Assignment 1

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

In [2]: ```python
import numpy as np
```

In [3]: ```python
data=pd.read_csv('vehicles.csv')
```

In [4]: ```python
data.head()
```

Out[4]:

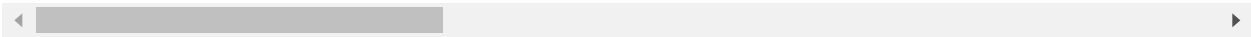|	id	url	region	region_url	pr
0	7222695916	https://prescott.craigslist.org/cto/d/prescott...	prescott	https://prescott.craigslist.org	6(
1	7218891961	https://fayar.craigslist.org/ctd/d/bentonville...	fayetteville	https://fayar.craigslist.org	11!
2	7221797935	https://keys.craigslist.org/cto/d/summerland-k...	florida keys	https://keys.craigslist.org	21(
3	7222270760	https://worcester.craigslist.org/cto/d/west-br...	worcester / central MA	https://worcester.craigslist.org	1!
4	7210384030	https://greensboro.craigslist.org/cto/d/trinit...	greensboro	https://greensboro.craigslist.org	4!

5 rows × 26 columns

In [5]: `data.tail()`

Out[5]:

	id	url	region	region_url	price	
1996	7308972392	https://dothan.craigslist.org/ctd/d/alachua-20...	dothan	https://dothan.craigslist.org	17860	2(
1997	7308971633	https://dothan.craigslist.org/ctd/d/alachua-20...	dothan	https://dothan.craigslist.org	54488	2(
1998	7308831024	https://dothan.craigslist.org/ctd/d/dothan-201...	dothan	https://dothan.craigslist.org	33990	2(
1999	7308801755	https://dothan.craigslist.org/ctd/d/alachua-20...	dothan	https://dothan.craigslist.org	3	2(
2000	7308801682	https://dothan.craigslist.org/ctd/d/alachua-20...	dothan	https://dothan.craigslist.org	18681	2(

5 rows × 26 columns

In [6]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2001 entries, 0 to 2000
Data columns (total 26 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   id             2001 non-null   int64
 1   url            2001 non-null   object
 2   region         2001 non-null   object
 3   region_url     2001 non-null   object
 4   price          2001 non-null   int64
 5   year           1952 non-null   float64
 6   manufacturer   1831 non-null   object
 7   model          1961 non-null   object
 8   condition      1122 non-null   object
 9   cylinders      1258 non-null   object
 10  fuel           1940 non-null   object
 11  odometer       1962 non-null   float64
 12  title_status   1871 non-null   object
 13  transmission   1973 non-null   object
 14  VIN            1268 non-null   object
 15  drive          1363 non-null   object
 16  size           493 non-null    object
 17  type           1380 non-null   object
 18  paint_color    1534 non-null   object
 19  image_url      1974 non-null   object
 20  description    1974 non-null   object
 21  county         0 non-null      float64
 22  state          2001 non-null   object
 23  lat            1962 non-null   float64
 24  long           1962 non-null   float64
 25  posting_date   1974 non-null   object
dtypes: float64(5), int64(2), object(19)
memory usage: 406.6+ KB
```

In [7]: `data.describe()`

Out[7]:

	id	price	year	odometer	county	lat	long
count	2.001000e+03	2.001000e+03	1952.000000	1.962000e+03	0.0	1962.000000	1962.000000
mean	7.308532e+09	5.625151e+05	2011.201332	2.038947e+05	NaN	33.735365	-85.549365
std	1.150144e+07	2.219042e+07	9.704075	9.915782e+05	NaN	2.167945	3.375666
min	7.208550e+09	0.000000e+00	1903.000000	0.000000e+00	NaN	26.021800	-122.693000
25%	7.305754e+09	5.500000e+03	2008.000000	2.967350e+04	NaN	32.922900	-86.817617
50%	7.310397e+09	1.720000e+04	2013.000000	8.799700e+04	NaN	33.455361	-86.737847
75%	7.313504e+09	2.998000e+04	2017.000000	1.572288e+05	NaN	33.736288	-85.480000
max	7.316878e+09	9.876543e+08	2021.000000	9.999999e+06	NaN	44.120200	-73.572300

In [8]: `data.isnull`

Out[8]: <bound method DataFrame.isnull of   id
         url   \
         0   7222695916   https://prescott.craigslist.org/cto/d/prescott... (https://pr
         escott.craigslist.org/cto/d/prescott...)
         1   7218891961   https://fayar.craigslist.org/ctd/d/bentonville... (https://fa
         yar.craigslist.org/ctd/d/bentonville...)
         2   7221797935   https://keys.craigslist.org/cto/d/summerland-k... (https://ke
         ys.craigslist.org/cto/d/summerland-k...)
         3   7222270760   https://worcester.craigslist.org/cto/d/west-br... (https://wo
         rcester.craigslist.org/cto/d/west-br...)
         4   7210384030   https://greensboro.craigslist.org/cto/d/trinit... (https://gr
         eensboro.craigslist.org/cto/d/trinit...)
         ...   .
         1996   7308972392   https://dothan.craigslist.org/ctd/d/alachua-20... (https://do
         than.craigslist.org/ctd/d/alachua-20...)
         1997   7308971633   https://dothan.craigslist.org/ctd/d/alachua-20... (https://do
         than.craigslist.org/ctd/d/alachua-20...)
         1998   7308831024   https://dothan.craigslist.org/ctd/d/dothan-201... (https://do
         than.craigslist.org/ctd/d/dothan-201...)
         1999   7308801755   https://dothan.craigslist.org/ctd/d/alachua-20... (https://do
         than.craigslist.org/ctd/d/alachua-20...)
         2000   7308801682   https://dothan.craigslist.org/ctd/d/alachua-20... (https://do
         than.craigslist.org/ctd/d/alachua-20...)

                               region   region_url   price   \
         0   prescott   https://prescott.craigslist.org (https://presco
         tt.craigslist.org)   6000
         1   fayetteville   https://fayar.craigslist.org (https://fayar.
         craigslist.org)   11900
         2   florida keys   https://keys.craigslist.org (https://keys.c
         raigslist.org)   21000
         3   worcester / central MA   https://worcester.craigslist.org (https://worces
         ter.craigslist.org)   1500
         4   greensboro   https://greensboro.craigslist.org (https://greens
         boro.craigslist.org)   4900
         ...   .
         1996   dothan   https://dothan.craigslist.org (https://dotha
         n.craigslist.org)   17860
         1997   dothan   https://dothan.craigslist.org (https://dotha
         n.craigslist.org)   54488
         1998   dothan   https://dothan.craigslist.org (https://dotha
         n.craigslist.org)   33990
         1999   dothan   https://dothan.craigslist.org (https://dotha
         n.craigslist.org)   3
         2000   dothan   https://dothan.craigslist.org (https://dotha
         n.craigslist.org)   18681

                 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
         ...   .
         1996 2014.0   jeep   grand cherokee limited   NaN   NaN

```
1997  2020.0        ford        f-150 lariat 4wd    NaN         NaN
1998  2017.0         ram  1500 crew cab tradesman   good  8 cylinders
1999  2020.0        ford         fusion se sedan    NaN         NaN
2000  2019.0      nissan     altima 2.5 sr sedan    NaN         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
...   ...  ...     ...         ...
1996  ...  NaN     NaN         NaN
1997  ...  NaN     NaN         NaN
1998  ...  NaN  pickup        blue
1999  ...  NaN     NaN         NaN
2000  ...  NaN     NaN         NaN

                                        image_url  \
0                                             NaN
1                                             NaN
2                                             NaN
3                                             NaN
4                                             NaN
...                                           ...
1996  https://images.craigslist.org/00O0O_ijP2Qx9vMi... (https://images.craigsl
ist.org/00O0O_ijP2Qx9vMi...)
1997  https://images.craigslist.org/00U0U_dTOexEOB2l... (https://images.craigsl
ist.org/00U0U_dTOexEOB2l...)
1998  https://images.craigslist.org/00Y0Y_jRbwm0dSjS... (https://images.craigsl
ist.org/00Y0Y_jRbwm0dSjS...)
1999  https://images.craigslist.org/00E0E_b3RHuW4jYv... (https://images.craigsl
ist.org/00E0E_b3RHuW4jYv...)
2000  https://images.craigslist.org/00101_lBMiBbuKBr... (https://images.craigsl
ist.org/00101_lBMiBbuKBr...)

                                      description county state  \
0                                             NaN    NaN    az
1                                             NaN    NaN    ar
2                                             NaN    NaN    fl
3                                             NaN    NaN    ma
4                                             NaN    NaN    nc
...                                           ...    ...   ...
1996  2014 JEEP GRAND CHEROKEE LIMITED ~ Hundreds of...    NaN    al
1997  2020 FORD F-150 LARIAT 4WD / F150 4X4 TRUCK ~ ...    NaN    al
1998  Carvana is the safer way to buy a car During t...    NaN    al
1999  2020 FORD FUSION SE SEDAN ~ Hundreds of NEW & ...    NaN    al
2000  2019 NISSAN ALTIMA 2.5 SR SEDAN ~ Hundreds of ...    NaN    al

            lat       long           posting_date
0           NaN        NaN                    NaN
1           NaN        NaN                    NaN
2           NaN        NaN                    NaN
3           NaN        NaN                    NaN
4           NaN        NaN                    NaN
...         ...        ...                    ...
1996  29.801374 -82.531052  2021-04-18T17:15:59-0500
```

```
1997   29.802119  -82.529850   2021-04-18T17:14:22-0500
1998   31.230000  -85.400000   2021-04-18T12:50:51-0500
1999   29.803337  -82.528074   2021-04-18T11:59:35-0500
2000   29.802566  -82.529678   2021-04-18T11:59:27-0500

[2001 rows x 26 columns]>
```

In [9]: `data.shape`

Out[9]: `(2001, 26)`

In [10]: `data.columns`

Out[10]:
```
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 [11]: `type("size")`

Out[11]: `str`

In [12]: `type("long")`

Out[12]: `str`

In [13]: `type("year")`

Out[13]: `str`

In [14]: `df =['year', 'price', 'size']`

In [15]:
```
data[df]
```

Out[15]:

	year	price	size
0	NaN	6000	NaN
1	NaN	11900	NaN
2	NaN	21000	NaN
3	NaN	1500	NaN
4	NaN	4900	NaN
...	...	...	...
1996	2014.0	17860	NaN
1997	2020.0	54488	NaN
1998	2017.0	33990	NaN
1999	2020.0	3	NaN
2000	2019.0	18681	NaN

2001 rows × 3 columns

In [16]:
```
data2 = data[df].head(10)
```

In [17]:
```
data2
```

Out[17]:

	year	price	size
0	NaN	6000	NaN
1	NaN	11900	NaN
2	NaN	21000	NaN
3	NaN	1500	NaN
4	NaN	4900	NaN
5	NaN	1600	NaN
6	NaN	1000	NaN
7	NaN	15995	NaN
8	NaN	5000	NaN
9	NaN	3000	NaN

In [18]:
```
caqu = pd.DataFrame(data, columns = ['state'])
```

```
In [19]: caqu_state = pd.get_dummies(caqu['state'])
         caqu_new = pd.concat([caqu, caqu_state], axis=1)
         print(caqu_new)
```

```
        state  al  ar  az  fl  ma  nc  ny  or  pa  tx  wa  wi
0          az   0   0   1   0   0   0   0   0   0   0   0   0
1          ar   0   1   0   0   0   0   0   0   0   0   0   0
2          fl   0   0   0   1   0   0   0   0   0   0   0   0
3          ma   0   0   0   0   1   0   0   0   0   0   0   0
4          nc   0   0   0   0   0   1   0   0   0   0   0   0
...       ...  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..
1996       al   1   0   0   0   0   0   0   0   0   0   0   0
1997       al   1   0   0   0   0   0   0   0   0   0   0   0
1998       al   1   0   0   0   0   0   0   0   0   0   0   0
1999       al   1   0   0   0   0   0   0   0   0   0   0   0
2000       al   1   0   0   0   0   0   0   0   0   0   0   0

[2001 rows x 13 columns]
```