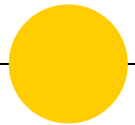
Pandas



History: Panel Data

Developer Wes McKinney started working on pandas in 2008 while at AQR Capital Management out of the need for a high performance, flexible tool to perform quantitative analysis on financial data. Before leaving AQR he was able to convince management to allow him to open source the library.

Another AQR employee, Chang She, joined the effort in 2012 as the second major contributor to the library.

In 2015, pandas signed on as a fiscally sponsored project of NumFOCUS, a 501(c)(3) nonprofit charity in the United States.[6]

What is it?

- As a data analyst with Python in your tool kit, one of the libraries you will use most often will be Pandas.

- Pandas is a useful library that makes data wrangling, transformation, and analysis easier and more intuitive

- You have to import the library :

```
>>>  import pandas as pd
```

Data Structures

- The primary data structures in Pandas are Series and DataFrames

```
>>> a = pd.Series(np.random.random(10))
>>> print(a)

>>> colnames = ['Column1','Column2','Column3','Column4','Column5']
>>> df = pd.DataFrame(np.random.random((10,5)), columns=colnames)
>>> df.head(5)
```

Convert data structures to df

```
>>> lst = [208500, 181500, 223500, 140000, 250000, 143000, 307000, 200000, 129900, 118000]

>>> price_df = pd.DataFrame(lst, columns=['SalePrice'])

>>> colnames = ['LotSize','Neighborhood','YearBuilt','Quality','SalePrice']
>>> pd.DataFrame(lst_lst, columns=colnames)

>>> pd.Dataframe(house_dict)

>>> house_df = pd.DataFrame.from_dict(house_dict, orient='index')
```

Math functions

```
# Sum

>>> df['Column'].sum()

# Mean

>>> df['Column'].mean()


# Max

>>> df['Column'].max()
```

Importing / Exporting data