



Department of Statistics
MALAYSIA



Make Sure You're
COUNTED



20 OCT



2016 - 2030

Data Scientist Stream Part 1

Day 3

DATA EXPLORATION WITH PYTHON

By

Ms. Nur Hurriyatul Huda Abdullah Sani and Mr. Ahmad Najmi Ariffin



@StatsMalaysia
@MyCensus2020



www.dosm.gov.my



www.mycensus.gov.my



What Can I Do With Python?



Manipulating data using Pandas

Series, Data Frame

03

04

Hands-on

<http://bit.ly/dstreamDay3>

Introduction to pandas

What is Pandas?
Basic syntax in pandas






02

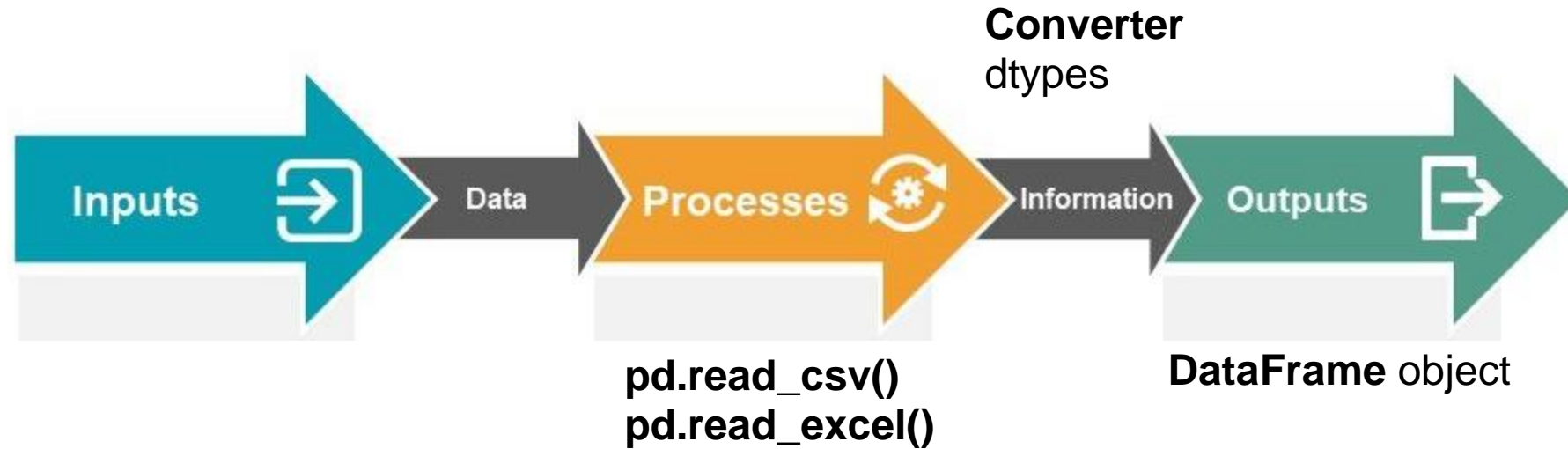
Loading data from a variety of common formats

csv, xlsx, html, json, sql.

01



NumPy	provides efficient storage and computation for multi-dimensional data arrays	
SciPy	contains a wide array of numerical tools such as numerical integration and interpolation	
Pandas	provides a DataFrame object along with a powerful set of methods to manipulate, filter, group, and transform data	
Matplotlib	provides a useful interface for creation of publication-quality plots and figures	
Scikit-Learn	provides a uniform toolkit for applying common machine learning algorithms to data	





Preparing Data : Scenario

Hands-on:

<http://bit.ly/dstreamDay03>

Country

Population

Area

Gross domestic
product

Continent

Independence
day



Loading data from a variety of common formats



www.dosm.gov.my



   
Stats Malaysia



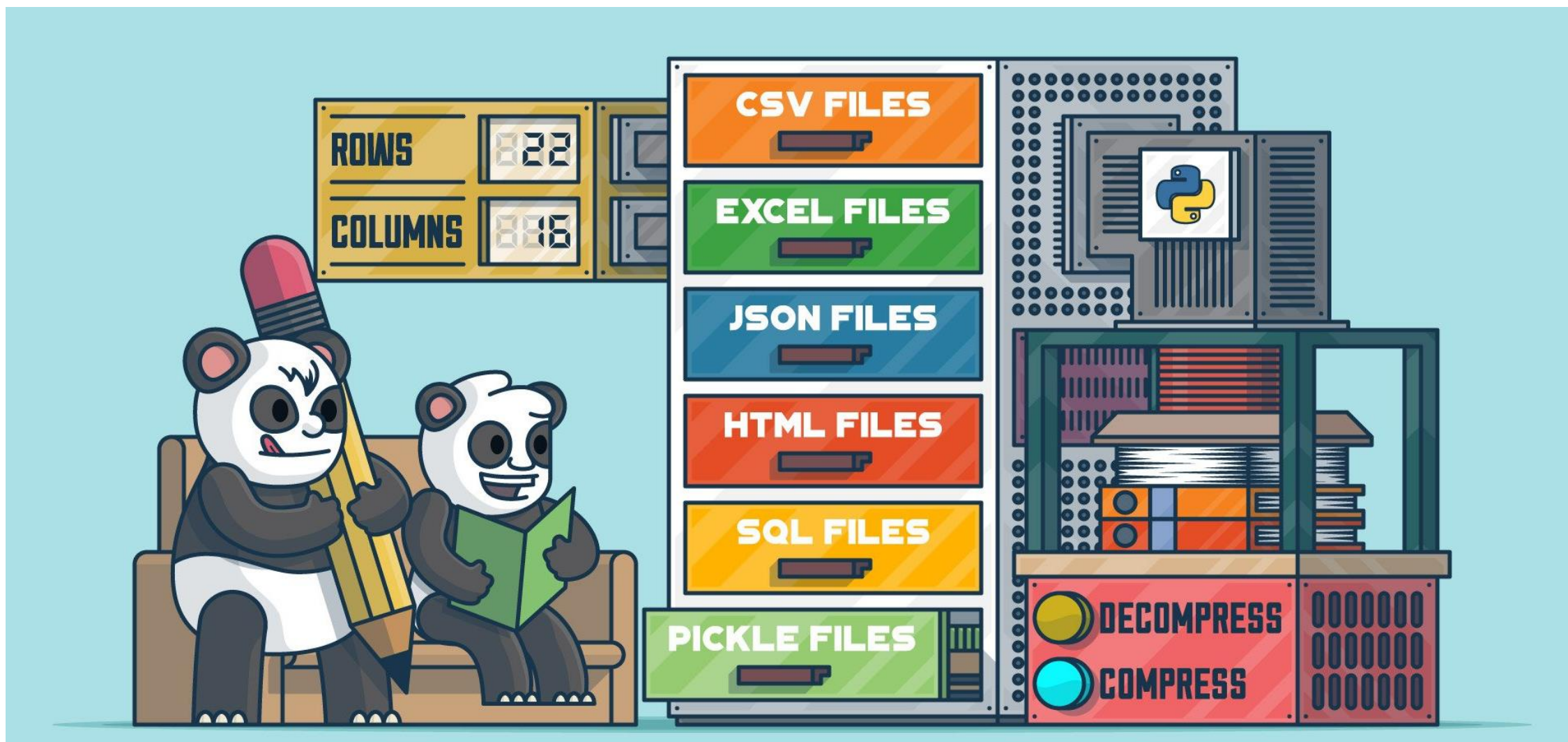
www.mycensus.gov.my





Working With Different File Types

<http://bit.ly/dstreamDay03>



- Load data from your local drive



- Load data from your Google Drive



- Load data from Github (Files < 25MB)





CREATE

UPDATE

Manipulating data using Pandas



www.dosm.gov.my




Stats Malaysia



www.mycensus.gov.my



01 Series

The first main data type we will learn about for pandas is the Series data type.

A Series is very similar to a NumPy array (it is built on top of the NumPy array object).

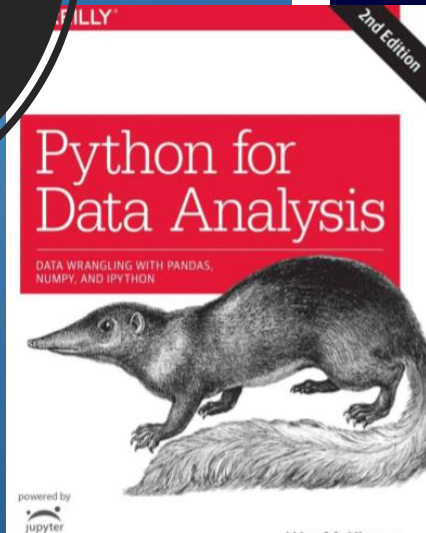
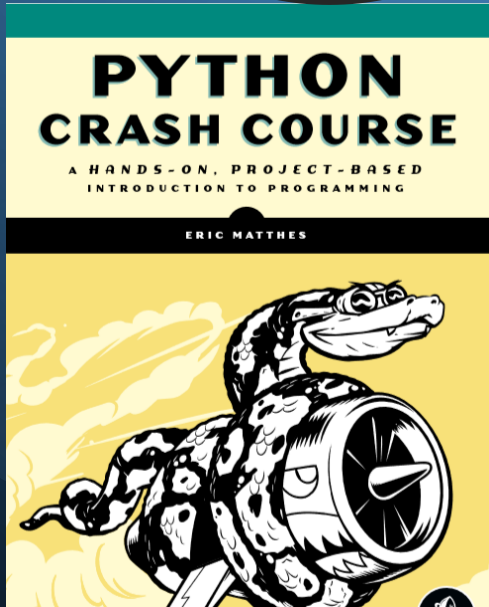
What differentiates the NumPy array from a Series?

1. is that a Series can have axis labels, meaning it can be indexed by a label, instead of just a number location.
2. It also doesn't need to hold numeric data, it can hold any arbitrary Python Object

02 Data Frame

- One basic structure that you get with pandas is a data frame.
- A data frame is a two dimensional grid, rather similar to a relational database table except in memory.
- Data Frames are the workhorse of pandas and are directly inspired by the R programming language.
- We can think of a Data Frame as a bunch of Series objects put together to share the same index

Some useful
references



coreteambda.wixsite.com/blog/learn-and-lead


Home Learn & Lead Best Practices Activities Our Team

CORE TEAM ANALITIK DATA RAYA

Better Data, Better Lives

Core Team Analitik Data Raya was established to modernise public sector services through managing and implementing big data analysis with the collaboration of other government agencies.

Are you ready to learn the foundational coding and statistics skills for a career in data science?



©2020 by Core Team Analitik Data Raya.

<https://coreteambda.wixsite.com/blog/learn-and-lead>

Let's explore this concept through some examples:

THANK YOU

BANCI MALAYSIA

POPULATION & HOUSING
**CENSUS
MALAYSIA 2020**

YOUR DATA IS OUR FUTURE



Make Sure You're
COUNTED

For More Info : www.dosm.gov.my
www.mycensus.gov.my



20 OCT



2016-2030

