Basics in Data-Science:

Tools to use:

1) Pandas

→ Python library that makes handling tabular data easier

→ Introduces 'data frames' & 'series' that allows you to slice up & dice columns of Info.

2) Numpy

→ We can use "NumPy arrays"
↳ multidimentional array objects

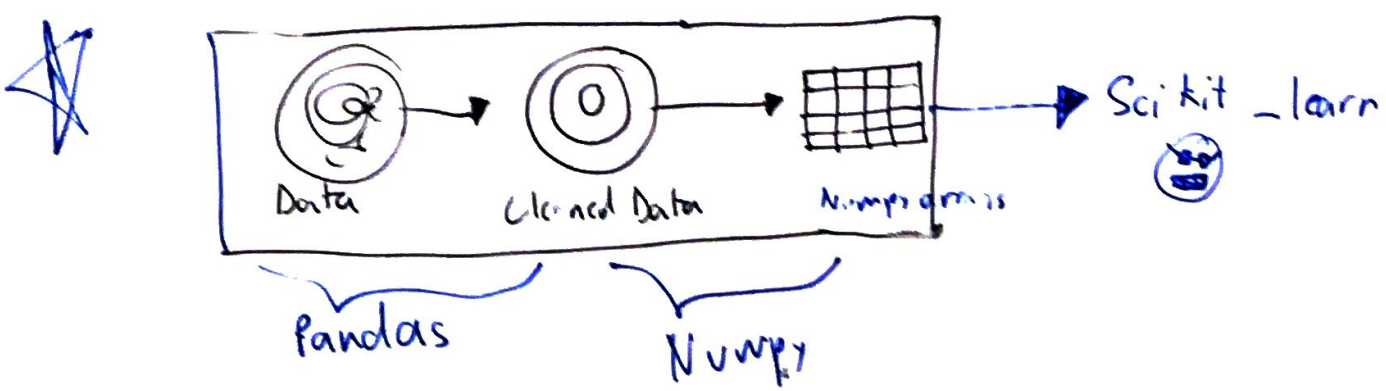
→ easy to create Pandas DataFrame from Numpy array

→ Pandas DataFrames can be cast as Numpy Arrays.

3) Scikit Learn

→ ML library.

→ takes Numpy arrays as its input.



Pandas Data frame functions:

1) head

eg df.head(10)
↑ gives first 10 rows
→ helps visualize data

2) tail(x)
→ last x

3) shape
output
eg df.shape (13, 7)
rows cols
→ gives dimentions

4) size
eg 91
→ rows * Cols count

5) len(df)

eg = 13 no. of rows

6) df.Columns.
eg => gives out named Cols
Index(['Count', 'Yrs of exp', 'Employed'], dtype='object')

7) df['x']
→ gives data in 'x' Col

8) [:x]

df['Hired'][:5]
→ gives first 5 cols

9) `df[['x', 'y']]`

→ basically extention of (7), with 2 cols

10) `value_counts()`

→ breaks down no. of unique values in a given column into series using `value_counts()`

eg `degree_counts = df['Level of Education'].value_counts`

output:

BS:	7
phd	4
MS	2

11) `sort_values(['x'])`

→ sorts data by 'x' coln

12) `.plot(kind='bar')`

→ draws a bar graph of data

Types of Data

1) Numerical 2) Categorical 3) Ordinal

1) Numerical

→ Quantitative measurement

eg heights of people, stock prices, etc

→ Discrete data

- Integer based ; often of same event.

eg 'heads' flip.

→ Continuous Data

- Infinit no. of possible values . ∞ amt of precision

eg → time taken to check out

→ rain on a day

2) Categorical data:

→ Quantitative data that has no inherent mathematical meaning

eg Gender, Yes/No (Binary) data, Race, Country of citizenship, etc.

→ We can assign no. to categories in order to represent them more compactly, but no.s don't have mathematical meaning.

3) Ordinal data.

→ mix of Numerical & Categorical data.

→ Categorical data that has mathematical meaning.

eg: movie ratings 1-5 stars