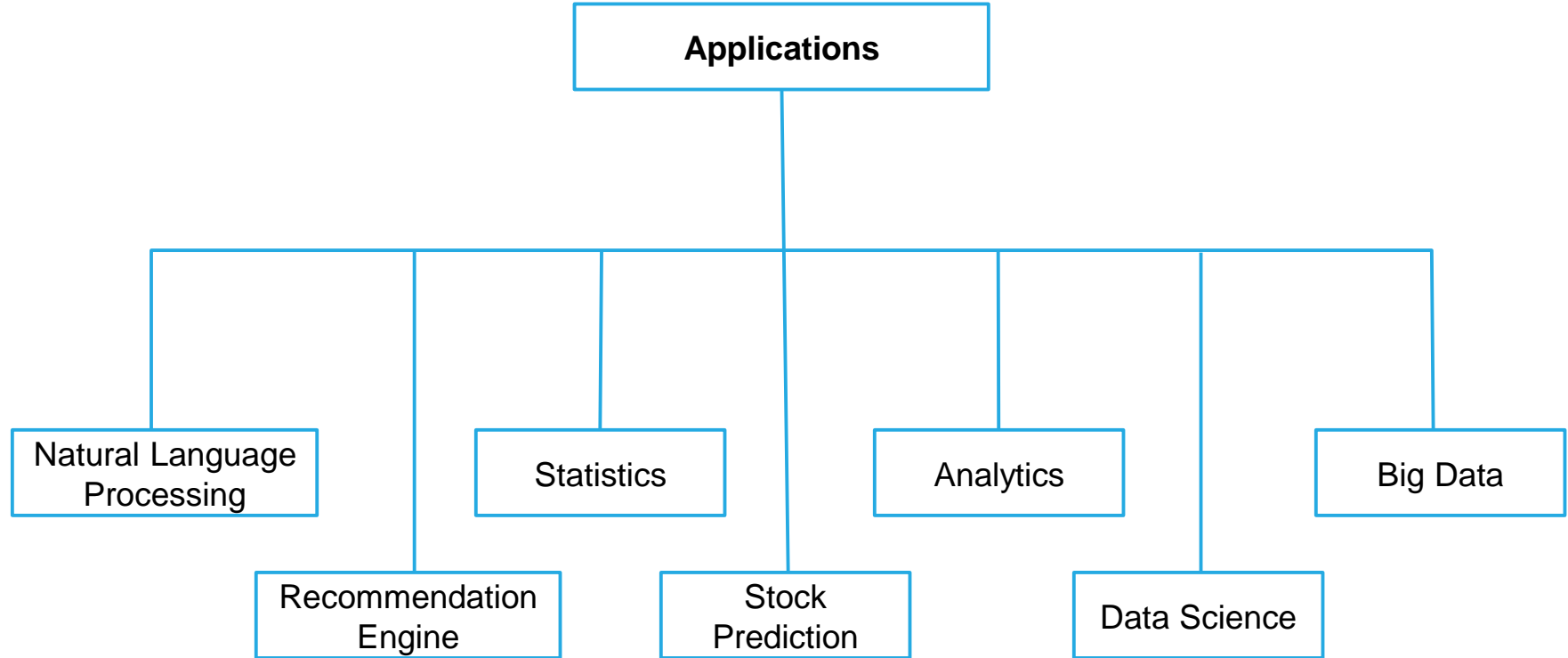# Python Pandas

Introduction

Introduction

- Pandas is a simple yet powerful and expressive tool

- It is an open source library in python

- It is useful in data manipulation and analysis

Applications of pandas

```
                        ┌─────────────────┐
                        │  Applications   │
                        └─────────────────┘

┌──────────────┐   ┌────────────┐   ┌────────────┐   ┌──────────┐
│   Natural    │   │ Statistics │   │ Analytics  │   │ Big Data │
│   Language   │   └────────────┘   └────────────┘   └──────────┘
│  Processing  │
└──────────────┘

        ┌──────────────┐   ┌────────────┐   ┌──────────────┐
        │Recommendation│   │    Stock   │   │ Data Science │
        │    Engine    │   │ Prediction │   └──────────────┘
        └──────────────┘   └────────────┘
```

Pandas vs. Numpy

Pandas	Numpy
High level data structures. It provides in-memory 2D table object called data frame.	Low level data structure. (np.array)
More streamlined handling of tabular data, and rich time series functionality.	Supports large multidimensional arrays and matrices.
Data alignment, handling missing data, groupby, merge, and, join methods.	A wide range of mathematical array operations.

Installation

Open terminal program (for Mac user) or command line (for Windows) and install it using following command:

```
conda install pandas
```

Or

```
pip install pandas
```

Installation

- Alternatively, you can install pandas in a jupyter notebook using below code:

```
!pip install pandas
```

- To import pandas we import it with a shorter name:

```
import pandas as pd
```

Components of pandas

- Series and dataframe are two primary components of pandas

- Series is a one-dimensional labeled array; typically a column, and the DataFrame is a two-dimensional table made up of a group of Series