
7252   7252  Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...       Kochi

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

       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
7252  2148 CC    170 bhp    5.0        NaN    NaN

[7253 rows x 14 columns]
```

In [2]:

```
df.describe()
```

Out[2]:

	S.No.	Year	Kilometers_Driven	Seats	Price
count	7253.000000	7253.000000	7.253000e+03	7200.000000	6019.000000
mean	3626.000000	2013.365366	5.869906e+04	5.279722	9.479468
std	2093.905084	3.254421	8.442772e+04	0.811660	11.187917
min	0.000000	1996.000000	1.710000e+02	0.000000	0.440000
25%	1813.000000	2011.000000	3.400000e+04	5.000000	3.500000
50%	3626.000000	2014.000000	5.341600e+04	5.000000	5.640000
75%	5439.000000	2016.000000	7.300000e+04	5.000000	9.950000
max	7252.000000	2019.000000	6.500000e+06	10.000000	160.000000

In [3]:

```
df.shape
```

Out[3]:

```
(7253, 14)
```

In [4]:

```
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 [5]:

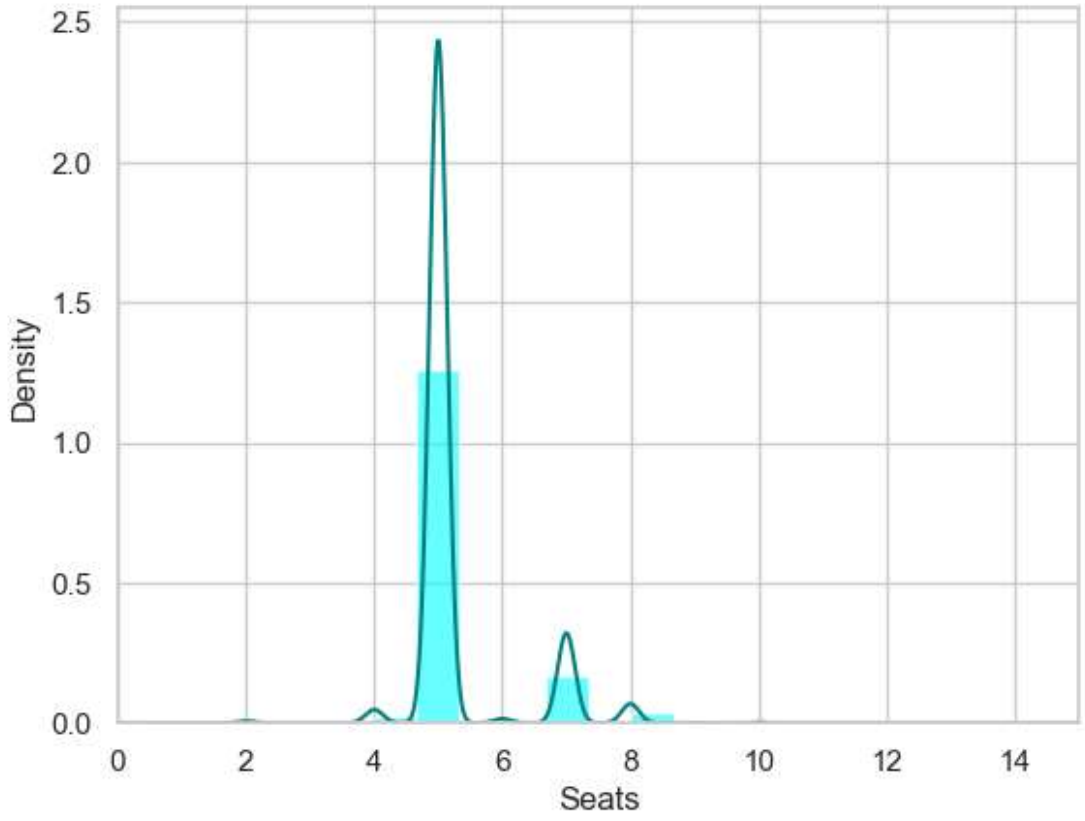
```
df.isnull().sum()
```

Out[5]:

```
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
```

In [6]:

```
ax = df["Seats"].hist(bins=15, density=True, stacked=True, color='cyan', alpha=0.6)
df["Seats"].plot(kind='density', color='teal')
ax.set(xlabel='Seats')
plt.xlim(-0,15)
plt.show()
```



In [7]:

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

```
5.279722222222222
5.0
```

In [8]:

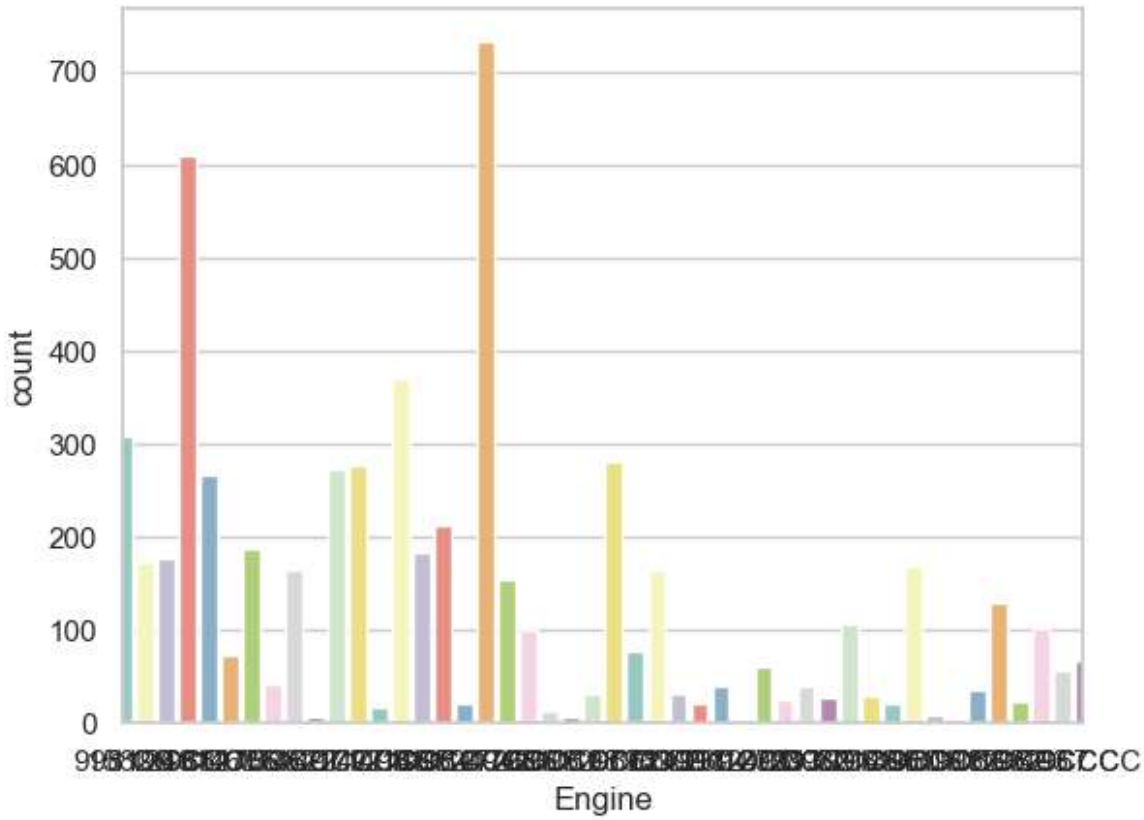
```
print(df["New_Price"].isnull().sum()/df.shape[0]*100)
print(df["Price"].isnull().sum()/df.shape[0]*100)
print(df["Mileage"].isnull().sum()/df.shape[0]*100)
print(df["Engine"].isnull().sum()/df.shape[0]*100)
print(df["Power"].isnull().sum()/df.shape[0]*100)
```

```
86.12987729215497
17.01364952433476
0.02757479663587481
0.6342203226251206
0.6342203226251206
```

In [9]:

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

```
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
```



In [10]:

```
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(),inplace=True)
data.drop('Engine',axis=1,inplace=True)
data.drop('Power',axis=1,inplace=True)
```

In [11]:

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

Out[11]:

```
S.No.                0
Name                 0
Location             0
Year                 0
Kilometers_Driven    0
Fuel_Type            0
Transmission         0
Owner_Type           0
Mileage              2
Seats                0
Price                0
dtype: int64
```

In [12]:

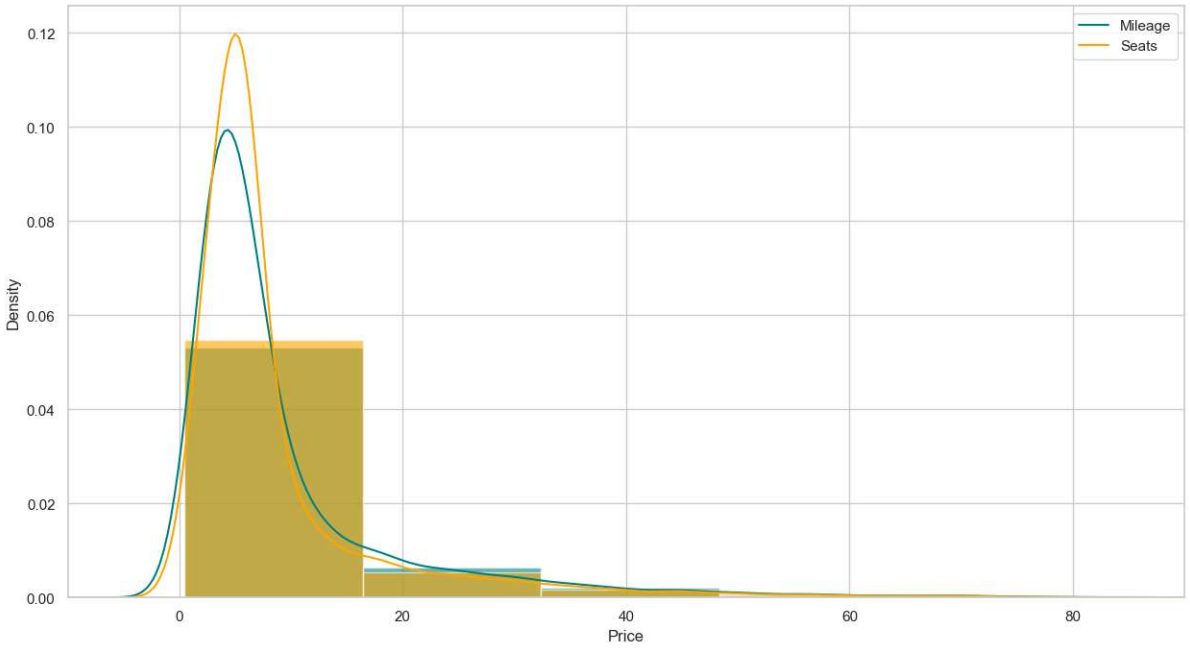
```
data.head()
```

Out[12]:

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

In [13]:

```
plt.figure(figsize=(15,8))
ax=df["Price"].hist(bins=10,density=True,stacked=True,color='teal',alpha=0.6)
df["Price"].plot(kind='density',color='teal')
ax=data["Price"].hist(bins=10,density=True,stacked=True,color='orange',alpha=0.6)
data["Price"].plot(kind='density',color='orange')
ax.legend(['Mileage','Seats'])
ax.set(xlabel='Price')
plt.xlim(-10,90)
plt.show()
```



In [14]:

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

Out[14]:

	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	S
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	
1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	
2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	
3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	
4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	

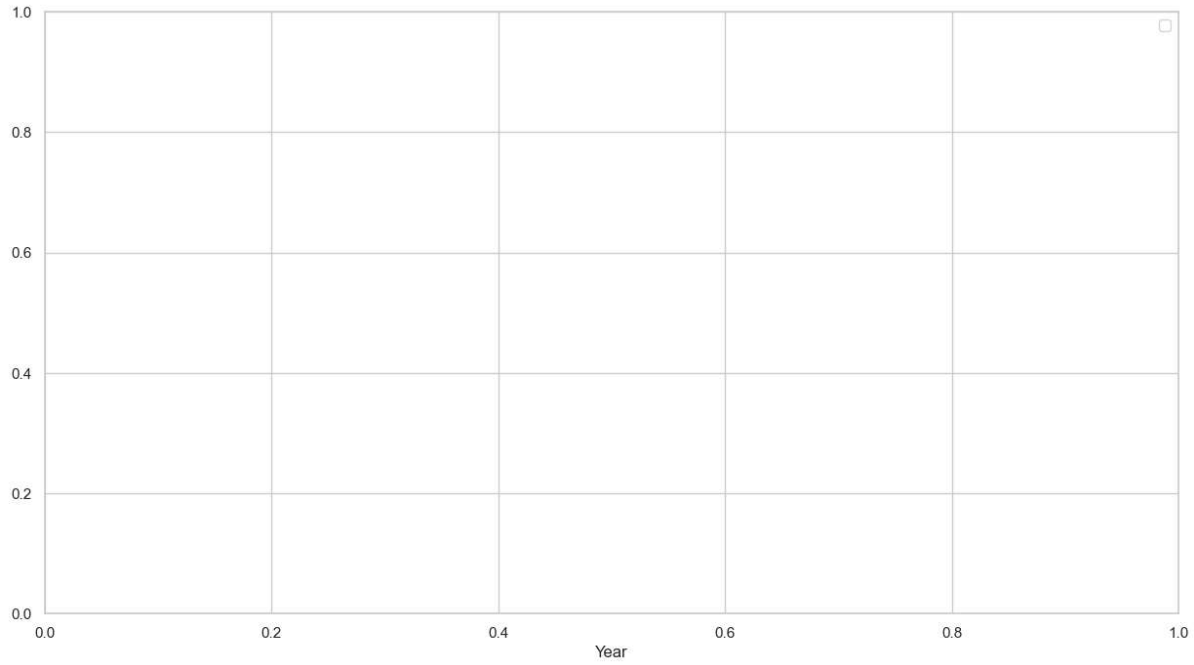
5 rows × 7263 columns

In [15]:

```
plt.figure(figsize=(15,8))
ax=sns.kdeplot(final_train["Price"][final_train.Year==1],color='darkturquoise',alpha=0.6)
sns.kdeplot(final_train["Kilometers_Driven"][final_train.Year==0],color="lightgreen",alpha=0.6
plt.legend(['Cars','density'])
ax.set(xlabel='Year')
```

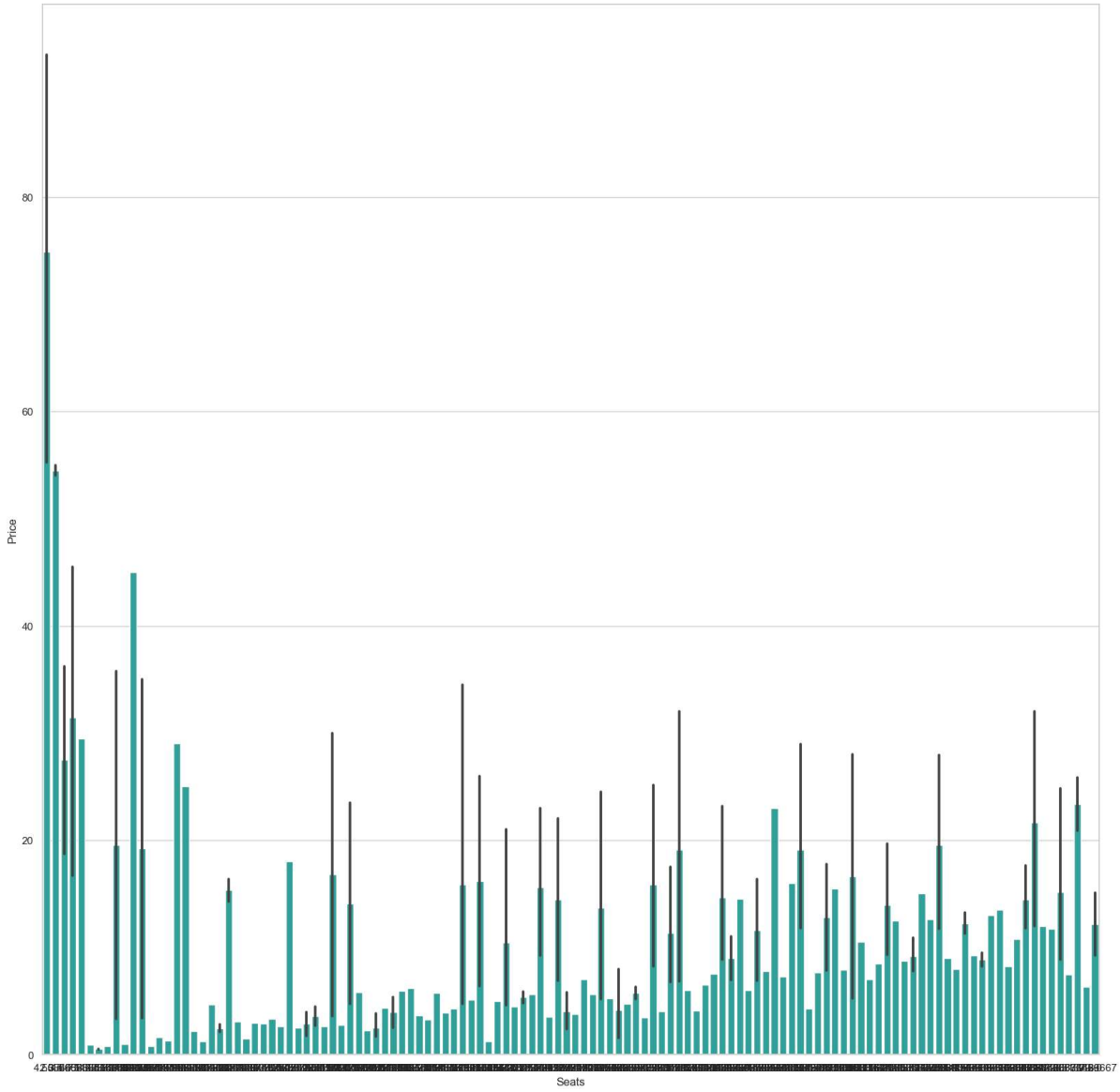
Out[15]:

```
[Text(0.5, 0, 'Year')]
```



In [16]:

```
plt.figure(figsize=(20,20))
avg_survival_byage=final_train[['Seats','Price']].groupby(['Price'],as_index=False).mean()
g=sns.barplot(x='Seats',y='Price',data=avg_survival_byage,color="LightSeaGreen")
plt.show()
```



In [17]:

```
sns.barplot(x='Price',y='Year',data=final_train,color="mediumturquoise")
plt.show()
```

In [18]:

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