In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv("C:/Users/user/Desktop/My learning/ClinSoft/merc.csv")
df
```

Out[2]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5
3	G Class	2016	61948	Automatic	16000	Petrol	325	30.4	4.0
4	G Class	2016	73948	Automatic	4000	Petrol	325	30.1	4.0
...	...	...	...	...	...	...	...	...	...
13114	C Class	2020	35999	Automatic	500	Diesel	145	55.4	2.0
13115	B Class	2020	24699	Automatic	2500	Diesel	145	55.4	2.0
13116	GLC Class	2019	30999	Automatic	11612	Diesel	145	41.5	2.1
13117	CLS Class	2019	37990	Automatic	2426	Diesel	145	45.6	2.0
13118	S Class	2019	54999	Automatic	2075	Diesel	145	52.3	2.9

13119 rows × 9 columns

In [3]:
```
print(df.head())
```

```
      model  year  price transmission  mileage fuelType  tax   mpg  \
0       SLK  2005   5200    Automatic    63000   Petrol  325  32.1
1   S Class  2017  34948    Automatic    27000   Hybrid   20  61.4
2  SL CLASS  2016  49948    Automatic     6200   Petrol  555  28.0
3   G Class  2016  61948    Automatic    16000   Petrol  325  30.4
4   G Class  2016  73948    Automatic     4000   Petrol  325  30.1

   engineSize
0         1.8
1         2.1
2         5.5
3         4.0
4         4.0
```

In [4]:
```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13119 entries, 0 to 13118
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   model         13119 non-null  object 
 1   year          13119 non-null  int64  
 2   price         13119 non-null  int64  
 3   transmission  13119 non-null  object 
 4   mileage       13119 non-null  int64  
 5   fuelType      13119 non-null  object 
 6   tax           13119 non-null  int64  
 7   mpg           13119 non-null  float64
 8   engineSize    13119 non-null  float64
dtypes: float64(2), int64(4), object(3)
memory usage: 922.6+ KB
None
```

In [5]:
```
print(df.shape)
```

```
(13119, 9)
```

In [6]:
```
print(df.describe())
```

```
               year          price        mileage           tax  \
mpg  \
count  13119.000000   13119.000000   13119.000000  13119.000000  13119.00
0000
mean    2017.296288   24698.596920   21949.559037    129.972178     55.15
5843
std        2.224709   11842.675542   21176.512267     65.260286     15.22
0082
min     1970.000000     650.000000       1.000000      0.000000      1.10
0000
25%     2016.000000   17450.000000    6097.500000    125.000000     45.60
0000
50%     2018.000000   22480.000000   15189.000000    145.000000     56.50
0000
75%     2019.000000   28980.000000   31779.500000    145.000000     64.20
0000
max     2020.000000  159999.000000  259000.000000    580.000000    217.30
0000

         engineSize
count  13119.000000
mean       2.071530
std        0.572426
min        0.000000
25%        1.800000
50%        2.000000
75%        2.100000
max        6.200000
```

In [7]:
```
print(df.values)
```

```
[[' SLK' 2005 5200 ... 325 32.1 1.8]
 [' S Class' 2017 34948 ... 20 61.4 2.1]
 [' SL CLASS' 2016 49948 ... 555 28.0 5.5]
 ...
 [' GLC Class' 2019 30999 ... 145 41.5 2.1]
 [' CLS Class' 2019 37990 ... 145 45.6 2.0]
 [' S Class' 2019 54999 ... 145 52.3 2.9]]
```

In [8]:
```
print(df.columns)
```

```
Index(['model', 'year', 'price', 'transmission', 'mileage', 'fuelType',
'tax',
       'mpg', 'engineSize'],
      dtype='object')
```

In [9]:
```
df_sorted = df.sort_values("model", ascending = True)
print(df_sorted.head())
```

```
          model  year  price transmission  mileage fuelType  tax   mpg  \
12289   A Class  2014  11200    Automatic    40685   Diesel   20  68.9
1900    A Class  2016  16120       Manual    22125   Diesel   20  68.9
6753    A Class  2019  22899    Semi-Auto     4264   Diesel  145  56.5
6747    A Class  2019  24199    Semi-Auto     2885   Diesel  145  67.3
6746    A Class  2020  25729    Semi-Auto     1000   Petrol  145  53.3

       engineSize
12289         2.1
1900          1.5
6753          2.0
6747          1.5
6746          1.3
```

In [10]:
```
df_sorted = df.sort_values(['year', 'model'], ascending =[False, True])
print(df_sorted.head())
```

```
        model  year  price transmission  mileage fuelType  tax   mpg  \
10    A Class  2020  32980    Automatic      606   Petrol  145  35.8
11    A Class  2020  26980    Automatic     1000   Diesel  145  62.8
33    A Class  2020  28128    Automatic     2563   Petrol  145  47.9
102   A Class  2020  24107    Automatic     3500   Diesel  145  62.8
622   A Class  2020  22490    Automatic     3521   Diesel  145  56.5

     engineSize
10          2.0
11          1.5
33          1.3
102         1.5
622         2.0
```

In [11]:
```
models = df[['model', 'year']]
models
```

Out[11]:

	model	year
0	SLK	2005
1	S Class	2017
2	SL CLASS	2016
3	G Class	2016
4	G Class	2016
...	...	...
13114	C Class	2020
13115	B Class	2020
13116	GLC Class	2019
13117	CLS Class	2019
13118	S Class	2019

13119 rows × 2 columns

In [12]:
```
df[df['price'] > 140000]
```

Out[12]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
5	SL CLASS	2011	149948	Automatic	3000	Petrol	570	21.4	6.2
6199	G Class	2020	159999	Semi-Auto	1350	Petrol	145	21.4	4.0
8737	A Class	2019	140319	Semi-Auto	785	Petrol	150	22.1	4.0
10044	G Class	2020	154998	Automatic	3000	Petrol	150	21.4	4.0

In [13]:
```
df[df['transmission'] == 'Automatic']
```

Out[13]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5
3	G Class	2016	61948	Automatic	16000	Petrol	325	30.4	4.0
4	G Class	2016	73948	Automatic	4000	Petrol	325	30.1	4.0
...	...	...	...	...	...	...	...	...	...
13114	C Class	2020	35999	Automatic	500	Diesel	145	55.4	2.0
13115	B Class	2020	24699	Automatic	2500	Diesel	145	55.4	2.0
13116	GLC Class	2019	30999	Automatic	11612	Diesel	145	41.5	2.1
13117	CLS Class	2019	37990	Automatic	2426	Diesel	145	45.6	2.0
13118	S Class	2019	54999	Automatic	2075	Diesel	145	52.3	2.9

4825 rows × 9 columns

In [14]:
```
df[(df['price'] < 10000) & (df['mileage'] < 20000)]
```

Out[14]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
611	B Class	2014	9998	Semi-Auto	16378	Diesel	20	68.9	1.5
11549	C Class	2002	1495	Automatic	13800	Diesel	305	39.8	2.7

In [15]:
```
df[(df['transmission'] == 'Automatic') & (df['model'] == ' S Class')]
```

Out[15]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1
7	S Class	2012	10948	Automatic	107000	Petrol	265	36.7	3.5
73	S Class	2014	41980	Automatic	26000	Petrol	555	28.0	5.5
77	S Class	2016	24880	Automatic	75000	Hybrid	190	42.4	3.5
80	S Class	2015	38980	Automatic	11000	Petrol	270	33.2	4.7
...	...	...	...	...	...	...	...	...	...
12880	S Class	2011	10950	Automatic	66000	Diesel	265	41.5	3.0
12943	S Class	2015	25460	Automatic	46757	Diesel	205	47.9	3.0
13047	S Class	2015	28086	Automatic	38347	Diesel	205	48.7	3.0
13058	S Class	2015	51990	Automatic	23156	Petrol	555	28.0	5.5
13118	S Class	2019	54999	Automatic	2075	Diesel	145	52.3	2.9

106 rows × 9 columns

In [16]:
```
mod = [' SLK', ' S Class', ' SL CLASS',]
```

In [17]:
```
df[df['model'].isin(mod)]
```

Out[17]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5
5	SL CLASS	2011	149948	Automatic	3000	Petrol	570	21.4	6.2
7	S Class	2012	10948	Automatic	107000	Petrol	265	36.7	3.5
...	...	...	...	...	...	...	...	...	...
13078	SL CLASS	2017	19889	Automatic	17564	Diesel	30	70.6	2.1
13088	SL CLASS	2017	18048	Automatic	24529	Diesel	30	70.6	2.1
13089	SL CLASS	2017	20659	Automatic	6573	Diesel	150	70.6	2.1
13108	SLK	2016	17499	Automatic	30128	Diesel	30	70.6	2.1
13118	S Class	2019	54999	Automatic	2075	Diesel	145	52.3	2.9

552 rows × 9 columns

In [18]:
```
df['prop'] = df['price'] / df['mileage']
df
```

Out[18]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize	p
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8	0.082
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1	1.294
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5	8.056
3	G Class	2016	61948	Automatic	16000	Petrol	325	30.4	4.0	3.871
4	G Class	2016	73948	Automatic	4000	Petrol	325	30.1	4.0	18.487
...	...	...	...	...	...	...	...	...	...	
13114	C Class	2020	35999	Automatic	500	Diesel	145	55.4	2.0	71.998
13115	B Class	2020	24699	Automatic	2500	Diesel	145	55.4	2.0	9.879
13116	GLC Class	2019	30999	Automatic	11612	Diesel	145	41.5	2.1	2.669
13117	CLS Class	2019	37990	Automatic	2426	Diesel	145	45.6	2.0	15.659
13118	S Class	2019	54999	Automatic	2075	Diesel	145	52.3	2.9	26.505

13119 rows × 10 columns

In [19]:
```
df['mileage_km'] = df['mileage']*1.61
df
```

Out[19]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize	p
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8	0.082
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1	1.294
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5	8.056
3	G Class	2016	61948	Automatic	16000	Petrol	325	30.4	4.0	3.871
4	G Class	2016	73948	Automatic	4000	Petrol	325	30.1	4.0	18.487
...	...	...	...	...	...	...	...	...	...	
13114	C Class	2020	35999	Automatic	500	Diesel	145	55.4	2.0	71.998
13115	B Class	2020	24699	Automatic	2500	Diesel	145	55.4	2.0	9.879
13116	GLC Class	2019	30999	Automatic	11612	Diesel	145	41.5	2.1	2.669
13117	CLS Class	2019	37990	Automatic	2426	Diesel	145	45.6	2.0	15.659
13118	S Class	2019	54999	Automatic	2075	Diesel	145	52.3	2.9	26.505

13119 rows × 11 columns

In [20]:
```
high_km = df[df['mileage_km'] > 250000]
high_km_sort = high_km.sort_values('mileage_km', ascending  =False)
result = high_km_sort[['mileage_km', 'model', 'year', 'price']]
print(result)
```

```
       mileage_km     model  year  price
11344   416990.00   V Class  2010   6949
12904   387195.34   A Class  2016  16249
11581   297850.00       CLK  2003   3495
11382   273700.00   C Class  2012   5250
11928   267398.46   E Class  2015   6775
11378   265650.00   M Class  2004   1995
12648   257600.00   C Class  2007   3695
11252   250960.36   M Class  2010   7449
```

In [21]:
```
print(df['price'].mean())
```

```
24698.596920496988
```

In [22]:
```
print(df['year'].min())
```

```
1970
```

In [23]:
```
sorted_year = df.sort_values("year")
sorted_year['price'].cummax()
```

Out[23]:
```
12072     24999
11912     24999
11794     24999
11808     24999
11860     24999
          ...
2789     159999
2787     159999
9318     159999
4983     159999
7437     159999
Name: price, Length: 13119, dtype: int64
```

In [24]:
```
df_drop_dup = df.drop_duplicates(subset =["model", "year"])
print(df_drop_dup.head())
```

```
       model  year   price transmission  mileage fuelType  tax   mpg  \
0        SLK  2005    5200    Automatic    63000   Petrol  325  32.1
1    S Class  2017   34948    Automatic    27000   Hybrid   20  61.4
2   SL CLASS  2016   49948    Automatic     6200   Petrol  555  28.0
3    G Class  2016   61948    Automatic    16000   Petrol  325  30.4
5   SL CLASS  2011  149948    Automatic     3000   Petrol  570  21.4

   engineSize       prop  mileage_km
0         1.8   0.082540    101430.0
1         2.1   1.294370     43470.0
2         5.5   8.056129      9982.0
3         4.0   3.871750     25760.0
5         6.2  49.982667      4830.0
```

In [25]:
```
df['transmission'].value_counts(normalize = True)
```

Out[25]:
```
transmission
Semi-Auto    0.521991
Automatic    0.367787
Manual       0.110069
Other        0.000152
Name: proportion, dtype: float64
```

In [26]:
```
df['model'].value_counts(normalize = True)
```

Out[26]:
```
model
 C Class      0.285616
 A Class      0.195213
 E Class      0.148868
 GLC Class    0.073176
 GLA Class    0.064563
 B Class      0.045049
 CL Class     0.038951
 GLE Class    0.035140
 SL CLASS     0.019819
 CLS Class    0.018065
 V Class      0.015779
 S Class      0.015016
 GL Class     0.009223
 SLK          0.007241
 CLA Class    0.006555
 X-CLASS      0.006250
 M Class      0.006022
 GLS Class    0.005641
 GLB Class    0.001448
 G Class      0.001143
 CLK          0.000534
 CLC Class    0.000229
 R Class      0.000152
180           0.000076
230           0.000076
220           0.000076
200           0.000076
Name: proportion, dtype: float64
```

In [27]:
```
df['price'].sum()
```

Out[27]:
```
324020893
```

In [28]:
```
df[df['transmission'] == "Manual"]["price"].mean()
```

Out[28]:
```
15442.434903047091
```

In [29]:
```
df[df['transmission'] == "Automatic"]["price"].mean()
```

Out[29]:
```
24406.959585492226
```

In [30]:
```
df.groupby('transmission')['price'].mean()
```

Out[30]:
```
transmission
Automatic    24406.959585
Manual       15442.434903
Other        15847.500000
Semi-Auto    26858.461157
Name: price, dtype: float64
```

In [31]:
```
df.groupby(['transmission', 'fuelType'])["price"].mean()
```

Out[31]:
```
transmission  fuelType
Automatic     Diesel      22902.562607
              Hybrid      21260.788462
              Other       22539.833333
              Petrol      29009.083813
Manual        Diesel      14139.761146
              Petrol      17886.894422
Other         Diesel      12995.000000
              Other       18700.000000
Semi-Auto     Diesel      25365.094395
              Hybrid      25851.014493
              Petrol      30378.890310
Name: price, dtype: float64
```

In [32]:
```
import numpy as np
```

In [33]:
```
df.groupby(['transmission', 'fuelType'])["price"].agg([np.mean, np.min, np
```

Out[33]:

		mean	amin	amax
transmission	**fuelType**			
Automatic	**Diesel**	22902.562607	1495	69990
	Hybrid	21260.788462	10500	56900
	Other	22539.833333	13850	40999
	Petrol	29009.083813	1490	154998
Manual	**Diesel**	14139.761146	1350	36999
	Petrol	17886.894422	650	29495
Other	**Diesel**	12995.000000	12995	12995
	Other	18700.000000	18700	18700
Semi-Auto	**Diesel**	25365.094395	5995	102502
	Hybrid	25851.014493	12995	45599
	Petrol	30378.890310	1995	159999

In [34]:
```
df.pivot_table("price", index = "fuelType")
```

Out[34]:

	price
fuelType	
Diesel	23275.122673
Hybrid	23091.572254
Other	21991.285714
Petrol	28263.208955

In [35]:
```
df.pivot_table("mileage", index = "transmission", aggfunc=[np.mean, np.min
```

Out[35]:

	mean	amin	amax
	mileage	mileage	mileage
transmission			
Automatic	25939.244145	9	259000
Manual	32311.330332	9	150000
Other	34947.500000	30895	39000
Semi-Auto	16949.759930	1	165000

In [36]:
```
df.pivot_table("mileage", index = "transmission", columns='fuelType', fill
               margins = True, aggfunc=[np.mean, np.min, np.max])
```

Out[36]:

	mean					amin
fuelType	Diesel	Hybrid	Other	Petrol	All	Diese
transmission						
Automatic	27969.069182	43718.413462	47420.333333	18479.714051	25939.244145	9
Manual	38964.178344	0.000000	0.000000	19827.300797	32311.330332	10
Other	39000.000000	0.000000	30895.000000	0.000000	34947.500000	39000
Semi-Auto	19437.694269	28631.565217	0.000000	10745.243974	16949.759930	1
All	24690.362142	37701.115607	45059.571429	14469.135928	21949.559037	1

In [ ]:
```

```

In [37]:
```
df.head()
```

Out[37]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize	prop
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8	0.082540
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1	1.294370
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5	8.056129
3	G Class	2016	61948	Automatic	16000	Petrol	325	30.4	4.0	3.871750
4	G Class	2016	73948	Automatic	4000	Petrol	325	30.1	4.0	18.487000

In [38]:
```
df_index = df.set_index("model")
print(df_index)
```

```
            year  price transmission  mileage fuelType  tax   mpg  engine
Size  \
model
 SLK        2005   5200    Automatic    63000   Petrol  325  32.1
1.8
 S Class    2017  34948    Automatic    27000   Hybrid   20  61.4
2.1
 SL CLASS   2016  49948    Automatic     6200   Petrol  555  28.0
5.5
 G Class    2016  61948    Automatic    16000   Petrol  325  30.4
4.0
 G Class    2016  73948    Automatic     4000   Petrol  325  30.1
4.0
...          ...    ...          ...      ...      ...  ...   ...
...
 C Class    2020  35999    Automatic      500   Diesel  145  55.4
2.0
 B Class    2020  24699    Automatic     2500   Diesel  145  55.4
2.0
 GLC Class  2019  30999    Automatic    11612   Diesel  145  41.5
2.1
 CLS Class  2019  37990    Automatic     2426   Diesel  145  45.6
2.0
 S Class    2019  54999    Automatic     2075   Diesel  145  52.3
2.9

                prop  mileage_km
model
 SLK         0.082540   101430.00
 S Class     1.294370    43470.00
 SL CLASS    8.056129     9982.00
 G Class     3.871750    25760.00
 G Class    18.487000     6440.00
...               ...         ...
 C Class    71.998000      805.00
 B Class     9.879600     4025.00
 GLC Class   2.669566    18695.32
 CLS Class  15.659522     3905.86
 S Class    26.505542     3340.75

[13119 rows x 10 columns]
```

In [39]:
```
df_index = df.reset_index(drop = True)
print(df_index)
```

```
            model  year  price transmission  mileage fuelType  tax   mpg
\
0             SLK  2005   5200    Automatic    63000   Petrol  325  32.1
1         S Class  2017  34948    Automatic    27000   Hybrid   20  61.4
2        SL CLASS  2016  49948    Automatic     6200   Petrol  555  28.0
3         G Class  2016  61948    Automatic    16000   Petrol  325  30.4
4         G Class  2016  73948    Automatic     4000   Petrol  325  30.1
...           ...   ...    ...          ...      ...      ...  ...   ...
13114     C Class  2020  35999    Automatic      500   Diesel  145  55.4
13115     B Class  2020  24699    Automatic     2500   Diesel  145  55.4
13116   GLC Class  2019  30999    Automatic    11612   Diesel  145  41.5
13117   CLS Class  2019  37990    Automatic     2426   Diesel  145  45.6
13118     S Class  2019  54999    Automatic     2075   Diesel  145  52.3

       engineSize       prop  mileage_km
0             1.8   0.082540   101430.00
1             2.1   1.294370    43470.00
2             5.5   8.056129     9982.00
3             4.0   3.871750    25760.00
4             4.0  18.487000     6440.00
...           ...        ...         ...
13114         2.0  71.998000      805.00
13115         2.0   9.879600     4025.00
13116         2.1   2.669566    18695.32
13117         2.0  15.659522     3905.86
13118         2.9  26.505542     3340.75

[13119 rows x 11 columns]
```

In [40]:
```
df_index.head()
```

Out[40]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize	prop
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8	0.082540
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1	1.294370
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5	8.056129
3	G Class	2016	61948	Automatic	16000	Petrol	325	30.4	4.0	3.871750
4	G Class	2016	73948	Automatic	4000	Petrol	325	30.1	4.0	18.487000

In [41]:
```
mods = [' B Class', ' A Class']

print(df[df['model'].isin(mods)])
```

```
          model  year  price transmission  mileage fuelType  tax   mpg  \
10      A Class  2020  32980    Automatic      606   Petrol  145  35.8
11      A Class  2020  26980    Automatic     1000   Diesel  145  62.8
12      B Class  2019  23750    Automatic      278   Diesel  145  55.4
15      B Class  2017  15890    Automatic    24841   Diesel  150  68.9
18      B Class  2017  17513    Automatic    14128   Diesel  150  68.9
...         ...   ...    ...          ...      ...      ...  ...   ...
13073   B Class  2019  19350    Automatic    13674   Petrol  145  45.6
13090   A Class  2015  12995    Automatic    34665   Diesel   20  70.6
13109   A Class  2017  16499    Automatic    20477   Diesel  145  68.9
13111   B Class  2019  19344    Automatic    15257   Petrol  145  45.6
13115   B Class  2020  24699    Automatic     2500   Diesel  145  55.4

       engineSize       prop  mileage_km
10            2.0  54.422442      975.66
11            1.5  26.980000     1610.00
12            2.0  85.431655      447.58
15            1.5   0.639668    39994.01
18            2.1   1.239595    22746.08
...           ...        ...         ...
13073         1.3   1.415094    22015.14
13090         1.5   0.374874    55810.65
13109         2.1   0.805733    32967.97
13111         1.3   1.267877    24563.77
13115         2.0   9.879600     4025.00

[3152 rows x 11 columns]
```

In [42]:
```
df_index = df.set_index(["model", "price"])
```

In [43]:
```
rows_to_keep = [(' A Class', 32980), (" B Class", 19350)]
```

In [44]:
```
print(df_index.loc[rows_to_keep])
```

```
                year transmission  mileage fuelType  tax   mpg  engineSiz
e  \
model    price
 A Class 32980  2020    Automatic      606   Petrol  145  35.8         2.
0
 B Class 19350  2019    Automatic    13674   Petrol  145  45.6         1.
3

                     prop  mileage_km
model    price
 A Class 32980  54.422442      975.66
 B Class 19350   1.415094    22015.14
```

In [45]:
```
print(df_index.sort_index(level = "model"))
```

```
                year transmission  mileage fuelType  tax   mpg  engineSiz
e  \
model    price
 A Class 650    2003       Manual   109090   Petrol  235  40.0         1.
4
         1350   2010       Manual   116126   Diesel  145  54.3         2.
0
         1800   2007    Automatic    84000   Petrol  200  42.8         1.
5
         2478   2007    Automatic    81000   Diesel  160  49.6         2.
0
         2795   2007       Manual    79485   Petrol  200  45.6         1.
5
...              ...          ...      ...      ...  ...   ...
...
 X-CLASS 47400  2019    Automatic     3789   Diesel  260  31.4         3.
0
180      10799  2014       Manual    39240   Diesel   20  72.4         1.
5
200      19495  2018       Manual    15113   Diesel  145  64.2         2.
2
220      19995  2018    Automatic     9814   Diesel  145  61.4         2.
0
230      4500   2007    Automatic    94000   Petrol  520  29.4         0.
0

                     prop  mileage_km
model    price
 A Class 650     0.005958   175634.90
         1350    0.011625   186962.86
         1800    0.021429   135240.00
         2478    0.030593   130410.00
         2795    0.035164   127970.85
...                   ...         ...
 X-CLASS 47400  12.509897     6100.29
180      10799   0.275204    63176.40
200      19495   1.289949    24331.93
220      19995   2.037396    15800.54
230      4500    0.047872   151340.00

[13119 rows x 9 columns]
```

In [46]:
```
print(df_index.sort_index(level = ["model", "price"],
     ascending = [True, False]))
```

```
                year transmission  mileage fuelType  tax   mpg  engineSi
ze  \
model    price
 A Class 140319  2019    Semi-Auto      785   Petrol  150  22.1
4.0
         139559  2019    Semi-Auto     1000   Petrol  145  22.1
4.0
         138439  2020    Semi-Auto     1000   Petrol  145  22.1
4.0
         134219  2019    Semi-Auto     1000   Petrol  145  24.8
4.0
         129990  2019    Automatic     1000   Petrol  145  24.8
4.0
...               ...          ...      ...      ...  ...   ...
...
 X-CLASS 18991   2018       Manual    14243   Diesel  260  37.2
2.3
180      10799   2014       Manual    39240   Diesel   20  72.4
1.5
200      19495   2018       Manual    15113   Diesel  145  64.2
2.2
220      19995   2018    Automatic     9814   Diesel  145  61.4
2.0
230      4500    2007    Automatic    94000   Petrol  520  29.4
0.0

                      prop  mileage_km
model    price
 A Class 140319  178.750318     1263.85
         139559  139.559000     1610.00
         138439  138.439000     1610.00
         134219  134.219000     1610.00
         129990  129.990000     1610.00
...                     ...         ...
 X-CLASS 18991     1.333357    22931.23
180      10799     0.275204    63176.40
200      19495     1.289949    24331.93
220      19995     2.037396    15800.54
230      4500      0.047872   151340.00

[13119 rows x 9 columns]
```

In [47]:
```
df_sorted = df.sort_index()
df_sorted
```

Out[47]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize	p
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8	0.082
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1	1.294
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5	8.056
3	G Class	2016	61948	Automatic	16000	Petrol	325	30.4	4.0	3.871
4	G Class	2016	73948	Automatic	4000	Petrol	325	30.1	4.0	18.487
...	...	...	...	...	...	...	...	...	...	
13114	C Class	2020	35999	Automatic	500	Diesel	145	55.4	2.0	71.998
13115	B Class	2020	24699	Automatic	2500	Diesel	145	55.4	2.0	9.879
13116	GLC Class	2019	30999	Automatic	11612	Diesel	145	41.5	2.1	2.669
13117	CLS Class	2019	37990	Automatic	2426	Diesel	145	45.6	2.0	15.659
13118	S Class	2019	54999	Automatic	2075	Diesel	145	52.3	2.9	26.505

13119 rows × 11 columns

In [50]:
```
print(df_sorted.loc["2005":"2009"])
```

```
         model  year  price transmission  mileage fuelType  tax   mpg  \
2005   C Class  2015  16592    Semi-Auto    25230   Diesel   20  64.2
2006   A Class  2019  22702    Semi-Auto    11615   Diesel  145  67.3
2007   C Class  2018  25882    Automatic    12145   Diesel  145  58.9
2008   C Class  2019  29046    Automatic     5971   Diesel  145  61.4
2009   A Class  2019  20096    Automatic     7032   Diesel  145  68.9

      engineSize      prop  mileage_km
2005         2.1  0.657630    40620.30
2006         1.5  1.954542    18700.15
2007         2.1  2.131083    19553.45
2008         2.0  4.864512     9613.31
2009         1.5  2.857793    11321.52
```

In [58]:
```
print(df.loc[(df['price'] > 150000) | (df['price'] < 1000)])
```

```
          model  year   price transmission  mileage fuelType  tax   mpg
\
6199    G Class  2020  159999    Semi-Auto     1350   Petrol  145  21.4
10044   G Class  2020  154998    Automatic     3000   Petrol  150  21.4
11816   A Class  2003     650       Manual   109090   Petrol  235  40.0

       engineSize        prop  mileage_km
6199          4.0  118.517778      2173.5
10044         4.0   51.666000      4830.0
11816         1.4    0.005958    175634.9
```

In [60]:
```
type(df['year'])
```

Out[60]: pandas.core.series.Series

In [68]:
```
df[(df['year'] >= 2020) | (df['year'] < 2010)]
```

Out[68]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize	p
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8	0.082
10	A Class	2020	32980	Automatic	606	Petrol	145	35.8	2.0	54.422
11	A Class	2020	26980	Automatic	1000	Diesel	145	62.8	1.5	26.980
33	A Class	2020	28128	Automatic	2563	Petrol	145	47.9	1.3	10.974
38	C Class	2020	31790	Automatic	3500	Petrol	145	64.2	2.0	9.082
...	...	...	...	...	...	...	...	...	...	
13053	A Class	2020	27995	Automatic	2500	Diesel	145	67.3	1.5	11.198
13056	GLE Class	2020	58450	Automatic	1159	Diesel	150	32.8	2.9	50.431
13113	C Class	2020	30999	Automatic	1000	Diesel	145	64.2	2.0	30.999
13114	C Class	2020	35999	Automatic	500	Diesel	145	55.4	2.0	71.998
13115	B Class	2020	24699	Automatic	2500	Diesel	145	55.4	2.0	9.879

840 rows × 11 columns

In [70]:
```
df.set_index('year').sort_index()
```

Out[70]:

year	model	price	transmission	mileage	fuelType	tax	mpg	engineSize	prop	m
1970	M Class	24999	Automatic	14000	Diesel	305	39.2	0.0	1.785643	
1997	SL CLASS	9995	Automatic	76000	Petrol	255	28.2	3.2	0.131513	
1998	S Class	19990	Automatic	43534	Petrol	265	23.3	6.0	0.459181	
1998	SLK	1990	Automatic	113557	Petrol	265	32.1	2.3	0.017524	
1998	SL CLASS	10450	Automatic	49000	Petrol	255	28.2	3.2	0.213265	
...	...	...	...	...	...	...	...	...	...	
2020	GLA Class	21991	Semi-Auto	1695	Petrol	145	44.8	1.6	12.974041	
2020	C Class	28590	Semi-Auto	1501	Diesel	145	54.3	1.6	19.047302	
2020	A Class	24109	Semi-Auto	1000	Diesel	145	67.3	1.5	24.109000	
2020	GLC Class	49980	Semi-Auto	3999	Petrol	145	26.4	3.0	12.498125	
2020	C Class	33800	Semi-Auto	2406	Petrol	145	47.1	1.5	14.048213	

13119 rows × 10 columns

In [73]:
```
df.reset_index(drop = 1True )
```

Out[73]:

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize	p
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8	0.082
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1	1.294
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5	8.056
3	G Class	2016	61948	Automatic	16000	Petrol	325	30.4	4.0	3.871
4	G Class	2016	73948	Automatic	4000	Petrol	325	30.1	4.0	18.487
...	...	...	...	...	...	...	...	...	...	
13114	C Class	2020	35999	Automatic	500	Diesel	145	55.4	2.0	71.998
13115	B Class	2020	24699	Automatic	2500	Diesel	145	55.4	2.0	9.879
13116	GLC Class	2019	30999	Automatic	11612	Diesel	145	41.5	2.1	2.669
13117	CLS Class	2019	37990	Automatic	2426	Diesel	145	45.6	2.0	15.659
13118	S Class	2019	54999	Automatic	2075	Diesel	145	52.3	2.9	26.505

13119 rows × 11 columns

In [74]:
```
print(df.loc[2018:2078])
```

```
          model  year  price transmission  mileage fuelType  tax   mpg
\
2018   GLC Class  2017  26000    Semi-Auto    16895   Diesel  145  56.5
2019    CL Class  2017  18300    Automatic    39962   Diesel  125  58.9
2020     C Class  2017  25232    Automatic    15104   Diesel  145  58.9
2021     E Class  2017  23072    Semi-Auto    16931   Diesel  145  65.7
2022    SL CLASS  2019  30182    Automatic     3988   Petrol  145  47.1
...          ...   ...    ...          ...      ...      ...  ...   ...
2074     C Class  2016  18211    Semi-Auto    16206   Diesel   30  64.2
2075     A Class  2016  13891    Semi-Auto    36046   Petrol  125  53.3
2076     E Class  2018  28372    Semi-Auto     7410   Diesel  145  70.6
2077   GLA Class  2017  21182    Automatic    10827   Diesel  125  56.5
2078     C Class  2019  49780    Automatic     5654   Petrol  145  28.5

      engineSize      prop  mileage_km
2018         2.1  1.538917    27200.95
2019         2.1  0.457935    64338.82
2020         2.1  1.670551    24317.44
2021         2.0  1.362707    27258.91
2022         2.0  7.568205     6420.68
...          ...       ...         ...
2074         2.1  1.123720    26091.66
2075         1.6  0.385369    58034.06
2076         2.0  3.828880    11930.10
2077         2.1  1.956405    17431.47
2078         4.0  8.804386     9102.94

[61 rows x 11 columns]
```

In [75]:
```
print(df.iloc[1, 2])
```

```
34948
```

In [76]:
```
print(df.iloc[:5])
```

```
       model  year  price transmission  mileage fuelType  tax   mpg  \
0        SLK  2005   5200    Automatic    63000   Petrol  325  32.1
1    S Class  2017  34948    Automatic    27000   Hybrid   20  61.4
2   SL CLASS  2016  49948    Automatic     6200   Petrol  555  28.0
3    G Class  2016  61948    Automatic    16000   Petrol  325  30.4
4    G Class  2016  73948    Automatic     4000   Petrol  325  30.1

   engineSize      prop  mileage_km
0         1.8  0.082540    101430.0
1         2.1  1.294370     43470.0
2         5.5  8.056129      9982.0
3         4.0  3.871750     25760.0
4         4.0 18.487000      6440.0
```

In [78]:
```
print(df.iloc[:10, 1:3])
```

```
   year   price
0  2005    5200
1  2017   34948
2  2016   49948
3  2016   61948
4  2016   73948
5  2011  149948
6  2018   30948
7  2012   10948
8  2019  139948
9  2017   19750
```

In [90]:
```
df_piv_tab = df.pivot_table("price", index = ['transmission', 'fuelType'],
```

In [93]:
```
df_piv_tab.loc[:, 2012:2015]
```

Out[93]:

	year	2012	2013	2014	2015
transmission	**fuelType**				
Automatic	**Diesel**	10983.894737	12459.371901	13782.443114	16318.797872
	Hybrid	NaN	16472.666667	10500.000000	15928.045455
	Other	NaN	NaN	NaN	15000.000000
	Petrol	11376.285714	12674.807692	17145.057143	23738.627119
Manual	**Diesel**	8648.000000	9748.468085	11057.000000	12269.543307
	Petrol	7786.500000	10458.294118	12092.000000	13981.638889
Other	**Diesel**	NaN	NaN	NaN	12995.000000
	Other	NaN	NaN	NaN	NaN
Semi-Auto	**Diesel**	NaN	11931.101695	14397.318750	17528.307692
	Hybrid	NaN	NaN	12995.000000	17405.333333
	Petrol	NaN	13683.315789	18889.718750	21148.210526

In [95]:
```
mean_price_by_year = df_piv_tab.mean()
mean_price_by_year
```

Out[95]:
```
year
1970    24999.000000
1997     9995.000000
1998     8605.000000
1999     5995.000000
2000     5743.333333
2001     4015.583333
2002     3928.062500
2003     3350.000000
2004     4102.750000
2005     4420.250000
2006     4454.444444
2007     4488.382222
2008     8489.541667
2009     5556.743590
2010     7372.537500
2011    11419.052174
2012     9698.670113
2013    12489.717992
2014    13857.317220
2015    16631.350419
2016    19625.718292
2017    20988.225112
2018    24011.888717
2019    31748.876171
2020    34886.544375
dtype: float64
```

In [96]:
```
mean_price_by_year[mean_price_by_year == mean_price_by_year.max()]
```

Out[96]:
```
year
2020    34886.544375
dtype: float64
```

In [97]:
```
mean_price_by_year.max()
```

Out[97]:
```
34886.544374583755
```

In [ ]:
```

```

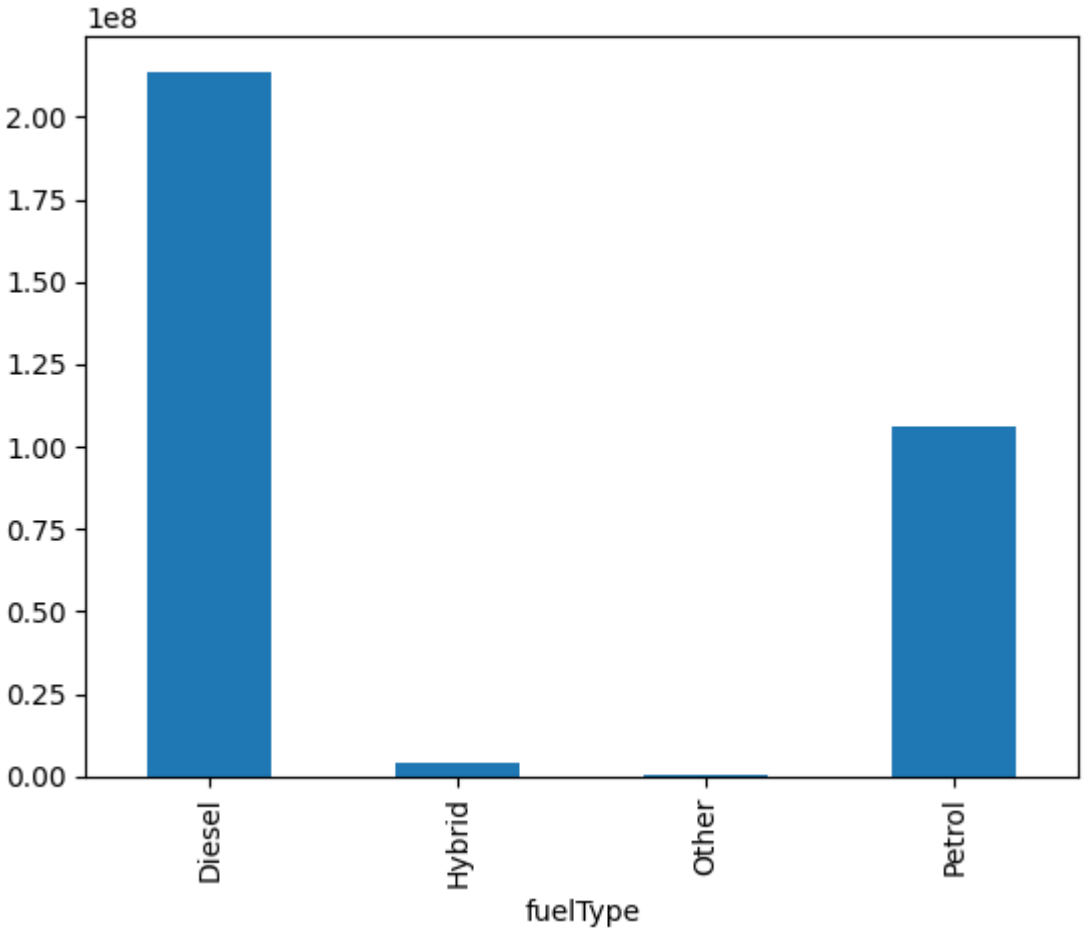
In [ ]:
```

```

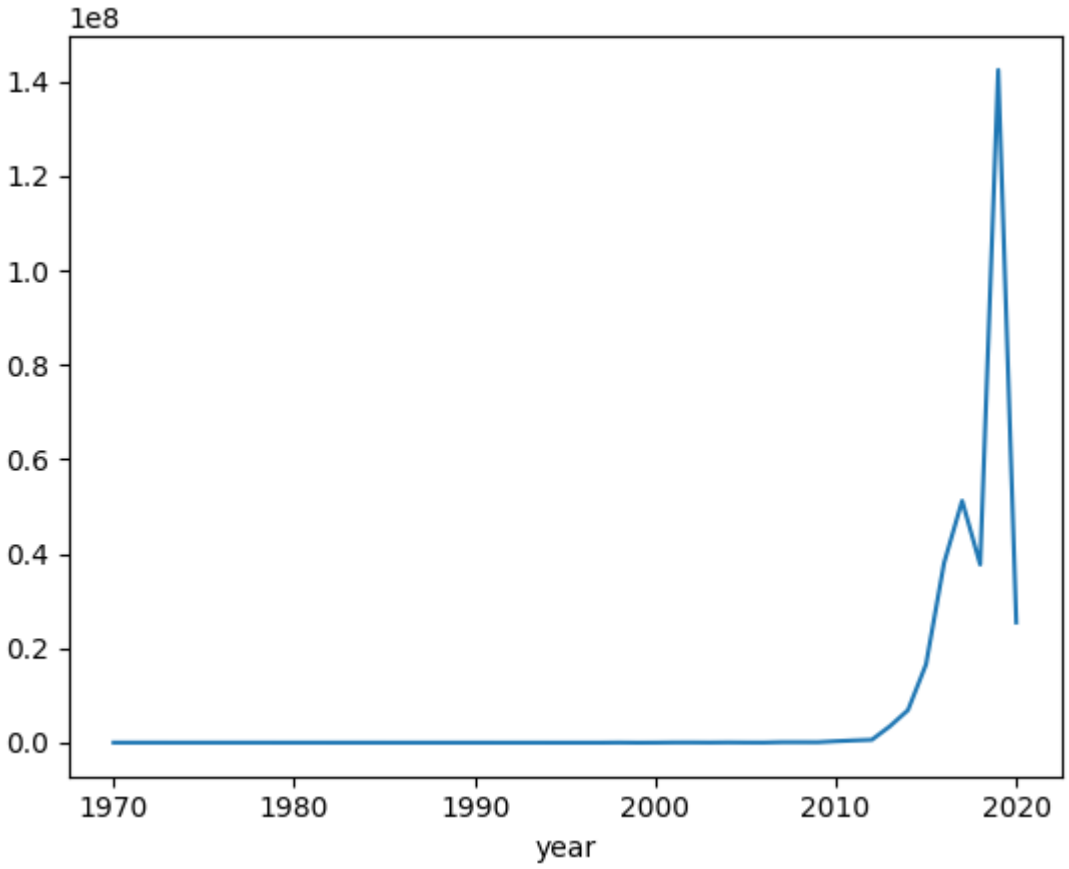
In [100]:
```
import matplotlib.pyplot as plt
```

In [102]:
```
plot1 = df.groupby('fuelType')['price'].sum()
```

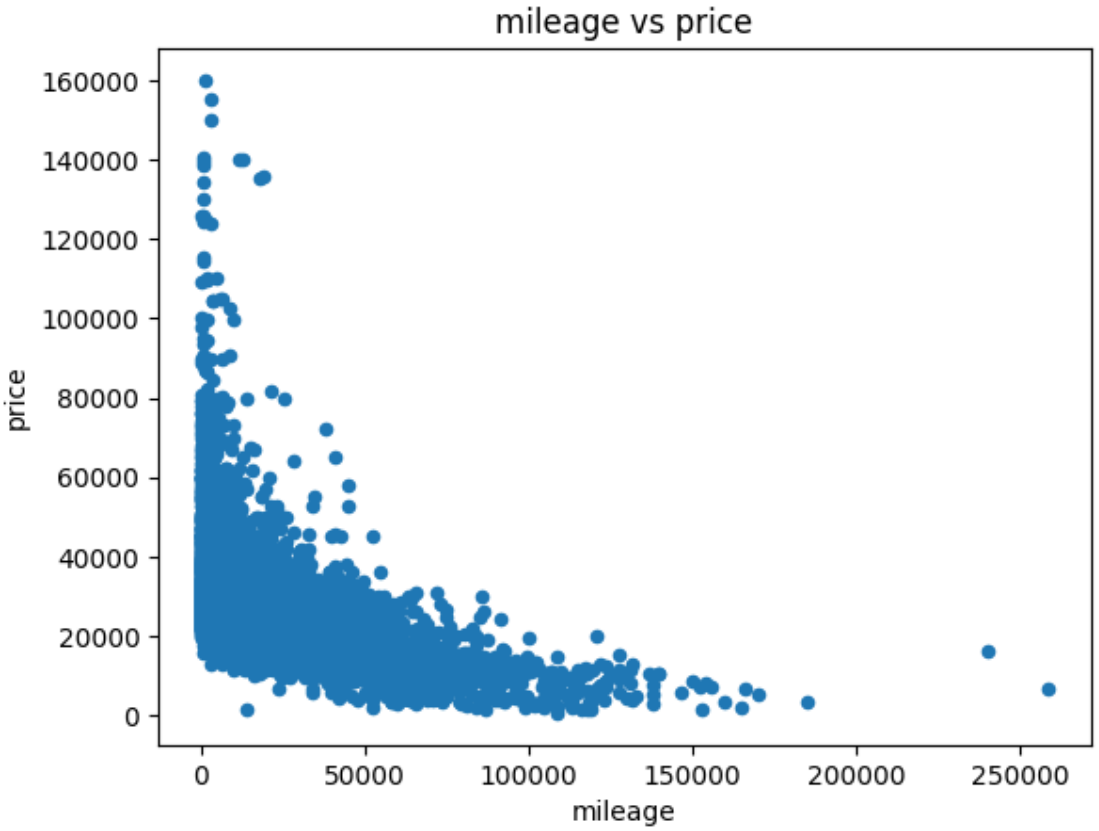
In [103]:
```
plot1.plot(kind = 'bar')
plt.show()
```



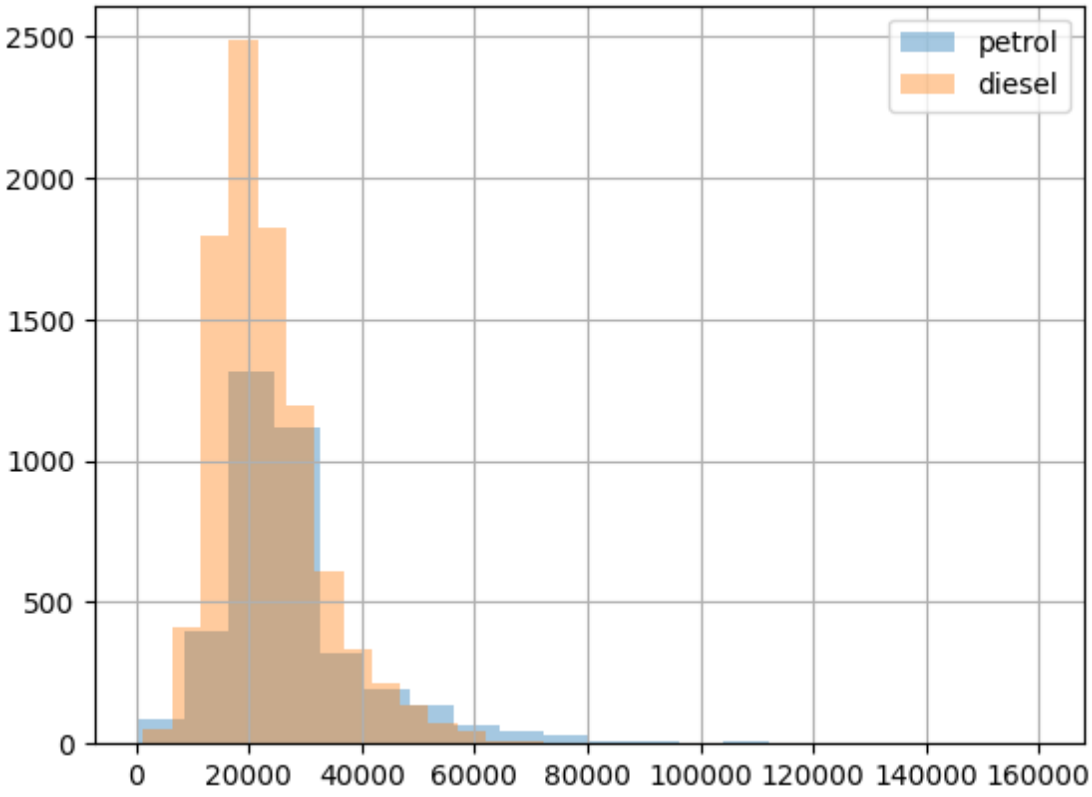
In [104]:
```
plot2 = df.groupby('year')['price'].sum()
plot2.plot(kind = 'line')
plt.show()
```



In [121]:
```
df.plot(x = 'mileage', y='price', kind = 'scatter',title = 'mileage vs pri
plt.show()
```



In [114]:
```
df[df['fuelType'] == 'Petrol']['price'].hist(alpha=0.4, bins = 20)
df[df['fuelType'] == 'Diesel']['price'].hist(alpha=0.4, bins = 20)
plt.legend(['petrol', 'diesel'])
plt.show()
```



In [ ]:
```
```