

DSBDA1

January 30, 2025

```
[100]: import numpy as np
import pandas as pd
```

```
[101]: data=pd.read_csv("vehicles.csv")
```

```
[102]: data
```

```
[102]:
```

	id	url \
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...
2	7221797935	https://keys.craigslist.org/cto/d/summerland-k...
3	7222270760	https://worcester.craigslist.org/cto/d/west-br...
4	7210384030	https://greensboro.craigslist.org/cto/d/trinit...
...
426875	7301591192	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426876	7301591187	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426877	7301591147	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426878	7301591140	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426879	7301591129	https://wyoming.craigslist.org/ctd/d/atlanta-2...

	region	region_url	price \
0	prescott	https://prescott.craigslist.org	6000
1	fayetteville	https://fayar.craigslist.org	11900
2	florida keys	https://keys.craigslist.org	21000
3	worcester / central MA	https://worcester.craigslist.org	1500
4	greensboro	https://greensboro.craigslist.org	4900
...
426875	wyoming	https://wyoming.craigslist.org	23590
426876	wyoming	https://wyoming.craigslist.org	30590
426877	wyoming	https://wyoming.craigslist.org	34990
426878	wyoming	https://wyoming.craigslist.org	28990
426879	wyoming	https://wyoming.craigslist.org	30590

	year	manufacturer	model	condition	cylinders \
0	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN

3	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN
...
426875	2019.0	nissan	maxima s sedan 4d	good	6 cylinders	
426876	2020.0	volvo	s60 t5 momentum sedan 4d	good		NaN
426877	2020.0	cadillac	xt4 sport suv 4d	good		NaN
426878	2018.0	lexus	es 350 sedan 4d	good	6 cylinders	
426879	2019.0	bmw	4 series 430i gran coupe	good		NaN

	size	type	paint_color	\
0	NaN	NaN	NaN	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	NaN	NaN	NaN	
4	NaN	NaN	NaN	
...	
426875	NaN	sedan	NaN	
426876	NaN	sedan	red	
426877	NaN	hatchback	white	
426878	NaN	sedan	silver	
426879	NaN	coupe	NaN	

	image_url	\
0	NaN	
1	NaN	
2	NaN	
3	NaN	
4	NaN	
...	...	
426875	https://images.craigslist.org/00o0o_iiraFnHg8q...	
426876	https://images.craigslist.org/00x0x_15sbgnxCIS...	
426877	https://images.craigslist.org/00L0L_farM7bxnxR...	
426878	https://images.craigslist.org/00z0z_bKnIVGLkDT...	
426879	https://images.craigslist.org/00Y0Y_lEUocjyRxa...	

	description	county	state	\
0		NaN	NaN	az
1		NaN	NaN	ar
2		NaN	NaN	fl
3		NaN	NaN	ma
4		NaN	NaN	nc
...	
426875	Carvana is the safer way to buy a car During t...	NaN	wy	
426876	Carvana is the safer way to buy a car During t...	NaN	wy	
426877	Carvana is the safer way to buy a car During t...	NaN	wy	
426878	Carvana is the safer way to buy a car During t...	NaN	wy	
426879	Carvana is the safer way to buy a car During t...	NaN	wy	

	lat	long	posting_date
0	NaN	NaN	NaN
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN
...
426875	33.786500	-84.445400	2021-04-04T03:21:31-0600
426876	33.786500	-84.445400	2021-04-04T03:21:29-0600
426877	33.779214	-84.411811	2021-04-04T03:21:17-0600
426878	33.786500	-84.445400	2021-04-04T03:21:11-0600
426879	33.779214	-84.411811	2021-04-04T03:21:07-0600

[426880 rows x 26 columns]

```
[103]: data.head()
```

```
[103]:
```

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

	region	region_url	price	year \
0	prescott	https://prescott.craigslist.org	6000	NaN
1	fayetteville	https://fayar.craigslist.org	11900	NaN
2	florida keys	https://keys.craigslist.org	21000	NaN
3	worcester / central MA	https://worcester.craigslist.org	1500	NaN
4	greensboro	https://greensboro.craigslist.org	4900	NaN

	manufacturer	model	condition	cylinders	... size	type	paint_color \
0	NaN	NaN	NaN	NaN	... NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	... NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	... NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	... NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	... NaN	NaN	NaN

	image_url	description	county	state	lat	long	posting_date
0	NaN	NaN	NaN	az	NaN	NaN	NaN
1	NaN	NaN	NaN	ar	NaN	NaN	NaN
2	NaN	NaN	NaN	fl	NaN	NaN	NaN
3	NaN	NaN	NaN	ma	NaN	NaN	NaN
4	NaN	NaN	NaN	nc	NaN	NaN	NaN

[5 rows x 26 columns]

```
[104]: data.tail()
```

```
[104]:
```

	id	url \
426875	7301591192	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426876	7301591187	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426877	7301591147	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426878	7301591140	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426879	7301591129	https://wyoming.craigslist.org/ctd/d/atlanta-2...

	region	region_url	price	year	manufacturer \
426875	wyoming	https://wyoming.craigslist.org	23590	2019.0	nissan
426876	wyoming	https://wyoming.craigslist.org	30590	2020.0	volvo
426877	wyoming	https://wyoming.craigslist.org	34990	2020.0	cadillac
426878	wyoming	https://wyoming.craigslist.org	28990	2018.0	lexus
426879	wyoming	https://wyoming.craigslist.org	30590	2019.0	bmw

	model	condition	cylinders	... size	type \
426875	maxima s sedan 4d	good	6 cylinders	... NaN	sedan
426876	s60 t5 momentum sedan 4d	good	NaN	... NaN	sedan
426877	xt4 sport suv 4d	good	NaN	... NaN	hatchback
426878	es 350 sedan 4d	good	6 cylinders	... NaN	sedan
426879	4 series 430i gran coupe	good	NaN	... NaN	coupe

	paint_color	image_url \
426875	NaN	https://images.craigslist.org/00o0o_iiraFnHg8q...
426876	red	https://images.craigslist.org/00x0x_15sbgnxCIS...
426877	white	https://images.craigslist.org/00L0L_farM7bxnXR...
426878	silver	https://images.craigslist.org/00z0z_bKnIVGLkDT...
426879	NaN	https://images.craigslist.org/00Y0Y_lEUocjyRxa...

	description	county	state \
426875	Carvana is the safer way to buy a car During t...	NaN	wy
426876	Carvana is the safer way to buy a car During t...	NaN	wy
426877	Carvana is the safer way to buy a car During t...	NaN	wy
426878	Carvana is the safer way to buy a car During t...	NaN	wy
426879	Carvana is the safer way to buy a car During t...	NaN	wy

	lat	long	posting_date
426875	33.786500	-84.445400	2021-04-04T03:21:31-0600
426876	33.786500	-84.445400	2021-04-04T03:21:29-0600
426877	33.779214	-84.411811	2021-04-04T03:21:17-0600
426878	33.786500	-84.445400	2021-04-04T03:21:11-0600
426879	33.779214	-84.411811	2021-04-04T03:21:07-0600

```
[5 rows x 26 columns]
```

```
[105]: data.shape
```

[105]: (426880, 26)

[106]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 426880 entries, 0 to 426879
Data columns (total 26 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    426880 non-null  int64
1   url                   426880 non-null  object
2   region                426880 non-null  object
3   region_url            426880 non-null  object
4   price                 426880 non-null  int64
5   year                  425675 non-null  float64
6   manufacturer          409234 non-null  object
7   model                 421603 non-null  object
8   condition             252776 non-null  object
9   cylinders             249202 non-null  object
10  fuel                  423867 non-null  object
11  odometer              422480 non-null  float64
12  title_status          418638 non-null  object
13  transmission          424324 non-null  object
14  VIN                   265838 non-null  object
15  drive                 296313 non-null  object
16  size                  120519 non-null  object
17  type                  334022 non-null  object
18  paint_color           296677 non-null  object
19  image_url             426812 non-null  object
20  description            426810 non-null  object
21  county                0 non-null       float64
22  state                 426880 non-null  object
23  lat                   420331 non-null  float64
24  long                  420331 non-null  float64
25  posting_date          426812 non-null  object
dtypes: float64(5), int64(2), object(19)
memory usage: 84.7+ MB
```

[107]: data.shape[0]

[107]: 426880

[108]: data.shape[1]

[108]: 26

[109]: data.columns

```
[109]: 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')
```

```
[110]: data.columns[0]
```

```
[110]: 'id'
```

```
[111]: data.columns[5]
```

```
[111]: 'year'
```

```
[112]: data.describe()
```

```
[112]:
```

	id	price	year	odometer	county \
count	4.268800e+05	4.268800e+05	425675.000000	4.224800e+05	0.0
mean	7.311487e+09	7.519874e+04	2011.235191	9.804333e+04	NaN
std	4.473170e+06	1.218228e+07	9.452120	2.138815e+05	NaN
min	7.207408e+09	0.000000e+00	1900.000000	0.000000e+00	NaN
25%	7.308143e+09	5.900000e+03	2008.000000	3.770400e+04	NaN
50%	7.312621e+09	1.395000e+04	2013.000000	8.554800e+04	NaN
75%	7.315254e+09	2.648125e+04	2017.000000	1.335425e+05	NaN
max	7.317101e+09	3.736929e+09	2022.000000	1.000000e+07	NaN

	lat	long
count	420331.000000	420331.000000
mean	38.493940	-94.748599
std	5.841533	18.365462
min	-84.122245	-159.827728
25%	34.601900	-111.939847
50%	39.150100	-88.432600
75%	42.398900	-80.832039
max	82.390818	173.885502

```
[113]: data.sample()
```

```
[113]:
```

	id	url \
331447	7312569190	https://allentown.craigslist.org/cto/d/stewart...

	region	region_url	price	year \
331447	lehigh valley	https://allentown.craigslist.org	2950	2006.0

	manufacturer	model	condition	cylinders	... size	type \
331447	mazda	6	excellent	4 cylinders	... NaN	sedan

```

        paint_color                                image_url \
331447          red  https://images.craigslist.org/00h0h_gsK9PLy9f8...

                                description county state \
331447  2006 MAZDA MAZDA6 In excellent condition. With...   NaN   pa

        lat    long                posting_date
331447  40.6937 -75.111  2021-04-25T18:53:11-0400

[1 rows x 26 columns]

```

```
[114]: data.iloc[0]
```

```

[114]: id                                7222695916
url                https://prescott.craigslist.org/cto/d/prescott...
region                                prescott
region_url          https://prescott.craigslist.org
price                                6000
year                                NaN
manufacturer          NaN
model                                NaN
condition              NaN
cylinders              NaN
fuel                  NaN
odometer              NaN
title_status          NaN
transmission          NaN
VIN                  NaN
drive                NaN
size                NaN
type                NaN
paint_color          NaN
image_url            NaN
description            NaN
county                NaN
state                az
lat                  NaN
long                NaN
posting_date          NaN
Name: 0, dtype: object

```

```
[115]: data.iloc[0:5]
```

```

[115]:          id                                url \
0  7222695916  https://prescott.craigslist.org/cto/d/prescott...
1  7218891961  https://fayar.craigslist.org/ctd/d/bentonville...

```

```

2 7221797935 https://keys.craigslist.org/cto/d/summerland-k...
3 7222270760 https://worcester.craigslist.org/cto/d/west-br...
4 7210384030 https://greensboro.craigslist.org/cto/d/trinit...

```

	region	region_url	price	year	\
0	prescott	https://prescott.craigslist.org	6000	NaN	
1	fayetteville	https://fayar.craigslist.org	11900	NaN	
2	florida keys	https://keys.craigslist.org	21000	NaN	
3	worcester / central MA	https://worcester.craigslist.org	1500	NaN	
4	greensboro	https://greensboro.craigslist.org	4900	NaN	

	manufacturer	model	condition	cylinders	...	size	type	paint_color	\
0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	

	image_url	description	county	state	lat	long	posting_date
0	NaN	NaN	NaN	az	NaN	NaN	NaN
1	NaN	NaN	NaN	ar	NaN	NaN	NaN
2	NaN	NaN	NaN	fl	NaN	NaN	NaN
3	NaN	NaN	NaN	ma	NaN	NaN	NaN
4	NaN	NaN	NaN	nc	NaN	NaN	NaN

[5 rows x 26 columns]

```
[116]: data['id']
```

```

[116]: 0      7222695916
      1      7218891961
      2      7221797935
      3      7222270760
      4      7210384030
      ...
      426875    7301591192
      426876    7301591187
      426877    7301591147
      426878    7301591140
      426879    7301591129
      Name: id, Length: 426880, dtype: int64

```

```
[117]: list1=['id','region']
```

```
[118]: data[list1]
```



```
[118]:
```

	id	region
0	7222695916	prescott
1	7218891961	fayetteville
2	7221797935	florida keys
3	7222270760	worcester / central MA
4	7210384030	greensboro
...
426875	7301591192	wyoming
426876	7301591187	wyoming
426877	7301591147	wyoming
426878	7301591140	wyoming
426879	7301591129	wyoming

[426880 rows x 2 columns]

```
[119]: data.iloc[0:5,0]
```

```
[119]:
```

0	7222695916
1	7218891961
2	7221797935
3	7222270760
4	7210384030

Name: id, dtype: int64

```
[120]: data.loc[0:5,'id']
```

```
[120]:
```

0	7222695916
1	7218891961
2	7221797935
3	7222270760
4	7210384030
5	7222379453

Name: id, dtype: int64

```
[121]: data.iloc[0:5,0:4]
```

```
[121]:
```

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

	region	region_url
0	prescott	https://prescott.craigslist.org
1	fayetteville	https://fayar.craigslist.org
2	florida keys	https://keys.craigslist.org

```
3 worcester / central MA https://worcester.craigslist.org
4 greensboro https://greensboro.craigslist.org
```

```
[122]: data.loc[0:5,['id','url','region','region_url']]
```

```
[122]:      id      url \
0  7222695916 https://prescott.craigslist.org/cto/d/prescott...
1  7218891961 https://fayar.craigslist.org/ctd/d/bentonville...
2  7221797935 https://keys.craigslist.org/cto/d/summerland-k...
3  7222270760 https://worcester.craigslist.org/cto/d/west-br...
4  7210384030 https://greensboro.craigslist.org/cto/d/trinit...
5  7222379453 https://hudsonvalley.craigslist.org/cto/d/west...
```

```
      region      region_url
0      prescott https://prescott.craigslist.org
1  fayetteville https://fayar.craigslist.org
2  florida keys https://keys.craigslist.org
3 worcester / central MA https://worcester.craigslist.org
4      greensboro https://greensboro.craigslist.org
5  hudson valley https://hudsonvalley.craigslist.org
```

```
[123]: newdata=data.iloc[0:500,:]
```

```
[124]: newdata
```

```
[124]:      id      url \
0  7222695916 https://prescott.craigslist.org/cto/d/prescott...
1  7218891961 https://fayar.craigslist.org/ctd/d/bentonville...
2  7221797935 https://keys.craigslist.org/cto/d/summerland-k...
3  7222270760 https://worcester.craigslist.org/cto/d/west-br...
4  7210384030 https://greensboro.craigslist.org/cto/d/trinit...
...      ...      ...
495  7313806066 https://bham.craigslist.org/cto/d/bessemer-200...
496  7313801999 https://bham.craigslist.org/cto/d/douglasville...
497  7313798038 https://bham.craigslist.org/ctd/d/clanton-2020...
498  7313798048 https://bham.craigslist.org/ctd/d/clanton-2018...
499  7313798006 https://bham.craigslist.org/ctd/d/clanton-2021...
```

```
      region      region_url price year \
0      prescott https://prescott.craigslist.org 6000 NaN
1  fayetteville https://fayar.craigslist.org 11900 NaN
2  florida keys https://keys.craigslist.org 21000 NaN
3 worcester / central MA https://worcester.craigslist.org 1500 NaN
4      greensboro https://greensboro.craigslist.org 4900 NaN
...      ...      ...      ...
495      birmingham https://bham.craigslist.org 2450 2000.0
496      birmingham https://bham.craigslist.org 12950 2013.0
```

497	birmingham	https://bham.craigslist.org	30488	2020.0
498	birmingham	https://bham.craigslist.org	17989	2018.0
499	birmingham	https://bham.craigslist.org	22602	2021.0

	manufacturer	model	condition	cylinders	...	size	type	\
0	NaN	NaN	NaN	NaN	...	NaN	NaN	
1	NaN	NaN	NaN	NaN	...	NaN	NaN	
2	NaN	NaN	NaN	NaN	...	NaN	NaN	
3	NaN	NaN	NaN	NaN	...	NaN	NaN	
4	NaN	NaN	NaN	NaN	...	NaN	NaN	
..	
495	infiniti	qx4	good	6 cylinders	...	full-size	SUV	
496	lexus	ct200h hybrid	good	4 cylinders	...	NaN	NaN	
497	infiniti	q50	NaN	NaN	...	NaN	NaN	
498	toyota	camry	NaN	NaN	...	NaN	NaN	
499	toyota	corolla	NaN	NaN	...	NaN	NaN	

	paint_color	image_url	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
..	
495	silver	https://images.craigslist.org/00H0H_fKm0syAjYk...	
496	black	https://images.craigslist.org/00404_2ShXAZAmgO...	
497	black	https://images.craigslist.org/00o0o_15ygWi1J2M...	
498	silver	https://images.craigslist.org/00F0F_7DqNJmRVDQ...	
499	red	https://images.craigslist.org/00t0t_59ahErd0Bd...	

	description	county	state	lat	\
0	NaN	NaN	az	NaN	
1	NaN	NaN	ar	NaN	
2	NaN	NaN	fl	NaN	
3	NaN	NaN	ma	NaN	
4	NaN	NaN	nc	NaN	
..	
495	2000 Infinity QX4 SUV, Automatic, V6, AWD, lea...	NaN	al	33.4680	
496	Selling a 2013 Lexus CT 200H Hybrid with 77k m...	NaN	al	33.6989	
497	phone: âŸŒ (205) 862-8003 text: Text ...	NaN	al	32.9229	
498	phone: âŸŒ (205) 862-8003 text: Text ...	NaN	al	32.9229	
499	phone: âŸŒ (205) 862-8003 text: Text ...	NaN	al	32.9229	

	long	posting_date
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN

```

3      NaN      NaN
4      NaN      NaN
..      ...      ...
495 -87.0924 2021-04-28T09:31:58-0500
496 -84.7454 2021-04-28T09:24:13-0500
497 -86.5450 2021-04-28T09:16:39-0500
498 -86.5450 2021-04-28T09:16:39-0500
499 -86.5450 2021-04-28T09:16:36-0500

```

[500 rows x 26 columns]

```
[125]: newdata=data.copy()
newdata['vage']=2025-newdata['year']
```

```
[126]: newdata
```

```
[126]:
```

	id	url \
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...
2	7221797935	https://keys.craigslist.org/cto/d/summerland-k...
3	7222270760	https://worcester.craigslist.org/cto/d/west-br...
4	7210384030	https://greensboro.craigslist.org/cto/d/trinit...
...
426875	7301591192	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426876	7301591187	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426877	7301591147	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426878	7301591140	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426879	7301591129	https://wyoming.craigslist.org/ctd/d/atlanta-2...

	region	region_url	price \
0	prescott	https://prescott.craigslist.org	6000
1	fayetteville	https://fayar.craigslist.org	11900
2	florida keys	https://keys.craigslist.org	21000
3	worcester / central MA	https://worcester.craigslist.org	1500
4	greensboro	https://greensboro.craigslist.org	4900
...
426875	wyoming	https://wyoming.craigslist.org	23590
426876	wyoming	https://wyoming.craigslist.org	30590
426877	wyoming	https://wyoming.craigslist.org	34990
426878	wyoming	https://wyoming.craigslist.org	28990
426879	wyoming	https://wyoming.craigslist.org	30590

	year	manufacturer	model	condition	cylinders \
0	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN

4	NaN	NaN	NaN	NaN	NaN	NaN
...
426875	2019.0	nissan	maxima s sedan 4d	good	6 cylinders	
426876	2020.0	volvo	s60 t5 momentum sedan 4d	good		NaN
426877	2020.0	cadillac	xt4 sport suv 4d	good		NaN
426878	2018.0	lexus	es 350 sedan 4d	good	6 cylinders	
426879	2019.0	bmw	4 series 430i gran coupe	good		NaN

	...	type	paint_color \
0	...	NaN	NaN
1	...	NaN	NaN
2	...	NaN	NaN
3	...	NaN	NaN
4	...	NaN	NaN
...
426875	...	sedan	NaN
426876	...	sedan	red
426877	...	hatchback	white
426878	...	sedan	silver
426879	...	coupe	NaN

		image_url \
0		NaN
1		NaN
2		NaN
3		NaN
4		NaN
...		...
426875		https://images.craigslist.org/00o0o_iiraFnHg8q...
426876		https://images.craigslist.org/00x0x_15sbgnxCIS...
426877		https://images.craigslist.org/00L0L_farM7bxnxR...
426878		https://images.craigslist.org/00z0z_bKnIVGLkDT...
426879		https://images.craigslist.org/00Y0Y_1EUocjyRxa...

		description	county	state \
0		NaN	NaN	az
1		NaN	NaN	ar
2		NaN	NaN	fl
3		NaN	NaN	ma
4		NaN	NaN	nc
...	
426875	Carvana is the safer way to buy a car During t...		NaN	wy
426876	Carvana is the safer way to buy a car During t...		NaN	wy
426877	Carvana is the safer way to buy a car During t...		NaN	wy
426878	Carvana is the safer way to buy a car During t...		NaN	wy
426879	Carvana is the safer way to buy a car During t...		NaN	wy

	lat	long	posting_date	vage
0	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN
...
426875	33.786500	-84.445400	2021-04-04T03:21:31-0600	6.0
426876	33.786500	-84.445400	2021-04-04T03:21:29-0600	5.0
426877	33.779214	-84.411811	2021-04-04T03:21:17-0600	5.0
426878	33.786500	-84.445400	2021-04-04T03:21:11-0600	7.0
426879	33.779214	-84.411811	2021-04-04T03:21:07-0600	6.0

[426880 rows x 27 columns]

```
[127]: #newdata.loc['vage']=2025-newdata.loc['year']
```

```
[128]: del newdata['vage']
```

```
[129]: newdata
```

```
[129]:
```

	id	url \
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...
2	7221797935	https://keys.craigslist.org/cto/d/summerland-k...
3	7222270760	https://worcester.craigslist.org/cto/d/west-br...
4	7210384030	https://greensboro.craigslist.org/cto/d/trinit...
...
426875	7301591192	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426876	7301591187	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426877	7301591147	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426878	7301591140	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426879	7301591129	https://wyoming.craigslist.org/ctd/d/atlanta-2...

	region	region_url	price \
0	prescott	https://prescott.craigslist.org	6000
1	fayetteville	https://fayar.craigslist.org	11900
2	florida keys	https://keys.craigslist.org	21000
3	worcester / central MA	https://worcester.craigslist.org	1500
4	greensboro	https://greensboro.craigslist.org	4900
...
426875	wyoming	https://wyoming.craigslist.org	23590
426876	wyoming	https://wyoming.craigslist.org	30590
426877	wyoming	https://wyoming.craigslist.org	34990
426878	wyoming	https://wyoming.craigslist.org	28990
426879	wyoming	https://wyoming.craigslist.org	30590

	year	manufacturer	model	condition	cylinders	\
0	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	
...	
426875	2019.0	nissan	maxima s sedan 4d	good	6 cylinders	
426876	2020.0	volvo	s60 t5 momentum sedan 4d	good	NaN	
426877	2020.0	cadillac	xt4 sport suv 4d	good	NaN	
426878	2018.0	lexus	es 350 sedan 4d	good	6 cylinders	
426879	2019.0	bmw	4 series 430i gran coupe	good	NaN	

	size	type	paint_color	\
0	NaN	NaN	NaN	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	NaN	NaN	NaN	
4	NaN	NaN	NaN	
...	
426875	NaN	sedan	NaN	
426876	NaN	sedan	red	
426877	NaN	hatchback	white	
426878	NaN	sedan	silver	
426879	NaN	coupe	NaN	

	image_url	\
0	NaN	
1	NaN	
2	NaN	
3	NaN	
4	NaN	
...	...	
426875	https://images.craigslist.org/00o0o_iiraFnHg8q...	
426876	https://images.craigslist.org/00x0x_15sbgnxCIS...	
426877	https://images.craigslist.org/00L0L_farM7bxnxR...	
426878	https://images.craigslist.org/00z0z_bKnIVGLkDT...	
426879	https://images.craigslist.org/00Y0Y_1EUocjyRxa...	

	description	county	state	\
0	NaN	NaN	az	
1	NaN	NaN	ar	
2	NaN	NaN	fl	
3	NaN	NaN	ma	
4	NaN	NaN	nc	
...	
426875	Carvana is the safer way to buy a car During t...	NaN	wy	

```

426876 Carvana is the safer way to buy a car During t... NaN wy
426877 Carvana is the safer way to buy a car During t... NaN wy
426878 Carvana is the safer way to buy a car During t... NaN wy
426879 Carvana is the safer way to buy a car During t... NaN wy

```

```

      lat      long      posting_date
0      NaN      NaN      NaN
1      NaN      NaN      NaN
2      NaN      NaN      NaN
3      NaN      NaN      NaN
4      NaN      NaN      NaN
...      ...      ...
426875  33.786500 -84.445400 2021-04-04T03:21:31-0600
426876  33.786500 -84.445400 2021-04-04T03:21:29-0600
426877  33.779214 -84.411811 2021-04-04T03:21:17-0600
426878  33.786500 -84.445400 2021-04-04T03:21:11-0600
426879  33.779214 -84.411811 2021-04-04T03:21:07-0600

```

```
[426880 rows x 26 columns]
```

```
[130]: newdata.columns
```

```
[130]: 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')
```

```
[131]: newdata.drop('url', axis=1)
```

```

[131]:
      id      region      region_url \
0  7222695916      prescott  https://prescott.craigslist.org
1  7218891961  fayetteville  https://fayar.craigslist.org
2  7221797935  florida keys  https://keys.craigslist.org
3  7222270760 worcester / central MA  https://worcester.craigslist.org
4  7210384030  greensboro  https://greensboro.craigslist.org
...      ...      ...
426875  7301591192      wyoming  https://wyoming.craigslist.org
426876  7301591187      wyoming  https://wyoming.craigslist.org
426877  7301591147      wyoming  https://wyoming.craigslist.org
426878  7301591140      wyoming  https://wyoming.craigslist.org
426879  7301591129      wyoming  https://wyoming.craigslist.org

      price  year  manufacturer      model  condition \
0      6000   NaN      NaN      NaN      NaN
1     11900   NaN      NaN      NaN      NaN

```


2	21000	NaN	NaN	NaN	NaN
3	1500	NaN	NaN	NaN	NaN
4	4900	NaN	NaN	NaN	NaN
...
426875	23590	2019.0	nissan	maxima s sedan 4d	good
426876	30590	2020.0	volvo	s60 t5 momentum sedan 4d	good
426877	34990	2020.0	cadillac	xt4 sport suv 4d	good
426878	28990	2018.0	lexus	es 350 sedan 4d	good
426879	30590	2019.0	bmw	4 series 430i gran coupe	good

	cylinders	fuel	...	size	type	paint_color	\
0	NaN	NaN	...	NaN	NaN	NaN	
1	NaN	NaN	...	NaN	NaN	NaN	
2	NaN	NaN	...	NaN	NaN	NaN	
3	NaN	NaN	...	NaN	NaN	NaN	
4	NaN	NaN	...	NaN	NaN	NaN	
...	
426875	6 cylinders	gas	...	NaN	sedan	NaN	
426876	NaN	gas	...	NaN	sedan	red	
426877	NaN	diesel	...	NaN	hatchback	white	
426878	6 cylinders	gas	...	NaN	sedan	silver	
426879	NaN	gas	...	NaN	coupe	NaN	

	image_url	\
0	NaN	
1	NaN	
2	NaN	
3	NaN	
4	NaN	
...	...	
426875	https://images.craigslist.org/00o0o_iiraFnHg8q...	
426876	https://images.craigslist.org/00x0x_15sbgnxCIS...	
426877	https://images.craigslist.org/00L0L_farM7bxnxR...	
426878	https://images.craigslist.org/00z0z_bKnIVGLkDT...	
426879	https://images.craigslist.org/00Y0Y_lEUocjyRxa...	

	description	county	state	\
0	NaN	NaN	az	
1	NaN	NaN	ar	
2	NaN	NaN	fl	
3	NaN	NaN	ma	
4	NaN	NaN	nc	
...	
426875	Carvana is the safer way to buy a car During t...	NaN	wy	
426876	Carvana is the safer way to buy a car During t...	NaN	wy	
426877	Carvana is the safer way to buy a car During t...	NaN	wy	
426878	Carvana is the safer way to buy a car During t...	NaN	wy	

426879 Carvana is the safer way to buy a car During t... NaN wy

	lat	long	posting_date
0	NaN	NaN	NaN
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN
...
426875	33.786500	-84.445400	2021-04-04T03:21:31-0600
426876	33.786500	-84.445400	2021-04-04T03:21:29-0600
426877	33.779214	-84.411811	2021-04-04T03:21:17-0600
426878	33.786500	-84.445400	2021-04-04T03:21:11-0600
426879	33.779214	-84.411811	2021-04-04T03:21:07-0600

[426880 rows x 25 columns]

```
[132]: newdata.drop(['price', 'year', 'model'], axis=1)
```

```
[132]:
```

	id	url \
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...
2	7221797935	https://keys.craigslist.org/cto/d/summerland-k...
3	7222270760	https://worcester.craigslist.org/cto/d/west-br...
4	7210384030	https://greensboro.craigslist.org/cto/d/trinit...
...
426875	7301591192	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426876	7301591187	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426877	7301591147	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426878	7301591140	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426879	7301591129	https://wyoming.craigslist.org/ctd/d/atlanta-2...

	region	region_url \
0	prescott	https://prescott.craigslist.org
1	fayetteville	https://fayar.craigslist.org
2	florida keys	https://keys.craigslist.org
3	worcester / central MA	https://worcester.craigslist.org
4	greensboro	https://greensboro.craigslist.org
...
426875	wyoming	https://wyoming.craigslist.org
426876	wyoming	https://wyoming.craigslist.org
426877	wyoming	https://wyoming.craigslist.org
426878	wyoming	https://wyoming.craigslist.org
426879	wyoming	https://wyoming.craigslist.org

	manufacturer	condition	cylinders	fuel	odometer	title_status \
0	NaN	NaN	NaN	NaN	NaN	NaN

1	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN
...
426875	nissan	good	6 cylinders	gas	32226.0	clean
426876	volvo	good	NaN	gas	12029.0	clean
426877	cadillac	good	NaN	diesel	4174.0	clean
426878	lexus	good	6 cylinders	gas	30112.0	clean
426879	bmw	good	NaN	gas	22716.0	clean

	size	type	paint_color	\
0	NaN	NaN	NaN	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	NaN	NaN	NaN	
4	NaN	NaN	NaN	
...	
426875	NaN	sedan	NaN	
426876	NaN	sedan	red	
426877	NaN	hatchback	white	
426878	NaN	sedan	silver	
426879	NaN	coupe	NaN	

	image_url	\
0	NaN	
1	NaN	
2	NaN	
3	NaN	
4	NaN	
...	...	
426875	https://images.craigslist.org/00o0o_iiraFnHg8q...	
426876	https://images.craigslist.org/00x0x_15sbgnxCIS...	
426877	https://images.craigslist.org/00L0L_farM7bxnxR...	
426878	https://images.craigslist.org/00z0z_bKnIVGLkDT...	
426879	https://images.craigslist.org/00Y0Y_lEUocjyRxa...	

	description	county	state	\
0		NaN	NaN	az
1		NaN	NaN	ar
2		NaN	NaN	fl
3		NaN	NaN	ma
4		NaN	NaN	nc
...	
426875	Carvana is the safer way to buy a car During t...	NaN	NaN	wy
426876	Carvana is the safer way to buy a car During t...	NaN	NaN	wy
426877	Carvana is the safer way to buy a car During t...	NaN	NaN	wy

426878	Carvana is the safer way to buy a car During t...	NaN	wy
426879	Carvana is the safer way to buy a car During t...	NaN	wy

	lat	long	posting_date
0	NaN	NaN	NaN
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN
...
426875	33.786500	-84.445400	2021-04-04T03:21:31-0600
426876	33.786500	-84.445400	2021-04-04T03:21:29-0600
426877	33.779214	-84.411811	2021-04-04T03:21:17-0600
426878	33.786500	-84.445400	2021-04-04T03:21:11-0600
426879	33.779214	-84.411811	2021-04-04T03:21:07-0600

[426880 rows x 23 columns]

```
[133]: data[['year']]
```

```
[133]:
```

	year
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
...	...
426875	2019.0
426876	2020.0
426877	2020.0
426878	2018.0
426879	2019.0

[426880 rows x 1 columns]

```
[134]: data.year>2019
```

```
[134]:
```

0	False
1	False
2	False
3	False
4	False
...	...
426875	False
426876	True
426877	True
426878	False

```
426879    False
Name: year, Length: 426880, dtype: bool
```

```
[135]: data[data.year>2019]
```

```
[135]:
```

	id	url \
29	7316814989	https://auburn.craigslist.org/ctd/d/auburn-uni...
49	7314811916	https://auburn.craigslist.org/ctd/d/auburn-uni...
50	7314811909	https://auburn.craigslist.org/ctd/d/auburn-uni...
52	7314584315	https://auburn.craigslist.org/ctd/d/auburn-uni...
53	7314584322	https://auburn.craigslist.org/ctd/d/auburn-uni...
...
426865	7301949193	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426870	7301592119	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426872	7301591201	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426876	7301591187	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426877	7301591147	https://wyoming.craigslist.org/ctd/d/atlanta-2...

	region	region_url	price	year	manufacturer \
29	auburn	https://auburn.craigslist.org	39590	2020.0	chevrolet
49	auburn	https://auburn.craigslist.org	32990	2020.0	jeep
50	auburn	https://auburn.craigslist.org	38990	2020.0	ford
52	auburn	https://auburn.craigslist.org	31590	2020.0	mazda
53	auburn	https://auburn.craigslist.org	27990	2020.0	ford
...
426865	wyoming	https://wyoming.craigslist.org	34590	2020.0	alfa-romeo
426870	wyoming	https://wyoming.craigslist.org	22990	2020.0	hyundai
426872	wyoming	https://wyoming.craigslist.org	32590	2020.0	mercedes-benz
426876	wyoming	https://wyoming.craigslist.org	30590	2020.0	volvo
426877	wyoming	https://wyoming.craigslist.org	34990	2020.0	cadillac

	model	condition	cylinders	... size	type \
29	silverado 1500 crew	good	8 cylinders	... NaN	pickup
49	wrangler sport suv 2d	good	NaN	... NaN	SUV
50	f150 supercrew cab xlt	good	6 cylinders	... NaN	pickup
52	mx-5 miata club	good	NaN	... NaN	other
53	ranger supercab xl pickup	good	NaN	... NaN	pickup
...
426865	romeo giulia sedan 4d	good	NaN	... NaN	sedan
426870	sonata se sedan 4d	good	NaN	... NaN	sedan
426872	c-class c 300	good	NaN	... NaN	sedan
426876	s60 t5 momentum sedan 4d	good	NaN	... NaN	sedan
426877	xt4 sport suv 4d	good	NaN	... NaN	hatchback

	paint_color	image_url \
29	red	https://images.craigslist.org/01212_jjirIWa0y0...
49	silver	https://images.craigslist.org/00w0w_c22tgqPHhy...

```

50          NaN https://images.craigslist.org/00t0t_6t4cNaVZFh...
52          NaN https://images.craigslist.org/00w0w_d0WXY6wEb...
53         white https://images.craigslist.org/01313_9Zne268YnW...
...         ...
426865     black https://images.craigslist.org/01212_9YGV5SR0gp...
426870       blue https://images.craigslist.org/00u0u_eRGgkSql5l...
426872     white https://images.craigslist.org/00808_bYj0inoZ58...
426876       red https://images.craigslist.org/00x0x_15sbgnxCIS...
426877     white https://images.craigslist.org/00L0L_farM7bxnxR...

```

```

                                     description county state \
29      Carvana is the safer way to buy a car During t...   NaN   al
49      Carvana is the safer way to buy a car During t...   NaN   al
50      Carvana is the safer way to buy a car During t...   NaN   al
52      Carvana is the safer way to buy a car During t...   NaN   al
53      Carvana is the safer way to buy a car During t...   NaN   al
...         ...
426865  Carvana is the safer way to buy a car During t...   NaN   wy
426870  Carvana is the safer way to buy a car During t...   NaN   wy
426872  Carvana is the safer way to buy a car During t...   NaN   wy
426876  Carvana is the safer way to buy a car During t...   NaN   wy
426877  Carvana is the safer way to buy a car During t...   NaN   wy

```

```

               lat      long      posting_date
29      32.590000 -85.480000  2021-05-04T12:31:25-0500
49      32.590000 -85.480000  2021-04-30T10:11:01-0500
50      32.590000 -85.480000  2021-04-30T10:11:00-0500
52      32.590000 -85.480000  2021-04-29T18:11:10-0500
53      32.590000 -85.480000  2021-04-29T18:11:11-0500
...         ...
426865  33.779214 -84.411811  2021-04-05T02:41:27-0600
426870  33.786500 -84.445400  2021-04-04T03:40:17-0600
426872  33.779214 -84.411811  2021-04-04T03:21:35-0600
426876  33.786500 -84.445400  2021-04-04T03:21:29-0600
426877  33.779214 -84.411811  2021-04-04T03:21:17-0600

```

[21827 rows x 26 columns]

```
[136]: data[data.year>2019]['year']
```

```

[136]: 29      2020.0
        49      2020.0
        50      2020.0
        52      2020.0
        53      2020.0
        ...
426865  2020.0

```

```

426870    2020.0
426872    2020.0
426876    2020.0
426877    2020.0
Name: year, Length: 21827, dtype: float64

```

```
[137]: data[data.year>2019]['year'].count()
```

```
[137]: np.int64(21827)
```

```
[138]: data.loc[:, 'pricepermile']=data.loc[:, 'price']/data.loc[:, 'odometer']
```

```
[139]: data
```

```
[139]:
```

	id	url \
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...
2	7221797935	https://keys.craigslist.org/cto/d/summerland-k...
3	7222270760	https://worcester.craigslist.org/cto/d/west-br...
4	7210384030	https://greensboro.craigslist.org/cto/d/trinit...
...
426875	7301591192	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426876	7301591187	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426877	7301591147	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426878	7301591140	https://wyoming.craigslist.org/ctd/d/atlanta-2...
426879	7301591129	https://wyoming.craigslist.org/ctd/d/atlanta-2...

	region	region_url	price \
0	prescott	https://prescott.craigslist.org	6000
1	fayetteville	https://fayar.craigslist.org	11900
2	florida keys	https://keys.craigslist.org	21000
3	worcester / central MA	https://worcester.craigslist.org	1500
4	greensboro	https://greensboro.craigslist.org	4900
...
426875	wyoming	https://wyoming.craigslist.org	23590
426876	wyoming	https://wyoming.craigslist.org	30590
426877	wyoming	https://wyoming.craigslist.org	34990
426878	wyoming	https://wyoming.craigslist.org	28990
426879	wyoming	https://wyoming.craigslist.org	30590

	year	manufacturer	model	condition	cylinders \
0	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN
...

426875	2019.0	nissan	maxima s sedan 4d	good	6 cylinders
426876	2020.0	volvo	s60 t5 momentum sedan 4d	good	NaN
426877	2020.0	cadillac	xt4 sport suv 4d	good	NaN
426878	2018.0	lexus	es 350 sedan 4d	good	6 cylinders
426879	2019.0	bmw	4 series 430i gran coupe	good	NaN

	...	type	paint_color	\
--	-----	------	-------------	---

0	...	NaN	NaN	
1	...	NaN	NaN	
2	...	NaN	NaN	
3	...	NaN	NaN	
4	...	NaN	NaN	

...	
426875	...	sedan	NaN	
426876	...	sedan	red	
426877	...	hatchback	white	
426878	...	sedan	silver	
426879	...	coupe	NaN	

		image_url	\
--	--	-----------	---

0		NaN	
1		NaN	
2		NaN	
3		NaN	
4		NaN	

...
426875	https://images.craigslist.org/00o0o_iiraFnHg8q...	
426876	https://images.craigslist.org/00x0x_15sbgnxCIS...	
426877	https://images.craigslist.org/00L0L_farM7bxnxR...	
426878	https://images.craigslist.org/00z0z_bKnIVGLkDT...	
426879	https://images.craigslist.org/00Y0Y_lEUocjyRxa...	

		description	county	state	\
--	--	-------------	--------	-------	---

0		NaN	NaN	az	
1		NaN	NaN	ar	
2		NaN	NaN	fl	
3		NaN	NaN	ma	
4		NaN	NaN	nc	

...
426875	Carvana is the safer way to buy a car During t...	NaN	wy	
426876	Carvana is the safer way to buy a car During t...	NaN	wy	
426877	Carvana is the safer way to buy a car During t...	NaN	wy	
426878	Carvana is the safer way to buy a car During t...	NaN	wy	
426879	Carvana is the safer way to buy a car During t...	NaN	wy	

	lat	long	posting_date	pricepermine
0	NaN	NaN	NaN	NaN

1		NaN	NaN		NaN	NaN
2		NaN	NaN		NaN	NaN
3		NaN	NaN		NaN	NaN
4		NaN	NaN		NaN	NaN
...
426875	33.786500	-84.445400	2021-04-04T03:21:31-0600			0.732018
426876	33.786500	-84.445400	2021-04-04T03:21:29-0600			2.543021
426877	33.779214	-84.411811	2021-04-04T03:21:17-0600			8.382846
426878	33.786500	-84.445400	2021-04-04T03:21:11-0600			0.962739
426879	33.779214	-84.411811	2021-04-04T03:21:07-0600			1.346628

[426880 rows x 27 columns]

```
[140]: data['pricepermine']
```

```
[140]: 0          NaN
      1          NaN
      2          NaN
      3          NaN
      4          NaN
      ...
      426875    0.732018
      426876    2.543021
      426877    8.382846
      426878    0.962739
      426879    1.346628
      Name: pricepermine, Length: 426880, dtype: float64
```

```
[141]: data['region'].unique()
```

```
[141]: array(['prescott', 'fayetteville', 'florida keys',
      'worcester / central MA', 'greensboro', 'hudson valley',
      'medford-ashland', 'erie', 'el paso', 'bellingham',
      'skagit / island / SJI', 'la crosse', 'auburn', 'birmingham',
      'dothan', 'florence / muscle shoals', 'gadsden-anniston',
      'huntsville / decatur', 'mobile', 'montgomery', 'tuscaloosa',
      'anchorage / mat-su', 'fairbanks', 'kenai peninsula',
      'southeast alaska', 'flagstaff / sedona', 'mohave county',
      'phoenix', 'show low', 'sierra vista', 'tucson', 'yuma',
      'fort smith', 'jonesboro', 'little rock', 'texarkana',
      'bakersfield', 'chico', 'fresno / madera', 'gold country',
      'hanford-corcoran', 'humboldt county', 'imperial county',
      'inland empire', 'los angeles', 'mendocino county', 'merced',
      'modesto', 'monterey bay', 'orange county', 'palm springs',
      'redding', 'reno / tahoe', 'sacramento', 'san diego',
      'san luis obispo', 'santa barbara', 'santa maria', 'SF bay area',
      'siskiyou county', 'stockton', 'susenville', 'ventura county',
```

'visalia-tulare', 'yuba-sutter', 'boulder', 'colorado springs',
 'denver', 'eastern CO', 'fort collins / north CO', 'high rockies',
 'pueblo', 'western slope', 'eastern CT', 'hartford', 'new haven',
 'northwest CT', 'washington, DC', 'delaware', 'daytona beach',
 'ft myers / SW florida', 'gainesville', 'heartland florida',
 'jacksonville', 'lakeland', 'north central FL', 'ocala',
 'okaloosa / walton', 'orlando', 'panama city', 'pensacola',
 'sarasota-bradenton', 'south florida', 'space coast',
 'st augustine', 'tallahassee', 'tampa bay area', 'treasure coast',
 'albany', 'athens', 'atlanta', 'augusta', 'brunswick', 'columbus',
 'macon / warner robins', 'northwest GA', 'savannah / hinesville',
 'statesboro', 'valdosta', 'hawaii', 'boise', 'east idaho',
 'lewiston / clarkston', 'pullman / moscow',
 'spokane / coeur d'alene', 'twin falls', 'bloomington-normal',
 'champaign urbana', 'chicago', 'decatur', 'la salle co',
 'mattoon-charleston', 'peoria', 'quad cities, IA/IL', 'rockford',
 'southern illinois', 'springfield', 'st louis, MO', 'western IL',
 'bloomington', 'evansville', 'fort wayne', 'indianapolis',
 'kokomo', 'lafayette / west lafayette', 'muncie / anderson',
 'richmond', 'south bend / michiana', 'terre haute', 'ames',
 'cedar rapids', 'des moines', 'dubuque', 'fort dodge', 'iowa city',
 'mason city', 'omaha / council bluffs', 'sioux city',
 'southeast IA', 'waterloo / cedar falls', 'kansas city, MO',
 'lawrence', 'manhattan', 'northwest KS', 'salina', 'southeast KS',
 'southwest KS', 'topeka', 'wichita', 'bowling green',
 'eastern kentucky', 'huntington-ashland', 'lexington',
 'louisville', 'owensboro', 'western KY', 'baton rouge',
 'central louisiana', 'houma', 'lafayette', 'lake charles',
 'monroe', 'new orleans', 'shreveport', 'maine', 'annapolis',
 'baltimore', 'cumberland valley', 'eastern shore', 'frederick',
 'southern maryland', 'western maryland', 'boston',
 'cape cod / islands', 'south coast', 'western massachusetts',
 'ann arbor', 'battle creek', 'central michigan', 'detroit metro',
 'flint', 'grand rapids', 'holland', 'jackson', 'kalamazoo',
 'lansing', 'muskegon', 'northern michigan', 'port huron',
 'saginaw-midland-baycity', 'southwest michigan', 'the thumb',
 'upper peninsula', 'bemidji', 'brainerd', 'duluth / superior',
 'fargo / moorhead', 'mankato', 'minneapolis / st paul',
 'rochester', 'southwest MN', 'st cloud', 'gulfport / biloxi',
 'hattiesburg', 'meridian', 'north mississippi', 'southwest MS',
 'columbia / jeff city', 'joplin', 'kansas city', 'kirksville',
 'lake of the ozarks', 'southeast missouri', 'st joseph',
 'st louis', 'billings', 'bozeman', 'butte', 'eastern montana',
 'great falls', 'helena', 'kalispell', 'missoula', 'asheville',
 'boone', 'charlotte', 'eastern NC', 'hickory / lenoir',
 'outer banks', 'raleigh / durham / CH', 'wilmington',
 'winston-salem', 'grand island', 'lincoln', 'north platte',

```
'scottsbluff / panhandle', 'elko', 'las vegas', 'central NJ',
'jersey shore', 'north jersey', 'south jersey', 'albuquerque',
'clovis / portales', 'farmington', 'las cruces',
'roswell / carlsbad', 'santa fe / taos', 'binghamton', 'buffalo',
'catskills', 'chautauqua', 'elmira-corning', 'finger lakes',
'glens falls', 'ithaca', 'long island', 'new york city', 'oneonta',
'plattsburgh-adirondacks', 'potsdam-canton-massena', 'syracuse',
'twin tiers NY/PA', 'utica-rome-oneida', 'watertown',
'new hampshire', 'bismarck', 'grand forks', 'north dakota',
'akron / canton', 'ashtabula', 'chillicothe', 'cincinnati',
'cleveland', 'dayton / springfield', 'lima / findlay', 'mansfield',
'northern panhandle', 'parkersburg-marietta', 'sandusky', 'toledo',
'tuscarawas co', 'youngstown', 'zanesville / cambridge',
'fort smith, AR', 'lawton', 'northwest OK', 'oklahoma city',
'stillwater', 'texoma', 'tulsa', 'bend', 'corvallis/albany',
'east oregon', 'eugene', 'klamath falls', 'oregon coast',
'portland', 'roseburg', 'salem', 'altoona-johnstown', 'harrisburg',
'lanaster', 'lehigh valley', 'meadville', 'philadelphia',
'pittsburgh', 'poconos', 'reading', 'scranton / wilkes-barre',
'state college', 'williamsport', 'york', 'rhode island',
'charleston', 'columbia', 'florence', 'greenville / upstate',
'hilton head', 'myrtle beach', 'northeast SD',
'pierre / central SD', 'rapid city / west SD',
'sioux falls / SE SD', 'south dakota', 'chattanooga',
'clarksville', 'cookeville', 'knoxville', 'memphis', 'nashville',
'tri-cities', 'abilene', 'amarillo', 'austin',
'beaumont / port arthur', 'brownsville', 'college station',
'corpus christi', 'dallas / fort worth', 'deep east texas',
'del rio / eagle pass', 'galveston', 'houston',
'killean / temple / ft hood', 'laredo', 'lubbock',
'mcallen / edinburg', 'odessa / midland', 'san angelo',
'san antonio', 'san marcos', 'southwest TX', 'tyler / east TX',
'victoria', 'waco', 'wichita falls', 'logan', 'ogden-clearfield',
'provo / orem', 'salt lake city', 'st george', 'vermont',
'charlottesville', 'danville', 'fredericksburg', 'harrisonburg',
'lynchburg', 'new river valley', 'norfolk / hampton roads',
'roanoke', 'southwest VA', 'winchester',
'kennewick-pasco-richland', 'moses lake', 'olympic peninsula',
'seattle-tacoma', 'wenatchee', 'yakima', 'eastern panhandle',
'morgantown', 'southern WV', 'west virginia (old)',
'appleton-oshkosh-FDL', 'eau claire', 'green bay', 'janesville',
'kenosha-racine', 'madison', 'milwaukee', 'northern WI',
'sheboygan', 'wausau', 'wyoming'], dtype=object)
```

```
[142]: data['region'].nunique()
```

```
[142]: 404
```

```
[143]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 426880 entries, 0 to 426879
Data columns (total 27 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   id                    426880 non-null  int64  
 1   url                   426880 non-null  object  
 2   region                426880 non-null  object  
 3   region_url            426880 non-null  object  
 4   price                 426880 non-null  int64  
 5   year                  425675 non-null  float64 
 6   manufacturer          409234 non-null  object  
 7   model                 421603 non-null  object  
 8   condition             252776 non-null  object  
 9   cylinders             249202 non-null  object  
10   fuel                  423867 non-null  object  
11   odometer              422480 non-null  float64 
12   title_status          418638 non-null  object  
13   transmission          424324 non-null  object  
14   VIN                   265838 non-null  object  
15   drive                 296313 non-null  object  
16   size                  120519 non-null  object  
17   type                  334022 non-null  object  
18   paint_color           296677 non-null  object  
19   image_url             426812 non-null  object  
20   description           426810 non-null  object  
21   county                0 non-null       float64 
22   state                 426880 non-null  object  
23   lat                   420331 non-null  float64 
24   long                  420331 non-null  float64 
25   posting_date          426812 non-null  object  
26   pricepermine          421629 non-null  float64 
dtypes: float64(6), int64(2), object(19)
memory usage: 87.9+ MB
```

```
[144]: data['size'].unique()
```

```
[144]: array([nan, 'full-size', 'mid-size', 'compact', 'sub-compact'],
      dtype=object)
```

```
[145]: data['size'].nunique()
```

```
[145]: 4
```

```
[146]: data['size']=data['size'].map({'full-size':0,'mid-size':1,'compact':  
    ↪2,'sub-compact':3})
```

```
[147]: data.loc[50:100,'size']
```

```
[147]: 50      NaN  
      51      NaN  
      52      NaN  
      53      NaN  
      54      NaN  
      55      0.0  
      56      NaN  
      57      NaN  
      58      NaN  
      59      0.0  
      60      NaN  
      61      NaN  
      62      NaN  
      63      NaN  
      64      NaN  
      65      0.0  
      66      NaN  
      67      NaN  
      68      NaN  
      69      NaN  
      70      NaN  
      71      NaN  
      72      NaN  
      73      1.0  
      74      NaN  
      75      NaN  
      76      NaN  
      77      NaN  
      78      NaN  
      79      NaN  
      80      NaN  
      81      NaN  
      82      NaN  
      83      NaN  
      84      NaN  
      85      NaN  
      86      1.0  
      87      NaN  
      88      NaN  
      89      NaN  
      90      NaN  
      91      NaN
```

```

92      NaN
93      NaN
94      NaN
95      NaN
96      NaN
97      NaN
98      NaN
99      NaN
100     NaN
Name: size, dtype: float64

```

```
[148]: data
```

```

[148]:
      id                                     url \
0      7222695916 https://prescott.craigslist.org/cto/d/prescott...
1      7218891961 https://fayar.craigslist.org/ctd/d/bentonville...
2      7221797935 https://keys.craigslist.org/cto/d/summerland-k...
3      7222270760 https://worcester.craigslist.org/cto/d/west-br...
4      7210384030 https://greensboro.craigslist.org/cto/d/trinit...
...      ...
426875  7301591192 https://wyoming.craigslist.org/ctd/d/atlanta-2...
426876  7301591187 https://wyoming.craigslist.org/ctd/d/atlanta-2...
426877  7301591147 https://wyoming.craigslist.org/ctd/d/atlanta-2...
426878  7301591140 https://wyoming.craigslist.org/ctd/d/atlanta-2...
426879  7301591129 https://wyoming.craigslist.org/ctd/d/atlanta-2...

      region      region_url  price \
0      prescott https://prescott.craigslist.org 6000
1      fayetteville https://fayar.craigslist.org 11900
2      florida keys https://keys.craigslist.org 21000
3      worcester / central MA https://worcester.craigslist.org 1500
4      greensboro https://greensboro.craigslist.org 4900
...      ...
426875      wyoming https://wyoming.craigslist.org 23590
426876      wyoming https://wyoming.craigslist.org 30590
426877      wyoming https://wyoming.craigslist.org 34990
426878      wyoming https://wyoming.craigslist.org 28990
426879      wyoming https://wyoming.craigslist.org 30590

      year manufacturer      model condition  cylinders \
0      NaN      NaN      NaN      NaN      NaN
1      NaN      NaN      NaN      NaN      NaN
2      NaN      NaN      NaN      NaN      NaN
3      NaN      NaN      NaN      NaN      NaN
4      NaN      NaN      NaN      NaN      NaN
...      ...      ...      ...      ...
426875  2019.0      nissan      maxima s sedan 4d      good 6 cylinders

```

426876	2020.0	volvo	s60 t5 momentum sedan 4d	good	NaN
426877	2020.0	cadillac	xt4 sport suv 4d	good	NaN
426878	2018.0	lexus	es 350 sedan 4d	good	6 cylinders
426879	2019.0	bmw	4 series 430i gran coupe	good	NaN

	...	type	paint_color	\
0	...	NaN	NaN	
1	...	NaN	NaN	
2	...	NaN	NaN	
3	...	NaN	NaN	
4	...	NaN	NaN	
...	
426875	...	sedan	NaN	
426876	...	sedan	red	
426877	...	hatchback	white	
426878	...	sedan	silver	
426879	...	coupe	NaN	

	image_url	\
0	NaN	
1	NaN	
2	NaN	
3	NaN	
4	NaN	
...	...	
426875	https://images.craigslist.org/00o0o_iiraFnHg8q...	
426876	https://images.craigslist.org/00x0x_15sbgnxCIS...	
426877	https://images.craigslist.org/00L0L_farM7bxnxR...	
426878	https://images.craigslist.org/00z0z_bKnIVGLkDT...	
426879	https://images.craigslist.org/00Y0Y_lEUocjyRxa...	

	description	county	state	\
0		NaN	NaN	az
1		NaN	NaN	ar
2		NaN	NaN	fl
3		NaN	NaN	ma
4		NaN	NaN	nc
...	
426875	Carvana is the safer way to buy a car During t...	NaN	NaN	wy
426876	Carvana is the safer way to buy a car During t...	NaN	NaN	wy
426877	Carvana is the safer way to buy a car During t...	NaN	NaN	wy
426878	Carvana is the safer way to buy a car During t...	NaN	NaN	wy
426879	Carvana is the safer way to buy a car During t...	NaN	NaN	wy

	lat	long	posting_date	pricepermine
0	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN

2		NaN	NaN		NaN	NaN
3		NaN	NaN		NaN	NaN
4		NaN	NaN		NaN	NaN
...
426875	33.786500	-84.445400	2021-04-04T03:21:31-0600		0.732018	
426876	33.786500	-84.445400	2021-04-04T03:21:29-0600		2.543021	
426877	33.779214	-84.411811	2021-04-04T03:21:17-0600		8.382846	
426878	33.786500	-84.445400	2021-04-04T03:21:11-0600		0.962739	
426879	33.779214	-84.411811	2021-04-04T03:21:07-0600		1.346628	

[426880 rows x 27 columns]

[149]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 426880 entries, 0 to 426879
Data columns (total 27 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    426880 non-null  int64
1   url                   426880 non-null  object
2   region               426880 non-null  object
3   region_url           426880 non-null  object
4   price                426880 non-null  int64
5   year                 425675 non-null  float64
6   manufacturer          409234 non-null  object
7   model                421603 non-null  object
8   condition            252776 non-null  object
9   cylinders             249202 non-null  object
10  fuel                 423867 non-null  object
11  odometer             422480 non-null  float64
12  title_status         418638 non-null  object
13  transmission         424324 non-null  object
14  VIN                  265838 non-null  object
15  drive                296313 non-null  object
16  size                 120519 non-null  float64
17  type                 334022 non-null  object
18  paint_color          296677 non-null  object
19  image_url            426812 non-null  object
20  description           426810 non-null  object
21  county               0 non-null       float64
22  state                426880 non-null  object
23  lat                  420331 non-null  float64
24  long                 420331 non-null  float64
25  posting_date         426812 non-null  object
26  pricepermine         421629 non-null  float64
dtypes: float64(7), int64(2), object(18)
```


memory usage: 87.9+ MB

```
[150]: data.isnull()
```

```
[150]:
```

	id	url	region	region_url	price	year	manufacturer	model	\
0	False	False	False	False	False	True	True	True	
1	False	False	False	False	False	True	True	True	
2	False	False	False	False	False	True	True	True	
3	False	False	False	False	False	True	True	True	
4	False	False	False	False	False	True	True	True	
...	
426875	False	False	False	False	False	False	False	False	
426876	False	False	False	False	False	False	False	False	
426877	False	False	False	False	False	False	False	False	
426878	False	False	False	False	False	False	False	False	
426879	False	False	False	False	False	False	False	False	

	condition	cylinders	...	type	paint_color	image_url	description	\
0	True	True	...	True	True	True	True	
1	True	True	...	True	True	True	True	
2	True	True	...	True	True	True	True	
3	True	True	...	True	True	True	True	
4	True	True	...	True	True	True	True	
...	
426875	False	False	...	False	True	False	False	
426876	False	True	...	False	False	False	False	
426877	False	True	...	False	False	False	False	
426878	False	False	...	False	False	False	False	
426879	False	True	...	False	True	False	False	

	county	state	lat	long	posting_date	pricepermine
0	True	False	True	True	True	True
1	True	False	True	True	True	True
2	True	False	True	True	True	True
3	True	False	True	True	True	True
4	True	False	True	True	True	True
...
426875	True	False	False	False	False	False
426876	True	False	False	False	False	False
426877	True	False	False	False	False	False
426878	True	False	False	False	False	False
426879	True	False	False	False	False	False

[426880 rows x 27 columns]

```
[166]: data.isnull().sum()
```

```
[166]: id          0
      url          0
      region      0
      region_url  0
      price       0
      year        1205
      manufacturer 17646
      model        5277
      condition    174104
      cylinders    177678
      fuel         3013
      odometer     4400
      title_status 8242
      transmission 2556
      VIN          161042
      drive        130567
      size         306361
      type         92858
      paint_color  130203
      image_url    68
      description  70
      county       426880
      state        0
      lat          6549
      long         6549
      posting_date 68
      dtype: int64
```

```
[175]: data=pd.read_csv("vehicles.csv")
```

```
[176]: data.isnull().sum()
```

```
[176]: id          0
      url          0
      region      0
      region_url  0
      price       9
      year        1205
      manufacturer 17646
      model        5277
      condition    174104
      cylinders    177678
      fuel         3013
      odometer     4400
      title_status 8242
      transmission 2556
      VIN          161042
```

```
drive      130567
size       306361
type       92858
paint_color 130203
image_url   68
description 70
county     426880
state       0
lat        6549
long       6549
posting_date 68
dtype: int64
```

```
[ ]:
```

```
[ ]:
```