DATA INGESTION AND INSPECTION

REVIEW OF PANDAS DATAFRAMES

Date	Open	High	Low	Close	Volume	Adj Close
2014-09-16	99.80	101.26	98.89	100.86	66818200	100.86
2014-09-15	102.81	103.05	101.44	101.63	61216500	101.63
2014-09-12	101.21	102.19	101.08	101.66	62626100	101.66
...	...	...	...	...	...	...

type(AAPL)
pandas.core.frame.DataFrame

AAPL**.shape**
(8514, 6)

INDEXES AND COLUMNS

AAPL**.index**
DatetimeIndex(["2014-09-16", "2014-09-15", ... "1980-12-12"], dtype="datetime64[ns]", name="Date", length=8514, freq=None)
type(AAPL.index)
pandas.tseries.index.DatetimeIndex

AAPL**.columns**
Index(["Open", "High", "Low", "Close", "Volume", "Adj Close"], dtype="object")
type(AAPL.columns)
pandas.indexes.base.Index

SLICING

AAPL**.iloc[:5, :]**
AAPL**.head(5)**

AAPL**.iloc[-5:, :]**
AAPL**.tail(5)**

BROADCASTING

AAPL**.iloc[::3, -1] = np.nan**

Date	Open	High	Low	Close	Volume	Adj Close
2014-09-16	99.80	101.26	98.89	100.86	66818200	NaN
2014-09-15	102.81	103.05	101.44	101.63	61216500	101.63
2014-09-12	101.21	102.19	101.08	101.66	62626100	101.66
2014-09-11	100.41	101.44	99.62	101.43	62353100	NaN
2014-09-10	98.01	101.11	97.76	101.00	100741900	101.00
2014-09-09	99.08	103.08	96.14	97.99	189560600	97.99
2014-09-08	99.30	99.31	98.05	98.36	46277800	NaN

SERIES

low = AAPL["Low"]
type(low)
pandas.core.series.Series

lows = low.values
type(lows)
numpy.ndarray

BUILDING DATAFRAMES FROM SCRATCH

	weekday	city	visitors	signups
0	Sun	Austin	139	7
1	Sun	Dallas	237	12
2	Mon	Austin	326	3
3	Mon	Dallas	456	5

DATAFRAMES FROM CSV FILES

users = pd.**read_csv**("datasets/users.csv"**, index_col=0)**

DATAFRAMES FROM DICT

data = {"weekday": ["Sun", "Sun", "Mon", "Mon"],
"city": ["Austin", "Dallas", "Austin", "Dallas",
"visitors": [139, 237, 326, 456],
"signups": [7, 12, 3, 5]}
users = pd.DataFrame(data**)**

cities = ["Austin", "Dallas", "Austin", "Dallas"]
signups = [7, 12, 3, 5]
visitors = [139, 237, 326, 456]
weekdays = ["Sun", "Sun", "Mon", "Mon"]

list_labels = ["city", "signups", "visitors", "weekday"**]**
list_cols = [cities, signups, visitors, weekdays**]**

zipped = list(zip(list_labels, list_cols**))**
data = dict(zipped**)**
users = pd.DataFrame(data**)**

BROADCASTING

users**["fees"] = 0**

	city	signups	visitors	weekday	fees
0	Austin	7	139	Sun	0
1	Dallas	12	237	Sun	0
2	Austin	3	326	Mon	0
3	Dallas	5	456	Mon	0

heights = [59.0, 65.2, 62.9, 65.4, 63.7, 65.7, 64.1]
data = {"height" : heights, "sex" : "M"**}**
results = pd.DataFrame(data**)**

	height	sex
0	59.0	M
1	65.2	M
2	62.9	M
3	65.4	M
4	63.7	M
5	65.7	M
6	64.1	M

results**.columns** = ["height (in)", "sex"]
results**.index** = ["A", "B", "C", "D", "E", "F", "G"]

	height (in)	sex
A	59.0	M
B	65.2	M
C	62.9	M
D	65.4	M
E	63.7	M
F	65.7	M
G	64.1	M