GUIDE

Table of Contents

- introducing csv and pd.read_csv
- how to get first and last few rows of dataframe
- iloc
- very basic loc

- Assignment at end

In [1]:
```
import numpy as np
import pandas as pd

# we be only working w csvs in this one
```

In [4]:
```
"""
the csv farm_mkt.csv is a file containing 5 rows and
4 columns:
Location
Intersection
Day
Type
"""
farm = pd.read_csv('farm_mkt.csv')
"""
this line opens the csv and keeps it as a dataframe
called farm. As soon as u use this command, u can apply whichever pandas dataframe
commands you'd like to use on this data. Because farm is now a dataframe.
"""
```

Out[4]:
```
"\nthis line opens the csv and keeps it as a dataframe\ncalled farm. As soon as
u use this command, u can apply whichever pandas dataframe commands you'd like t
o use on this data. Because farm is now a dataframe.\n"
```

In [11]:
```
# for instance, if you want to see the first few (by default, its 5) lines of the d
ata: use df.head()
farm.head()
# if you want to see a certain number of the first lines:
```

Out[11]:

	LOCATION	INTERSECTION	DAY	TYPE
0	Bridgeport	35th & Wallace	Saturday	Weekly
1	Beverly	W 95th St & S Longwood Dr	Sunday	Weekly
2	Daley Plaza	W Washington St & N Dearborn St	Thursday	Downtown
3	Federal Plaza	W Adams St & S Dearborn St	Tuesday	Downtown
4	Pullman	E 111th Pl & S Cottage Grove Ave	Wednesday	Weekly

In [12]:
```
print('Print the first 7 lines')
farm.head(7)
```

```
Print the first 7 lines
```

Out[12]:

	LOCATION	INTERSECTION	DAY	TYPE
0	Bridgeport	35th & Wallace	Saturday	Weekly
1	Beverly	W 95th St & S Longwood Dr	Sunday	Weekly
2	Daley Plaza	W Washington St & N Dearborn St	Thursday	Downtown
3	Federal Plaza	W Adams St & S Dearborn St	Tuesday	Downtown
4	Pullman	E 111th Pl & S Cottage Grove Ave	Wednesday	Weekly
5	Hyde Park	53rd & Hyde Park	Thursday	Weekly
6	Lincoln Park	W Armitage Ave & N Orchard St	Saturday	Weekly

In [13]:
```
# if u want to see the last few (by default, 5) lines of the data:
print('Print the last 5 lines')
farm.tail()
```

```
Print the last 5 lines
```

Out[13]:

	LOCATION	INTERSECTION	DAY	TYPE
19	La Follette Park	1333 N Laramie Ave	Wednesday	Weekly
20	Columbus Park	Harrison & Central	Tuesday	Weekly
21	Austin Town Center	Lake & Central	Saturday	Weekly
22	Argyle Market	Argyle and Broadway	Thursday	Weekly
23	Devon Community Monthly Market	2720 W Devon	Varies	Monthly

In [14]:
```
# if you want to see a certain number of the last lines:
print('Print the last 7 lines')
farm.tail(7)
```

```
Print the last 7 lines
```

Out[14]:

	LOCATION	INTERSECTION	DAY	TYPE
17	Southport Market	1420 West Grace (Blaine School)	Saturday	Weekly
18	Logan Square Night Market	3107 W Logan Blvd	Wednesday	Independent
19	La Follette Park	1333 N Laramie Ave	Wednesday	Weekly
20	Columbus Park	Harrison & Central	Tuesday	Weekly
21	Austin Town Center	Lake & Central	Saturday	Weekly
22	Argyle Market	Argyle and Broadway	Thursday	Weekly
23	Devon Community Monthly Market	2720 W Devon	Varies	Monthly

In [20]:
```
# if you'd like to see certain lines, you have to use loc or iloc methods, a concep
t we'll work on later:
print('Here are the columns')
farm.columns
# this command above will tell you what columns are there are:
```

```
Here are the columns
```

Out[20]:
```
Index(['LOCATION', 'INTERSECTION', 'DAY', 'TYPE'], dtype='object')
```

In [21]:
```
print('The datatype info of the columns')
farm.info()
# this one tells you what type of columns you have
```

```
The datatype info of the columns
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24 entries, 0 to 23
Data columns (total 4 columns):
LOCATION        24 non-null object
INTERSECTION    24 non-null object
DAY             24 non-null object
TYPE            24 non-null object
dtypes: object(4)
memory usage: 848.0+ bytes
```

In [19]:
```
print('Statistics of the data')
farm.describe()
# this command gives descriptive statistics that summarize the central tendency, di
spersion and shape of a dataset's distribution, excluding NaN values.
```

```
Statistics of the data
```

Out[19]:

	LOCATION	INTERSECTION	DAY	TYPE
count	24	24	24	24
unique	24	24	6	4
top	Devon Community Monthly Market	Lake & Central	Saturday	Weekly
freq	1	1	8	18

In [22]:
```
# Sometimes you need to select certain things, or find them somehow. You can use lo
cation commands for that
# they're called iloc and loc. They have quite a learning curve, so you'll have to
mess with them a lot

# how to select a column:
print('One way of selecting it')
farm['LOCATION']
```

```
One way of selecting it
```

Out[22]:
```
0                                      Bridgeport
1                                         Beverly
2                                     Daley Plaza
3                                  Federal Plaza
4                                         Pullman
5                                     Hyde Park
6                                   Lincoln Park
7                                 Lincoln Square
8       Museum of Contemporary Art/Streeterville
9                                  Division Street
10                    Garfield Park Conservatory
11                                    Northcenter
12                                 Printer's Row
13                                   South Shore
14                            Willis Tower Plaza
15                         Wicker Park & Bucktown
16    West Humboldt Park Farmers Market & Bazaar
17                               Southport Market
18                     Logan Square Night Market
19                               La Follette Park
20                                  Columbus Park
21                              Austin Town Center
22                                   Argyle Market
23                Devon Community Monthly Market
Name: LOCATION, dtype: object
```

In [23]:
```
# another way of selecting it:
print('Another way of Selecting it')
farm.LOCATION
```

```
Another way of Selecting it
```

Out[23]:
```
0                                      Bridgeport
1                                         Beverly
2                                     Daley Plaza
3                                  Federal Plaza
4                                         Pullman
5                                      Hyde Park
6                                    Lincoln Park
7                                  Lincoln Square
8       Museum of Contemporary Art/Streeterville
9                                  Division Street
10                    Garfield Park Conservatory
11                                   Northcenter
12                                 Printer's Row
13                                   South Shore
14                            Willis Tower Plaza
15                         Wicker Park & Bucktown
16    West Humboldt Park Farmers Market & Bazaar
17                               Southport Market
18                     Logan Square Night Market
19                               La Follette Park
20                                  Columbus Park
21                               Austin Town Center
22                                   Argyle Market
23                 Devon Community Monthly Market
Name: LOCATION, dtype: object
```

Look at the following image to understand iloc

In [28]:
```
# now suppose you wanted a specific record, with all the columns:
print('One specific record using iloc')
farm.iloc[1:2]
# iloc chooses a specific row, and the num after the colon tells how many more rows
after the previous number to print:
# it gets rows (or columns) at particular positions in the index (so it only takes
integers).
```

```
One specific record using iloc
```

Out[28]:

	LOCATION	INTERSECTION	DAY	TYPE
1	Beverly	W 95th St & S Longwood Dr	Sunday	Weekly

In [29]:
```
# let us start from beginning: print the second row onwards:
print('2nd record onwards')
farm.iloc[2:,]
```

```
2nd record onwards
```

Out[29]:

	LOCATION	INTERSECTION	DAY	TYPE
2	Daley Plaza	W Washington St & N Dearborn St	Thursday	Downtown
3	Federal Plaza	W Adams St & S Dearborn St	Tuesday	Downtown
4	Pullman	E 111th Pl & S Cottage Grove Ave	Wednesday	Weekly
5	Hyde Park	53rd & Hyde Park	Thursday	Weekly
6	Lincoln Park	W Armitage Ave & N Orchard St	Saturday	Weekly
7	Lincoln Square	W Leland Ave & N Western Ave	Tuesday	Weekly
8	Museum of Contemporary Art/Streeterville	E Chicago Ave & N Mies Van Der Rohe Way	Tuesday	Weekly
9	Division Street	W Division St & N Dearborn St	Saturday	Weekly
10	Garfield Park Conservatory	300 N. Central Park	Sunday	Weekly
11	Northcenter	W Belle Plaine Ave & N Damen Ave	Saturday	Weekly
12	Printer's Row	S Dearborn St & W Polk St	Saturday	Weekly
13	South Shore	77th and Lake Michigan	Sunday	Weekly
14	Willis Tower Plaza	W Jackson Blvd & S Wacker Dr	Thursday	Downtown
15	Wicker Park & Bucktown	N Wicker Park & Damen Ave	Sunday	Weekly
16	West Humboldt Park Farmers Market & Bazaar	3601 W Chicago Ave	Saturday	Independent
17	Southport Market	1420 West Grace (Blaine School)	Saturday	Weekly
18	Logan Square Night Market	3107 W Logan Blvd	Wednesday	Independent
19	La Follette Park	1333 N Laramie Ave	Wednesday	Weekly
20	Columbus Park	Harrison & Central	Tuesday	Weekly
21	Austin Town Center	Lake & Central	Saturday	Weekly
22	Argyle Market	Argyle and Broadway	Thursday	Weekly
23	Devon Community Monthly Market	2720 W Devon	Varies	Monthly

In [30]:
```
# print the second row to 15th row
print('Row 1 to Row 15')
farm.iloc[2:16]
```

```
Row 1 to Row 15
```

Out[30]:

	LOCATION	INTERSECTION	DAY	TYPE
2	Daley Plaza	W Washington St & N Dearborn St	Thursday	Downtown
3	Federal Plaza	W Adams St & S Dearborn St	Tuesday	Downtown
4	Pullman	E 111th Pl & S Cottage Grove Ave	Wednesday	Weekly
5	Hyde Park	53rd & Hyde Park	Thursday	Weekly
6	Lincoln Park	W Armitage Ave & N Orchard St	Saturday	Weekly
7	Lincoln Square	W Leland Ave & N Western Ave	Tuesday	Weekly
8	Museum of Contemporary Art/Streeterville	E Chicago Ave & N Mies Van Der Rohe Way	Tuesday	Weekly
9	Division Street	W Division St & N Dearborn St	Saturday	Weekly
10	Garfield Park Conservatory	300 N. Central Park	Sunday	Weekly
11	Northcenter	W Belle Plaine Ave & N Damen Ave	Saturday	Weekly
12	Printer's Row	S Dearborn St & W Polk St	Saturday	Weekly
13	South Shore	77th and Lake Michigan	Sunday	Weekly
14	Willis Tower Plaza	W Jackson Blvd & S Wacker Dr	Thursday	Downtown
15	Wicker Park & Bucktown	N Wicker Park & Damen Ave	Sunday	Weekly

In [26]:
```
# print up till the 5th row
print(farm.iloc[:6])
```

```
        LOCATION                      INTERSECTION        DAY      TYPE
0     Bridgeport                    35th & Wallace   Saturday    Weekly
1        Beverly         W 95th St & S Longwood Dr     Sunday    Weekly
2    Daley Plaza   W Washington St & N Dearborn St   Thursday  Downtown
3  Federal Plaza        W Adams St & S Dearborn St    Tuesday  Downtown
4        Pullman  E 111th Pl & S Cottage Grove Ave  Wednesday    Weekly
5      Hyde Park                  53rd & Hyde Park   Thursday    Weekly
```

In [32]:
```
# now suppose you want to load a specific row from a specific column:
# choose column as farm['somethin'] or as farm.somethin
print("Specific Column's record")
farm.TYPE.iloc[2:3]
# this above command (the iloc one specifically) prints only one commmand
```

```
Specific Column's record
```

Out[32]:
```
2    Downtown
Name: TYPE, dtype: object
```

In [33]:
```
# now we do loc:

# this one should just print one record, #4
print('One Record, num #4')
farm.loc[4]
```

```
One Record, num #4
```

Out[33]:
```
LOCATION                                  Pullman
INTERSECTION    E 111th Pl & S Cottage Grove Ave
DAY                                     Wednesday
TYPE                                       Weekly
Name: 4, dtype: object
```

In [36]:
```
print('Location of the 4th record')
farm.LOCATION.loc[4]
# you can get just the location of the 4th record w this
```

```
Location of the 4th record
```

Out[36]:
```
'Pullman'
```

Assignment:

1. Find and download a csv for analysis. (It can b p much w.e u want, don't get a dataset that's too big though, you don't want anything to crash. Keep it at a limit of 20 rows, and max 10 columns.
2. Get the head, tail and 5 random rows in between of the dataframe
3. Get a set of 3 rows from a specific column in your dataset