

```
In [47]: #To import pandas library
import pandas as pd
import numpy as np
```

```
In [60]: # To access dataset into dataframe
datav1= pd.read_csv('vehicles.csv')
```

```
In [61]: #Display column names
datav1.columns
```

```
Out[61]: Index(['id', 'url', 'region', 'region_url', 'price', 'year', 'manufacturer',
   ... 'model', 'condition', 'cylinders', 'fuel', 'odometer', 'title_status',
   ... 'transmission', 'VIN', 'drive', 'size', 'type', 'paint_color',
   ... 'image_url', 'description', 'county', 'state', 'lat', 'long',
   ... 'posting_date'],
   ... dtype='object')
```

```
In [62]: #copy first 50 rows into new dataframe datav2
datav2= datav1.head(50)
```

```
In [63]: # Display data of dataframe
datav2
```

	id	url	region	region_1
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...	prescott	https://prescott.craigslist.c
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...	fayetteville	https://fayar.craigslist.c
2	7221797935	https://keys.craigslist.org/cto/d/summerland-k...	florida keys	https://keys.craigslist.c
3	7222270760	https://worcester.craigslist.org/cto/d/west-br...	worcester / central MA	https://worcester.craigslist.c
4	7210384030	https://greensboro.craigslist.org/cto/d/trinit...	greensboro	https://greensboro.craigslist.c
5	7222379453	https://hudsonvalley.craigslist.org/cto/d/west...	hudson valley	https://hudsonvalley.craigslist.c
6	7221952215	https://hudsonvalley.craigslist.org/cto/d/west...	hudson valley	https://hudsonvalley.craigslist.c
7	7220195662	https://hudsonvalley.craigslist.org/cto/d/pough...	hudson valley	https://hudsonvalley.craigslist.c
8	7209064557	https://medford.craigslist.org/cto/d/grants-pa...	medford-ashland	https://medford.craigslist.c
9	7219485069	https://erie.craigslist.org/cto/d/erie-2012-su...	erie	https://erie.craigslist.c
10	7218893038	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso	https://elpaso.craigslist.c
11	7218325704	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso	https://elpaso.craigslist.c
12	7217788283	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso	https://elpaso.craigslist.c
13	7217147606	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso	https://elpaso.craigslist.c
14	7209027818	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso	https://elpaso.craigslist.c
15	7223509794	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham	https://bellingham.craigslist.c

id		url	region	region_i
16	7222753076	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham	https://bellingham.craigslist.c
17	7222206015	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham	https://bellingham.craigslist.c
18	7220030122	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham	https://bellingham.craigslist.c
19	7218423006	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham	https://bellingham.craigslist.c
20	7216672204	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham	https://bellingham.craigslist.c
21	7215617048	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham	https://bellingham.craigslist.c
22	7213839225	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham	https://bellingham.craigslist.c
23	7208549803	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham	https://bellingham.craigslist.c
24	7213843538	https://skagit.craigslist.org/ctd/d/bellingham...	skagit / island / SJI	https://skagit.craigslist.c
25	7212631321	https://skagit.craigslist.org/ctd/d/bellingham...	skagit / island / SJI	https://skagit.craigslist.c
26	7219973522	https://lacrosse.craigslist.org/cto/d/la-cross...	la crosse	https://lacrosse.craigslist.c
27	7316814884	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn	https://auburn.craigslist.c
28	7316814758	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn	https://auburn.craigslist.c
29	7316814989	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn	https://auburn.craigslist.c
30	7316743432	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn	https://auburn.craigslist.c
31	7316356412	https://auburn.craigslist.org/cto/d/auburn-uni...	auburn	https://auburn.craigslist.c
32	7316343444	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn	https://auburn.craigslist.c
33	7316304717	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn	https://auburn.craigslist.c

	id		url	region	region_i
34	7316285779		https://auburn.craigslist.org/cto/d/aubum-201...	auburn	https://auburn.craigslist.c
35	7316257769	https://auburn.craigslist.org/ctd/d/aubum-uni...		auburn	https://aubum.craigslist.c
36	7316133914	https://auburn.craigslist.org/ctd/d/aubum-uni...		auburn	https://aubum.craigslist.c
37	7316130053	https://auburn.craigslist.org/cto/d/opelika-19...		auburn	https://aubum.craigslist.c
38	7315816316	https://auburn.craigslist.org/ctd/d/aubum-uni...		auburn	https://aubum.craigslist.c
39	7315770394	https://auburn.craigslist.org/ctd/d/aubum-uni...		auburn	https://aubum.craigslist.c
40	7315756348	https://auburn.craigslist.org/ctd/d/aubum-uni...		auburn	https://aubum.craigslist.c
41	7315715960	https://auburn.craigslist.org/ctd/d/aubum-uni...		auburn	https://aubum.craigslist.c
42	7315379459	https://auburn.craigslist.org/ctd/d/aubum-uni...		auburn	https://aubum.craigslist.c
43	7315275975	https://auburn.craigslist.org/ctd/d/aubum-uni...		auburn	https://aubum.craigslist.c
44	7315275942	https://auburn.craigslist.org/ctd/d/aubum-uni...		auburn	https://aubum.craigslist.c

	id		url	region	region_url
45	7315270785	https://auburn.craigslist.org/ctd/d/auburn-university... 999...		auburn	https://auburn.craigslist.c...
46	7315039307	https://auburn.craigslist.org/ctd/d/auburn-university... 999...		auburn	https://auburn.craigslist.c...
47	7314910156	https://auburn.craigslist.org/ctd/d/auburn-university... 999...		auburn	https://auburn.craigslist.c...
48	7314854462	https://auburn.craigslist.org/ctd/d/auburn-university... 999...		auburn	https://auburn.craigslist.c...
49	7314811916	https://auburn.craigslist.org/ctd/d/auburn-university... 999...		auburn	https://auburn.craigslist.c...

50 rows × 26 columns

◀	▶
---	---

In [65]:

```
## Multiple row and column selections using iloc and DataFrame
#The iloc indexer syntax is data.iloc[<row selection>, <column selection>]
datav2.iloc[0:5] # first five rows of dataframe
```

Out[65]:

	id		url	region	region_url	
0	7222695916	https://prescott.craigslist.org/cto/d/prescott... br...		prescott	https://prescott.craigslist.org	
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville... br...		fayetteville	https://fayar.craigslist.org	1
2	7221797935	https://keys.craigslist.org/cto/d/summerland-... k...		florida keys	https://keys.craigslist.org	2
3	7222270760	https://worcester.craigslist.org/cto/d/west-... br...	worcester / central MA		https://worcester.craigslist.org	
4	7210384030	https://greensboro.craigslist.org/cto/d/trinit... br...	greensboro		https://greensboro.craigslist.org	

5 rows × 26 columns

◀	▶
---	---

In [7]:

```
datav2.iloc[:, 0:2] # first two columns of data frame with all rows
```

Out[7]:

	id		url
0	7222695916		https://prescott.craigslist.org/cto/d/prescott... br...
1	7218891961		https://fayar.craigslist.org/ctd/d/bentonville... br...

id	url
2	https://keys.craigslist.org/cto/d/summerland-k...
3	https://worcester.craigslist.org/cto/d/west-br...
4	https://greensboro.craigslist.org/cto/d/trinit...
5	https://hudsonvalley.craigslist.org/cto/d/west...
6	https://hudsonvalley.craigslist.org/cto/d/west...
7	https://hudsonvalley.craigslist.org/cto/d/poug...
8	https://medford.craigslist.org/cto/d/grants-pa...
9	https://erie.craigslist.org/cto/d/erie-2012-su...
10	https://elpaso.craigslist.org/ctd/d/el-paso-20...
11	https://elpaso.craigslist.org/ctd/d/el-paso-20...
12	https://elpaso.craigslist.org/ctd/d/el-paso-20...
13	https://elpaso.craigslist.org/ctd/d/el-paso-20...
14	https://elpaso.craigslist.org/ctd/d/el-paso-20...
15	https://bellingham.craigslist.org/ctd/d/bellin...
16	https://bellingham.craigslist.org/ctd/d/bellin...
17	https://bellingham.craigslist.org/ctd/d/bellin...
18	https://bellingham.craigslist.org/ctd/d/bellin...
19	https://bellingham.craigslist.org/ctd/d/bellin...
20	https://bellingham.craigslist.org/ctd/d/bellin...
21	https://bellingham.craigslist.org/ctd/d/bellin...
22	https://bellingham.craigslist.org/ctd/d/bellin...
23	https://bellingham.craigslist.org/ctd/d/bellin...
24	https://skagit.craigslist.org/ctd/d/bellingham...
25	https://skagit.craigslist.org/ctd/d/bellingham...
26	https://lacrosse.craigslist.org/cto/d/la-cross...
27	https://auburn.craigslist.org/ctd/d/auburn-uni...
28	https://auburn.craigslist.org/ctd/d/auburn-uni...
29	https://auburn.craigslist.org/ctd/d/auburn-uni...
30	https://auburn.craigslist.org/ctd/d/auburn-uni...
31	https://auburn.craigslist.org/cto/d/auburn-uni...
32	https://auburn.craigslist.org/ctd/d/auburn-uni...
33	https://auburn.craigslist.org/ctd/d/auburn-uni...
34	https://auburn.craigslist.org/cto/d/auburn-201...
35	https://auburn.craigslist.org/ctd/d/auburn-uni...
36	https://auburn.craigslist.org/ctd/d/auburn-uni...
37	https://auburn.craigslist.org/cto/d/opelika-19...
38	https://auburn.craigslist.org/ctd/d/auburn-uni...

	id	url
39	7315770394	https://auburn.craigslist.org/ctd/d/auburn-university-area/7315770394.html
40	7315756348	https://auburn.craigslist.org/ctd/d/auburn-university-area/7315756348.html
41	7315715960	https://auburn.craigslist.org/ctd/d/auburn-university-area/7315715960.html
42	7315379459	https://auburn.craigslist.org/ctd/d/auburn-university-area/7315379459.html
43	7315275975	https://auburn.craigslist.org/ctd/d/auburn-university-area/7315275975.html
44	7315275942	https://auburn.craigslist.org/ctd/d/auburn-university-area/7315275942.html
45	7315270785	https://auburn.craigslist.org/ctd/d/auburn-university-area/7315270785.html
46	7315039307	https://auburn.craigslist.org/ctd/d/auburn-999.../7315039307.html
47	7314910156	https://auburn.craigslist.org/ctd/d/auburn-university-area/7314910156.html
48	7314854462	https://auburn.craigslist.org/ctd/d/auburn-university-area/7314854462.html
49	7314811916	https://auburn.craigslist.org/ctd/d/auburn-university-area/7314811916.html

In [8]: `datav2.iloc[[0,3,6,24], [0,5,6]] # 1st, 4th, 7th, 25th row + 1st 6th 7th columns.`

Out[8]:

	id	year	manufacturer
0	7222695916	NaN	NaN
3	7222270760	NaN	NaN
6	7221952215	NaN	NaN
24	7213843538	NaN	NaN

In [9]: `datav2.iloc[0:5, 5:8] # first 5 rows and 5th, 6th, 7th columns of data frame`

Out[9]:

	year	manufacturer	model
0	NaN	NaN	NaN
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN

In [10]: `#Display single column or to access single column
datav2['region']`

Out[10]:

0	prescott
1	fayetteville
2	florida keys
3	worcester / central MA
4	greensboro
5	hudson valley
6	hudson valley
7	hudson valley
8	medford-ashland
9	erie
10	el paso

```

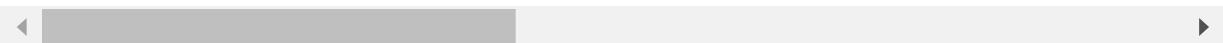
11          el paso
12          el paso
13          el paso
14          el paso
15          bellingham
16          bellingham
17          bellingham
18          bellingham
19          bellingham
20          bellingham
21          bellingham
22          bellingham
23          bellingham
24  skagit / island / SJI
25  skagit / island / SJI
26          la crosse
27          auburn
28          auburn
29          auburn
30          auburn
31          auburn
32          auburn
33          auburn
34          auburn
35          auburn
36          auburn
37          auburn
38          auburn
39          auburn
40          auburn
41          auburn
42          auburn
43          auburn
44          auburn
45          auburn
46          auburn
47          auburn
48          auburn
49          auburn
Name: region, dtype: object

```

In [11]: *#display single row using slicing*
datav2[0:2]

	id	url	region	region_url	price
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...	prescott	https://prescott.craigslist.org	600
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...	fayetteville	https://fayar.craigslist.org	1190

2 rows × 26 columns



In [52]: *# display or access contents of more than one column*
datav2[['id','url','region']]

	id	url	region
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...	prescott
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...	fayetteville

id	url	region
2	https://keys.craigslist.org/cto/d/summerland-k...	florida keys
3	https://worcester.craigslist.org/cto/d/west-br...	worcester / central MA
4	https://greensboro.craigslist.org/cto/d/trinit...	greensboro
5	https://hudsonvalley.craigslist.org/cto/d/west...	hudson valley
6	https://hudsonvalley.craigslist.org/cto/d/west...	hudson valley
7	https://hudsonvalley.craigslist.org/cto/d/poug...	hudson valley
8	https://medford.craigslist.org/cto/d/grants-pa...	medford-ashland
9	https://erie.craigslist.org/cto/d/erie-2012-su...	erie
10	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
11	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
12	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
13	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
14	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
15	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
16	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
17	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
18	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
19	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
20	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
21	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
22	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
23	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
24	https://skagit.craigslist.org/ctd/d/bellingham...	skagit / island / SJI
25	https://skagit.craigslist.org/ctd/d/bellingham...	skagit / island / SJI
26	https://lacrosse.craigslist.org/cto/d/la-cross...	la crosse
27	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
28	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
29	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
30	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
31	https://auburn.craigslist.org/cto/d/auburn-uni...	auburn
32	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
33	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
34	https://auburn.craigslist.org/cto/d/auburn-201...	auburn
35	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
36	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
37	https://auburn.craigslist.org/cto/d/opelika-19...	auburn
38	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn

	id	url	region
39	7315770394	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
40	7315756348	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
41	7315715960	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
42	7315379459	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
43	7315275975	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
44	7315275942	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
45	7315270785	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
46	7315039307	https://auburn.craigslist.org/ctd/d/auburn-999...	auburn
47	7314910156	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
48	7314854462	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
49	7314811916	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn

In [12]:

```
# another way to access contents of more than one column
list1 = ['id', 'url', 'region']
datav2[list1]
```

Out[12]:

	id	url	region
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...	prescott
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...	fayetteville
2	7221797935	https://keys.craigslist.org/cto/d/summerland-k...	florida keys
3	7222270760	https://worcester.craigslist.org/cto/d/west-br... worcester / central MA	worcester / central MA
4	7210384030	https://greensboro.craigslist.org/cto/d/trinit...	greensboro
5	7222379453	https://hudsonvalley.craigslist.org/cto/d/west...	hudson valley
6	7221952215	https://hudsonvalley.craigslist.org/cto/d/west...	hudson valley
7	7220195662	https://hudsonvalley.craigslist.org/cto/d/poug...	hudson valley
8	7209064557	https://medford.craigslist.org/cto/d/grants-pa...	medford-ashland
9	7219485069	https://erie.craigslist.org/cto/d/erie-2012-su...	erie
10	7218893038	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
11	7218325704	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
12	7217788283	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
13	7217147606	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
14	7209027818	https://elpaso.craigslist.org/ctd/d/el-paso-20...	el paso
15	7223509794	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
16	7222753076	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
17	7222206015	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
18	7220030122	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
19	7218423006	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
20	7216672204	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham

	id	url	region
21	7215617048	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
22	7213839225	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
23	7208549803	https://bellingham.craigslist.org/ctd/d/bellin...	bellingham
24	7213843538	https://skagit.craigslist.org/ctd/d/bellingham...	skagit / island / SJI
25	7212631321	https://skagit.craigslist.org/ctd/d/bellingham...	skagit / island / SJI
26	7219973522	https://lacrosse.craigslist.org/cto/d/la-cross...	la crosse
27	7316814884	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
28	7316814758	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
29	7316814989	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
30	7316743432	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
31	7316356412	https://auburn.craigslist.org/cto/d/auburn-uni...	auburn
32	7316343444	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
33	7316304717	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
34	7316285779	https://auburn.craigslist.org/cto/d/auburn-201...	auburn
35	7316257769	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
36	7316133914	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
37	7316130053	https://auburn.craigslist.org/cto/d/opelika-19...	auburn
38	7315816316	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
39	7315770394	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
40	7315756348	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
41	7315715960	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
42	7315379459	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
43	7315275975	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
44	7315275942	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
45	7315270785	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
46	7315039307	https://auburn.craigslist.org/ctd/d/auburn-999...	auburn
47	7314910156	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
48	7314854462	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn
49	7314811916	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn

In [13]:

```
# create new dataframe of 50 rows and three columns
datav3=datav2.iloc[0:50,0:3]
```

In [14]:

```
# enables us to select a particular cell of the dataset
datav2.iloc[0:8,0:4]
```

Out[14]:

	id	url	region	region_ue
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...	prescott	https://prescott.craigslist.or

	id		url	region	region_ui
1	7218891961		https://fayar.craigslist.org/ctd/d/bentonville...	fayetteville	https://fayar.craigslist.or...
2	7221797935		https://keys.craigslist.org/cto/d/summerland-...	florida keys	https://keys.craigslist.or...
3	7222270760		https://worcester.craigslist.org/cto/d/west-br...	worcester / central MA	https://worcester.craigslist.or...
4	7210384030		https://greensboro.craigslist.org/cto/d/trinit...	greensboro	https://greensboro.craigslist.or...
5	7222379453		https://hudsonvalley.craigslist.org/cto/d/west...	hudson valley	https://hudsonvalley.craigslist.or...
6	7221952215		https://hudsonvalley.craigslist.org/cto/d/west...	hudson valley	https://hudsonvalley.craigslist.or...
7	7220195662		https://hudsonvalley.craigslist.org/cto/d/poug...	hudson valley	https://hudsonvalley.craigslist.or...

```
In [15]: datay2 = datay2.drop('url', axis=1) #drop particular column
```

In [16]: `datav2`

Out[16]:	id	region	region_url	price	year	manufacturer	model
0	7222695916	prescott	https://prescott.craigslist.org	6000	NaN	NaN	NaN
1	7218891961	fayetteville	https://fayar.craigslist.org	11900	NaN	NaN	NaN
2	7221797935	florida keys	https://keys.craigslist.org	21000	NaN	NaN	NaN
3	7222270760	worcester / central MA	https://worcester.craigslist.org	1500	NaN	NaN	NaN
4	7210384030	greensboro	https://greensboro.craigslist.org	4900	NaN	NaN	NaN
5	7222379453	hudson valley	https://hudsonvalley.craigslist.org	1600	NaN	NaN	NaN
6	7221952215	hudson valley	https://hudsonvalley.craigslist.org	1000	NaN	NaN	NaN
7	7220195662	hudson valley	https://hudsonvalley.craigslist.org	15995	NaN	NaN	NaN
8	7209064557	medford-ashland	https://medford.craigslist.org	5000	NaN	NaN	NaN
9	7219485069	erie	https://erie.craigslist.org	3000	NaN	NaN	NaN
10	7218893038	el paso	https://elpaso.craigslist.org	0	NaN	NaN	NaN
11	7218325704	el paso	https://elpaso.craigslist.org	0	NaN	NaN	NaN
12	7217788283	el paso	https://elpaso.craigslist.org	0	NaN	NaN	NaN
13	7217147606	el paso	https://elpaso.craigslist.org	0	NaN	NaN	NaN
14	7209027818	el paso	https://elpaso.craigslist.org	0	NaN	NaN	NaN
15	7223509794	bellingham	https://bellingham.craigslist.org	13995	NaN	NaN	NaN
16	7222753076	bellingham	https://bellingham.craigslist.org	24999	NaN	NaN	NaN
17	7222206015	bellingham	https://bellingham.craigslist.org	21850	NaN	NaN	NaN

	id	region		region_url	price	year	manufacturer	model
18	7220030122	bellingham		https://bellingham.craigslist.org	26850	NaN	NaN	NaN
19	7218423006	bellingham		https://bellingham.craigslist.org	11999	NaN	NaN	NaN
20	7216672204	bellingham		https://bellingham.craigslist.org	24999	NaN	NaN	NaN
21	7215617048	bellingham		https://bellingham.craigslist.org	21850	NaN	NaN	NaN
22	7213839225	bellingham		https://bellingham.craigslist.org	26850	NaN	NaN	NaN
23	7208549803	bellingham		https://bellingham.craigslist.org	11999	NaN	NaN	NaN
24	7213843538	skagit / island / SJI		https://skagit.craigslist.org	24999	NaN	NaN	NaN
25	7212631321	skagit / island / SJI		https://skagit.craigslist.org	21850	NaN	NaN	NaN
26	7219973522	la crosse		https://lacrosse.craigslist.org	500	NaN	NaN	NaN
27	7316814884	auburn		https://auburn.craigslist.org	33590	2014.0	gmc	sierra 1500 crew cab slt
28	7316814758	auburn		https://auburn.craigslist.org	22590	2010.0	chevrolet	silverado 1500
29	7316814989	auburn		https://auburn.craigslist.org	39590	2020.0	chevrolet	silverado 1500 crew
30	7316743432	auburn		https://auburn.craigslist.org	30990	2017.0	toyota	tundra double cab sr
31	7316356412	auburn		https://auburn.craigslist.org	15000	2013.0	ford	f-150 xlt
32	7316343444	auburn		https://auburn.craigslist.org	27990	2012.0	gmc	sierra 2500 hd extended cab
33	7316304717	auburn		https://auburn.craigslist.org	34590	2016.0	chevrolet	silverado 1500 double
34	7316285779	auburn		https://auburn.craigslist.org	35000	2019.0	toyota	tacoma

	id	region		region_url	price	year	manufacturer	model
35	7316257769	auburn		https://auburn.craigslist.org	29990	2016.0	chevrolet	colorado extended cab
36	7316133914	auburn		https://auburn.craigslist.org	38590	2011.0	chevrolet	corvette grand sport
37	7316130053	auburn		https://auburn.craigslist.org	4500	1992.0	jeep	cherokee
38	7315816316	auburn		https://auburn.craigslist.org	32990	2017.0	jeep	wrangler unlimited sport
39	7315770394	auburn		https://auburn.craigslist.org	24590	2017.0	chevrolet	silverado 1500 regular
40	7315756348	auburn		https://auburn.craigslist.org	30990	2016.0	chevrolet	colorado crew cab z71
41	7315715960	auburn		https://auburn.craigslist.org	27990	2014.0	toyota	tacoma access cab pickup
42	7315379459	auburn		https://auburn.craigslist.org	37990	2016.0	chevrolet	camaro ss coupe 2d
43	7315275975	auburn		https://auburn.craigslist.org	33590	2014.0	toyota	tundra crewmax sr5 pickup
44	7315275942	auburn		https://auburn.craigslist.org	30990	2019.0	ford	ranger supercrew xl pickup
45	7315270785	auburn		https://auburn.craigslist.org	27990	2018.0	nissan	frontier crew cab pro-4x

	id	region		region_url	price	year	manufacturer	model
46	7315039307	auburn		https://auburn.craigslist.org	0	2011.0	jeep	compass
47	7314910156	auburn		https://auburn.craigslist.org	34590	2018.0	ford	f150 super cab xl pickup 4d
48	7314854462	auburn		https://auburn.craigslist.org	30590	2016.0	toyota	tacoma double cab sr5
49	7314811916	auburn		https://auburn.craigslist.org	32990	2020.0	jeep	wrangler sport suv 2d

50 rows × 25 columns

In [17]:	#datav2['age']= 2022-datav2['year'] #creating new column in dataframe
In [18]:	#datav2['age']
In [19]:	#datav2['age'].dtype # knowing datatype of variable
In [20]:	#datav2['region']
In [21]:	#datav2['region'].unique #displays unique entries of the particular column
In [22]:	#datav2[datav2.age<5] #displays data with condition
In [23]:	#datav4 = datav2[(datav2.age<5)&(datav2.price>50000)] #store data into new dataframe b
In [24]:	datav2['price_mile'] = datav2['price']/datav2['odometer'] #new feature creation with
In [25]:	datav2['price_mile']
Out[25]:	0 .. NaN 1 .. NaN 2 .. NaN 3 .. NaN 4 .. NaN 5 .. NaN

```

6      NaN
7      NaN
8      NaN
9      NaN
10     NaN
11     NaN
12     NaN
13     NaN
14     NaN
15     NaN
16     NaN
17     NaN
18     NaN
19     NaN
20     NaN
21     NaN
22     NaN
23     NaN
24     NaN
25     NaN
26     NaN
27    0.579908
28    0.317146
29    2.066284
30    0.753575
31    0.117188
32    0.407447
33    1.172582
34    0.813953
35    1.733326
36    1.276251
37    0.023438
38    1.098166
39    0.602933
40    0.886949
41    1.572030
42    3.914880
43    0.607953
44    16.897492
45    0.749759
46    0.000000
47    1.658516
48    1.013720
49    1.602935
Name: price_mile, dtype: float64

```

In [26]:

```

# function or defination to double the price of particular column
def pricedouble(x):
    return x*2
datav2['pricedouble'] = datav2['price'].apply(pricedouble)

```

In [27]:

```
datav2[['price','pricedouble']]
```

Out[27]:

	price	pricedouble
0	6000	12000
1	11900	23800
2	21000	42000
3	1500	3000

	price	pricedouble
4	4900	9800
5	1600	3200
6	1000	2000
7	15995	31990
8	5000	10000
9	3000	6000
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	13995	27990
16	24999	49998
17	21850	43700
18	26850	53700
19	11999	23998
20	24999	49998
21	21850	43700
22	26850	53700
23	11999	23998
24	24999	49998
25	21850	43700
26	500	1000
27	33590	67180
28	22590	45180
29	39590	79180
30	30990	61980
31	15000	30000
32	27990	55980
33	34590	69180
34	35000	70000
35	29990	59980
36	38590	77180
37	4500	9000
38	32990	65980
39	24590	49180
40	30990	61980

	price	pricedouble
41	27990	55980
42	37990	75980
43	33590	67180
44	30990	61980
45	27990	55980
46	0	0
47	34590	69180
48	30590	61180
49	32990	65980

In [28]: `datav2['pr'] = datav2.price.apply(lambda x:x*4) #function can be written and applied`

In [29]: `datav2[['price','pr']]`

Out[29]:

	price	pr
0	6000	24000
1	11900	47600
2	21000	84000
3	1500	6000
4	4900	19600
5	1600	6400
6	1000	4000
7	15995	63980
8	5000	20000
9	3000	12000
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	13995	55980
16	24999	99996
17	21850	87400
18	26850	107400
19	11999	47996
20	24999	99996
21	21850	87400
22	26850	107400

	price	pr
23	11999	47996
24	24999	99996
25	21850	87400
26	500	2000
27	33590	134360
28	22590	90360
29	39590	158360
30	30990	123960
31	15000	60000
32	27990	111960
33	34590	138360
34	35000	140000
35	29990	119960
36	38590	154360
37	4500	18000
38	32990	131960
39	24590	98360
40	30990	123960
41	27990	111960
42	37990	151960
43	33590	134360
44	30990	123960
45	27990	111960
46	0	0
47	34590	138360
48	30590	122360
49	32990	131960

```
In [30]: mylist = ["a", "b", "a", "c", "c"]
mylist = list(dict.fromkeys(mylist))
print(mylist)
```

```
['a', 'b', 'c']
```

```
In [31]: datav2['region']
```

```
Out[31]: 0      prescott
1      fayetteville
2      florida keys
3      worcester / central MA
4      greensboro
5      hudson valley
```

```

6      hudson valley
7      hudson valley
8      medford-ashland
9      erie
10     el paso
11     el paso
12     el paso
13     el paso
14     el paso
15     bellingham
16     bellingham
17     bellingham
18     bellingham
19     bellingham
20     bellingham
21     bellingham
22     bellingham
23     bellingham
24     skagit / island / SJI
25     skagit / island / SJI
26     la crosse
27     auburn
28     auburn
29     auburn
30     auburn
31     auburn
32     auburn
33     auburn
34     auburn
35     auburn
36     auburn
37     auburn
38     auburn
39     auburn
40     auburn
41     auburn
42     auburn
43     auburn
44     auburn
45     auburn
46     auburn
47     auburn
48     auburn
49     auburn
Name: region, dtype: object

```

In [32]: #REMOVING DUPLICATES

In [33]:

```

datav2['region']
datav5= list(dict.fromkeys(datav2['region']))
print(datav5)

```

```

['prescott', 'fayetteville', 'florida keys', 'worcester / central MA', 'greensboro',
'hudson valley', 'medford-ashland', 'erie', 'el paso', 'bellingham', 'skagit / island
/ SJI', 'la crosse', 'auburn']

```

In [34]:

```

test_list = [1, 5, 3, 6, 3, 5, 6, 1]
print ("The original list is : " + str(test_list))

# using set()
# to remove duplicated

```

```
# from list
test_list = list(set(test_list))

# printing list after removal
# distorted ordering
print ("The list after removing duplicates : " + str(test_list))
```

The original list is : [1, 5, 3, 6, 3, 5, 6, 1]
The list after removing duplicates : [1, 3, 5, 6]

In [35]:

```
datav2['region']
print ("The original list is : " + str(datav2['region']))

# using set()
# to remove duplicated
# from list
test_list = list(set(datav2['region']))

# printing list after removal
# distorted ordering
print ("The list after removing duplicates : " + str(test_list))
```

The original list is : 0 prescott
1 fayetteville
2 florida keys
3 worcester / central MA
4 greensboro
5 hudson valley
6 hudson valley
7 hudson valley
8 medford-ashland
9 erie
10 el paso
11 el paso
12 el paso
13 el paso
14 el paso
15 bellingham
16 bellingham
17 bellingham
18 bellingham
19 bellingham
20 bellingham
21 bellingham
22 bellingham
23 bellingham
24 skagit / island / SJI
25 skagit / island / SJI
26 la crosse
27 auburn
28 auburn
29 auburn
30 auburn
31 auburn
32 auburn
33 auburn
34 auburn
35 auburn
36 auburn
37 auburn
38 auburn
39 auburn
40 auburn

```

41 ..... auburn
42 ..... auburn
43 ..... auburn
44 ..... auburn
45 ..... auburn
46 ..... auburn
47 ..... auburn
48 ..... auburn
49 ..... auburn
Name: region, dtype: object

```

The list after removing duplicates : ['florida keys', 'fayetteville', 'prescott', 'med ford-ashland', 'erie', 'skagit / island / SJI', 'greensboro', 'el paso', 'la crosse', 'worcester / central MA', 'bellingham', 'auburn', 'hudson valley']

In [36]: `#convert the categorical variable values into numerical (Label encoding)`
`datav2['size']=datav2['size'].map({'full-size':0,'mid-size':1,'compact':2,'sub-compact':3})`

In [37]: `datav2['size'].value_counts() #counts values of each categories`

Out[37]: `0.0 ... 2`
Name: size, dtype: int64

In [38]: `datav2.shape #display tuple of size in (rows, columns) or no of element in each dimension`

Out[38]: `(50, 28)`

In [39]: `datav2.describe()#is used to view some basic statistical details Like percentile, mean, median, standard deviation, minimum and maximum values in each column`

Out[39]:

	id	price	year	odometer	size	county	lat	long	price
count	5.000000e+01	50.000000	23.000000	23.000000	2.0	0.0	23.000000	23.000000	23.000000
mean	7.262865e+09	19246.900000	2014.608696	47699.521739	0.0	NaN	32.591139	-85.475190	1.0
std	4.952944e+07	13268.776049	5.702783	42850.728957	0.0	NaN	0.015041	0.024084	3.0
min	7.208550e+09	0.000000	1992.000000	1834.000000	0.0	NaN	32.547500	-85.518900	0.0
25%	7.218540e+09	4925.000000	2013.500000	20718.500000	0.0	NaN	32.590000	-85.480000	0.0
50%	7.222724e+09	22220.000000	2016.000000	34940.000000	0.0	NaN	32.590000	-85.480000	0.0
75%	7.315767e+09	30890.000000	2017.500000	56587.000000	0.0	NaN	32.590000	-85.480000	1.0
max	7.316815e+09	39590.000000	2020.000000	192000.000000	0.0	NaN	32.645400	-85.378300	16.0

In [40]: `datav2.isnull().sum() #display sum of missing values in dataframe`

Out[40]:

id	0
region	0
region_url	0
price	0
year	27
manufacturer	27
model	27
condition	27
cylinders	34
fuel	27
odometer	27

```

title_status    27
transmission    27
VIN             31
drive           38
size            48
type            28
paint_color     29
image_url       27
description     27
county          50
state           0
lat              27
long             27
posting_date    27
price_mile      27
pricedouble     0
pr...            0
dtype: int64

```

In [41]: `datav2.isnull()`

Out[41]:

	id	region	region_url	price	year	manufacturer	model	condition	cylinders	fuel	...	image
0	False	False	False	False	True		True	True	True	True	True	...
1	False	False	False	False	True		True	True	True	True	True	...
2	False	False	False	False	True		True	True	True	True	True	...
3	False	False	False	False	True		True	True	True	True	True	...
4	False	False	False	False	True		True	True	True	True	True	...
5	False	False	False	False	True		True	True	True	True	True	...
6	False	False	False	False	True		True	True	True	True	True	...
7	False	False	False	False	True		True	True	True	True	True	...
8	False	False	False	False	True		True	True	True	True	True	...
9	False	False	False	False	True		True	True	True	True	True	...
10	False	False	False	False	True		True	True	True	True	True	...
11	False	False	False	False	True		True	True	True	True	True	...
12	False	False	False	False	True		True	True	True	True	True	...
13	False	False	False	False	True		True	True	True	True	True	...
14	False	False	False	False	True		True	True	True	True	True	...
15	False	False	False	False	True		True	True	True	True	True	...
16	False	False	False	False	True		True	True	True	True	True	...
17	False	False	False	False	True		True	True	True	True	True	...
18	False	False	False	False	True		True	True	True	True	True	...
19	False	False	False	False	True		True	True	True	True	True	...
20	False	False	False	False	True		True	True	True	True	True	...
21	False	False	False	False	True		True	True	True	True	True	...
22	False	False	False	False	True		True	True	True	True	True	...
23	False	False	False	False	True		True	True	True	True	True	...

	id	region	region_url	price	year	manufacturer	model	condition	cylinders	fuel	...	image
24	False	False	False	False	True		True	True	True	True	True	...
25	False	False	False	False	True		True	True	True	True	True	...
26	False	False	False	False	True		True	True	True	True	True	...
27	False	False	False	False	False		False	False	False	False	False	...
28	False	False	False	False	False		False	False	False	False	False	...
29	False	False	False	False	False		False	False	False	False	False	...
30	False	False	False	False	False		False	False	False	False	False	...
31	False	False	False	False	False		False	False	False	False	False	...
32	False	False	False	False	False		False	False	False	False	False	...
33	False	False	False	False	False		False	False	False	False	False	...
34	False	False	False	False	False		False	False	False	False	False	...
35	False	False	False	False	False		False	False	False	False	False	...
36	False	False	False	False	False		False	False	False	False	False	...
37	False	False	False	False	False		False	False	False	False	False	...
38	False	False	False	False	False		False	False	False	False	False	...
39	False	False	False	False	False		False	False	False	False	False	...
40	False	False	False	False	False		False	False	False	True	False	...
41	False	False	False	False	False		False	False	False	True	False	...
42	False	False	False	False	False		False	False	False	False	False	...
43	False	False	False	False	False		False	False	False	True	False	...
44	False	False	False	False	False		False	False	False	True	False	...
45	False	False	False	False	False		False	False	False	False	False	...
46	False	False	False	False	False		False	False	False	True	False	...
47	False	False	False	False	False		False	False	False	False	False	...
48	False	False	False	False	False		False	False	False	True	False	...
49	False	False	False	False	False		False	False	False	True	False	...

50 rows × 28 columns



In [48]:

```
datav1 = datav2.replace("?", np.NaN)
datav1.head()
```

Out[48]:

	id	region	region_url	price	year	manufacturer	model	condition
0	7222695916	prescott	https://prescott.craigslist.org	6000	NaN		NaN	NaN
1	7218891961	fayetteville	https://fayar.craigslist.org	11900	NaN		NaN	NaN
2	7221797935	florida keys	https://keys.craigslist.org	21000	NaN		NaN	NaN
3	7222270760	worcester / central MA	https://worcester.craigslist.org	1500	NaN		NaN	NaN

	id	region		region_url	price	year	manufacturer	model	condition
4	7210384030	greensboro	https://greensboro.craigslist.org		4900	NaN		NaN	NaN

5 rows × 28 columns

