DA0101EN-Review-Introduction

July 21, 2019

```
< a \ href="http://cocl.us/DA0101EN_NotbookLink_Top"> \\ < img \ src="https://s3-api.us-geo.objectstorage.softlayer.net/cf-courses-data/CognitiveClass/DA0101EN/Images </a> <math display="block">< /a>
```

Data Analysis with Python

Introduction

Welcome!

In this section, you will learn how to approach data acquisition in various ways, and obtain necessary insights from a dataset. By the end of this lab, you will successfully load the data into Jupyter Notebook, and gain some fundamental insights via Pandas Library.

Table of Contents

Data Acquisition

Basic Insight of Dataset

Estimated Time Needed: 10 min

Data Acquisition

There are various formats for a dataset, .csv, .json, .xlsx etc. The dataset can be stored in different places, on your local machine or sometimes online. In this section, you will learn how to load a dataset into our Jupyter Notebook. In our case, the Automobile Dataset is an online source, and it is in CSV (comma separated value) format. Let's use this dataset as an example to practice data reading.

data source: https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.data

data type: csv

The Pandas Library is a useful tool that enables us to read various datasets into a data frame; our Jupyter notebook platforms have a built-in Pandas Library so that all we need to do is import Pandas without installing.

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

Read Data

We use pandas.read_csv() function to read the csv file. In the bracket, we put the file path along with a quotation mark, so that pandas will read the file into a data frame from that address. The file path can be either an URL or your local file address. Because the data does not include headers, we can add an argument headers = None inside the read_csv() method, so that pandas will not automatically set the first row as a header. You can also assign the dataset to any variable you create.

This dataset was hosted on IBM Cloud object click HERE for free storage.

```
[2]: # Import pandas library import pandas as pd

# Read the online file by the URL provides above, and assign it to variable "df" other_path = "https://s3-api.us-geo.objectstorage.softlayer.net/cf-courses-data/CognitiveClass/
→DA0101EN/auto.csv"

df = pd.read_csv(other_path, header=None)
```

After reading the dataset, we can use the dataframe.head(n) method to check the top n rows of the dataframe; where n is an integer. Contrary to dataframe.head(n), dataframe.tail(n) will show you the bottom n rows of the dataframe.

```
[3]: # show the first 5 rows using dataframe.head() method print("The first 5 rows of the dataframe") df.head(5)
```

The first 5 rows of the dataframe

```
[3]:
             alfa-romero gas std
                                   two convertible rwd front 88.6 ...
             alfa-romero gas std
                                  two convertible rwd front 88.6 ...
    2
           ? alfa-romero gas std
                                         hatchback rwd front 94.5 ...
                                  two
    3
       2 164
                   audi gas std four
                                          sedan fwd front 99.8 ...
       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 \text{ rows x } 26 \text{ columns}]
```

3 13 W.S. 11 **2**0 00141111115

Question #1:

check the bottom 10 rows of data frame "df".

```
[5]: # Write your code below and press Shift+Enter to execute

df.tail(10)
```

```
[5]:
       0
                2
                            4
                                5
                                     6
                                         7
                                              8
                                                    9
           1
                      3
                                                            16 \
    195 - 1
            74 volvo
                        gas
                              std four wagon rwd front 104.3 ... 141
    196 -2 103 volvo
                              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
                            turbo four wagon rwd front 104.3 ... 130
                        gas
    200 - 1
            95 volvo
                              std four sedan rwd front 109.1 ... 141
                        gas
                        gas turbo four sedan rwd front 109.1 ... 141
    201 - 1
            95 volvo
    202 - 1
            95 volvo
                              std four sedan rwd front 109.1 ... 173
                        gas
```

```
203 -1 95 volvo diesel turbo four sedan rwd front 109.1 ... 145
       95 volvo
                   gas turbo four sedan rwd front 109.1 ... 141
         18
              19
                      21
                           22 23 24
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
                   7.5 162 5100 17 22 18950
199 mpfi 3.62 3.15
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]

Question #1 Answer:

Run the code below for the solution!

Double-click here for the solution.

Add Headers

Take a look at our dataset; pandas automatically set the header by an integer from 0.

To better describe our data we can introduce a header, this information is available at: https://archive.ics.uci.edu/ml/datasets/Automobile

Thus, we have to add headers manually.

Firstly, we create a list "headers" that include all column names in order. Then, we use dataframe.columns = headers to replace the headers by the list we created.

```
[6]: # create headers list
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"]
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']

We replace headers and recheck our data frame

```
[7]: df.columns = headers \\ df.head(10)
```

```
[7]:
        symboling normalized-losses
                                               make fuel-type aspiration num-of-doors \
                               ? alfa-romero
     0
               3
                                                    gas
                                                               \operatorname{std}
                                                                           two
                               ? alfa-romero
     1
               3
                                                               \operatorname{std}
                                                                           two
                                                     gas
     2
               1
                               ? alfa-romero
                                                               \operatorname{std}
                                                                           two
                                                    gas
     3
               2
                             164
                                        audi
                                                                         four
                                                    gas
                                                              \operatorname{std}
               2
     4
                             164
                                        audi
                                                              \operatorname{std}
                                                                         four
                                                    gas
               2
                               ?
     5
                                        audi
                                                             \operatorname{std}
                                                                         two
                                                   gas
     6
               1
                             158
                                        audi
                                                              \operatorname{std}
                                                                         four
                                                   gas
     7
               1
                               ?
                                        audi
                                                                        four
                                                             \operatorname{std}
                                                   gas
     8
               1
                             158
                                        audi
                                                   gas
                                                            turbo
                                                                          four
     9
               0
                               ?
                                        audi
                                                           turbo
                                                                          two
                                                  gas
         body-style drive-wheels engine-location wheel-base ...
                                                                         engine-size \
        convertible
                             rwd
                                          front
                                                       88.6 ...
                                                                         130
        convertible
                             rwd
                                          front
                                                       88.6 ...
     1
                                                                         130
     2
          hatchback
                             rwd
                                           front
                                                       94.5 \dots
                                                                          152
     3
             sedan
                            \operatorname{fwd}
                                         front
                                                      99.8 ...
                                                                        109
     4
             sedan
                            4 wd
                                          front
                                                      99.4 \dots
                                                                         136
     5
             sedan
                            fwd
                                         front
                                                      99.8 ...
                                                                        136
     6
                                                     105.8 ...
             sedan
                            fwd
                                         front
                                                                         136
     7
                             fwd
                                                      105.8 \dots
                                                                         136
             wagon
                                          front
     8
             sedan
                            fwd
                                         front
                                                     105.8 ...
                                                                         131
     9
          hatchback
                              4 wd
                                           front
                                                        99.5 \dots
                                                                          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
                                              9.0
                                                                 5000
                                                                            21
                             2.68
                                                         111
     2
              mpfi 2.68
                             3.47
                                              9.0
                                                        154
                                                                 5000
                                                                             19
     3
                                                                             24
              mpfi 3.19
                             3.40
                                             10.0
                                                         102
                                                                  5500
     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
               22
     9
                      ?
```

[10 rows x 26 columns]

we can drop missing values along the column "price" as follows

[8]:	8]: df.dropna(subset=["price"], axis=0)										
[8]:		symboling normalized-losses make fuel-type aspiration \									
	0 3 ? alfa-romero gas std										
	1	3		lfa-romero	$_{ m gas}$	std					
	2	1		lfa-romero	$_{ m gas}$	std					
	3	2	164	audi	gas	std					
	4	2	164	audi	$_{ m gas}$	std					
	5	2	?	audi	gas	std					
	6	1	158	audi	gas	std					
	7	1	?	audi	gas	std					
	8	1	158	audi	gas	turbo					
	9	0	?	audi	gas	turbo					
	10	2	192	bmw	gas	std					
	11	0	192	bmw	gas	std					
	12	0	188	bmw	gas	std					
	13	0	188	bmw	gas	std					
	14	1	?	$_{ m bmw}$	gas	std					
	15	0	?	$_{ m bmw}$	gas	std					
	16	0	?	$_{ m bmw}$	gas	std					
	17	0	?	$_{ m bmw}$	gas	std					
	18	2	121	chevrolet	gas	std					
	19	1	98	chevrolet	${ m gas}$	std					
	20	0	81	chevrolet	${ m gas}$	std					
	21	1	118	dodge	$_{ m gas}$	std					
	22	1	118	dodge	gas	std					
	23	1	118	dodge	gas	turbo					
	24	1	148	dodge	$_{ m gas}$	std					
	25	1	148	dodge	${ m gas}$	std					
	26	1	148	dodge	gas	std					
	27	1	148	dodge	gas	turbo					
	28	-1	110	dodge	gas	std					
	29	3	145	dodge	gas	turbo					
	175	-1	65	toyota	gas	std					
	176	-1	65	toyota	gas	std					
	177	-1	65	toyota	gas	std					
	178		197	toyota	gas	std					
	179		197	toyota	gas	std					
	180		90	toyota	gas	std					
	181		?	toyota	gas	std					
	182		122	volkswagen	diesel	std					
	183		122	volkswagen	gas	std					
	184	2	94	volkswagen	diesel	std					

185	2	94	volkswagen	gas	std
186	2	94	volkswagen	gas	std
187	2	94	volkswagen	diesel	turbo
188	2	94	volkswagen	gas	std
189	3	?	volkswagen	gas	std
190	3	256	volkswagen	gas	std
191	0	?	volkswagen	gas	std
192	0	?	volkswagen	diesel	turbo
193	0	?	volkswagen	gas	std
194	-2	103	volvo	gas	std
195	-1	74	volvo	gas	std
196	-2	103	volvo	gas	std
197	-1	74	volvo	gas	std
198	-2	103	volvo	gas	turbo
199	-1	74	volvo	gas	turbo
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 front88.6 ... 88.6 ... 1 convertible rwd front two 2 hatchback rwd front $94.5 \dots$ two 3 four sedan fwd front 99.8 ... $99.4 \dots$ 4 four sedan 4 wdfront 5 fwd front 99.8 ... two sedan 6 four sedan fwd front $105.8 \dots$ 7 105.8 ... four wagon fwd front 105.8 ... 8 sedan four fwd front 9 hatchback 99.5 ... 4 wdfront two $101.2 \dots$ 10 sedan front two rwd sedan 101.2 ... 11 four rwd front12 sedan rwd front $101.2 \dots$ two 13 four sedan rwd front 101.2 ... 14 four sedan front $103.5 \dots$ rwd 15 four sedan rwd front $103.5 \dots$ $103.5 \dots$ 16 sedan rwd front two 17 four sedanrwd front 110.0 ... 18 hatchback fwd front 88.4 ... two 19 two hatchback fwd front $94.5 \dots$ 20 four sedanfwd front $94.5 \dots$ 21 hatchback fwd front $93.7 \dots$ two 22 hatchback fwd $93.7 \dots$ front two 23 $93.7 \dots$ hatchback fwd two front 24 hatchback $93.7 \dots$ four fwd front

```
93.7 \dots
25
          four
                     sedan
                                   fwd
                                               front
26
                                                           93.7 \dots
          four
                     sedan
                                  fwd
                                               front
            ?
                                              front
27
                    sedan
                                  fwd
                                                          93.7 \dots
28
                                                           103.3 ...
          four
                     wagon
                                   fwd
                                                front
29
                  hatchback
                                    fwd
                                                 front
                                                            95.9 \dots
           two
                                                            . . .
          . . .
                     . . .
                                                            102.4 ...
175
                  hatchback
          four
                                    fwd
                                                 front
                                                           102.4 ...
176
          four
                     sedan
                                   fwd
                                               front
                                                            102.4 ...
177
                                    fwd
          four
                  hatchback
                                                 front
178
                  hatchback
                                     rwd
                                                 front
                                                            102.9 ...
           two
179
                  hatchback
                                                 front
                                                            102.9 \dots
           two
                                     rwd
180
          four
                     sedan
                                   rwd
                                                front
                                                           104.5 \dots
181
          four
                     wagon
                                   rwd
                                                front
                                                           104.5 \dots
                                                           97.3 ...
182
           two
                     sedan
                                   fwd
                                                front
                                                front
                                                           97.3 ...
183
                     sedan
                                   fwd
           two
                                                           97.3 ...
184
          four
                     sedan
                                   fwd
                                                front
185
                                                           97.3 ...
          four
                     sedan
                                   fwd
                                                front
186
                                   fwd
                                                           97.3 ...
          four
                     sedan
                                               front
                                                           97.3 ...
187
          four
                     sedan
                                   fwd
                                                front
                                                           97.3 ...
188
          four
                     sedan
                                   fwd
                                                front
189
                                    fwd
                                                            94.5 \dots
           two
                convertible
                                                front
                                                             94.5 \dots
190
                  hatchback
                                    fwd
                                                 front
           two
191
          four
                     sedan
                                   fwd
                                                front
                                                           100.4 ...
                                                           100.4 ...
192
          four
                                   fwd
                                                front
                     sedan
193
          four
                                   fwd
                                                front
                                                           100.4 \dots
                     wagon
194
          four
                     sedan
                                   rwd
                                                front
                                                           104.3 \dots
195
          four
                     wagon
                                   rwd
                                                front
                                                           104.3 \dots
196
          four
                                   rwd
                                                front
                                                           104.3 \dots
                     sedan
                                                           104.3 ...
197
          four
                     wagon
                                   rwd
                                                front
198
                                                           104.3 \dots
          four
                     sedan
                                   rwd
                                                front
199
          four
                                                front
                                                           104.3 ...
                     wagon
                                   rwd
200
                                                           109.1 ...
          four
                     sedan
                                   rwd
                                                front
                                                           109.1 ...
201
          four
                     sedan
                                                front
                                   rwd
202
                                                           109.1 ...
          four
                     sedan
                                   rwd
                                                front
                                                           109.1 ...
203
          four
                     sedan
                                   rwd
                                                front
204
          four
                     sedan
                                   rwd
                                                front
                                                           109.1 ...
                fuel-system bore stroke compression-ratio horsepower \
    engine-size
0
          130
                     mpfi 3.47
                                   2.68
                                                  9.00
                                                             111
                     mpfi 3.47
1
          130
                                   2.68
                                                  9.00
                                                             111
2
          152
                     mpfi 2.68
                                                  9.00
                                                             154
                                   3.47
3
          109
                     mpfi 3.19
                                   3.40
                                                  10.00
                                                              102
4
          136
                     mpfi 3.19
                                                  8.00
                                   3.40
                                                             115
5
          136
                     mpfi 3.19
                                   3.40
                                                  8.50
                                                             110
6
          136
                     mpfi 3.19
                                                  8.50
                                   3.40
                                                             110
7
          136
                     mpfi 3.19
                                                  8.50
                                   3.40
                                                             110
8
                     mpfi 3.13
          131
                                   3.40
                                                  8.30
                                                             140
```

9	131	mpfi 3.13	3.40	7.00	160
10	108	mpfi 3.50	2.80	8.80	101
11	108	mpfi 3.50	2.80	8.80	101
12	164	mpfi 3.31	3.19	9.00	121
13	164	mpfi 3.31	3.19	9.00	121
		-			
14	164	mpfi 3.31	3.19	9.00	121
15	209	mpfi 3.62	3.39	8.00	182
16	209	mpfi 3.62	3.39	8.00	182
17	209	mpfi 3.62	3.39	8.00	182
18	61	2bbl 2.91	3.03	9.50	48
19	90	2bbl 3.03	3.11	9.60	70
20	90	$2bbl \ 3.03$	3.11	9.60	70
21	90	$2bbl \ \ 2.97$	3.23	9.41	68
22	90	$2bbl \ \ 2.97$	3.23	9.40	68
23	98	mpfi 3.03	3.39	7.60	102
24	90	2bbl 2.97	3.23	9.40	68
25	90		3.23		68
		2bbl 2.97		9.40	
26	90	2bbl 2.97	3.23	9.40	68
27	98	mpfi 3.03	3.39	7.60	102
28	122	2bbl 3.34	3.46	8.50	88
29	156	mfi 3.60	3.90	7.00	145
175	122	mpfi 3.31	3.54	8.70	92
176	122	mpfi 3.31	3.54	8.70	92
177	122	mpfi 3.31	3.54	8.70	92
178	171	mpfi 3.27	3.35	9.30	161
179	171	mpfi 3.27	3.35	9.30	161
180	171	mpfi 3.27	3.35	9.20	156
		•			
181	161	mpfi 3.27	3.35	9.20	156
182	97	idi 3.01	3.40	23.00	52
183	109	mpfi 3.19	3.40	9.00	85
184	97	idi 3.01	3.40	23.00	52
185	109	mpfi 3.19	3.40	9.00	85
186	109	mpfi 3.19	3.40	9.00	85
187	97	idi 3.01	3.40	23.00	68
188	109	mpfi 3.19	3.40	10.00	100
189	109	mpfi 3.19	3.40	8.50	90
190	109	mpfi 3.19	3.40	8.50	90
191	136	mpfi 3.19	3.40	8.50	110
192	97	idi 3.01	3.40	23.00	68
193	109	mpfi 3.19	3.40	9.00	88
193 194		=		9.50	
	141	mpfi 3.78	3.15		114
195	141	mpfi 3.78	3.15	9.50	114
196	141	mpfi 3.78	3.15	9.50	114
197	141	mpfi 3.78	3.15	9.50	114
198	130	mpfi 3.62	3.15	7.50	162
199	130	mpfi 3.62	3.15	7.50	162

200	141	mpfi 3	3.78	3.15	9.50	114
201	141	mpfi 3	3.78	3.15	8.70	160
202	173	mpfi 3	3.58	2.87	8.80	134
203	145	idi 3	.01	3.40	23.00	106
204	141	mpfi :	3.78	3.15	9.50	114

peak-rpm city-mpg highway-mpg price

	peak-rpm	city-mpg	highwa	y-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	
5	5500	19	25	15250	
6	5500	19	25	17710	
7	5500	19	25	18920	
8	5500	17	20	23875	
9	5500	16	22	?	
10	5800	23	29	16430	
11	5800	23	29	16925	
12	4250	21	28	20970	
13	4250	21	28	21105	
14	4250	20	25	24565	
15	5400	16	22	30760	
16	5400	16	22	41315	
17	5400	15	20	36880	
18	5100	47	53	5151	
19	5400	38	43	6295	
20	5400	38	43	6575	
21	5500	37	41	5572	
22	5500	31	38	6377	
23	5500	24	30	7957	
24	5500	31	38	6229	
25	5500	31	38	6692	
26	5500	31	38	7609	
27	5500	24	30	8558	
28	5000	24	30	8921	
29	5000	19	24	12964	
175	4200	27	32	9988	
176	4200	27	32	10898	
177	4200	27	32	11248	
178	5200	20	24	16558	
179	5200	19	24	15998	
180	5200	20	24	15690	
181	5200	19	24	15750	
182	4800	37	46	7775	
183	5250	27	34	7975	