**Python fundamentals:**

**Data Types:**

* ‘Hello world’: str,
* None: NoneType,
* 15: int,
* 20.14: float,
* function,
* (1, ‘b’, 3, 5): tuple

/\* immutable, can contain mixed data types, written between parentheses () \*/

* [1, ‘b’, 3, 5]: list

/\* mutable (add by append() function), can contain mixed data types, written between square brackets [] \*/

* {‘name’: ‘Asmaa’, ‘surname’: ‘Mirkhan’}: dictionary

/\* labeled list, no order, similar to JSON \*/

**Classes:**

Note: Objects do not have private or protected members.

Constructor: (Syntax: \_\_init\_\_) it is not necessary

**Map Function:**

Function signature: map(function, iterable , …)

Note: Useful for analysis.

**BASIC DATA PROCESSING WITH PANDAS:**

/\* loc and iloc for row based querying and square brackets [] for column based querying \*/

**Series in pandas:**

Between list and dictionary, sorted but has keys(labeled), labels are indices starting from 0 and the last key is dtype key (data type).

**Data Frames in pandas:**

Briefly, it is a two-dimensional series.

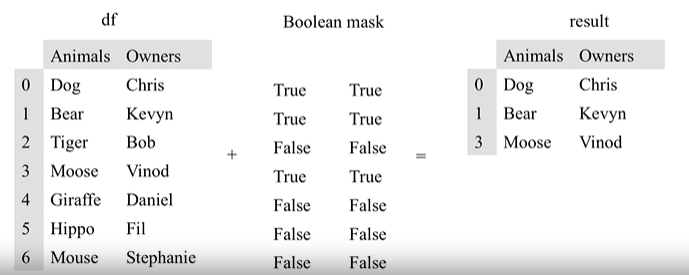
/\* like SQL database tables\*/

**BASIC DATA PROCESSING WITH PANDAS:**

**Boolean Masking:**

Useful for removing unwanted data

e.g



* Data frames offer functions like SQL queries such as join, groupby,…

**Scales in pandas:**

* **Ratio Scale**
  + Units are equally spaced
  + Math. ops. are valid
  + Ex: height, weight
* **Interval Scale**
  + Units are equally spaced
  + No true zero
  + Ex: temperature scale
* **Ordinal Scale**
  + Order is important
  + Not evenly spaced
  + Common in machine learning
  + Ex: letter grades (AA, BA, ..)
* **Nominal Scale**
  + Like categories, order is not important
  + Ex: Teams of sports

**Statistical analysis:**

**Distribution:** Set of all possible random variables (Flipping coins)

(a shape that describes the probability of a value being pulled when we sample a population)

/\*

* Binomial Distribution:

N trials of binary experiment

* Uniform Distribution:

Random variable takes same value (does not change)

* Normal (Gaussian) Distribution:

Has to symmetric curves \*/

/\* Expected value:

The mean of the random variable if we did an infinite number of trials\*/

Variance: a measure of how broadly values of samples are spread out from the mean.

**Characteristics of a distribution:**

* **Central tendency**

Measures: mode, median or mean.

About really where the bulk probability is in the distribution.

* **Variability:**

Measures: standard deviation (which is a measure of how different each item in our sample is from the mean), interquartile range.

**Kurtosis** (the shape of the tales): negative value means that the curve is slightly more flat than a normal distribution and positive value means that the curve is more peaky than a normal distribution.