Comparison of Prices through Web scraping

```python
In [1]:  import bs4, requests
         import re
         import pandas as pd
         import numpy as np
         import json

         from re import sub
         from decimal import Decimal
         import matplotlib.pyplot as plt
```

```python
In [2]:  url='https://www.cargurus.com/Cars/inventorylisting/viewDetailsFilterViewInventoryl
         res = requests.get(url)
```

```python
In [3]:  soup = bs4.BeautifulSoup(res.text)
         #print(soup.prettify())
```

```python
In [4]:  #find data in dictionary format
         dataset = soup.find_all("script")
         dictionary = []
         for data in dataset:
             #print(type(data))
             #print(data)
             if 'PREFLIGHT' in str(data):
                 dictionary.append(data)
                 #print('found')
                 break

         data_str = str(dictionary)
         dict_object = json.loads(re.search('({.+})', data_str).group(0).replace("'", '"'))
         print(type(dict_object))
         #print(dict_object)
```

```
<class 'dict'>
```

```python
In [5]:  #for key, value in dict_object.items():
             #print(key)

         listings = dict_object["listings"]
         #print(type(listings))
         #print(listings)
         output_frame = pd.DataFrame(columns = ['Title','Price','Mileage', 'Year'])
         for car in listings:
             year = car['carYear']
             title = car['listingTitle']
             price = car['price']
             mileage = car['mileage']
             output_frame.loc[len(output_frame.index)] = [title, price, mileage, year]

         output_frame
```

Out[5]:

	Title	Price	Mileage	Year
0	2018 Audi Q7 3.0 TFSI quattro Premium Plus AWD	25995.0	70362	2018
1	2019 Audi Q7 quattro Progressiv 45 TFSI AWD	34050.0	32160	2019
2	2019 Audi Q7 2.0T quattro Premium Plus AWD	41000.0	9957	2019
3	2019 Audi Q7 2.0T quattro Premium Plus AWD	36545.0	35338	2019
4	2019 Audi Q7 3.0T quattro Premium Plus AWD	38995.0	34021	2019
5	2015 Audi Q7 3.0T quattro Premium Plus AWD	17500.0	79679	2015
6	2019 Audi Q7 3.0T quattro Prestige AWD	44900.0	20464	2019
7	2018 Audi Q7 3.0 TFSI quattro Premium Plus AWD	26203.0	80105	2018
8	2018 Audi Q7 2.0 TFSI quattro Premium Plus AWD	32143.0	44143	2018
9	2015 Audi Q7 3.0T quattro Premium Plus AWD	16990.0	91012	2015
10	2015 Audi Q7 3.0T quattro Premium Plus AWD	14995.0	112347	2015
11	2019 Audi Q7 quattro Premium 55 TFSI AWD	36500.0	45260	2019
12	2019 Audi Q7 2.0T quattro Premium Plus AWD	38545.0	32550	2019
13	2019 Audi Q7 3.0T quattro Premium Plus AWD	42745.0	20672	2019
14	2020 Audi Q7 quattro Premium 45 TFSI AWD	42798.0	42223	2020

In [6]:
```python
Average_price_point_Boston= output_frame['Price'].mean()
Average_price_point_Boston
```

Out[6]:
```
32660.266666666666
```

In [7]:
```python
plt.scatter(output_frame['Year'], output_frame['Price'], color='red', marker='o')
plt.title('price of Audi Q7 in Boston', fontsize=14)
plt.xlabel('year', fontsize=14)
plt.ylabel('price', fontsize=14)
plt.grid(True)
plt.show()
```

price of Audi Q7 in Boston
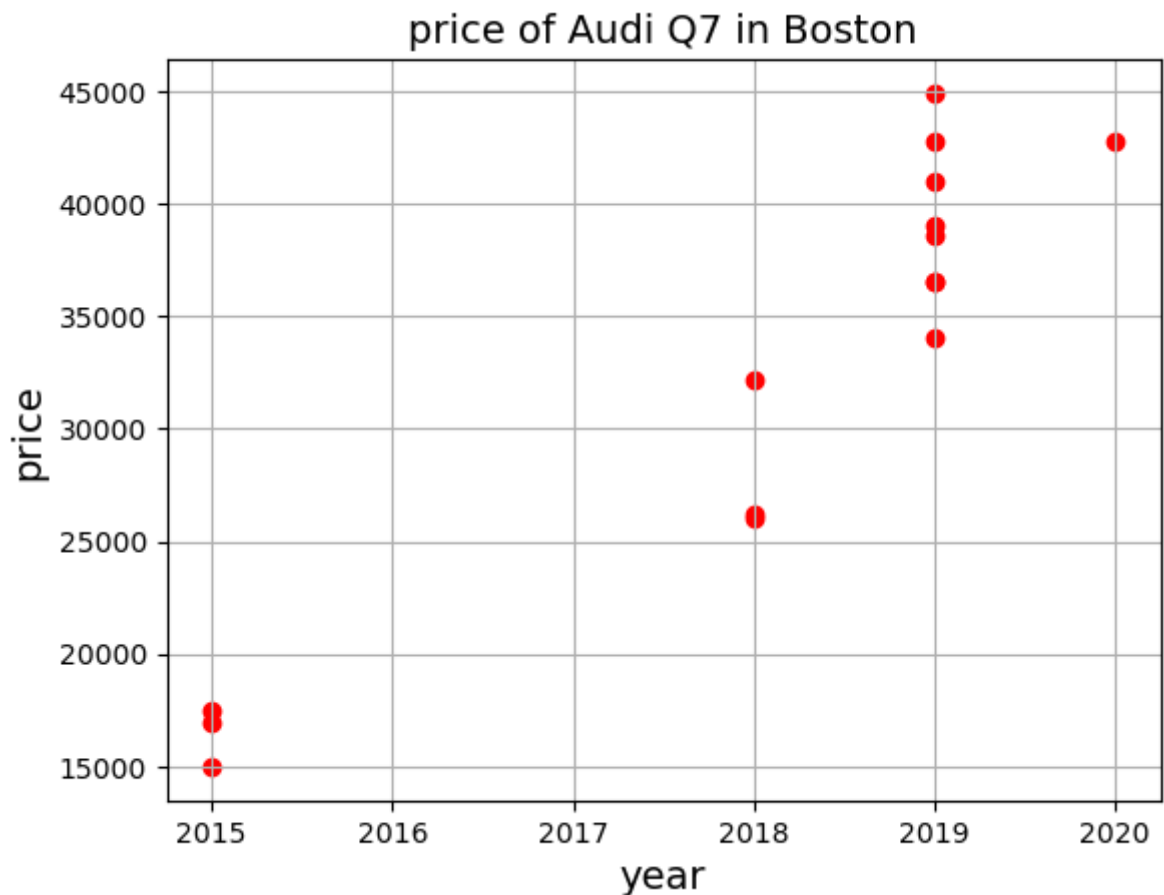

```
In [8]:  output_frame['Year'] = pd.to_numeric(output_frame['Year'])
         output_frame['Price'] = pd.to_numeric(output_frame['Price'])
         output_frame['Price'].corr(output_frame['Year'])
         output_frame.corr()
```

Out[8]:

	Price	Mileage	Year
Price	1.000000	-0.936563	0.931330
Mileage	-0.936563	1.000000	-0.843712
Year	0.931330	-0.843712	1.000000

```
In [9]:   url2='https://www.cargurus.com/Cars/inventorylisting/viewDetailsFilterViewInventory
          res = requests.get(url2)
```

```
In [10]:  soup = bs4.BeautifulSoup(res.text)
          #print(soup.prettify())
```

```
In [11]:  #find data in dictionary format
          dataset = soup.find_all("script")
          dictionary = []
          for data in dataset:
              #print(type(data))
              #print(data)
              if 'PREFLIGHT' in str(data):
                  dictionary.append(data)
                  #print('found')
                  break

          data_str = str(dictionary)
          dict_object = json.loads(re.search('({.+})', data_str).group(0).replace("'", '"'))
```

```
print(type(dict_object))
#print(dict_object)
```

```
<class 'dict'>
```

In [12]:
```
#for key, value in dict_object.items():
    #print(key)

listings = dict_object["listings"]
#print(type(listings))
#print(listings)
output_frame = pd.DataFrame(columns = ['Title','Price','Mileage', 'Year'])
for car in listings:
    year = car['carYear']
    title = car['listingTitle']
    price = car['price']
    mileage = car['mileage']
    output_frame.loc[len(output_frame.index)] = [title, price, mileage, year]

output_frame
```

Out[12]:

	Title	Price	Mileage	Year
0	2014 Audi Q7 3.0T quattro S-Line Prestige AWD	13995.0	114155	2014
1	2018 Audi Q7 2.0 TFSI quattro Premium Plus AWD	29372.0	52235	2018
2	2018 Audi Q7 3.0 TFSI quattro Prestige AWD	28000.0	81270	2018
3	2022 Audi Q7 quattro Premium Plus 45 TFSI AWD	55185.0	11629	2022
4	2017 Audi Q7 3.0T quattro Prestige AWD	28799.0	61768	2017
5	2017 Audi Q7 2.0T quattro Premium Plus AWD	24380.0	73144	2017
6	2019 Audi Q7 2.0T quattro SE Premium Plus AWD	32275.0	56497	2019
7	2019 Audi Q7 3.0T quattro SE Premium Plus AWD	34980.0	44336	2019
8	2013 Audi Q7 3.0T quattro S-Line Prestige AWD	13977.0	119478	2013
9	2019 Audi Q7 2.0T quattro SE Premium Plus AWD	32275.0	58166	2019
10	2017 Audi Q7 3.0T quattro Prestige AWD	25699.0	89123	2017
11	2011 Audi Q7 3.0 TDI quattro Premium Plus AWD	12790.5	108873	2011
12	2019 Audi Q7 2.0T quattro Premium Plus AWD	34334.6	51824	2019
13	2018 Audi Q7 3.0 TFSI quattro Prestige AWD	26989.0	88182	2018
14	2013 Audi Q7 3.0 TDI quattro Prestige AWD	16971.0	100918	2013

In [13]:
```
Average_price_point_Columbus=output_frame['Price'].mean()
Average_price_point_Columbus
```
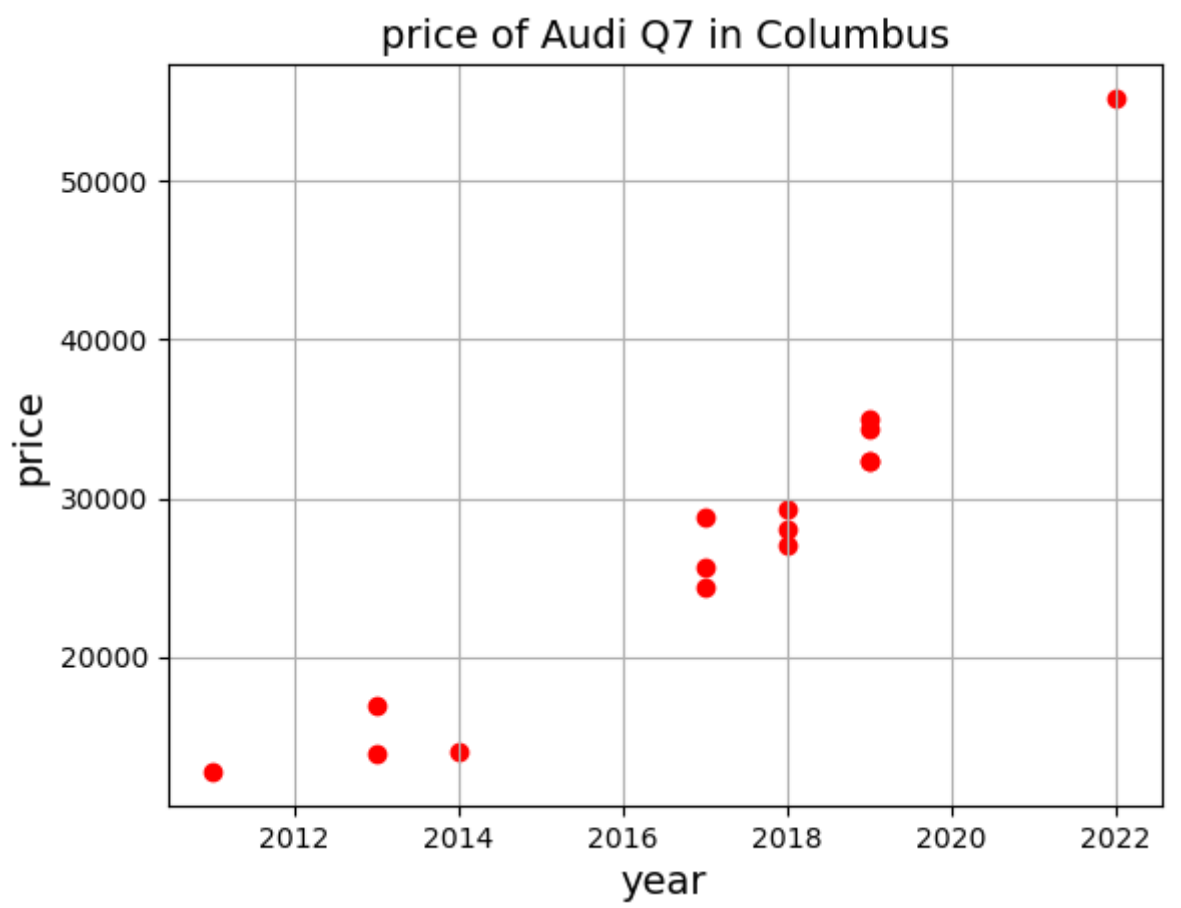
Out[13]:
```
27334.806666666664
```

In [14]:
```
output_frame['Year'] = pd.to_numeric(output_frame['Year'])
output_frame['Price'] = pd.to_numeric(output_frame['Price'])
output_frame['Price'].corr(output_frame['Year'])
output_frame.corr()
```

Out[14]:

	Price	Mileage	Year
Price	1.000000	-0.951627	0.931124
Mileage	-0.951627	1.000000	-0.894382
Year	0.931124	-0.894382	1.000000

In [15]:
```python
plt.scatter(output_frame['Year'], output_frame['Price'], color='red', marker='o')
plt.title('price of Audi Q7 in Columbus', fontsize=14)
plt.xlabel('year', fontsize=14)
plt.ylabel('price', fontsize=14)
plt.grid(True)
plt.show()
```


In [ ]: