```
import pandas as pd
import numpy as np

import seaborn as sns
import matplotlib as plt

data=pd.read_csv("quikr_car.csv")
data.head()
```

```
                              Name of car Company name  Year
Price  \
0    Hyundai Santro Xing XO eRLX Euro III      Hyundai  2007
80,000
1                 Mahindra Jeep CL550 MDI     Mahindra  2006
4,25,000
2              Maruti Suzuki Alto 800 Vxi       Maruti  2018  Ask For
Price
3  Hyundai Grand i10 Magna 1.2 Kappa VTVT      Hyundai  2014
3,25,000
4        Ford EcoSport Titanium 1.5L TDCi         Ford  2014
5,75,000

   Kms driven Fuel_type
0      45,000    Petrol
1          40    Diesel
2  22,000 kms    Petrol
3      28,000    Petrol
4      36,000    Diesel
```

```
data.shape
```

```
(892, 6)
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 892 entries, 0 to 891
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Name of car   892 non-null    object
 1   Company name  892 non-null    object
 2   Year          892 non-null    object
 3   Price         892 non-null    object
 4   Kms driven    840 non-null    object
 5   Fuel_type     837 non-null    object
dtypes: object(6)
memory usage: 41.9+ KB
```

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

```
Name of car      0
Company name     0
Year             0
Price            0
Kms driven      52
Fuel_type       55
dtype: int64
```

```
data=data[data['Year'].str.isnumeric()]
data['Year']=data['Year'].astype(int)
data=data[data['Price']!="Ask For Price"]
data['Price']=data['Price'].str.replace(',','').astype(int)
data['Kms driven']=data['Kms driven'].str.split('
').str.get(0).str.replace(',','')
data=data[data['Kms driven'].str.isnumeric()]
data['Kms driven']=data['Kms driven'].astype(int)
data=data[~data['Fuel_type'].isna()]
data['Name of car']=data['Name of car'].str.split('
').str.slice(0,3).str.join(' ')
data=data[data['Price']<6000000].reset_index(drop=True)
data.to_csv('Cleanedcar data.csv')
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 815 entries, 0 to 814
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Name of car   815 non-null    object
 1   Company name  815 non-null    object
 2   Year          815 non-null    int32
 3   Price         815 non-null    int32
 4   Kms driven    815 non-null    int32
 5   Fuel_type     815 non-null    object
dtypes: int32(3), object(3)
memory usage: 28.8+ KB
```

```
data.describe(include="all")
```

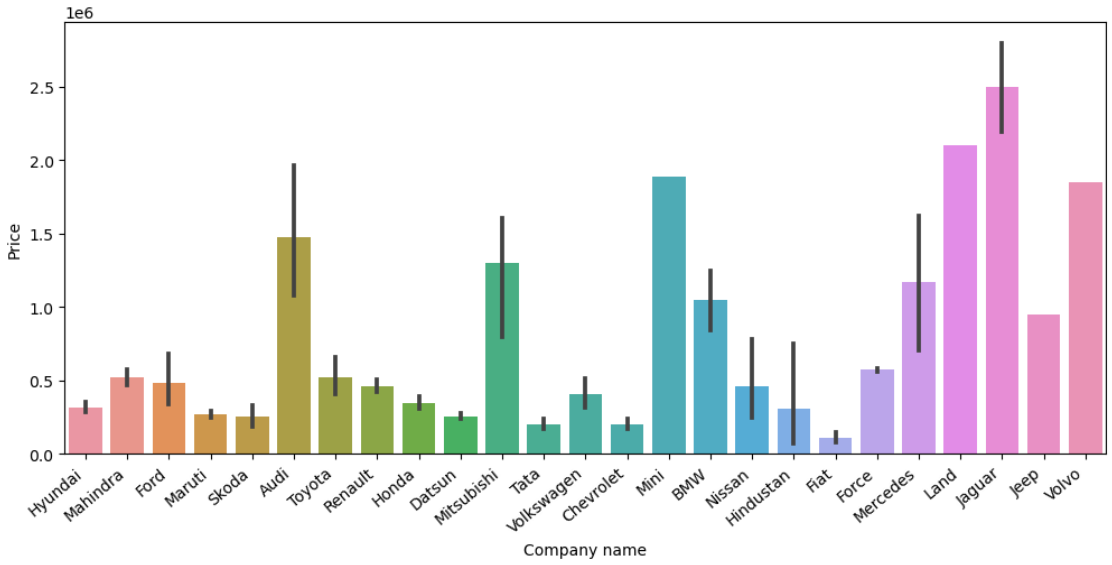
```
                Name of car Company name         Year         Price  \
count                   815          815   815.000000  8.150000e+02
unique                  254           25          NaN           NaN
top     Maruti Suzuki Swift       Maruti          NaN           NaN
freq                     51          221          NaN           NaN
mean                    NaN          NaN  2012.442945  4.017933e+05
std                     NaN          NaN     4.005079  3.815888e+05
min                     NaN          NaN  1995.000000  3.000000e+04
25%                     NaN          NaN  2010.000000  1.750000e+05
50%                     NaN          NaN  2013.000000  2.999990e+05
75%                     NaN          NaN  2015.000000  4.900000e+05
```

```
max                    NaN          NaN  2019.000000  3.100000e+06
```

```
           Kms driven Fuel_type
count      815.000000       815
unique            NaN         3
top               NaN    Petrol
freq              NaN       428
mean     46277.096933       NaN
std      34318.459638       NaN
min          0.000000       NaN
25%      27000.000000       NaN
50%      41000.000000       NaN
75%      56879.000000       NaN
max     400000.000000       NaN
```

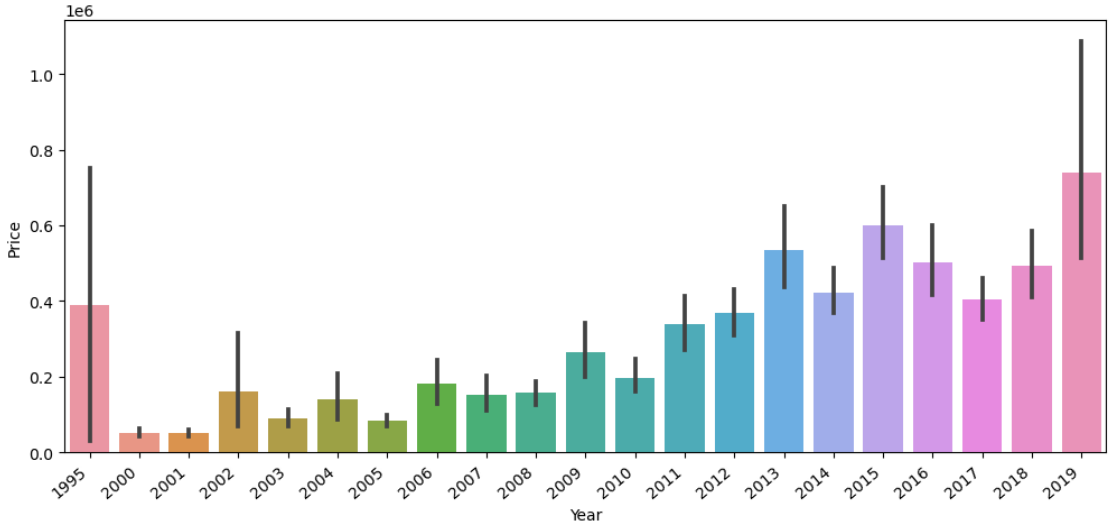
Relationshio of company with Price

```
import matplotlib.pyplot as plt
plt.subplots(figsize=(12,5))
ax=sns.barplot(x='Company name',y='Price',data=data)
ax.set_xticklabels(ax.get_xticklabels(),rotation=40,ha='right')
plt.show()
```



Relationship of year with Price

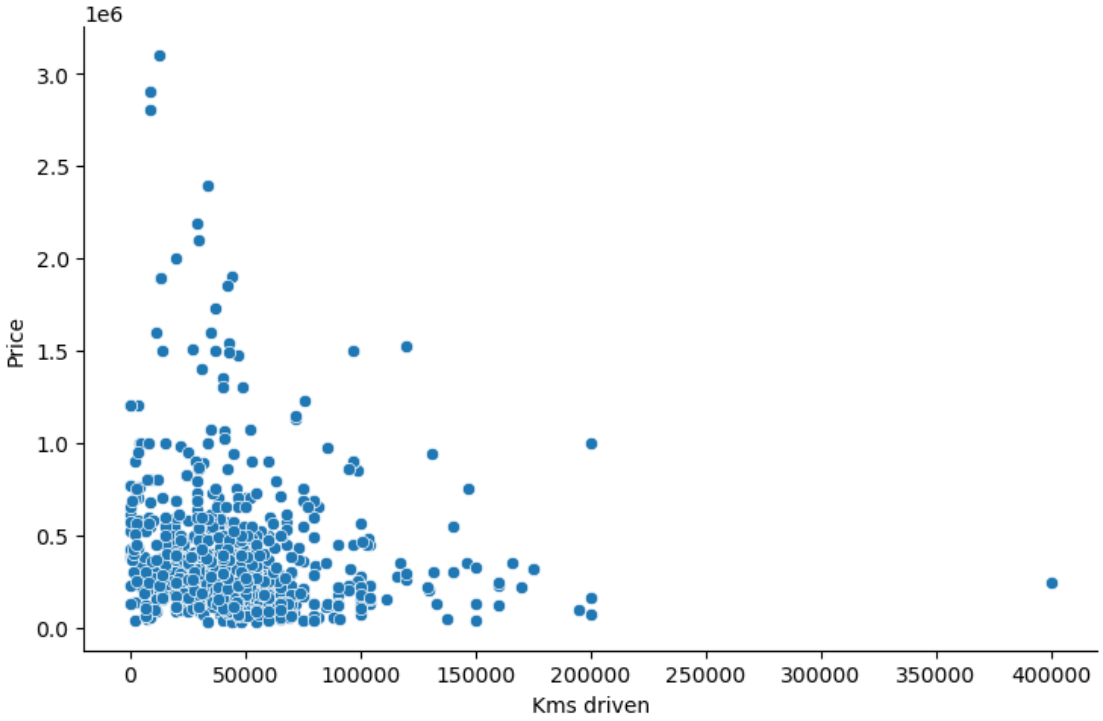
```
plt.subplots(figsize=(12,5))
ax=sns.barplot(x='Year',y='Price',data=data)
ax.set_xticklabels(ax.get_xticklabels(),rotation=40,ha='right')
plt.show()
```



Relationship of kms_driven with price

```
sns.relplot(x='Kms driven',y='Price',data=data,height=5,aspect=1.5)
```

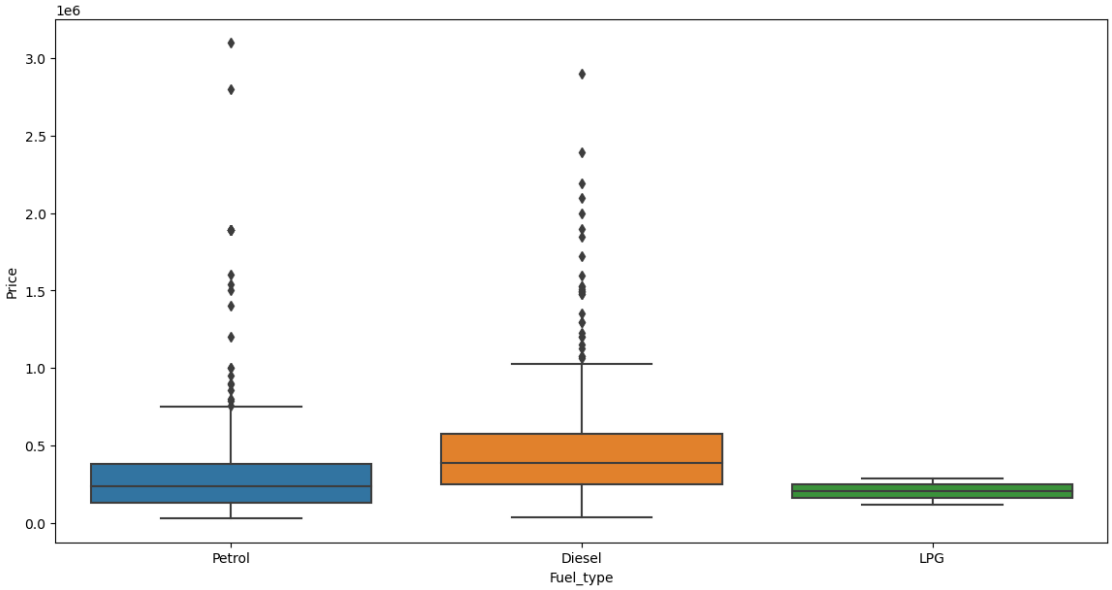
```
<seaborn.axisgrid.FacetGrid at 0x227d9ec7eb0>
```



relationship of fuel type with price

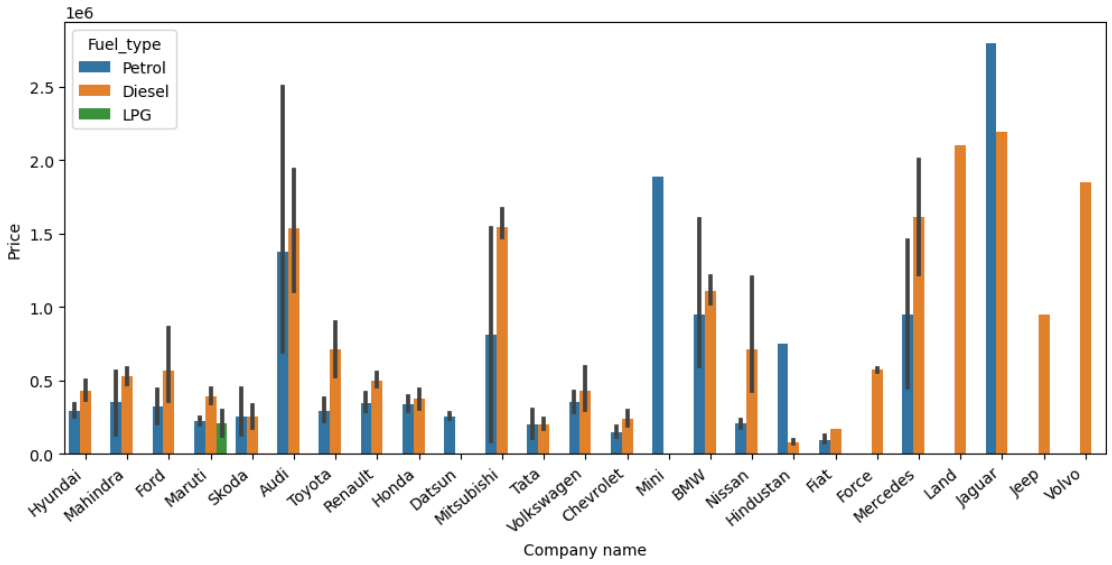
```
plt.subplots(figsize=(14,7))
sns.boxplot(x='Fuel_type',y='Price',data=data)
```

```
<AxesSubplot:xlabel='Fuel_type', ylabel='Price'>
```



##Relationship of price with year and company

```
plt.subplots(figsize=(12,5))
ax=sns.barplot(x='Company name',y='Price',data=data,hue='Fuel_type')
ax.set_xticklabels(ax.get_xticklabels(),rotation=40,ha='right')
plt.show()
```




```
data['Price'].unique()
```

```
array(['80,000', '4,25,000', 'Ask For Price', '3,25,000', '5,75,000',
       '1,75,000', '1,90,000', '8,30,000', '2,50,000', '1,82,000',
       '3,15,000', '4,15,000', '3,20,000', '10,00,000', '5,00,000',
       '3,50,000', '1,60,000', '3,10,000', '75,000', '1,00,000',
```

'2,90,000', '95,000', '1,80,000', '3,85,000', '1,05,000',
'6,50,000', '6,89,999', '4,48,000', '5,49,000', '5,01,000',
'4,89,999', '2,80,000', '3,49,999', '2,84,999', '3,45,000',
'4,99,999', '2,35,000', '2,49,999', '14,75,000', '3,95,000',
'2,20,000', '1,70,000', '85,000', '2,00,000', '5,70,000',
'1,10,000', '4,48,999', '18,91,111', '1,59,500', '3,44,999',
'4,49,999', '8,65,000', '6,99,000', '3,75,000', '2,24,999',
'12,00,000', '1,95,000', '3,51,000', '2,40,000', '90,000',
'1,55,000', '6,00,000', '1,89,500', '2,10,000', '3,90,000',
'1,35,000', '16,00,000', '7,01,000', '2,65,000', '5,25,000',
'3,72,000', '6,35,000', '5,50,000', '4,85,000', '3,29,500',
'2,51,111', '5,69,999', '69,999', '2,99,999', '3,99,999',
'4,50,000', '2,70,000', '1,58,400', '1,79,000', '1,25,000',
'2,99,000', '1,50,000', '2,75,000', '2,85,000', '3,40,000',
'70,000', '2,89,999', '8,49,999', '7,49,999', '2,74,999',
'9,84,999', '5,99,999', '2,44,999', '4,74,999', '2,45,000',
'1,69,500', '3,70,000', '1,68,000', '1,45,000', '98,500',
'2,09,000', '1,85,000', '9,00,000', '6,99,999', '1,99,999',
'5,44,999', '1,99,000', '5,40,000', '49,000', '7,00,000',
'55,000',
'8,95,000', '3,55,000', '5,65,000', '3,65,000', '40,000',
'4,00,000', '3,30,000', '5,80,000', '3,79,000', '2,19,000',
'5,19,000', '7,30,000', '20,00,000', '21,00,000', '14,00,000',
'3,11,000', '8,55,000', '5,35,000', '1,78,000', '3,00,000',
'2,55,000', '5,49,999', '3,80,000', '57,000', '4,10,000',
'2,25,000', '1,20,000', '59,000', '5,99,000', '6,75,000',
'72,500',
'6,10,000', '2,30,000', '5,20,000', '5,24,999', '4,24,999',
'6,44,999', '5,84,999', '7,99,999', '4,44,999', '6,49,999',
'9,44,999', '5,74,999', '3,74,999', '1,30,000', '4,01,000',
'13,50,000', '1,74,999', '2,39,999', '99,999', '3,24,999',
'10,74,999', '11,30,000', '1,49,000', '7,70,000', '30,000',
'3,35,000', '3,99,000', '65,000', '1,69,999', '1,65,000',
'5,60,000', '9,50,000', '7,15,000', '45,000', '9,40,000',
'1,55,555', '15,00,000', '4,95,000', '8,00,000', '12,99,000',
'5,30,000', '14,99,000', '32,000', '4,05,000', '7,60,000',
'7,50,000', '4,19,000', '1,40,000', '15,40,000', '1,23,000',
'4,98,000', '4,80,000', '4,88,000', '15,25,000', '5,48,900',
'7,25,000', '99,000', '52,000', '28,00,000', '4,99,000',
'3,81,000', '2,78,000', '6,90,000', '2,60,000', '90,001',
'1,15,000', '15,99,000', '1,59,000', '51,999', '2,15,000',
'35,000', '11,50,000', '2,69,000', '60,000', '4,30,000',
'85,00,003', '4,01,919', '4,90,000', '4,24,000', '2,05,000',
'5,49,900', '3,71,500', '4,35,000', '1,89,700', '3,89,700',
'3,60,000', '2,95,000', '1,14,990', '10,65,000', '4,70,000',
'48,000', '1,88,000', '4,65,000', '1,79,999', '21,90,000',
'23,90,000', '10,75,000', '4,75,000', '10,25,000', '6,15,000',
'19,00,000', '14,90,000', '15,10,000', '18,50,000', '7,90,000',
'17,25,000', '12,25,000', '68,000', '9,70,000', '31,00,000',
'8,99,000', '88,000', '53,000', '5,68,500', '71,000',

```
'5,90,000',
       '7,95,000', '42,000', '1,89,000', '1,62,000', '35,999',
       '29,00,000', '39,999', '50,500', '5,10,000', '8,60,000',
       '5,00,001'], dtype=object)
```