📝PANDAS THEORY AND EXAMPLE

PANDAS SERIES

* It is like column in table
* It is one dimensional array which holds any kind of datatypes, means, it may hold string, character, numerical values etc.

⚠️*Refer: 01\_simple example of pandas.py*

⭕ NB: The key of the dictionary become the label, while we change the dictionary into the pandas series.

* To select specified data from the dictionary, we can select is by mentioning as argument in the pandas series. Index argument is there to specify the key.
* Refer: *01\_simple example of pandas.py*

Data frame in pandas:

* It is like the whole table, while series is a column
* It is like two-dimensional data structure, like table with column and row.
* It is data set, multidimensional table is called data frame.
* NB: loc is used to locate one or more rows
  + When we select a single column using loc function, pandas series will be returned.

NB: JSON file have the same format with python dictionary.

Data cleaning:

* Fixing bad data in our dataset
* Bad data:
  + Empty cell
  + Data in wrong format
  + Wrong data
  + Duplicate data

Empty cell can potentially gives wrong output.

The panda’s package is the most important tool at the disposal of Data Scientists and Analysts working in Python today. The powerful machine learning and glamorous visualization tools may get all the attention, but pandas is the backbone of most data projects.

📝[pandas] is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. — Wikipedia

Pandas is built on top of the NumPy package, meaning a lot of the structure of NumPy is used or replicated in Pandas. Data in pandas is often used to feed statistical analysis in SciPy, plotting functions from Matplotlib, and machine learning algorithms in Scikit-learn.

**Core components of pandas: Series and Data Frames:**

The primary two components of pandas are the Series and DataFrame. A Series is essentially a column, and a DataFrame is a multi-dimensional table made up of a collection of Series.

Table

Description automatically generated