Universidad **Ricardo Palma**

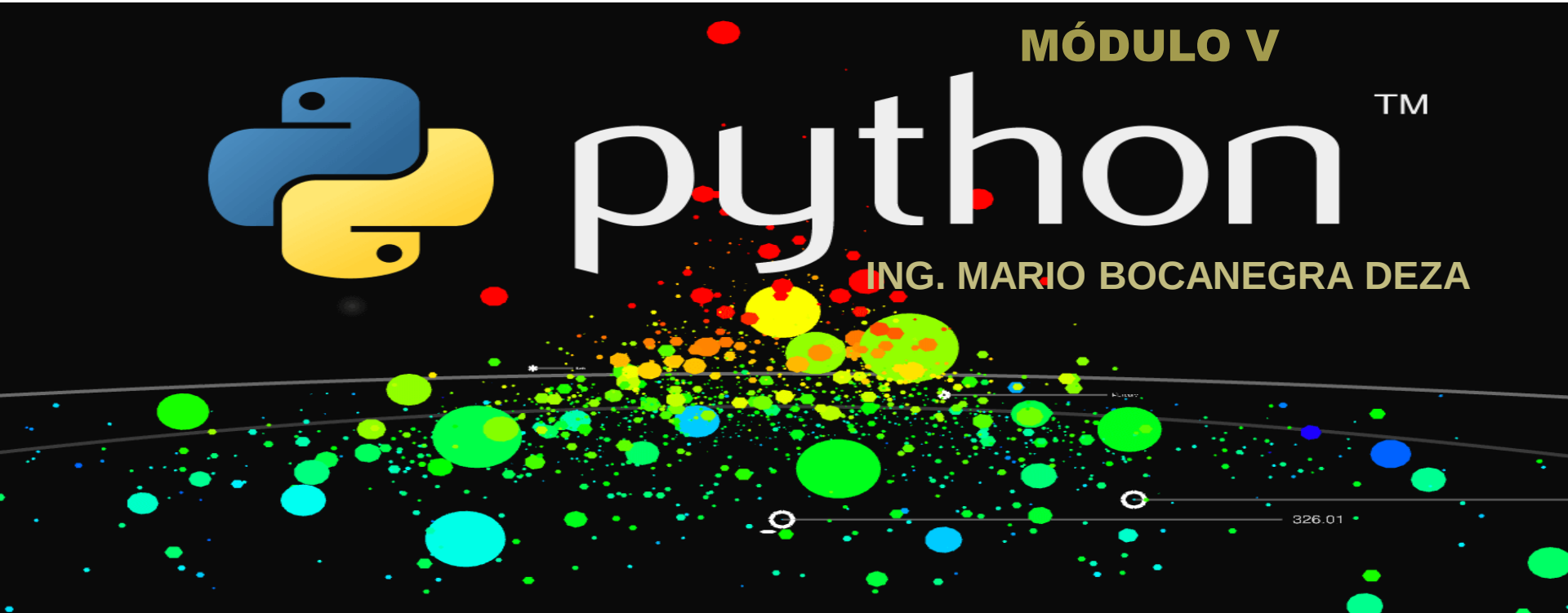
RECTORADO

Formamos seres humanos para una cultura de paz

Primer Programa de Especialización

INTRODUCCIÓN AL DATA SCIENCE

PROCESAMIENTO DE DATOS UTILIZANDO LA LIBRERÍA PANDAS

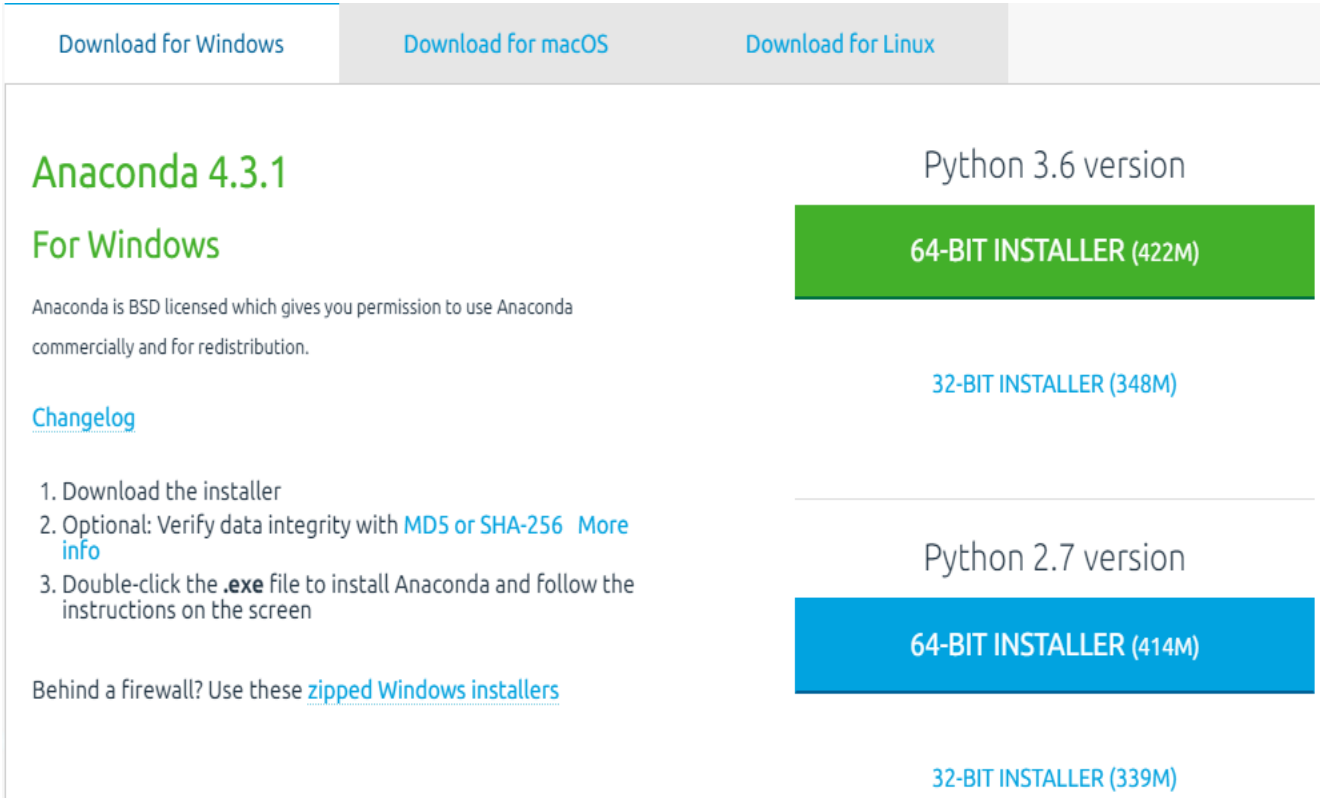


Contenido

- ✓Instalación de Anaconda
- ✓Manejo de datos de SQL SERVER 2014
- ✓Manejo de datos de archivos csv, txt
- ✓Manejo de datos de archivos SPSS

Instalación de Anaconda

https://www.continuum.io/downloads



ANACONDA MAKES...

DATA SCIENCE TEAMS HAPPIER

That means better and more results

ANACONDA EMPOWERS YOUR DATA SCIENTISTS

Analytic Workflows

- Explore data to reveal patterns
- Analyze data to discover insights and predict future outcomes
- Publish interactive visualizations and presentations
- Deploy predictive models into production systems

Analytics Interaction

- Empower everyone on the Data Science team with their favorite tools
- Use your favorite approach to building and deploying data science – Jupyter / IPython Notebooks, Visual Data Exploration, Interactive Visual Web Apps, Visual Programming, Interactive Development Environment, or CLI

High Performance Distributions

- Achieve faster analytics throughput with high performance Python and R
- Works across multiple platforms including Windows, Linux, OSX

ANACONDA.

Leading Open Data Science Platform Powered by Python

CONDA®

Leading Package and Environment Manager

OPEN DATA SCIENCE

DATA

COMPUTATION

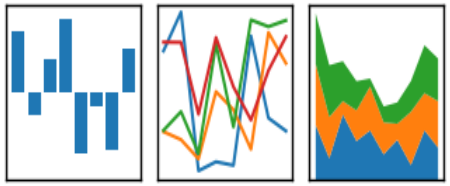
RESULTS

- Reproducibility for packages and environments
- Building interactive visualizations

pandas.pydata.org

pandas

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



Python Data Analysis Library

pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

pandas is a NUMFocus sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project.

A Fiscally Sponsored Project of

NUMFOCUS

OPEN CODE = BETTER SCIENCE

VERSIONS

Release
0.19.2 - December 2016
download // docs // pdf

Development
0.20.0 - 2017
github // docs

Previous Releases
0.19.1 - download // docs // pdf
0.19.0 - download // docs // pdf

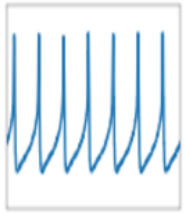
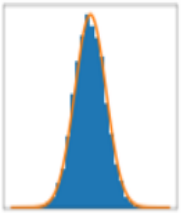
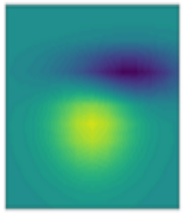
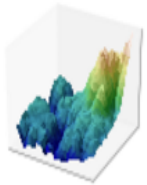
matplotlib.org

matplotlib

home | examples | gallery | pyplot | docs »

Introduction

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shell, the jupyter notebook, web application servers, and four graphical user interface toolkits.

Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc., with just a few lines of code. For a sampling, see the screenshots, thumbnail gallery, and examples directory

For simple plotting the `pyplot` module provides a MATLAB-like interface, particularly when combined with `IPython`. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

www.numpy.org

NumPy

Scipy.org

NumPy

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

NumPy is licensed under the BSD license, enabling reuse with few restrictions.

Manejo de datos de SQL SERVER 2014

```
import pyodbc
import pandas as pd
conn = pyodbc.connect(
    r'DRIVER={ODBC Driver 11 for SQL Server};'
    r'SERVER=BLACKCLOUD\SQLEXPRESS;'
    r'DATABASE=Ejemplo;'
    r'UID=sa;'
    r'PWD=123456'
    )
sqlquery="SELECT * FROM dbo.salario"
df=pd.read_sql_query(sqlquery,conn)
```

Manejo de datos de SQL SERVER 2014

Manejo de datos archivos csv, txt

```
In [2]: import pandas as pd
```

```
In [9]: df=pd.read_csv("C:/Users/Mario/Downloads/salario.csv")
```

```
In [3]: import pandas as pd
```

```
In [23]: df=pd.read_csv("C:/Users/Mario/Downloads/salario.txt",sep=',')
```

IO Tools (Text, CSV, HDF5, ...)

The pandas I/O API is a set of top level `reader` functions accessed like `pd.read_csv()` that generally return a `pandas` object.

- read_csv
- read_excel
- read_hdf
- read_sql
- read_json
- read_msgpack (experimental)
- read_html
- read_gbq (experimental)
- read_stata
- read_sas
- read_clipboard
- read_pickle

Manejo de datos archivos SPSS

In [1]:
```
import savReaderWriter as spss
import pandas as pd
```

In [12]:
```
with spss.SavReader("C:/Users/Mario/Downloads/salario.sav", returnHeader=True) as reader:
    raw_data= reader.all()
```

In [20]:
```
data=pd.DataFrame(raw_data)
data = data.rename(columns=data.loc[0]).iloc[1:]
```

Manejo de datos archivos SPSS

ation [US] | https://pypi.python.org/pypi/savReaderWriter/3.4.2

» Package Index > savReaderWriter > 3.4.2

savReaderWriter 3.4.2

Read and write SPSS files

Download
savReaderWriter-3.4.2.tar.gz

Package Documentation

A cross-platform Python interface to the IBM SPSS Statistics Input Output Module. Read or Write SPSS system files (.sav, .zsav). Works with Python 2.7, 3.3, 3.4, 3.5, pypy

installation:

```
python setup.py install
```

Or alternatively:

```
pip install savReaderWriter
```

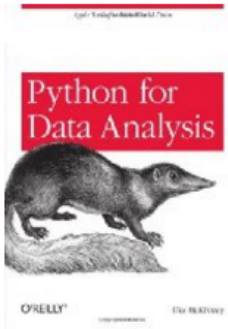
www.foxebook.net/python-for-data-analysis/

Fox eBook Books Top Sellers Most Reviews New Release Co

You are here: Books » O'Reilly Media » Python for Data Analysis

Python for Data Analysis

October 30, 2012 foxebook Computers & Internet 2662 0



23 Reviews

by Wes McKinney

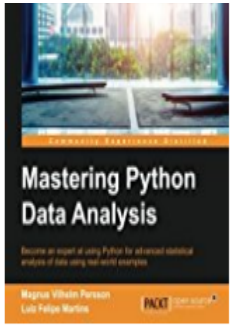
O'Reilly Media

2012-10-29

470 pages

www.foxebook.net/search/python+for+data+analysis

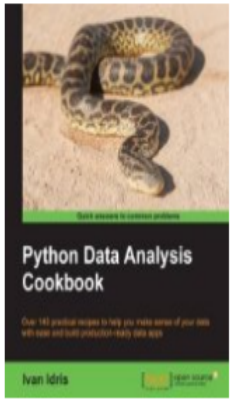
Search Results for: python for data analysis



by Luiz Felipe Martins, Magnus Vilhelm Persson

Packt Publishing, 2016-07-06, 284 pages, pdf, epub

Download



Python Data Analysis Cookbook

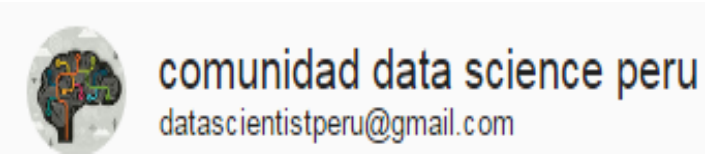
by Ivan Idris

Packt Publishing, 2016-08-04, 403 pages, pdf, ep

Download


fanpage
facebook


Comunidad
Data Science
Perú
@comunidaddatascience


comunidad data science peru
datascientistperu@gmail.com