Untitled

September 3, 2020

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

[2]:
```
other_path = "https://s3-api.us-geo.objectstorage.softlayer.net/cf-courses-data/
↪CognitiveClass/DA0101EN/auto.csv"
```

[4]:
```
df = pd.read_csv(other_path, header=None)
```

[5]:
```
print("The first 5 rows of the data frame")
```

```
The first 5 rows of the data frame
```

[6]:
```
df.head(5)
```

[6]:

	0	1	2	3	4	5	6	7	8	9	…	\
0	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	…	
1	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	…	
2	1	?	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	…	
3	2	164	audi	gas	std	four	sedan	fwd	front	99.8	…	
4	2	164	audi	gas	std	four	sedan	4wd	front	99.4	…	

	16	17	18	19	20	21	22	23	24	25
0	130	mpfi	3.47	2.68	9.0	111	5000	21	27	13495
1	130	mpfi	3.47	2.68	9.0	111	5000	21	27	16500
2	152	mpfi	2.68	3.47	9.0	154	5000	19	26	16500
3	109	mpfi	3.19	3.40	10.0	102	5500	24	30	13950
4	136	mpfi	3.19	3.40	8.0	115	5500	18	22	17450

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

[7]:
```
df.tail(10)
```

[7]:

	0	1	2	3	4	5	6	7	8	9	…	16	\
195	-1	74	volvo	gas	std	four	wagon	rwd	front	104.3	…	141	
196	-2	103	volvo	gas	std	four	sedan	rwd	front	104.3	…	141	
197	-1	74	volvo	gas	std	four	wagon	rwd	front	104.3	…	141	
198	-2	103	volvo	gas	turbo	four	sedan	rwd	front	104.3	…	130	
199	-1	74	volvo	gas	turbo	four	wagon	rwd	front	104.3	…	130	
200	-1	95	volvo	gas	std	four	sedan	rwd	front	109.1	…	141	

```
201  -1   95  volvo     gas  turbo  four  sedan  rwd  front  109.1  …  141
202  -1   95  volvo     gas    std  four  sedan  rwd  front  109.1  …  173
203  -1   95  volvo  diesel  turbo  four  sedan  rwd  front  109.1  …  145
204  -1   95  volvo     gas  turbo  four  sedan  rwd  front  109.1  …  141

       17    18    19    20   21    22  23  24     25
195  mpfi  3.78  3.15   9.5  114  5400  23  28  13415
196  mpfi  3.78  3.15   9.5  114  5400  24  28  15985
197  mpfi  3.78  3.15   9.5  114  5400  24  28  16515
198  mpfi  3.62  3.15   7.5  162  5100  17  22  18420
199  mpfi  3.62  3.15   7.5  162  5100  17  22  18950
200  mpfi  3.78  3.15   9.5  114  5400  23  28  16845
201  mpfi  3.78  3.15   8.7  160  5300  19  25  19045
202  mpfi  3.58  2.87   8.8  134  5500  18  23  21485
203   idi  3.01  3.40  23.0  106  4800  26  27  22470
204  mpfi  3.78  3.15   9.5  114  5400  19  25  22625

[10 rows x 26 columns]
```

[8]:
```
headers = ["symboling","normalized-losses","make","fuel-type","aspiration",␣
↪"num-of-doors","body-style",
        "drive-wheels","engine-location","wheel-base",␣
↪"length","width","height","curb-weight","engine-type",
        "num-of-cylinders",␣
↪"engine-size","fuel-system","bore","stroke","compression-ratio","horsepower",
        "peak-rpm","city-mpg","highway-mpg","price"]
```

[10]:
```
print("headers\n", headers)
```

```
headers
 ['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiration', 'num-of-
doors', 'body-style', 'drive-wheels', 'engine-location', 'wheel-base', 'length',
'width', 'height', 'curb-weight', 'engine-type', 'num-of-cylinders', 'engine-
size', 'fuel-system', 'bore', 'stroke', 'compression-ratio', 'horsepower',
'peak-rpm', 'city-mpg', 'highway-mpg', 'price']
```

[11]:
```
df.columns = headers
```

[12]:
```
df.head(10)
```

[12]:
```
   symboling normalized-losses         make fuel-type aspiration num-of-doors  \
0          3                 ?  alfa-romero       gas        std          two
1          3                 ?  alfa-romero       gas        std          two
2          1                 ?  alfa-romero       gas        std          two
3          2               164         audi       gas        std         four
4          2               164         audi       gas        std         four
5          2                 ?         audi       gas        std          two
```

```
6          1                158         audi       gas        std         four
7          1                  ?         audi       gas        std         four
8          1                158         audi       gas      turbo         four
9          0                  ?         audi       gas      turbo          two

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136
5        sedan          fwd           front        99.8  …          136
6        sedan          fwd           front       105.8  …          136
7        wagon          fwd           front       105.8  …          136
8        sedan          fwd           front       105.8  …          131
9    hatchback          4wd           front        99.5  …          131

   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm city-mpg  \
0         mpfi  3.47    2.68               9.0        111      5000       21
1         mpfi  3.47    2.68               9.0        111      5000       21
2         mpfi  2.68    3.47               9.0        154      5000       19
3         mpfi  3.19    3.40              10.0        102      5500       24
4         mpfi  3.19    3.40               8.0        115      5500       18
5         mpfi  3.19    3.40               8.5        110      5500       19
6         mpfi  3.19    3.40               8.5        110      5500       19
7         mpfi  3.19    3.40               8.5        110      5500       19
8         mpfi  3.13    3.40               8.3        140      5500       17
9         mpfi  3.13    3.40               7.0        160      5500       16

  highway-mpg  price
0          27  13495
1          27  16500
2          26  16500
3          30  13950
4          22  17450
5          25  15250
6          25  17710
7          25  18920
8          20  23875
9          22      ?

[10 rows x 26 columns]
```

[14]:
```
df.dropna(subset=["price"], axis=0)
```

[14]:
```
   symboling normalized-losses         make fuel-type aspiration  \
0          3                 ?  alfa-romero       gas        std
```

```
1            3                 ?  alfa-romero       gas        std
2            1                 ?  alfa-romero       gas        std
3            2               164         audi       gas        std
4            2               164         audi       gas        std
..         …                 …           …        …          …
200         -1                95        volvo       gas        std
201         -1                95        volvo       gas      turbo
202         -1                95        volvo       gas        std
203         -1                95        volvo    diesel      turbo
204         -1                95        volvo       gas      turbo

    num-of-doors   body-style drive-wheels engine-location  wheel-base  …  \
0            two  convertible          rwd           front        88.6  …
1            two  convertible          rwd           front        88.6  …
2            two    hatchback          rwd           front        94.5  …
3           four        sedan          fwd           front        99.8  …
4           four        sedan          4wd           front        99.4  …
..           …          …           …              …          …  …
200         four        sedan          rwd           front       109.1  …
201         four        sedan          rwd           front       109.1  …
202         four        sedan          rwd           front       109.1  …
203         four        sedan          rwd           front       109.1  …
204         four        sedan          rwd           front       109.1  …

     engine-size  fuel-system  bore  stroke compression-ratio horsepower  \
0            130         mpfi  3.47    2.68               9.0        111
1            130         mpfi  3.47    2.68               9.0        111
2            152         mpfi  2.68    3.47               9.0        154
3            109         mpfi  3.19    3.40              10.0        102
4            136         mpfi  3.19    3.40               8.0        115
..           …          …   …     …                 …         …
200          141         mpfi  3.78    3.15               9.5        114
201          141         mpfi  3.78    3.15               8.7        160
202          173         mpfi  3.58    2.87               8.8        134
203          145          idi  3.01    3.40              23.0        106
204          141         mpfi  3.78    3.15               9.5        114

     peak-rpm city-mpg highway-mpg  price
0        5000       21          27  13495
1        5000       21          27  16500
2        5000       19          26  16500
3        5500       24          30  13950
4        5500       18          22  17450
..        …       …          …    …
200      5400       23          28  16845
201      5300       19          25  19045
202      5500       18          23  21485
```

```
203      4800        26           27   22470
204      5400        19           25   22625

[205 rows x 26 columns]
```

[15]:
```
print(df.columns)
```

```
Index(['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiration',
       'num-of-doors', 'body-style', 'drive-wheels', 'engine-location',
       'wheel-base', 'length', 'width', 'height', 'curb-weight', 'engine-type',
       'num-of-cylinders', 'engine-size', 'fuel-system', 'bore', 'stroke',
       'compression-ratio', 'horsepower', 'peak-rpm', 'city-mpg',
       'highway-mpg', 'price'],
      dtype='object')
```

[16]:
```
df.to_csv("automobile.csv", index=False)
```

[17]:
```
df.dtypes
```

[17]:
```
symboling              int64
normalized-losses     object
make                  object
fuel-type             object
aspiration            object
num-of-doors          object
body-style            object
drive-wheels          object
engine-location       object
wheel-base           float64
length               float64
width                float64
height               float64
curb-weight            int64
engine-type           object
num-of-cylinders      object
engine-size            int64
fuel-system           object
bore                  object
stroke                object
compression-ratio    float64
horsepower            object
peak-rpm              object
city-mpg               int64
highway-mpg            int64
price                 object
dtype: object
```

[18]:
```
print(df.dtypes)
```

```
symboling              int64
normalized-losses     object
make                  object
fuel-type             object
aspiration            object
num-of-doors          object
body-style            object
drive-wheels          object
engine-location       object
wheel-base           float64
length               float64
width                float64
height               float64
curb-weight            int64
engine-type           object
num-of-cylinders      object
engine-size            int64
fuel-system           object
bore                  object
stroke                object
compression-ratio    float64
horsepower            object
peak-rpm              object
city-mpg               int64
highway-mpg            int64
price                 object
dtype: object
```

[20]:
```
df.describe()
```

[20]:

	symboling	wheel-base	length	width	height	\
count	205.000000	205.000000	205.000000	205.000000	205.000000	
mean	0.834146	98.756585	174.049268	65.907805	53.724878	
std	1.245307	6.021776	12.337289	2.145204	2.443522	
min	-2.000000	86.600000	141.100000	60.300000	47.800000	
25%	0.000000	94.500000	166.300000	64.100000	52.000000	
50%	1.000000	97.000000	173.200000	65.500000	54.100000	
75%	2.000000	102.400000	183.100000	66.900000	55.500000	
max	3.000000	120.900000	208.100000	72.300000	59.800000	

	curb-weight	engine-size	compression-ratio	city-mpg	highway-mpg
count	205.000000	205.000000	205.000000	205.000000	205.000000
mean	2555.565854	126.907317	10.142537	25.219512	30.751220
std	520.680204	41.642693	3.972040	6.542142	6.886443
min	1488.000000	61.000000	7.000000	13.000000	16.000000

```
25%    2145.000000    97.000000          8.600000   19.000000    25.000000
50%    2414.000000   120.000000          9.000000   24.000000    30.000000
75%    2935.000000   141.000000          9.400000   30.000000    34.000000
max    4066.000000   326.000000         23.000000   49.000000    54.000000
```

[21]:
```
df.describe(include = "all")
```

[21]:
```
         symboling normalized-losses    make fuel-type aspiration  \
count   205.000000               205     205       205        205
unique         NaN                52      22         2          2
top            NaN                 ?  toyota       gas        std
freq           NaN                41      32       185        168
mean      0.834146               NaN     NaN       NaN        NaN
std       1.245307               NaN     NaN       NaN        NaN
min      -2.000000               NaN     NaN       NaN        NaN
25%       0.000000               NaN     NaN       NaN        NaN
50%       1.000000               NaN     NaN       NaN        NaN
75%       2.000000               NaN     NaN       NaN        NaN
max       3.000000               NaN     NaN       NaN        NaN

       num-of-doors body-style drive-wheels engine-location  wheel-base  …  \
count           205        205          205             205  205.000000  …
unique            3          5            3               2         NaN  …
top            four      sedan          fwd           front         NaN  …
freq            114         96          120             202         NaN  …
mean            NaN        NaN          NaN             NaN   98.756585  …
std             NaN        NaN          NaN             NaN    6.021776  …
min             NaN        NaN          NaN             NaN   86.600000  …
25%             NaN        NaN          NaN             NaN   94.500000  …
50%             NaN        NaN          NaN             NaN   97.000000  …
75%             NaN        NaN          NaN             NaN  102.400000  …
max             NaN        NaN          NaN             NaN  120.900000  …

        engine-size  fuel-system  bore  stroke compression-ratio horsepower  \
count    205.000000          205   205     205        205.000000        205
unique          NaN            8    39      37               NaN         60
top             NaN         mpfi  3.62    3.40               NaN         68
freq            NaN           94    23      20               NaN         19
mean     126.907317          NaN   NaN     NaN         10.142537        NaN
std       41.642693          NaN   NaN     NaN          3.972040        NaN
min       61.000000          NaN   NaN     NaN          7.000000        NaN
25%       97.000000          NaN   NaN     NaN          8.600000        NaN
50%      120.000000          NaN   NaN     NaN          9.000000        NaN
75%      141.000000          NaN   NaN     NaN          9.400000        NaN
max      326.000000          NaN   NaN     NaN         23.000000        NaN

        peak-rpm    city-mpg highway-mpg price
```

```
count        205  205.000000  205.000000   205
unique        24         NaN         NaN   187
top         5500         NaN         NaN     ?
freq          37         NaN         NaN     4
mean         NaN   25.219512   30.751220   NaN
std          NaN    6.542142    6.886443   NaN
min          NaN   13.000000   16.000000   NaN
25%          NaN   19.000000   25.000000   NaN
50%          NaN   24.000000   30.000000   NaN
75%          NaN   30.000000   34.000000   NaN
max          NaN   49.000000   54.000000   NaN

[11 rows x 26 columns]
```

[22]:
```
df[['length', 'compression-ratio']].describe()
```

[22]:
```
           length  compression-ratio
count  205.000000         205.000000
mean   174.049268          10.142537
std     12.337289           3.972040
min    141.100000           7.000000
25%    166.300000           8.600000
50%    173.200000           9.000000
75%    183.100000           9.400000
max    208.100000          23.000000
```

[23]:
```
df.info
```

[23]:
```
<bound method DataFrame.info of      symboling normalized-losses         make
fuel-type aspiration  \
0            3                 ?  alfa-romero       gas        std
1            3                 ?  alfa-romero       gas        std
2            1                 ?  alfa-romero       gas        std
3            2               164         audi       gas        std
4            2               164         audi       gas        std
..         …               …           …       …        …
200         -1                95        volvo       gas        std
201         -1                95        volvo       gas      turbo
202         -1                95        volvo       gas        std
203         -1                95        volvo    diesel      turbo
204         -1                95        volvo       gas      turbo

    num-of-doors   body-style drive-wheels engine-location  wheel-base  …  \
0            two  convertible          rwd           front        88.6  …
1            two  convertible          rwd           front        88.6  …
2            two    hatchback          rwd           front        94.5  …
3           four        sedan          fwd           front        99.8  …
```

```
4           four        sedan          4wd           front        99.4  …
..           …          …          …              …          …  …
200         four        sedan          rwd           front       109.1  …
201         four        sedan          rwd           front       109.1  …
202         four        sedan          rwd           front       109.1  …
203         four        sedan          rwd           front       109.1  …
204         four        sedan          rwd           front       109.1  …

     engine-size  fuel-system  bore  stroke compression-ratio horsepower  \
0            130         mpfi  3.47    2.68               9.0        111
1            130         mpfi  3.47    2.68               9.0        111
2            152         mpfi  2.68    3.47               9.0        154
3            109         mpfi  3.19    3.40              10.0        102
4            136         mpfi  3.19    3.40               8.0        115
..           …          …   …     …                 …        …
200          141         mpfi  3.78    3.15               9.5        114
201          141         mpfi  3.78    3.15               8.7        160
202          173         mpfi  3.58    2.87               8.8        134
203          145          idi  3.01    3.40              23.0        106
204          141         mpfi  3.78    3.15               9.5        114

     peak-rpm city-mpg highway-mpg  price
0        5000       21          27  13495
1        5000       21          27  16500
2        5000       19          26  16500
3        5500       24          30  13950
4        5500       18          22  17450
..        …      …          …    …
200      5400       23          28  16845
201      5300       19          25  19045
202      5500       18          23  21485
203      4800       26          27  22470
204      5400       19          25  22625

[205 rows x 26 columns]>
```

[ ]: