# KOLT Python

## Introduction to Data Science with Python: Pandas and Matplotlib

Necla Mutlu

Tuesday 14th April, 2020

**KOÇ UNIVERSITY**

OFFICE OF LEARNING AND TEACHING

# Agenda

## 1. Recap

## 2. Pandas

## 3. Matplotlib

## 4. Case Study

# Python Modules

Modules are files that contains Python statements and definitions

# Python Modules

Modules are files that contains Python statements and definitions

Modules help you to break down large programs into small manageable and organized files

# Python Modules

Modules are files that contains Python statements and definitions

Modules help you to break down large programs into small manageable and organized files

Modules provides reusability of a code

# Python Modules

Modules are files that contains Python statements and definitions

Modules help you to break down large programs into small manageable and organized files

Modules provides reusability of a code

Most used functions can be defined in Modules and can be used in other programs without copying the definitions into every program

# Python Modules

Modules are files that contains Python statements and definitions

Modules help you to break down large programs into small manageable and organized files

Modules provides reusability of a code

Most used functions can be defined in Modules and can be used in other programs without copying the definitions into every program

To use definitions in modules, modules are needed to be **imported**

# Python Packages

- Packages can be thought as directories in which multiple modules are present

# Python Packages

- Packages can be thought as directories in which multiple modules are present
- Packages are organized hierarchically

# Python Packages

- Packages can be thought as directories in which multiple modules are present
- Packages are organized hierarchically
- Packages can contain subpackages as well as regular modules

# Python Packages

- Packages can be thought as directories in which multiple modules are present
- Packages are organized hierarchically
- Packages can contain subpackages as well as regular modules
- All packages are modules but not all modules are packages

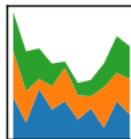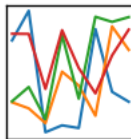# What is pandas?

# What is pandas?

- pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation package, built on top of the Python programming language.
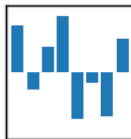
# What is pandas?

- pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation package, built on top of the Python programming language.
- It is used for working with structured data. For example tables, which consists of rows and columns can easily manipulated with pandas.



$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

# What can pandas do for you?

- Loading tabular data from different resources

# What can pandas do for you?

- Loading tabular data from different resources
- Search for particular rows or columns
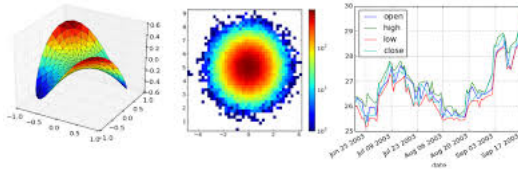
# What can pandas do for you?

- Loading tabular data from different resources
- Search for particular rows or columns
- Calculate aggregate statistics

# What can pandas do for you?

- Loading tabular data from different resources
- Search for particular rows or columns
- Calculate aggregate statistics
- Manipulate excel sheets, csv files

# What is matplotlib?

Matplotlib is a plotting library for the Python programming language. You can visualize various datasets in various ways by using the tools provided by Matplotlib.

# Case Study: Covid-19 USA Data

```
https://github.com/koltpython/python-workshops/tree/master/
4-Introduction-to-Pandas-and-Matplotlib
```

# Case Study: Covid-19 USA Data

```
https://github.com/koltpython/python-workshops/tree/master/
4-Introduction-to-Pandas-and-Matplotlib
```

To learn about pandas, we will make a case study. The dataset link is provided in the above pag

In this study, we will use Jupyter Notebook to code. The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. It is one of the powerful tools that allows easy data transformation, visualisation and statistical modelling.

# How to Install Jupyter Notebook?

## pip

If you use `pip`, you can install it with:

```
pip install notebook
```

Congratulations, you have installed Jupyter Notebook! To run the notebook, run the following command at the Terminal (Mac/Linux) or Command Prompt (Windows):

```
jupyter notebook
```

If "jupyter notebook" command doesn't work for you. You can try to execute ***python -m notebook*** or ***py -m notebook*** commands.

# Covid-19 US Data

From now on, we will continue our work in Jupyter Notebook. For the ones, who wants to check the project or review, you can find a well-documented version of the project in this link:

```
https://github.com/koltpython/python-workshops/tree/master/
4-Introduction-to-Pandas-and-Matplotlib
```