Lesson 3: Pandas

Course Agenda

- Python Crash Course
- Data Analysis:
  - NumPy
  - **Pandas**
- Data Visualization:
  - Matplotlib
  - Seaborn
  - Pandas
  - Plottly and Cufflinks
  - Geographical Plotting
- Machine Learning
  - Linear Regression
  - Logistic Regression
  - K Nearest Neighbors
  - Decision Trees and Random Forests
  - Support Vector Machines
  - K Means Clustering
  - Recommender Systems

Pandas

- Pandas is an open-source library built on top of NumPy
- It allows for fast analysis and data cleaning and preparation
- It excels in performance and productivity.
- It also has built-in visualization features.
- It can work with data from a wide variety of sources.

Pandas Installation

If you use Anaconda distribution, it is most likely you have pandas installed.

Otherwise, you'll need to install pandas by going to your command line or terminal and using either:

conda install pandas

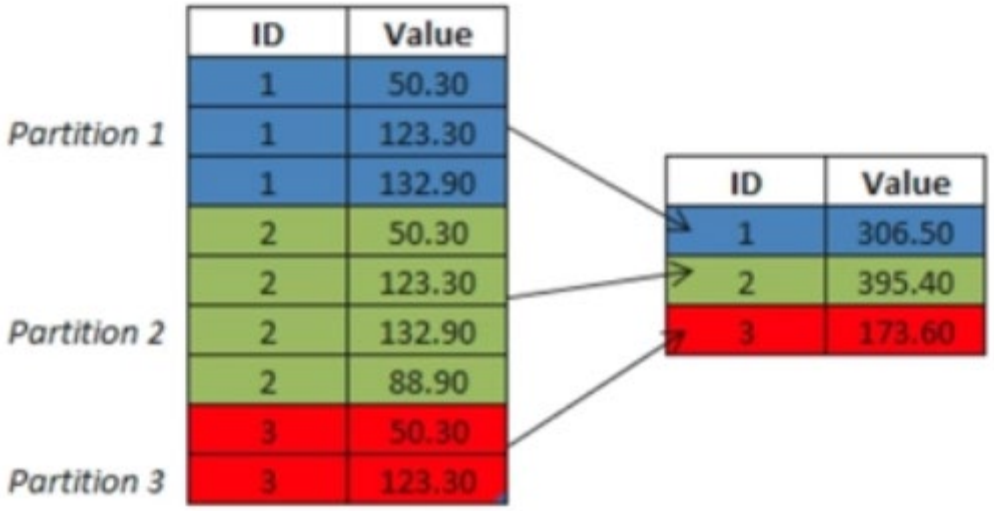
pip install pandas

Pandas Agenda

- Series
- DataFrames
- Missing Data
- GroupBy
- Merging, Joining, and Concatenating
- Operations
- Data Input and Output

Groupby

- Groupby allows you to group together rows based of of a column and perform an aggregate function on them.



Course Agenda

- Python Crash Course
- Data Analysis:
  - Numpy
  - Pandas
- Data Visualization:
  - **Matplotlib**
  - Seaborn
  - Pandas
  - Plottly and Cufflinks
  - Geographical Plotting
- Machine Learning
  - Linear Regression
  - Logistic Regression
  - K Nearest Neighbors
  - Decision Trees and Random Forests
  - Support Vector Machines
  - K Means Clustering
  - Recommender Systems