In [1]:
```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import make_column_transformer
from sklearn.pipeline import make_pipeline
```

In [2]:
```
df = pd.read_csv('D:\Machine Learning\quikr_car.csv')
df.head()
```

Out[2]:

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing XO eRLX Euro III	Hyundai	2007	80,000	45,000 kms	Petrol
1	Mahindra Jeep CL550 MDI	Mahindra	2006	4,25,000	40 kms	Diesel
2	Maruti Suzuki Alto 800 Vxi	Maruti	2018	Ask For Price	22,000 kms	Petrol
3	Hyundai Grand i10 Magna 1.2 Kappa VTVT	Hyundai	2014	3,25,000	28,000 kms	Petrol
4	Ford EcoSport Titanium 1.5L TDCi	Ford	2014	5,75,000	36,000 kms	Diesel

In [3]:
```
df.isnull().sum()
```

Out[3]:
```
name           0
company        0
year           0
Price          0
kms_driven    52
fuel_type     55
dtype: int64
```

In [4]:
```
df.shape
```

Out[4]:
```
(892, 6)
```

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 892 entries, 0 to 891
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   name        892 non-null    object
 1   company     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
```

In [6]:
```
df.Price.unique()
```

Out[6]:
```
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)
```

#Quality

- year has many non-year values
- year object to int
- price has Ask for Price value
- price object to int
- kms_driven has kms with integers
- kms_driven has nan values
- fuel_type has nan values
- keep first 3 words of name

In [7]:
```
df2 = df.copy()
```

In [8]:
```
df = df[df.year.str.isnumeric()]
```

In [9]:
```
df.year = df.year.astype(int)
```

In [10]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 842 entries, 0 to 891
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   name        842 non-null    object
 1   company     842 non-null    object
 2   year        842 non-null    int32 
 3   Price       842 non-null    object
 4   kms_driven  840 non-null    object
 5   fuel_type   837 non-null    object
dtypes: int32(1), object(5)
memory usage: 42.8+ KB
```

In [11]:
```
df = df[df.Price != 'Ask For Price']
```

In [12]:
```
df.Price = df.Price.str.replace(',','').astype(int)
```

In [13]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 819 entries, 0 to 891
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   name        819 non-null    object
 1   company     819 non-null    object
 2   year        819 non-null    int32 
 3   Price       819 non-null    int32 
 4   kms_driven  819 non-null    object
 5   fuel_type   816 non-null    object
dtypes: int32(2), object(4)
memory usage: 38.4+ KB
```

In [14]:
```
df['kms_driven']
```

Out[14]:
```
0        45,000 kms
1            40 kms
3        28,000 kms
4        36,000 kms
6        41,000 kms
           ...
886    1,32,000 kms
888      27,000 kms
889      40,000 kms
890          Petrol
891          Petrol
Name: kms_driven, Length: 819, dtype: object
```

In [15]:
```
# df['kms_driven'].str.split(' ').str[0].str.replace(',', '')
df.kms_driven = df['kms_driven'].str.split(' ').str.get(0).str.replace(',', '')
```

In [16]:
```
df = df[df['kms_driven'].str.isnumeric()]
```

In [17]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 817 entries, 0 to 889
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   name        817 non-null    object
 1   company     817 non-null    object
 2   year        817 non-null    int32 
 3   Price       817 non-null    int32 
 4   kms_driven  817 non-null    object
 5   fuel_type   816 non-null    object
dtypes: int32(2), object(4)
memory usage: 38.3+ KB
```

In [18]:
```
df['kms_driven'] = df['kms_driven'].astype(int)
```

In [19]:
```
df.isnull().sum()
```

Out[19]:
```
name          0
company       0
year          0
Price         0
kms_driven    0
fuel_type     1
dtype: int64
```

In [20]:
```
df = df[~df['fuel_type'].isna()]
```

In [21]:
```
# df.name
# df.name.str.split(' ')
# df.name.str.split(' ').str.slice(0,3)
df['name'] = df.name.str.split(' ').str.slice(0,3).str.join(' ')
```

In [22]:
```
df
```

Out[22]:

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing	Hyundai	2007	80000	45000	Petrol
1	Mahindra Jeep CL550	Mahindra	2006	425000	40	Diesel
3	Hyundai Grand i10	Hyundai	2014	325000	28000	Petrol
4	Ford EcoSport Titanium	Ford	2014	575000	36000	Diesel
6	Ford Figo	Ford	2012	175000	41000	Diesel
...	...	...	...	...	...	...
883	Maruti Suzuki Ritz	Maruti	2011	270000	50000	Petrol
885	Tata Indica V2	Tata	2009	110000	30000	Diesel
886	Toyota Corolla Altis	Toyota	2009	300000	132000	Petrol
888	Tata Zest XM	Tata	2018	260000	27000	Diesel
889	Mahindra Quanto C8	Mahindra	2013	390000	40000	Diesel

816 rows × 6 columns

In [23]:
```
# here the index are not proper so we do reset_index
df.reset_index(drop=True)
```

Out[23]:

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing	Hyundai	2007	80000	45000	Petrol
1	Mahindra Jeep CL550	Mahindra	2006	425000	40	Diesel
2	Hyundai Grand i10	Hyundai	2014	325000	28000	Petrol
3	Ford EcoSport Titanium	Ford	2014	575000	36000	Diesel
4	Ford Figo	Ford	2012	175000	41000	Diesel
...	...	...	...	...	...	...
811	Maruti Suzuki Ritz	Maruti	2011	270000	50000	Petrol
812	Tata Indica V2	Tata	2009	110000	30000	Diesel
813	Toyota Corolla Altis	Toyota	2009	300000	132000	Petrol
814	Tata Zest XM	Tata	2018	260000	27000	Diesel
815	Mahindra Quanto C8	Mahindra	2013	390000	40000	Diesel

816 rows × 6 columns

In [24]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 816 entries, 0 to 889
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   name        816 non-null    object
 1   company     816 non-null    object
 2   year        816 non-null    int32 
 3   Price       816 non-null    int32 
 4   kms_driven  816 non-null    int32 
 5   fuel_type   816 non-null    object
dtypes: int32(3), object(3)
memory usage: 35.1+ KB
```

In [25]:
```
df.describe()
```

Out[25]:

	year	Price	kms_driven
count	816.000000	8.160000e+02	816.000000
mean	2012.444853	4.117176e+05	46275.531863
std	4.002992	4.751844e+05	34297.428044
min	1995.000000	3.000000e+04	0.000000
25%	2010.000000	1.750000e+05	27000.000000
50%	2013.000000	2.999990e+05	41000.000000
75%	2015.000000	4.912500e+05	56818.500000
max	2019.000000	8.500003e+06	400000.000000

In [26]:
```
# max   2019.000000     8.500003e+06    400000.000000  this is outlier
# df[df['Price']>6e6]
df = df[df['Price']<6e6].reset_index(drop=True)
df
```

Out[26]:

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing	Hyundai	2007	80000	45000	Petrol
1	Mahindra Jeep CL550	Mahindra	2006	425000	40	Diesel
2	Hyundai Grand i10	Hyundai	2014	325000	28000	Petrol
3	Ford EcoSport Titanium	Ford	2014	575000	36000	Diesel
4	Ford Figo	Ford	2012	175000	41000	Diesel
...	...	...	...	...	...	...
810	Maruti Suzuki Ritz	Maruti	2011	270000	50000	Petrol
811	Tata Indica V2	Tata	2009	110000	30000	Diesel
812	Toyota Corolla Altis	Toyota	2009	300000	132000	Petrol
813	Tata Zest XM	Tata	2018	260000	27000	Diesel
814	Mahindra Quanto C8	Mahindra	2013	390000	40000	Diesel

815 rows × 6 columns

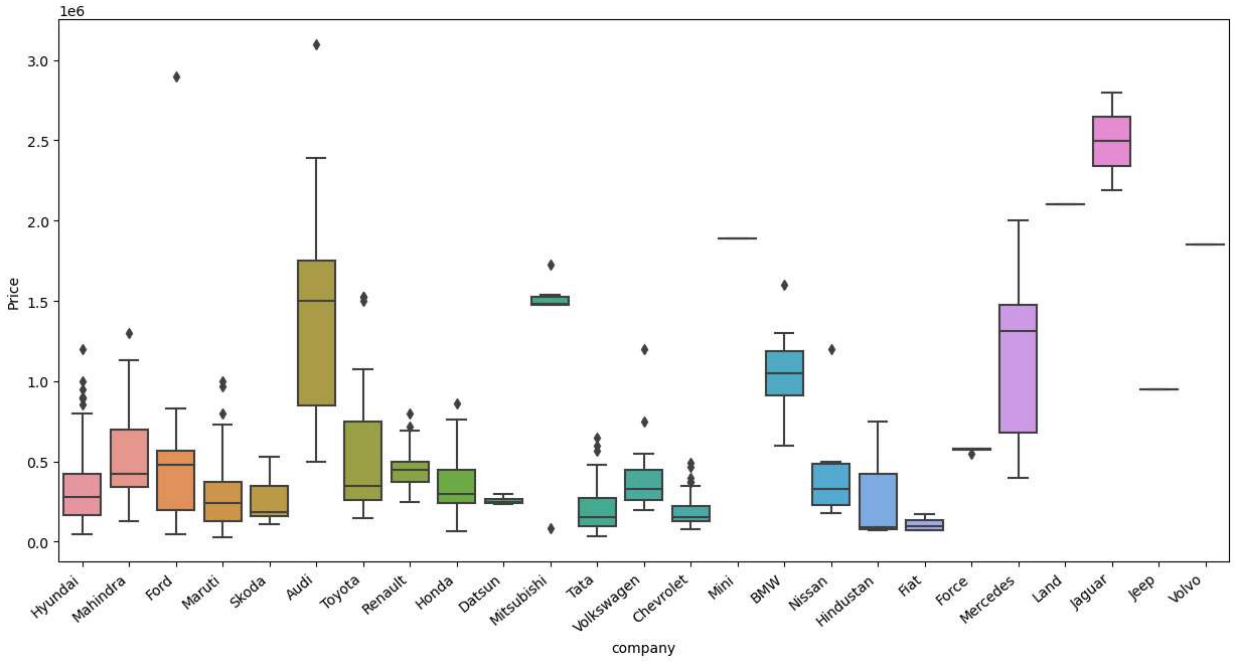
In [27]:
```
# df.to_csv('D:\Machine Learning\Cleaned_quikr_car.csv')
```

In [28]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [29]:
```
# Checking relationship of Company with Price
df['company'].unique()
```

Out[29]:
```
array(['Hyundai', 'Mahindra', 'Ford', 'Maruti', 'Skoda', 'Audi', 'Toyota',
       'Renault', 'Honda', 'Datsun', 'Mitsubishi', 'Tata', 'Volkswagen',
       'Chevrolet', 'Mini', 'BMW', 'Nissan', 'Hindustan', 'Fiat', 'Force',
       'Mercedes', 'Land', 'Jaguar', 'Jeep', 'Volvo'], dtype=object)
```

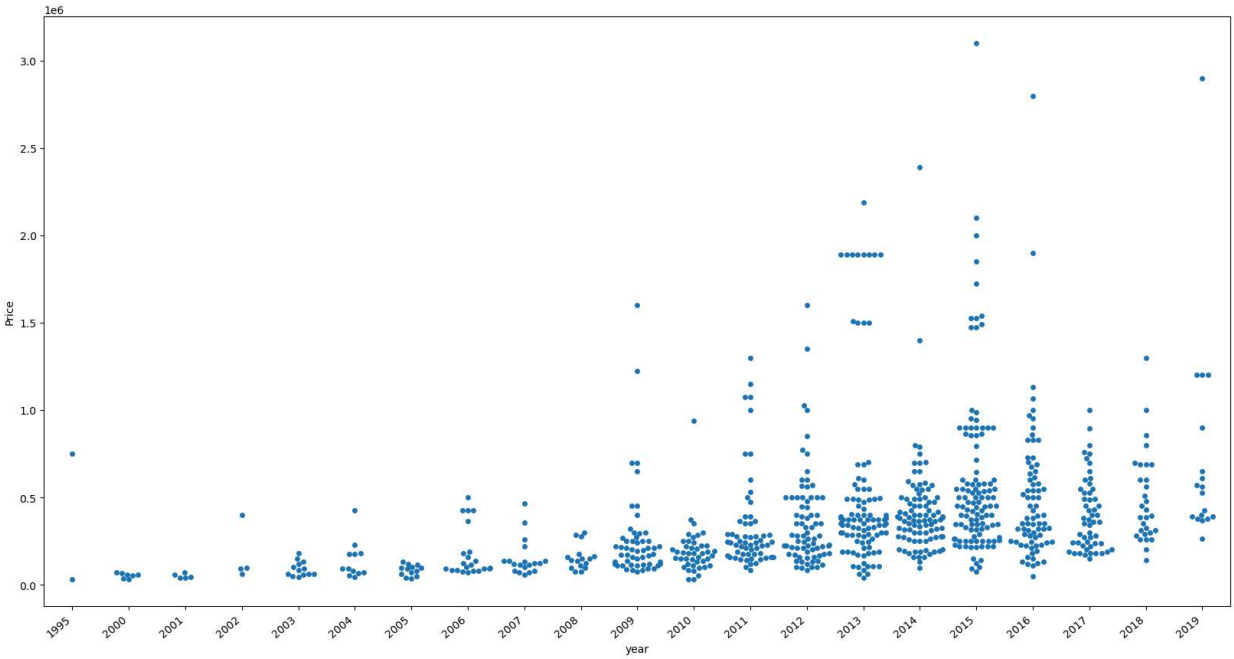
In [30]:
```
plt.subplots(figsize=(15,7))
ax=sns.boxplot(x='company',y='Price',data=df)
ax.set_xticklabels(ax.get_xticklabels(),rotation=40,ha='right') #just to rotate the ir
plt.show()
```



In [31]:
```
# Checking relationship of Year with Price

plt.subplots(figsize=(20,10))
ax=sns.swarmplot(x='year',y='Price',data=df)
ax.set_xticklabels(ax.get_xticklabels(),rotation=40,ha='right')
plt.show()
```

```
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning:
13.6% of the points cannot be placed; you may want to decrease the size of the marker
s or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning:
13.0% of the points cannot be placed; you may want to decrease the size of the marker
s or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning:
6.8% of the points cannot be placed; you may want to decrease the size of the markers
or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning:
10.6% of the points cannot be placed; you may want to decrease the size of the marker
s or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning:
7.7% of the points cannot be placed; you may want to decrease the size of the markers
or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\dhwan\AppData\Local\Temp\ipykernel_32720\954219204.py:5: UserWarning: FixedF
ormatter should only be used together with FixedLocator
  ax.set_xticklabels(ax.get_xticklabels(),rotation=40,ha='right')
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning:
9.3% of the points cannot be placed; you may want to decrease the size of the markers
or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning:
9.6% of the points cannot be placed; you may want to decrease the size of the markers
or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning:
5.5% of the points cannot be placed; you may want to decrease the size of the markers
or use stripplot.
  warnings.warn(msg, UserWarning)
```

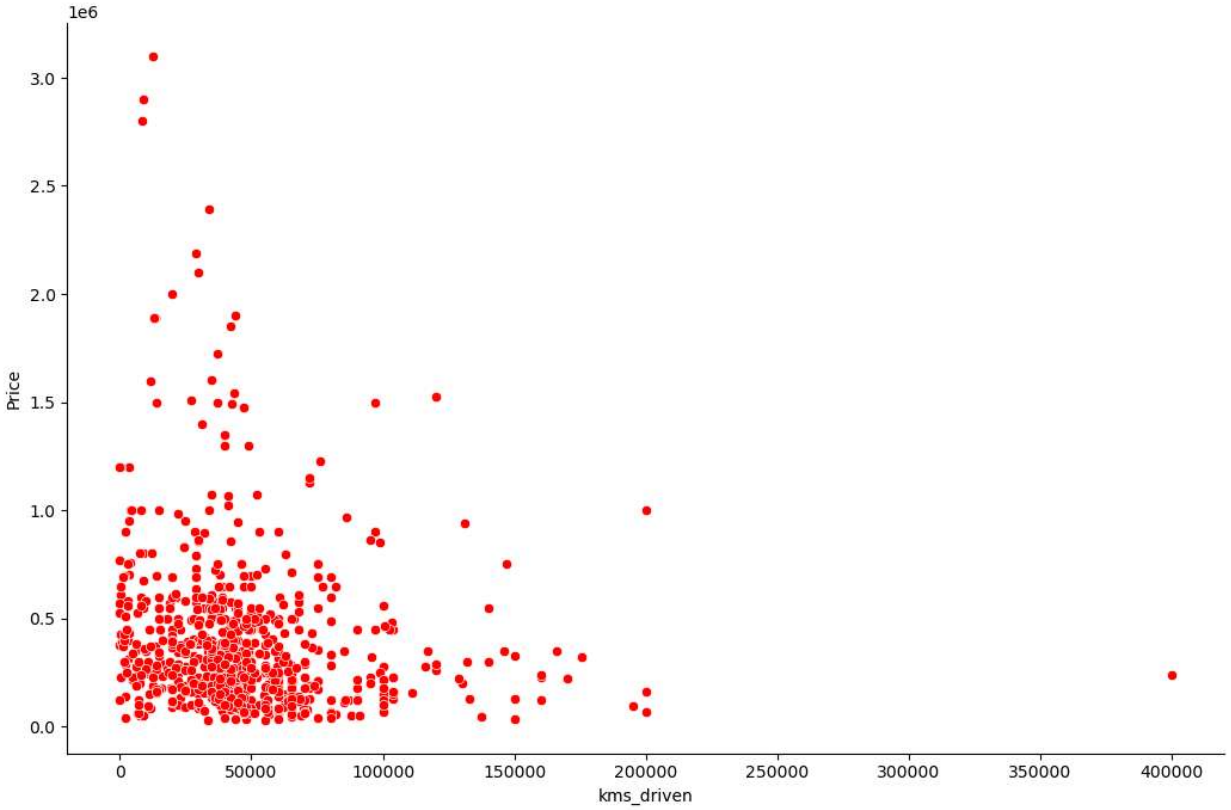


In [32]:
```
# Checking relationship of kms_driven with Price

sns.relplot(x='kms_driven',y='Price',data=df,height=7,aspect=1.5, color='red')
```

```
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The
figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

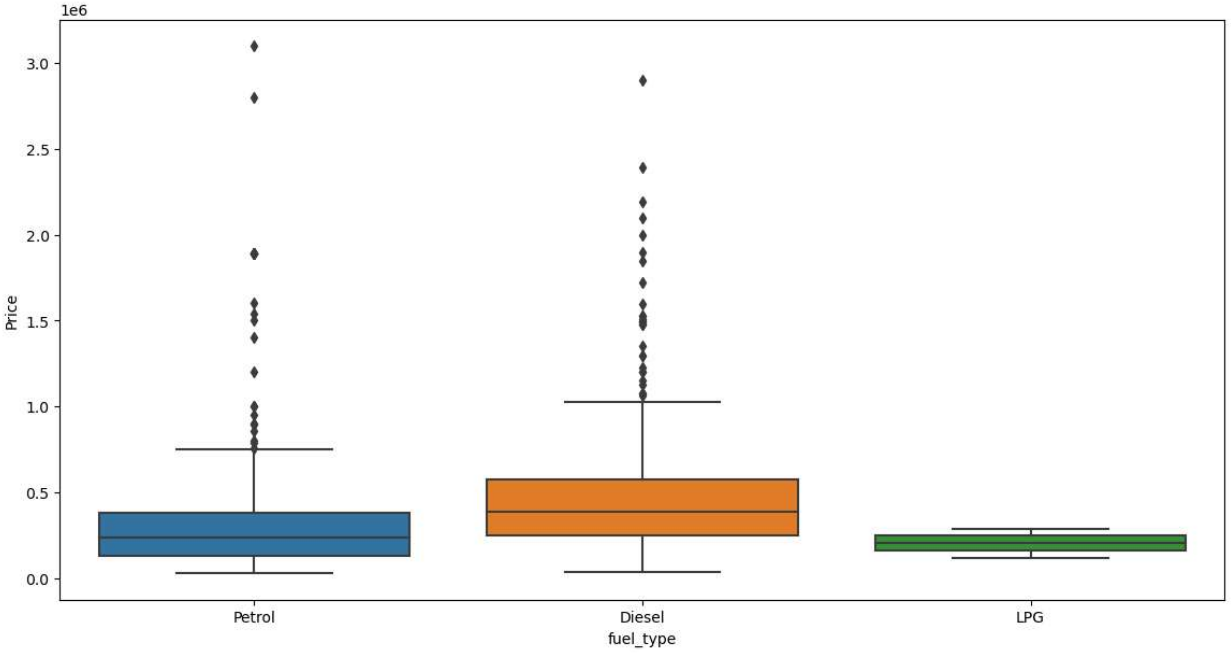
Out[32]:
```
<seaborn.axisgrid.FacetGrid at 0x2ebd1d1d110>
```



In [33]:
```
# Checking relationship of Fuel Type with Price

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

Out[33]:
```
<Axes: xlabel='fuel_type', ylabel='Price'>
```

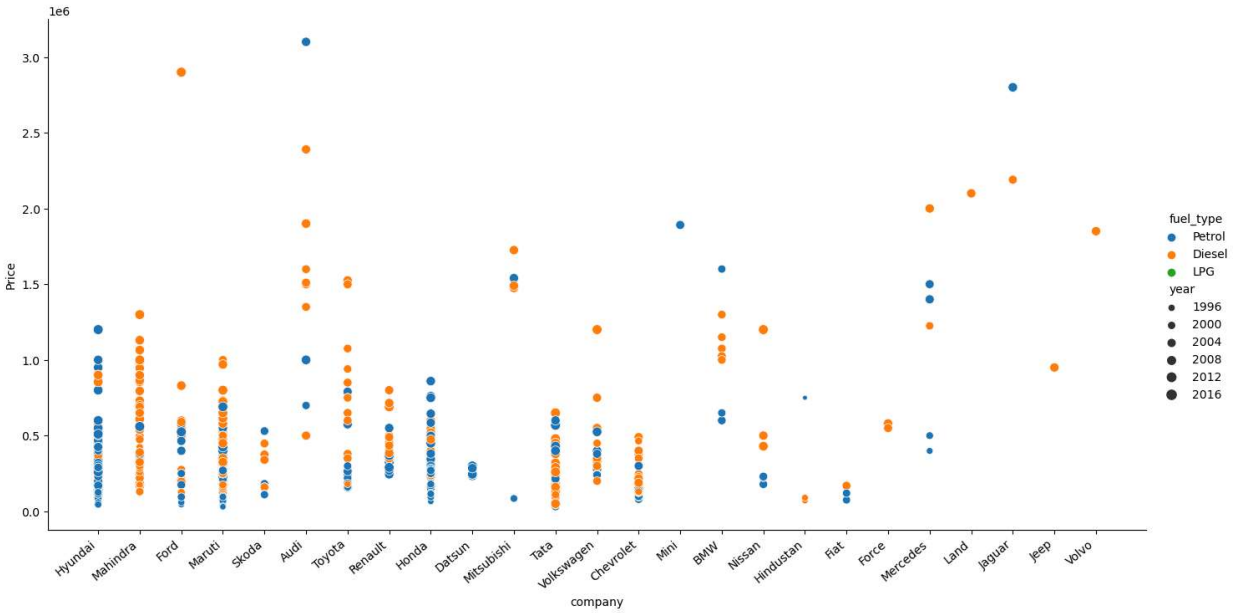


In [34]:
```
# Relationship of Price with FuelType, Year and Company mixed

ax=sns.relplot(x='company',y='Price',data=df,hue='fuel_type',size='year',height=7,aspe
ax.set_xticklabels(rotation=40,ha='right')
```

```
C:\Users\dhwan\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The
figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

Out[34]:
```
<seaborn.axisgrid.FacetGrid at 0x2ebd12bbb10>
```



In [ ]:

In [35]:
```
# Model
```

In [36]:
```
x = df.drop(columns=['Price'])
y = df.Price
```

In [37]:
```
x
```

Out[37]:

	name	company	year	kms_driven	fuel_type
0	Hyundai Santro Xing	Hyundai	2007	45000	Petrol
1	Mahindra Jeep CL550	Mahindra	2006	40	Diesel
2	Hyundai Grand i10	Hyundai	2014	28000	Petrol
3	Ford EcoSport Titanium	Ford	2014	36000	Diesel
4	Ford Figo	Ford	2012	41000	Diesel
...	...	...	...	...	...
810	Maruti Suzuki Ritz	Maruti	2011	50000	Petrol
811	Tata Indica V2	Tata	2009	30000	Diesel
812	Toyota Corolla Altis	Toyota	2009	132000	Petrol
813	Tata Zest XM	Tata	2018	27000	Diesel
814	Mahindra Quanto C8	Mahindra	2013	40000	Diesel

815 rows × 5 columns

In [38]:
```
y
```

Out[38]:
```
0       80000
1      425000
2      325000
3      575000
4      175000
        ...
810    270000
811    110000
812    300000
813    260000
814    390000
Name: Price, Length: 815, dtype: int32
```

In [39]:
```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)
```

In [58]:
```
ohe = OneHotEncoder(handle_unknown='ignore')
ohe.fit(x[['name', 'company', 'fuel_type', 'kms_driven']])
```

Out[58]:
```
OneHotEncoder(handle_unknown='ignore')
```

In [41]:
```
# ohe.categories_
```

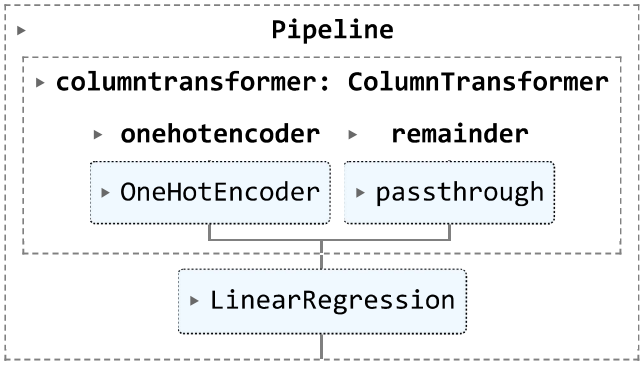
In [59]:
```
column_trans = make_column_transformer((OneHotEncoder(categories=ohe.categories_, hanc
                                                      remainder='passthrough')
#  ohe.categories_ -----> We are passing all the categories so OneHotEncoder knows all
```

In [73]:
```
lr = LinearRegression()
```

In [74]:
```
pipe = make_pipeline(column_trans, lr)
```

In [75]:
```
pipe.fit(x_train, y_train)
```

Out[75]:



In [76]:
```
y_pred = pipe.predict(x_test)
y_pred
```

Out[76]:
```
array([ 2.69536287e+05,  4.90173611e+05,  2.45707386e+05,  1.28043213e+06,
        1.02121224e+05,  2.70264840e+05,  4.52945111e+05,  5.00993523e+05,
        4.08889006e+04,  2.79994745e+05,  4.23654172e+05,  1.81999283e+05,
        4.18928332e+05,  6.06058896e+05,  5.60159943e+05,  4.54345138e+05,
        3.71993517e+05,  3.29566917e+05,  1.93012604e+05,  2.21213800e+05,
        4.59495334e+05,  1.56367842e+05,  3.08468146e+05,  2.36566235e+05,
       -8.04874969e+04,  1.32344131e+05,  2.62725485e+05,  6.28006545e+05,
        3.29483307e+05,  1.31957204e+04,  5.03635672e+05,  1.72542655e+05,
        4.03707526e+05,  2.26650902e+05,  1.43431313e+03,  1.56110052e+05,
        7.43898218e+05,  8.65459252e+04,  5.21717112e+05,  9.41034651e+05,
        4.83405856e+05,  8.18772914e+04,  2.10845824e+05,  3.86262008e+05,
        6.47223237e+05,  3.20011020e+05,  1.14249821e+05,  2.58853981e+05,
        2.24601254e+05,  4.28398043e+05,  1.89111649e+06,  2.79623650e+05,
        2.36566235e+05,  5.84691135e+04,  1.90283016e+05,  1.14322607e+06,
        2.66218890e+05,  3.35141473e+05,  5.50939623e+04,  4.28280615e+05,
        3.42402639e+05,  9.62085242e+05,  9.00112248e+04,  3.58013896e+05,
        8.30007957e+05,  2.72690969e+05,  3.31889717e+05,  5.61133440e+05,
        1.48703476e+06,  2.62164239e+05,  2.25015695e+05,  1.00564607e+06,
        3.90015360e+05,  5.08453759e+05,  4.49699818e+05,  3.56406119e+05,
        7.92029438e+05,  2.18042279e+05,  4.02744006e+05,  1.86916407e+05,
        1.50902076e+04,  5.09390831e+05])
```

In [77]:
```
r2_score(y_test, y_pred)
```

Out[77]:
```
0.7339126929475659
```

In [78]:
```
scores=[]
for i in range(1000):
    x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.1,random_state=i)
    lr=LinearRegression()
    pipe=make_pipeline(column_trans,lr)
    pipe.fit(x_train,y_train)
    y_pred=pipe.predict(x_test)
    scores.append(r2_score(y_test,y_pred))
```

In [79]:
```
np.argmax(scores)
```

Out[79]:
```
172
```

In [80]:
```
scores[np.argmax(scores)]
```

Out[80]:
```
0.9330650248016428
```

In [81]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.1,random_state=i)
lr=LinearRegression()
pipe=make_pipeline(column_trans,lr)
pipe.fit(x_train,y_train)
y_pred=pipe.predict(x_test)
r2_score(y_test, y_pred)
```

Out[81]:
```
0.7339126929475659
```

In [82]:
```
import pickle
```

In [83]:
```
pickle.dump(pipe,open('D:\Machine Learning\LinearRegressionModel.pkl','wb'))
```

In [84]:
```
pipe.predict(pd.DataFrame(columns=x_test.columns,data=np.array(['Maruti Suzuki Swift',
```

Out[84]:
```
array([406561.03575503])
```

In [85]:
```
pipe.steps[0][1].transformers[0][1].categories[0]
```

Out[85]:
```
array(['Audi A3 Cabriolet', 'Audi A4 1.8', 'Audi A4 2.0', 'Audi A6 2.0',
       'Audi A8', 'Audi Q3 2.0', 'Audi Q5 2.0', 'Audi Q7', 'BMW 3 Series',
       'BMW 5 Series', 'BMW 7 Series', 'BMW X1', 'BMW X1 sDrive20d',
       'BMW X1 xDrive20d', 'Chevrolet Beat', 'Chevrolet Beat Diesel',
       'Chevrolet Beat LS', 'Chevrolet Beat LT', 'Chevrolet Beat PS',
       'Chevrolet Cruze LTZ', 'Chevrolet Enjoy', 'Chevrolet Enjoy 1.4',
       'Chevrolet Sail 1.2', 'Chevrolet Sail UVA', 'Chevrolet Spark',
       'Chevrolet Spark 1.0', 'Chevrolet Spark LS', 'Chevrolet Spark LT',
       'Chevrolet Tavera LS', 'Chevrolet Tavera Neo', 'Datsun GO T',
       'Datsun Go Plus', 'Datsun Redi GO', 'Fiat Linea Emotion',
       'Fiat Petra ELX', 'Fiat Punto Emotion', 'Force Motors Force',
       'Force Motors One', 'Ford EcoSport', 'Ford EcoSport Ambiente',
       'Ford EcoSport Titanium', 'Ford EcoSport Trend',
       'Ford Endeavor 4x4', 'Ford Fiesta', 'Ford Fiesta SXi', 'Ford Figo',
       'Ford Figo Diesel', 'Ford Figo Duratorq', 'Ford Figo Petrol',
       'Ford Fusion 1.4', 'Ford Ikon 1.3', 'Ford Ikon 1.6',
       'Hindustan Motors Ambassador', 'Honda Accord', 'Honda Amaze',
       'Honda Amaze 1.2', 'Honda Amaze 1.5', 'Honda Brio', 'Honda Brio V',
       'Honda Brio VX', 'Honda City', 'Honda City 1.5', 'Honda City SV',
       'Honda City VX', 'Honda City ZX', 'Honda Jazz S', 'Honda Jazz VX',
       'Honda Mobilio', 'Honda Mobilio S', 'Honda WR V', 'Hyundai Accent',
       'Hyundai Accent Executive', 'Hyundai Accent GLE',
       'Hyundai Accent GLX', 'Hyundai Creta', 'Hyundai Creta 1.6',
       'Hyundai Elantra 1.8', 'Hyundai Elantra SX', 'Hyundai Elite i20',
       'Hyundai Eon', 'Hyundai Eon D', 'Hyundai Eon Era',
       'Hyundai Eon Magna', 'Hyundai Eon Sportz', 'Hyundai Fluidic Verna',
       'Hyundai Getz', 'Hyundai Getz GLE', 'Hyundai Getz Prime',
       'Hyundai Grand i10', 'Hyundai Santro', 'Hyundai Santro AE',
       'Hyundai Santro Xing', 'Hyundai Sonata Transform', 'Hyundai Verna',
       'Hyundai Verna 1.4', 'Hyundai Verna 1.6', 'Hyundai Verna Fluidic',
       'Hyundai Verna Transform', 'Hyundai Verna VGT',
       'Hyundai Xcent Base', 'Hyundai Xcent SX', 'Hyundai i10',
       'Hyundai i10 Era', 'Hyundai i10 Magna', 'Hyundai i10 Sportz',
       'Hyundai i20', 'Hyundai i20 Active', 'Hyundai i20 Asta',
       'Hyundai i20 Magna', 'Hyundai i20 Select', 'Hyundai i20 Sportz',
       'Jaguar XE XE', 'Jaguar XF 2.2', 'Jeep Wrangler Unlimited',
       'Land Rover Freelander', 'Mahindra Bolero DI',
       'Mahindra Bolero Power', 'Mahindra Bolero SLE',
       'Mahindra Jeep CL550', 'Mahindra Jeep MM', 'Mahindra KUV100',
       'Mahindra KUV100 K8', 'Mahindra Logan', 'Mahindra Logan Diesel',
       'Mahindra Quanto C4', 'Mahindra Quanto C8', 'Mahindra Scorpio',
       'Mahindra Scorpio 2.6', 'Mahindra Scorpio LX',
       'Mahindra Scorpio S10', 'Mahindra Scorpio S4',
       'Mahindra Scorpio SLE', 'Mahindra Scorpio SLX',
       'Mahindra Scorpio VLX', 'Mahindra Scorpio Vlx',
       'Mahindra Scorpio W', 'Mahindra TUV300 T4', 'Mahindra TUV300 T8',
       'Mahindra Thar CRDe', 'Mahindra XUV500', 'Mahindra XUV500 W10',
       'Mahindra XUV500 W6', 'Mahindra XUV500 W8', 'Mahindra Xylo D2',
       'Mahindra Xylo E4', 'Mahindra Xylo E8', 'Maruti Suzuki 800',
       'Maruti Suzuki A', 'Maruti Suzuki Alto', 'Maruti Suzuki Baleno',
       'Maruti Suzuki Celerio', 'Maruti Suzuki Ciaz',
       'Maruti Suzuki Dzire', 'Maruti Suzuki Eeco',
       'Maruti Suzuki Ertiga', 'Maruti Suzuki Esteem',
       'Maruti Suzuki Estilo', 'Maruti Suzuki Maruti',
       'Maruti Suzuki Omni', 'Maruti Suzuki Ritz', 'Maruti Suzuki S',
       'Maruti Suzuki SX4', 'Maruti Suzuki Stingray',
       'Maruti Suzuki Swift', 'Maruti Suzuki Versa',
       'Maruti Suzuki Vitara', 'Maruti Suzuki Wagon', 'Maruti Suzuki Zen',
       'Mercedes Benz A', 'Mercedes Benz B', 'Mercedes Benz C',
       'Mercedes Benz GLA', 'Mini Cooper S', 'Mitsubishi Lancer 1.8',
```

```
'Mitsubishi Pajero Sport', 'Nissan Micra XL', 'Nissan Micra XV',
'Nissan Sunny', 'Nissan Sunny XL', 'Nissan Terrano XL',
'Nissan X Trail', 'Renault Duster', 'Renault Duster 110',
'Renault Duster 110PS', 'Renault Duster 85', 'Renault Duster 85PS',
'Renault Duster RxL', 'Renault Kwid', 'Renault Kwid 1.0',
'Renault Kwid RXT', 'Renault Lodgy 85', 'Renault Scala RxL',
'Skoda Fabia', 'Skoda Fabia 1.2L', 'Skoda Fabia Classic',
'Skoda Laura', 'Skoda Octavia Classic', 'Skoda Rapid Elegance',
'Skoda Superb 1.8', 'Skoda Yeti Ambition', 'Tata Aria Pleasure',
'Tata Bolt XM', 'Tata Indica', 'Tata Indica V2', 'Tata Indica eV2',
'Tata Indigo CS', 'Tata Indigo LS', 'Tata Indigo LX',
'Tata Indigo Marina', 'Tata Indigo eCS', 'Tata Manza',
'Tata Manza Aqua', 'Tata Manza Aura', 'Tata Manza ELAN',
'Tata Nano', 'Tata Nano Cx', 'Tata Nano GenX', 'Tata Nano LX',
'Tata Nano Lx', 'Tata Sumo Gold', 'Tata Sumo Grande',
'Tata Sumo Victa', 'Tata Tiago Revotorq', 'Tata Tiago Revotron',
'Tata Tigor Revotron', 'Tata Venture EX', 'Tata Vista Quadrajet',
'Tata Zest Quadrajet', 'Tata Zest XE', 'Tata Zest XM',
'Toyota Corolla', 'Toyota Corolla Altis', 'Toyota Corolla H2',
'Toyota Etios', 'Toyota Etios G', 'Toyota Etios GD',
'Toyota Etios Liva', 'Toyota Fortuner', 'Toyota Fortuner 3.0',
'Toyota Innova 2.0', 'Toyota Innova 2.5', 'Toyota Qualis',
'Volkswagen Jetta Comfortline', 'Volkswagen Jetta Highline',
'Volkswagen Passat Diesel', 'Volkswagen Polo',
'Volkswagen Polo Comfortline', 'Volkswagen Polo Highline',
'Volkswagen Polo Highline1.2L', 'Volkswagen Polo Trendline',
'Volkswagen Vento Comfortline', 'Volkswagen Vento Highline',
'Volkswagen Vento Konekt', 'Volvo S80 Summum'], dtype=object)
```

In [ ]:

In [86]:
```
print(df.dtypes)
```

```
name          object
company       object
year           int32
Price          int32
kms_driven     int32
fuel_type     object
dtype: object
```

In [ ]: