

OPEN THREAT RESEARCH

EMPOWERING THE INFOSEC COMMUNITY

Análisis de Datos
en Seguridad
Defensiva via
Jupyter
Notebooks



Quiénes Somos?



Amantes de la Colaboración y Fuente Abierta

Roberto Rodriguez 

@Cyb3rWard0g

- Microsoft Threat Intelligence Center (MSTIC)

Jose Rodriguez 

@Cyb3rPandaH

- MITRE - ATT&CK

- Colaboración Abierta ❤
- Threat Hunter Playbook
[@HunterPlaybook](#)
- Mordor [@Mordor_Project](#)
- OSSEM [@OSSEM_Project](#)
- Blacksmith & more..

Agenda

- 1) Introducción a Jupyter Notebooks
 - Opciones de instalación
 - El proyecto Binder
- 2) Introducción a Pandas
 - Importing the Library
- 3) El proceso de Análisis de Datos
- 4) Necesitamos data?... Mordor
 - Descargando sets de datos
 - De datos a Dataframe
- 5) Algunos ejemplos de técnicas para Análisis de Datos

Espera...
Whát? 🤷

¿Qué
haremos?



La Estructura del Workshop:

Plataformas para Análisis



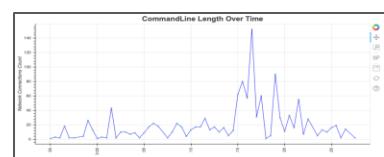
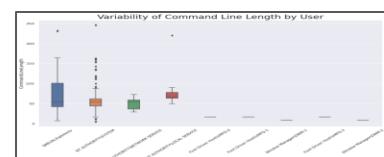
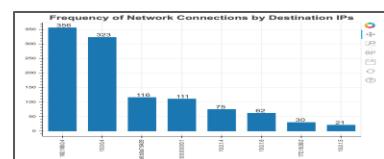
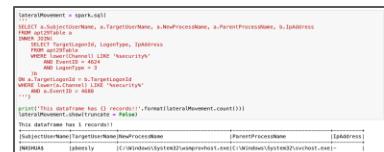
Set de Datos **MORDOR**



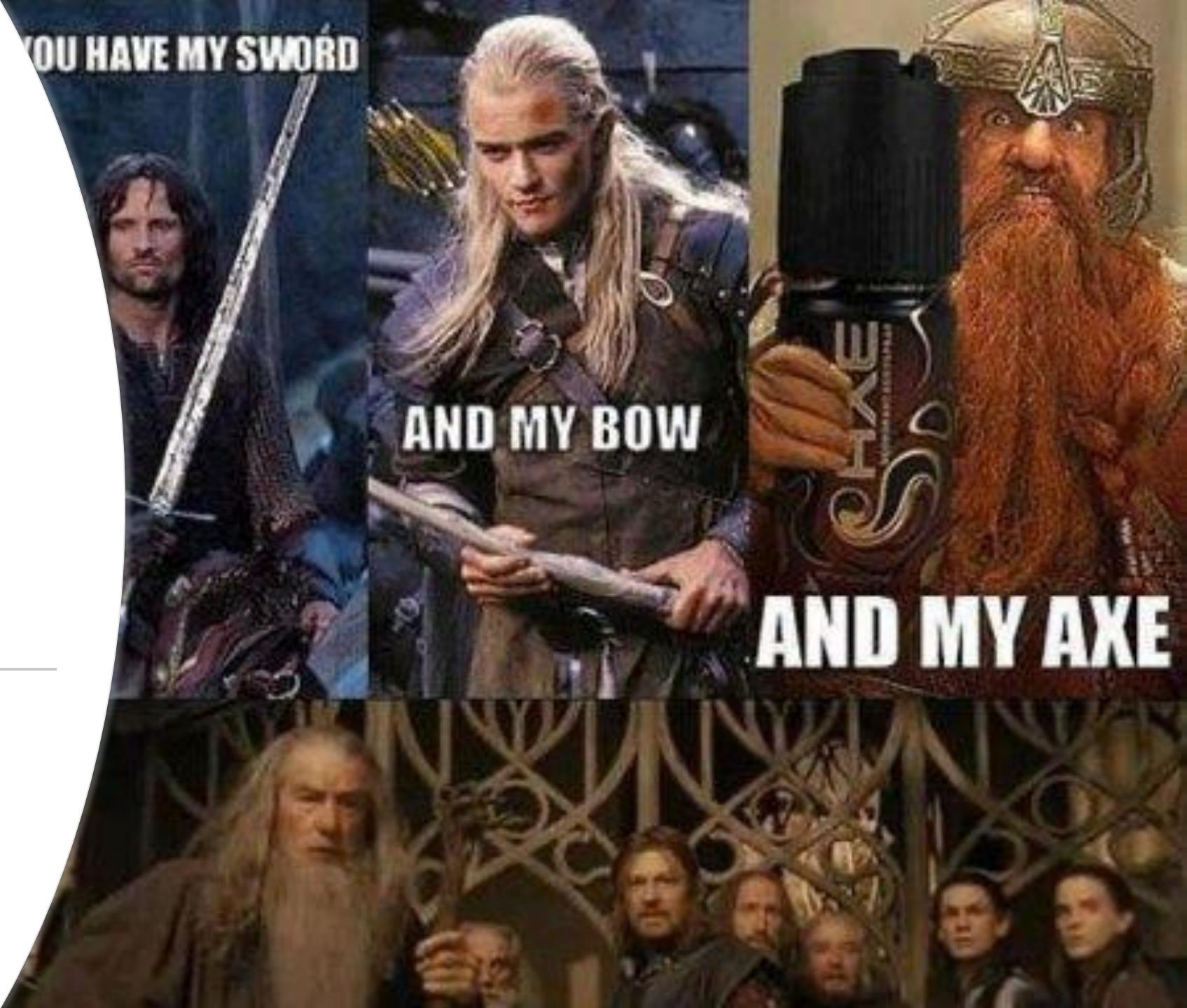
La Librería Pandas



Técnicas de Análisis

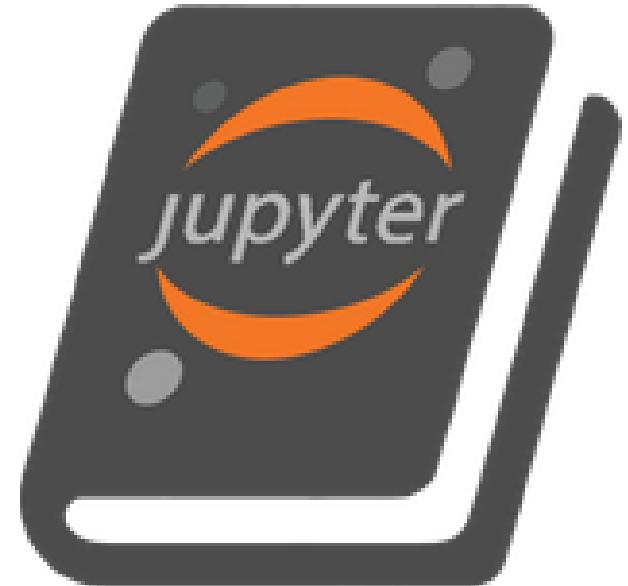


Pre - Requisitos



Todo lo que necesitas esta aqui:

[https://otrf.github.io/
workshop-ekoparty-
bluespace-2020](https://otrf.github.io/workshop-ekoparty-bluespace-2020)



Infosec Jupyter Book

Introducción a Jupyter Notebooks



¿Qué son Jupyter Notebooks?



- Son documentos que podemos accesar a través de una interface web. Nos permite gestionar y almacenar:
 - **Input:** Código (por ejemplo Python)
 - **Output:** Resultados de Código ejecutado
- Excelente para contar la historia de la investigacion desarrollada.

Python Interpreter -> IPython -> Jupyter

```
[Robertos-MacBook-Pro:~ wardog$ python3
Python 3.7.2 (default, Feb 12 2019, 08:16:38)
[Clang 10.0.0 (clang-1000.11.45.5)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
[>>> print('Hola Python!!!')
Hola Python!!
[>>> 12 * 2
24
>>> ]
```

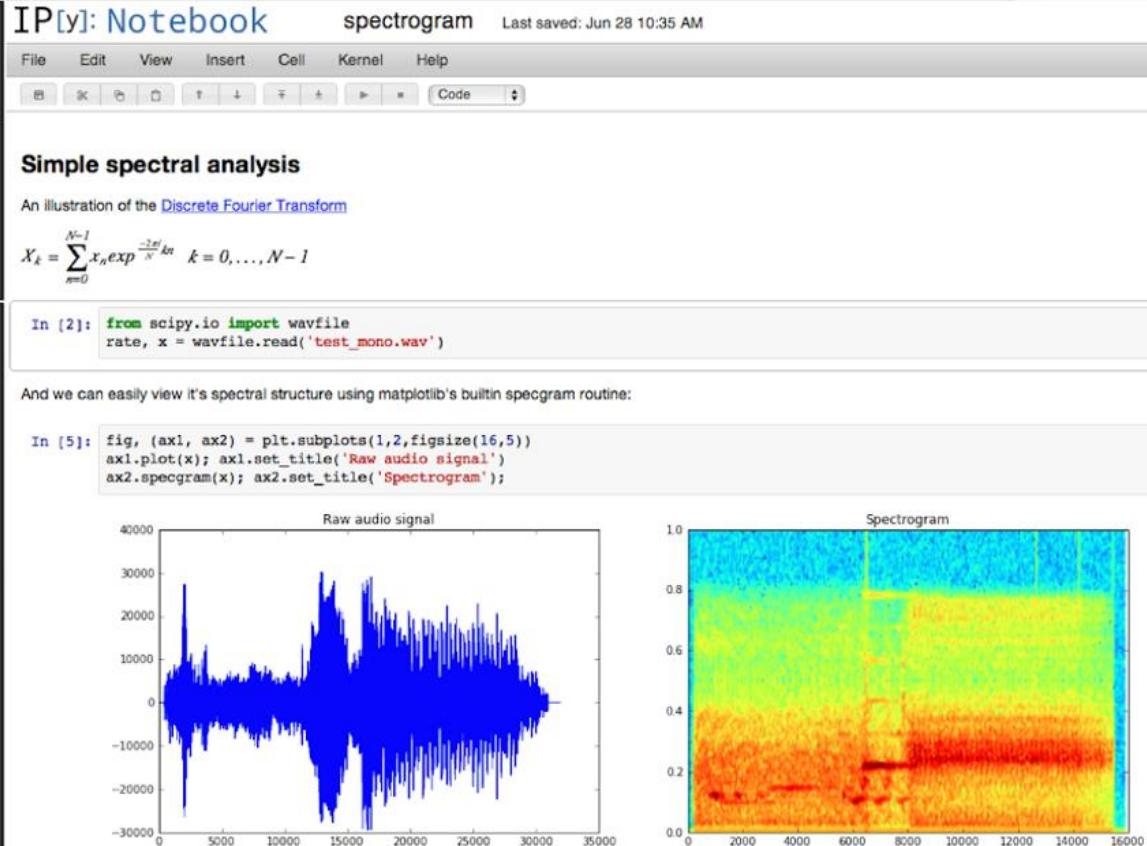
```
[Robertos-MacBook-Pro:GitHub wardog$ ipython
Python 2.7.10 (default, Oct  6 2017, 22:29:07)
Type "copyright", "credits" or "license" for more information.

IPython 5.7.0 -- An enhanced Interactive Python.
?          -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help       -> Python's own help system.
object?    -> Details about 'object', use 'object??' for extra details.

[In [1]: print('Hola IPython!!!')
Hola IPython!!

[In [2]: 12 * 2
Out[2]: 24

In [3]: ]
```



IPython -> Jupyter

Ada Lovelace
(First Computer
Programmer)
mid 19th Century

Guido Van Rossum
(Creator of Python)
1991

IIPython 0.12
Released (First
Notebook)
Dec 18, 2011

JupyterHub 0.1 is
released
March, 2015

Binder Project is
announced
May, 2016

JupyterLab Beta 1
Jan 11, 2018

Grace M. Hopper
(First Compiler)
1952

Fernando Perez
(Releases IIPython)
Dec 10, 2001

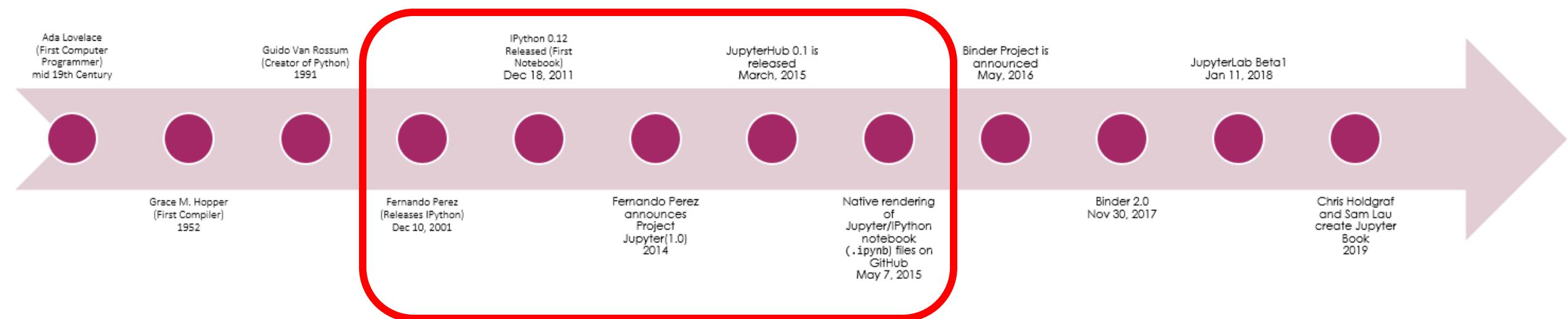
Fernando Perez
announces
Project
Jupyter(1.0)
2014

Native rendering
of
Jupyter/IIPython
notebook
(.ipynb) files on
GitHub
May 7, 2015

Binder 2.0
Nov 30, 2017

Chris Holdgraf
and Sam Lau
create Jupyter
Book
2019

IPython -> Jupyter



IPython -> Jupyter

IPython

- Interactive Python shell at the terminal
- Kernel for this protocol in Python
- Tools for Interactive Parallel computing

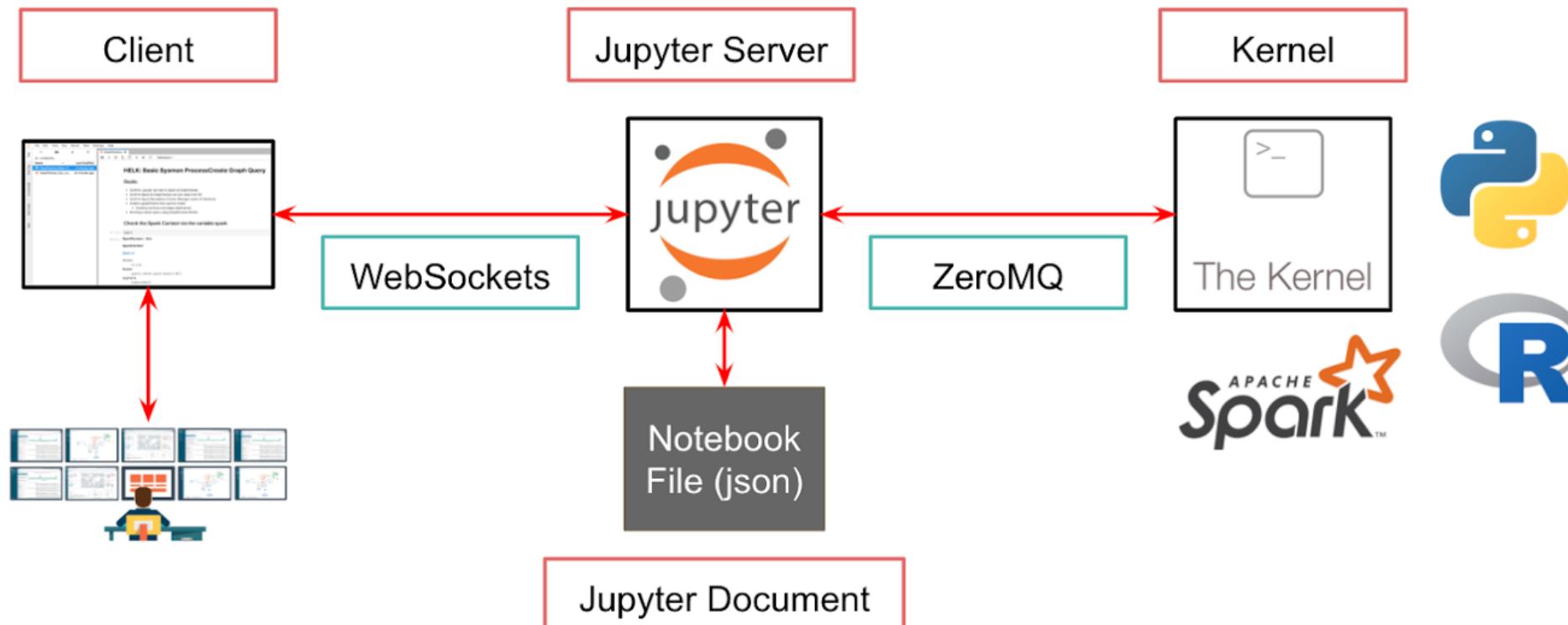


Jupyter

- Network protocol for interactive computing
- Clients for protocol
 - Console
 - Qt Console
 - Notebook
- Notebook file format & tools (nbconvert...)
- Nbviewer

Language Agnostic

La arquitectura básica de Jupyter Notebooks

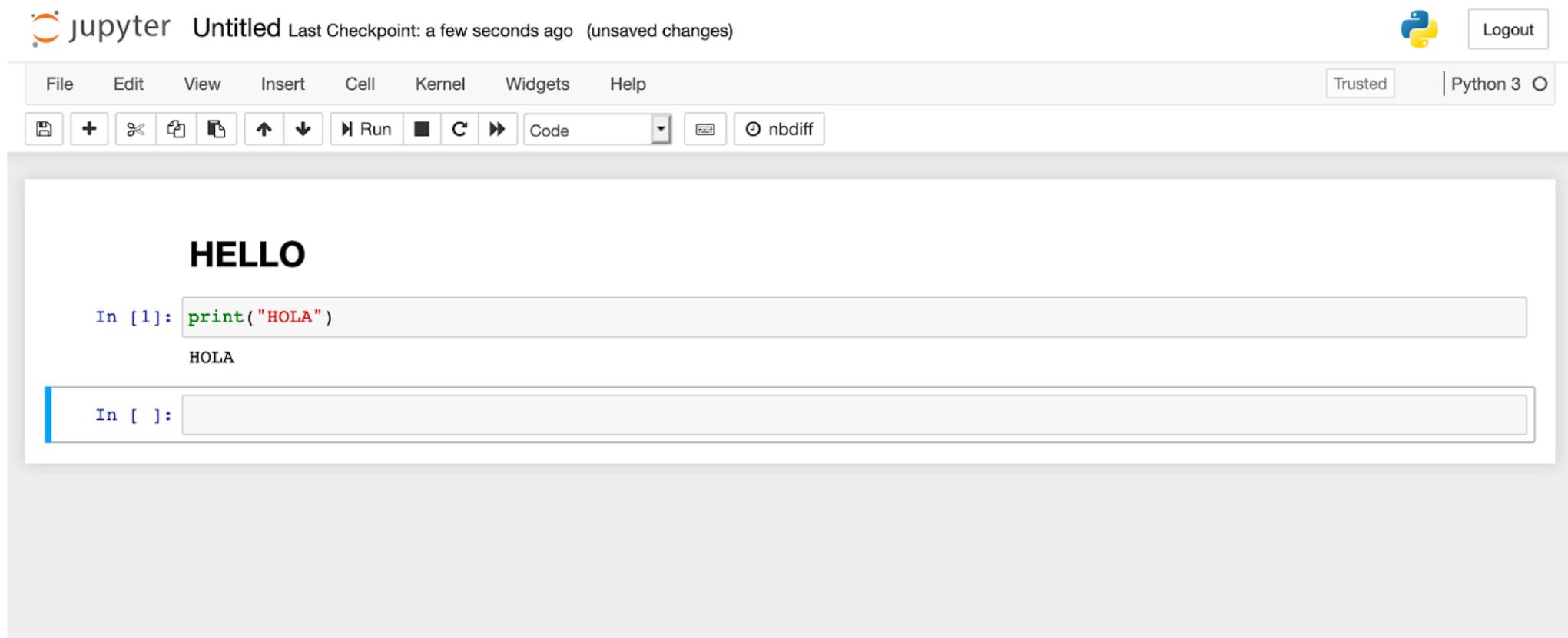


Algunos usos de Jupyter Notebooks



- Limpieza y transformación de data
- Visualización de data
- Modelamiento estadístico de datos
- Aplicaciones de machine learning y mas..

Cómo se ve un Jupyter Notebook?



Cómo podemos instalarlo localmente?

- La opción **Manual**
- Usando **Docker CE**
- Dando click con **Binder**

Implementándolo manualmente:

Prerequisite: Python

Así Jupyter pueda correr diferentes tipos de lenguaje, Python es un requerimiento (Python 3.3+, or Python 2.7) para instalar la aplicación de Jupyter Notebooks.

Using Conda : conda install -c conda-forge notebook

Using PIP : pip install notebook

Una vez que Jupyter Notebook este instalado, puedes correr lo siguiente en tu terminal para poder inicializar el servidor: **jupyter notebook**

Implementándolo con **Docker CE**:

Prerequisite: Docker CE

Te recomiendo que primero instales docker desktop y la version "Community Edition". Despues de eso vas a poder bajar y correr una imagen de docker que hemos creamos para este workshop con todo el material que vamos a usar por las siguientes 2 horas.

Tambien existen imagenes de docker "Ready-To-Go":

- Jupyter Docker Stacks: <https://github.com/jupyter/docker-stacks>
- Jupyter Docker Base Image: <https://hub.docker.com/r/jupyter/base-notebook/>

docker run -p 8888:8888 jupyter/minimal-notebook:3b1f4f5e6cc1

Implementándolo con **Docker CE**:

```
docker image pull cyb3rward0g/ekoparty-blue-  
2020:0.2
```

```
docker run --rm -ti -e  
JUPYTER_NOTEBOOKS_DIR=/home/jovyan/docs/co  
nceptos-basicos -p  
8888:8888 cyb3rward0g/ekoparty-blue-2020:0.2
```

Implementándolo con Docker CE:

```
cyb3rward0g@Robertos-MBP workshop-ekoparty-bluespace-2020 % docker run --rm -ti -e JUPYTER_NOTEBOOKS_DIR=/home/jovyan/docs/conceptos-basicos -p 8888:8888 ekoblue
[NOTEBOOK-JUPYTER-DOCKER-INSTALLATION-INFO] Running Jupyter Type: notebook..
[NOTEBOOK-JUPYTER-DOCKER-INSTALLATION-INFO] Running the following parameters --ip=0.0.0.0 --port=8888 --notebook-dir=/home/jovyan/docs/conceptos-basicos --no-browser --NotebookApp.max_buffer_size=536870912 --NotebookApp.base_url=/
[NOTEBOOK-JUPYTER-DOCKER-INSTALLATION-INFO] Starting Jupyter notebook..
[I 19:47:16.560 NotebookApp] Writing notebook server cookie secret to /home/jovyan/.local/share/jupyter/runtime/notebook_cookie_secret
[I 19:47:17.178 NotebookApp] JupyterLab extension loaded from /opt/conda/lib/python3.7/site-packages/jupyterlab
[I 19:47:17.178 NotebookApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 19:47:17.181 NotebookApp] Serving notebooks from local directory: /home/jovyan/docs/conceptos-basicos
[I 19:47:17.182 NotebookApp] Jupyter Notebook 6.1.4 is running at:
[I 19:47:17.182 NotebookApp] http://e9a832088e28:8888/?token=b4091cf00c2289d4b2a008192e25800075bfebdee8016bf2
[I 19:47:17.182 NotebookApp] or http://127.0.0.1:8888/?token=b4091cf00c2289d4b2a008192e25800075bfebdee8016bf2
[I 19:47:17.182 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 19:47:17.187 NotebookApp]

To access the notebook, open this file in a browser:
  file:///home/jovyan/.local/share/jupyter/runtime/nbserver-7-open.html
Or copy and paste one of these URLs:
  http://e9a832088e28:8888/?token=b4091cf00c2289d4b2a008192e25800075bfebdee8016bf2
  or http://127.0.0.1:8888/?token=b4091cf00c2289d4b2a008192e25800075bfebdee8016bf2
```

Implementándolo con Docker CE:

The screenshot shows a Jupyter Notebook interface running in a Docker container. The browser address bar displays the URL `http://127.0.0.1:8888/?token=b4091cf00c2289d4b2a008192e25800075bfebdee8016bf2`. The page title is "jupyter". The top navigation bar includes links for "Files", "Running", and "Clusters", along with "Quit" and "Logout" buttons. A search bar at the top right contains a downward arrow icon. Below the navigation, a message says "Select items to perform actions on them." On the right, there are buttons for "Upload", "New", and a refresh symbol. The main area is a file list table with columns for "Name", "Last Modified", and "File size". The table shows the following files:

	Name	Last Modified	File size
<input type="checkbox"/> 0	/		
<input type="checkbox"/>	sets_datos	an hour ago	
<input type="checkbox"/>	1_intro_numpy_arrays.ipynb	3 days ago	7.96 kB
<input type="checkbox"/>	2_dataframe_desde_Set_datos_Mordor.ipynb	3 days ago	30.9 kB
<input type="checkbox"/>	3_Analisis_Datos_Pandas_Filtrando_Resumiendo_Datos.ipynb	3 days ago	43.5 kB
<input type="checkbox"/>	4_Analisis_Datos_Pandas_Transformando_Datos.ipynb	an hour ago	34.2 kB
<input type="checkbox"/>	5_Analisis_Datos_Pandas_Correlacionando_Datos.ipynb	3 days ago	29.2 kB
<input type="checkbox"/>	6_Analisis_Datos_Pandas_Visualizando_Datos.ipynb	an hour ago	63.8 kB

Implementándolo con binder :

- El proyecto Binder es una comunidad abierta que hace posible la creación ambientes interactivos y facil de compartir con la comunidad.
- El producto principal de esta comunidad es el BinderHub. Lo cual es un proyecto que maneja, monitorea y ejecuta ambientes definidos por usuarios en la comunidad y guardados en repositorios de binder (Ejemplo: GitHub)
- Para quien: Desarrolladores, Profesores, Personas que quieren compartir su research y las que se quieren comunidad por medio de data

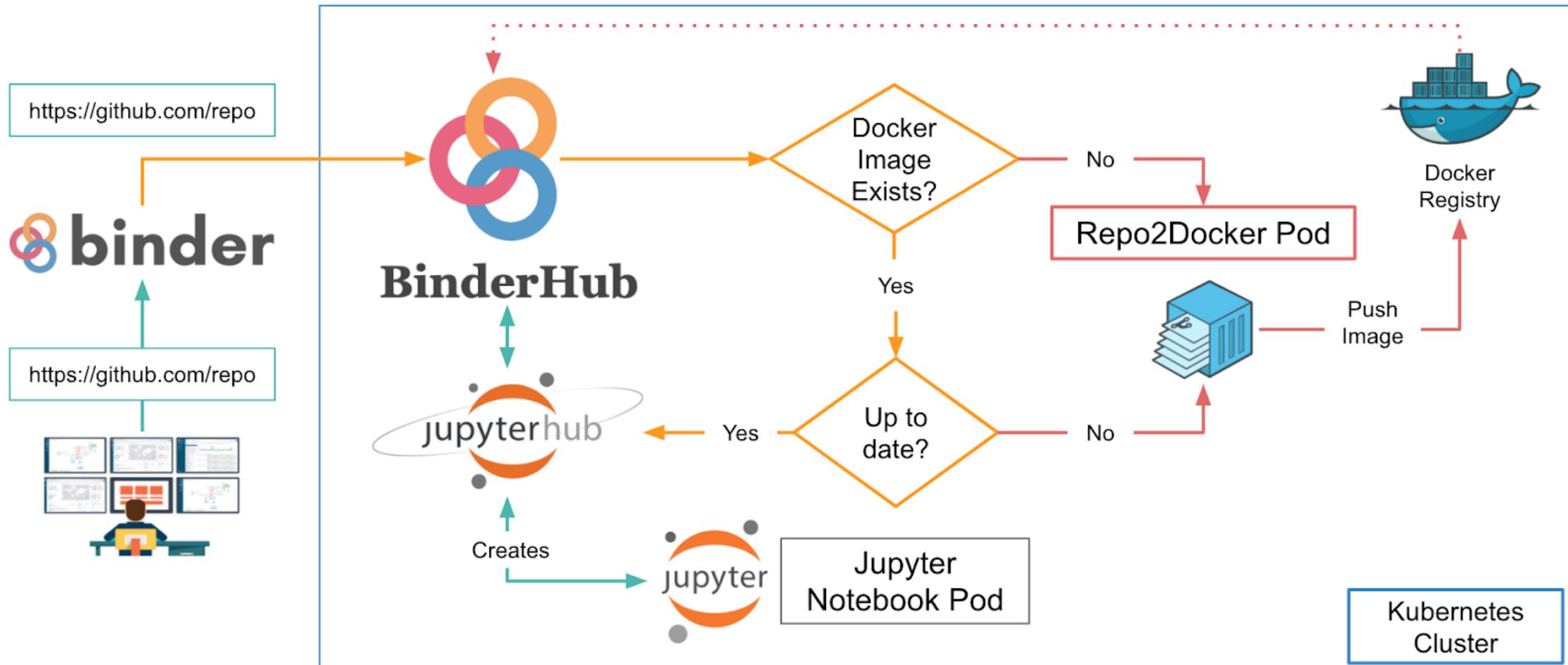


BinderHub conecta bastantes servicios para proveer un ambiente en la nube.

Utiliza los siguientes servicios:

- Un **proveedor de servicios en la nube** como Google Cloud, Microsoft Azure, Amazon EC2, and others
- **Kubernetes** para administrar los recursos en la nube
- **Helm** para configurar y administrar Kubernetes
- **Docker** para usar contenedores que estandarizan ambientes computacionales
- Una **BinderHub UI** que los usuarios pueden accessar para especificar repositorios Git que desean crear
- **BinderHub** para generar imagenes de Docker usando la URL de un repositorio GIT
- **A Docker registry** (por ejemplo gcr.io) que alberga las imagenes de los contenedores
- **JupyterHub** para implementar contenedores temporales para usuarios

El diseño de binder



Usando binder

Use Cases

- Data Analysis
- Data Connectors
- Data Visualizations

Community Projects

- Threat Hunter Playbook

Community Workshops

- Defcon BTV 2020

Basic Data Analysis Concepts

- Creating a Spark Dataframe

Creating a Spark SQL View from a Mordor Dataset

Data Analysis with Spark.SQL: Filtering & Summarizing

Data Analysis with Spark.SQL: Transforming

Data Analysis with Spark.SQL:

←

Creating a Spark Dataframe

 Binder
 Thebe Launch Binder

- Author: Jose Rodriguez (@Cyb3rPandah)
- Project: Infosec Jupyter Book
- Public Organization: [Open Threat Research](#)
- License: [Creative Commons Attribution-ShareAlike 4.0 International](#)
- Reference: <https://mordordatasets.com/introduction.html>

Importing Spark libraries

```
from pyspark.sql import SparkSession
```

Creating Spark session

```
spark = SparkSession \
    .builder \
    .appName("Spark_example") \
    .config("spark.sql.caseSensitive","True") \
    .getOrCreate()
```

On this page

- Importing Spark libraries
- Creating Spark session
- Creating a Spark Sample DataFrame
- Exposing Spark DataFrame as a SQL View
- Testing a SQL-like Query

Thank you! I hope you enjoyed it!

Usando binder

Thanks to [Google Cloud](#), [OVH](#), [GESIS Notebooks](#) and the [Turing Institute](#) for supporting us! 🎉





Starting repository: OTRF/infosec-jupyter-book/master

Take a look at our [gallery](#) of example repositories.

Build logs show

Here's a non-interactive preview on [nbviewer](#) while we start a server for you. Your binder will open automatically when it is ready.

 **jupyter**
nbviewer

infosec-jupyter-book master

Name

Usando binder



Starting repository: OTRF/infosec-jupyter-book/master

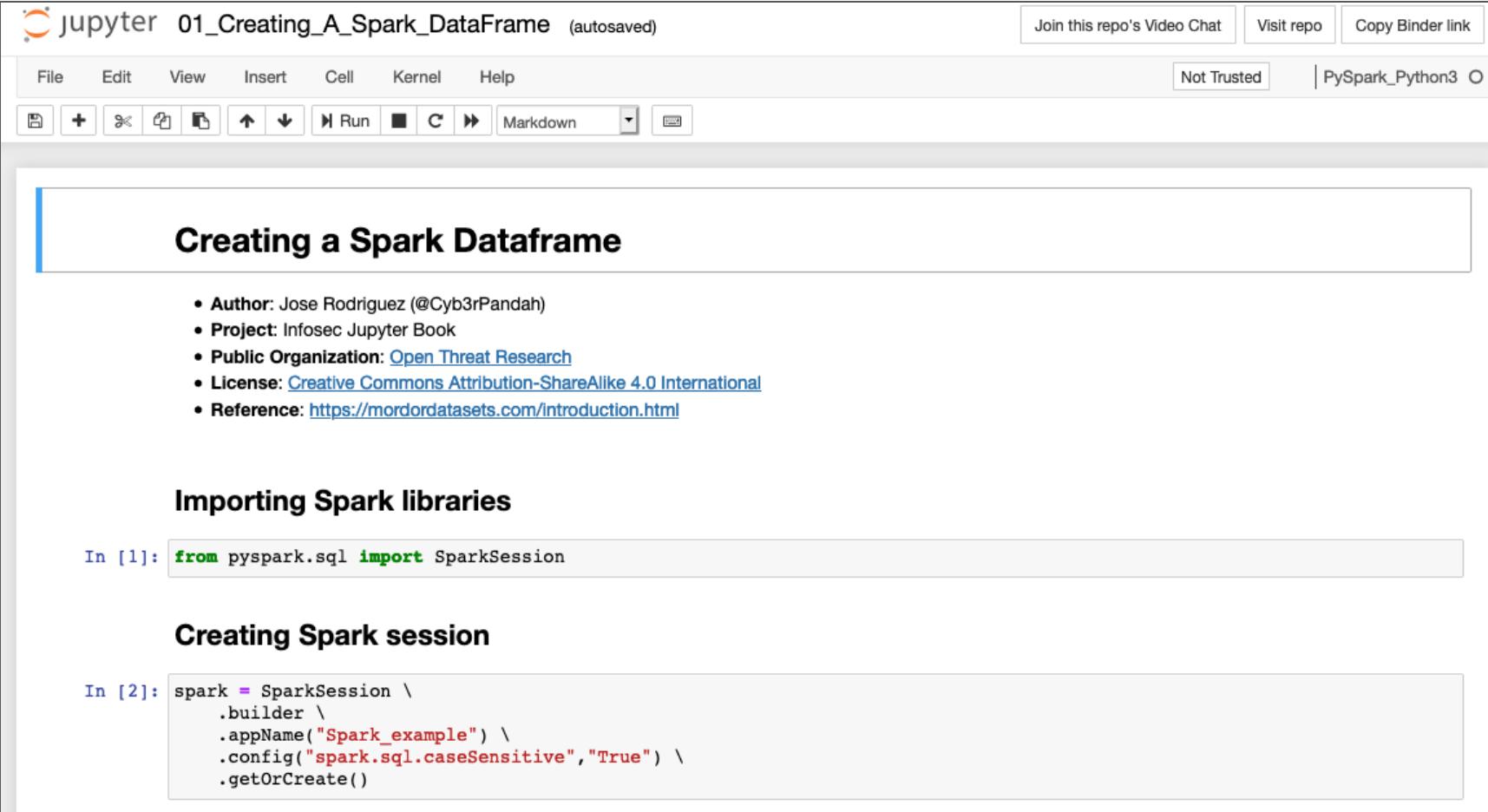
We use the [repo2docker](#) tool to automatically build the environment in which
to run your code.

Build logs

hide

```
----> adba5dal0ffd
Step 11/18 : RUN adduser --disabled-password      --gecos "Default user"      --uid ${NB_UID}      ${NB_USER}
    ---> Using cache
    ---> d35527f8072d
Step 12/18 : USER ${NB_USER}
    ---> Using cache
    ---> 417aea3e0599
Step 13/18 : RUN python3 -m pip install openhunt==1.6.8 bokeh==2.1.1 seaborn==0.10.1 --user
    ---> Using cache
    ---> 291ad28cac1b
Step 14/18 : COPY docs ${HOME}/docs
    ---> 398ec97b54a7
Step 15/18 : USER root
    ---> Running in 9c6945764ddb
Removing intermediate container 9c6945764ddb
```

Usando binder



jupyter 01_Creating_A_Spark_DataFrame (autosaved)

Join this repo's Video Chat | Visit repo | Copy Binder link

File Edit View Insert Cell Kernel Help

Not Trusted | PySpark_Python3 O

Creating a Spark Dataframe

- Author: Jose Rodriguez (@Cyb3rPandah)
- Project: Infosec Jupyter Book
- Public Organization: [Open Threat Research](#)
- License: [Creative Commons Attribution-ShareAlike 4.0 International](#)
- Reference: <https://mordordatasets.com/introduction.html>

Importing Spark libraries

In [1]: `from pyspark.sql import SparkSession`

Creating Spark session

In [2]: `spark = SparkSession \
 .builder \
 .appName("Spark_example") \
 .config("spark.sql.caseSensitive","True") \
 .getOrCreate()`

Introducción a Pandas



Una librería del lenguaje Python

- Pandas provee **estructuras de datos** que han sido diseñadas para que el trabajo con data relacionada o categorizada sea de una forma eficiente, fácil e intuitiva.
- A su vez, Pandas depende de la librería **Numpy** y sus conceptos de arrays/arreglos multidimensionales para ejecutar operaciones matemáticas de una manera eficiente.

Un Array?

- **Estructuras de Python similares a una lista de Python pero limitadas en los tipos de objetos que pueden ser guardados en el mismo array.**

```
import array

array_one = array.array('i',[1,2,3,4])
type(array_one)

type(array_one[0])

int
```

Type code	C Type	Python Type	Minimum size in bytes
'b'	signed char	int	1
'B'	unsigned char	int	1
'u'	Py_UNICODE	Unicode character	2
'h'	signed short	int	2
'H'	unsigned short	int	2
'i'	signed int	int	2
'I'	unsigned int	int	2
'l'	signed long	int	4
'L'	unsigned long	int	4
'f'	float	float	4
'd'	double	float	8

Quieres Seguir Los Demos? (Notebook #1)

https://otrf.github.io/workshop-ekoparty-bluespace-2020/conceptos-basicos/1_intro_numpy_arrays.html

NumPy N-Dimensional Array (ndarray)?

- El ndarray es un contenedor multidimensional de objetos del mismo tipo de data.
- Extiende el concepto de ejecucion de funciones matematicas en un array y permite la ejecución de tasks complejas en un array entero sin crear un Python For loop (Operaciones vectoriales eficientes).
- Numpy arrays, a su vez, usa menos memoria que una lista

```
import numpy as np  
np.__version__  
'1.19.2'
```

```
list_one = [1,2,3,4,5]
```

```
numpy_array = np.array(list_one)  
type(numpy_array)
```

```
numpy.ndarray
```

```
numpy_array
```

```
array([1, 2, 3, 4, 5])
```

NumPy N-Dimensional Array (ndarray)?

- El ndarray es un contenedor multidimensional de objetos del mismo tipo de data.
- Extiende el concepto de ejecucion de funciones matematicas en un array y permite la ejecución de tasks complejas en un array entero sin crear un Python For loop (Operaciones vectoriales eficientes).
- Numpy arrays, a su vez, usa menos memoria que una lista

```
: list_two = [1,2,3,4,5]
# The following will throw an error:
list_two + 2
```

```
-----  
TypeError                                         Traceback (most
<ipython-input-8-03923fe34c76> in <module>
      1 list_two = [1,2,3,4,5]
      2 # The following will throw an error:
----> 3 list_two + 2

TypeError: can only concatenate list (not "int") to list
```

- Performing a loop to add **2** to every integer in the list

NumPy N-Dimensional Array (ndarray)?

- El ndarray es un contenedor multidimensional de objetos del mismo tipo de data.
- Extiende el concepto de ejecucion de funciones matematicas en un array y permite la ejecución de tasks complejas en un array entero sin crear un Python For loop (Operaciones vectoriales eficientes).
- Numpy arrays, a su vez, usa menos memoria que una lista

```
for index, item in enumerate(list_two):
    list_two[index] = item + 2
list_two
[3, 4, 5, 6, 7]
```

NumPy N-Dimensional Array (ndarray)?

- El ndarray es un contenedor multidimensional de objetos del mismo tipo de data.
- Extiende el concepto de ejecucion de funciones matematicas en un array y permite la ejecución de tasks complejas en un array entero sin crear un Python For loop (Operaciones vectoriales eficientes).
- Numpy arrays, a su vez, usa menos memoria que una lista

numpy_array

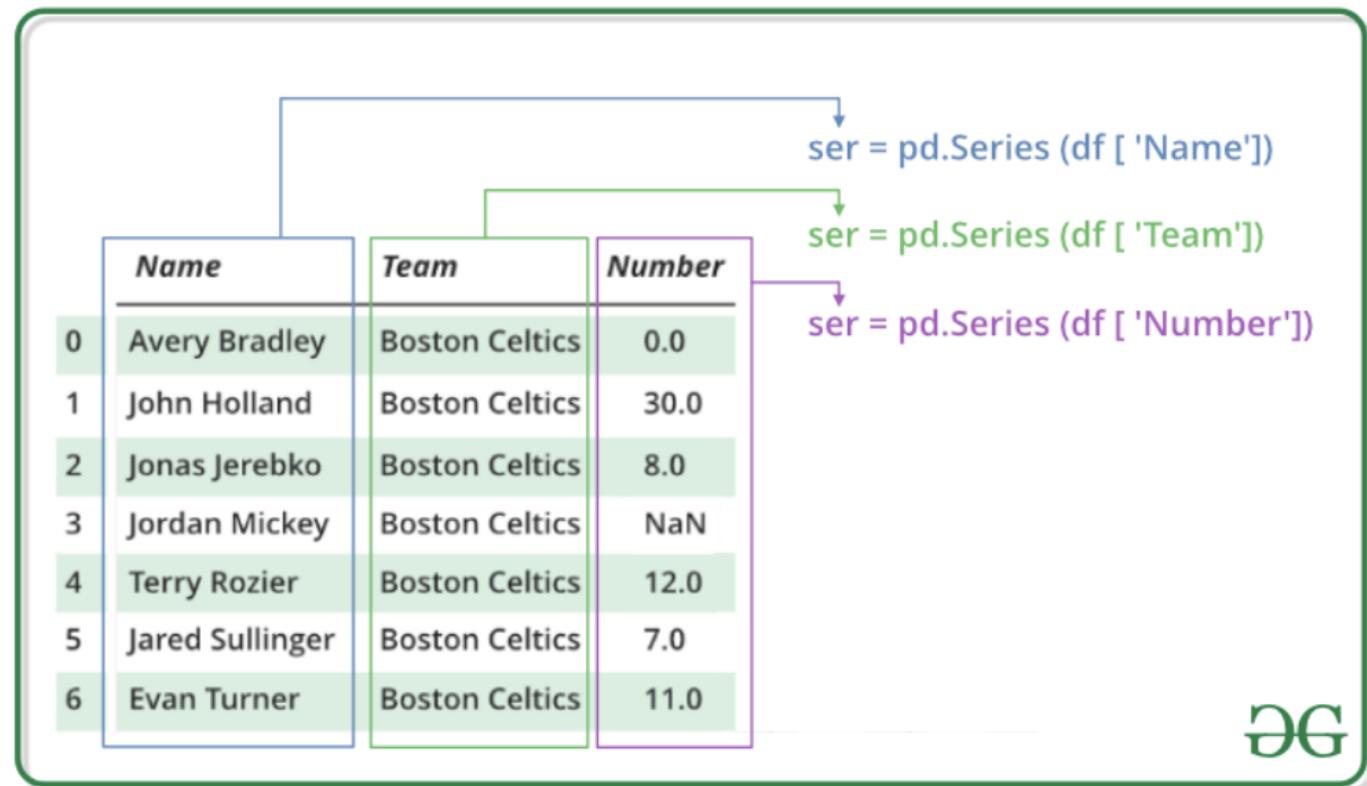
array([1, 2, 3, 4, 5])

numpy_array + 2

array([3, 4, 5, 6, 7])

Estructuras de data de Pandas

- Las dos principales estructuras de datos que utiliza Pandas son **Series** (1 dimensión) y **Dataframe** (2 dimensiones).
- Estas estructuras permiten a un usuario gestionar data en diversos casos de uso.



DEG

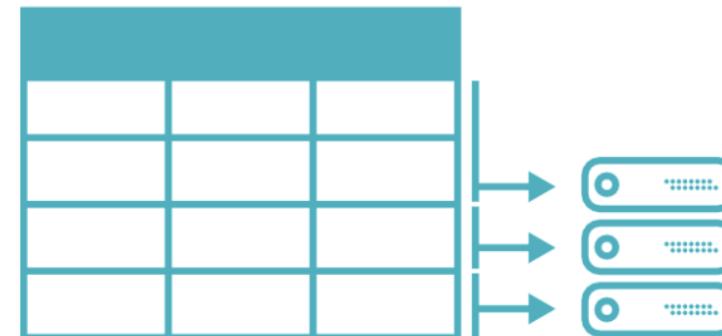
Que es una **Dataframe**?

Un dataframe es una estructura de datos con dos dimensiones que se encuentra categorizada por columnas que podrían incluir diferentes tipos de datos.

Spreadsheet on a single machine



Table or DataFrame partitioned across servers in a data center



Que podemos hacer con **Pandas**?

- Gestionar data faltante
- Agregar data (Group By)
- Segmentar, indexar y categorizar data
- Representar data como dataframe otras estructuras de datos
- Correlacionar data (Join, Merge)
- Análisis de series de tiempo
- Crear visualizaciones
- Y mucho más...

Cómo instalar **Pandas**?

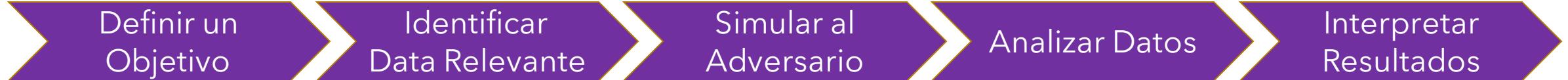
- A través de Anaconda
- A través de Miniconda
 - conda install pandas
 - conda install pandas=0.20.3
- Instalando desde PyPi (Python Package Index)
 - pip install pandas



El Proceso de Análisis de Datos



El proceso de Análisis de Datos

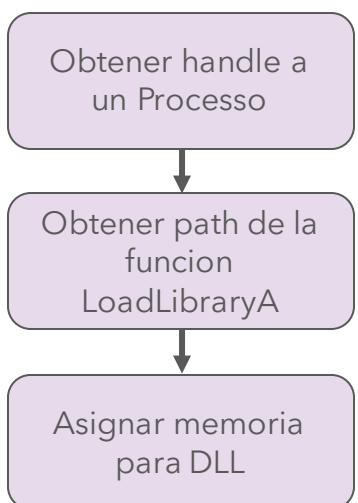


El proceso de Análisis de Datos



MITRE | ATT&CK®

Strategia del Adversario



El proceso de Análisis de Datos



MITRE | ATT&CK®

Strategia del Adversario

Obtener handle a un Processo

Obtener path de la función LoadLibraryA

Asignar memoria para DLL

Sysmon 10
Acceso a Proceso

Process A Process B

Sysmon 8
Creación de Thread remoto

El proceso de Análisis de Datos

Definir un
Objetivo

Identificar
Data Relevante

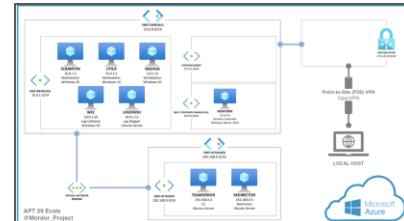
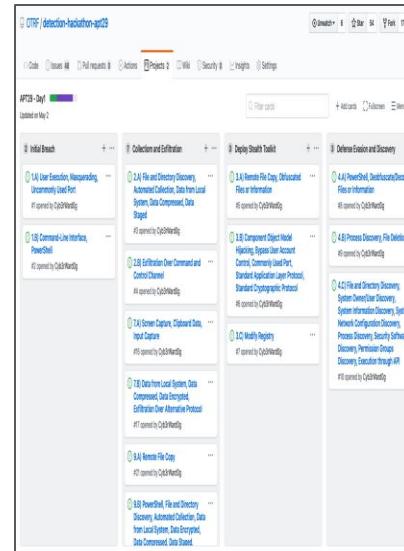
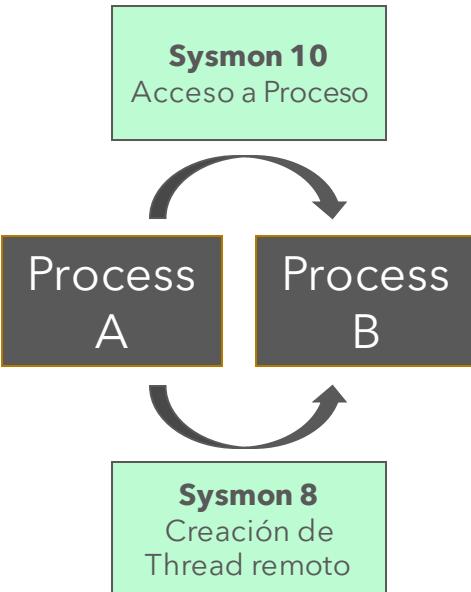
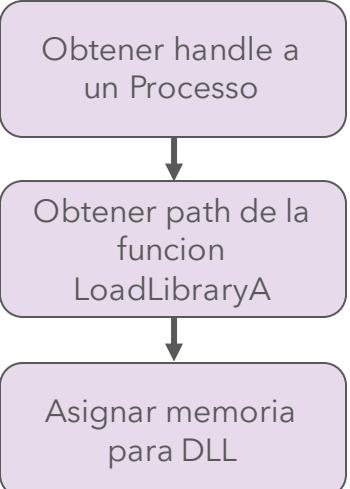
Simular al
Adversario

Analizar Datos

Interpretar
Resultados

MITRE | ATT&CK®

Strategia del
Adversario



El proceso de Análisis de Datos

Definir un
Objetivo

Identificar
Data Relevante

Simular al
Adversario

Analizar Datos

Interpretar
Resultados

MITRE | ATT&CK®

Strategia del
Adversario

Obtener handle a
un Processo

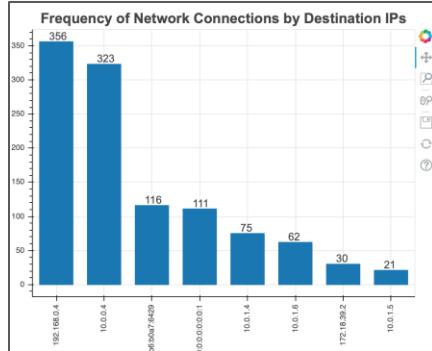
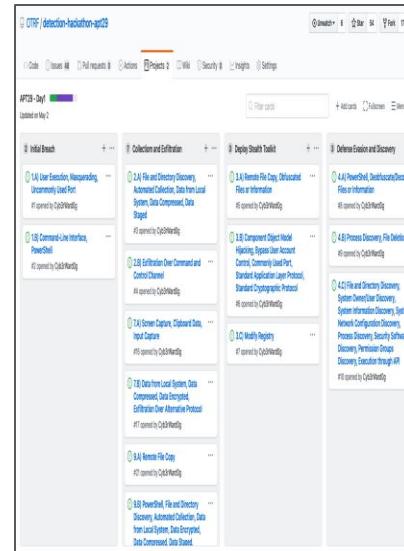
Obtener path de la
funcion
LoadLibraryA

Asignar memoria
para DLL

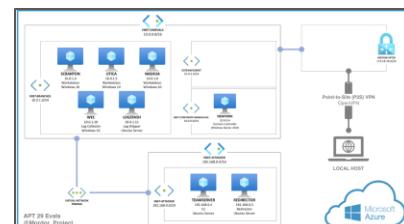
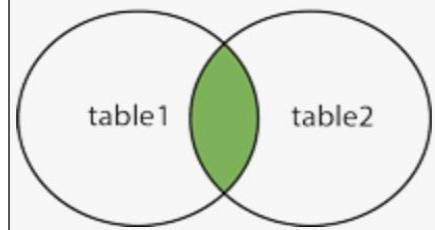
Sysmon 10
Acceso a Proceso



Sysmon 8
Creación de
Thread remoto



INNER JOIN



El proceso de Análisis de Datos

Definir un
Objetivo

Identificar
Data Relevante

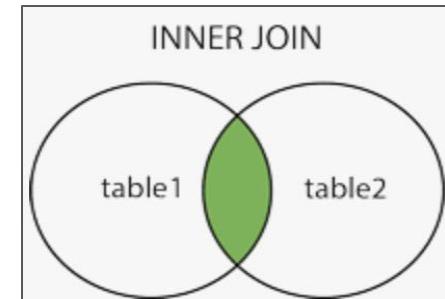
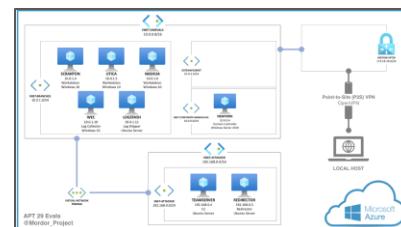
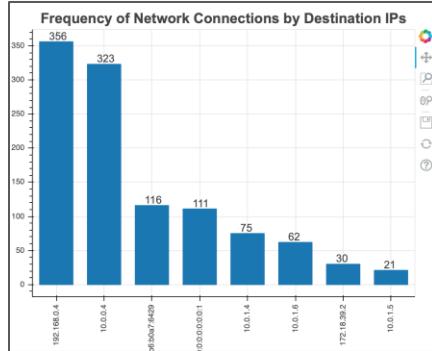
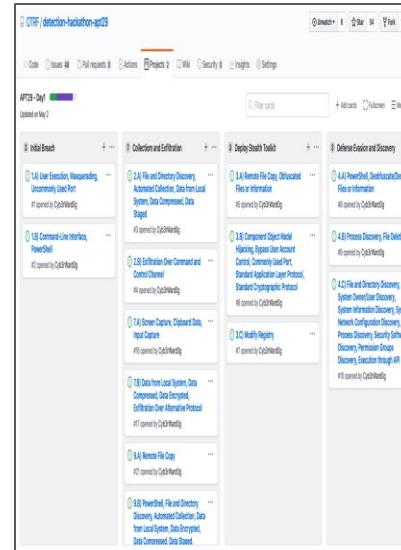
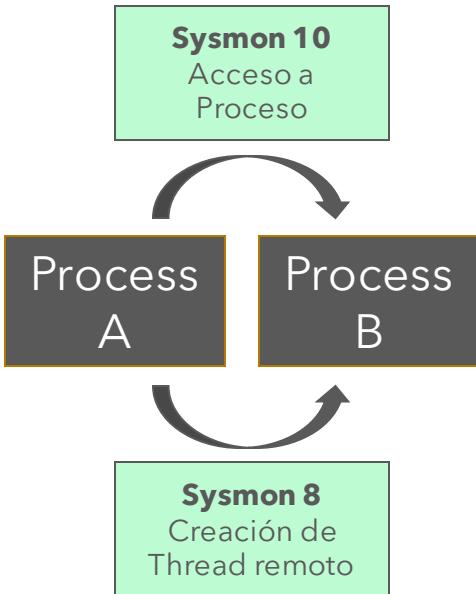
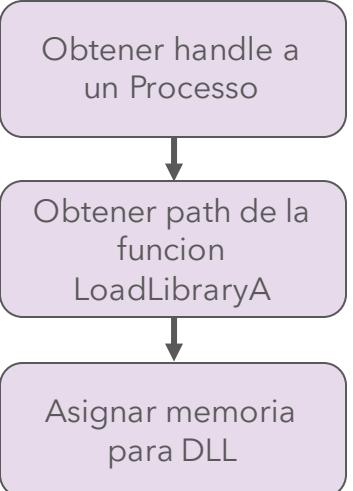
Simular al
Adversario

Analizar Datos

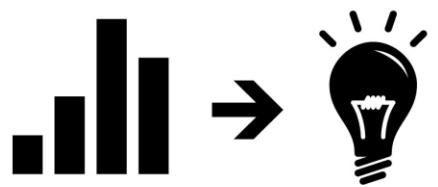
Interpretar
Resultados

MITRE | ATT&CK®

Strategia del
Adversario



Ideas para
comenzar a
definir
estrategias de
detección



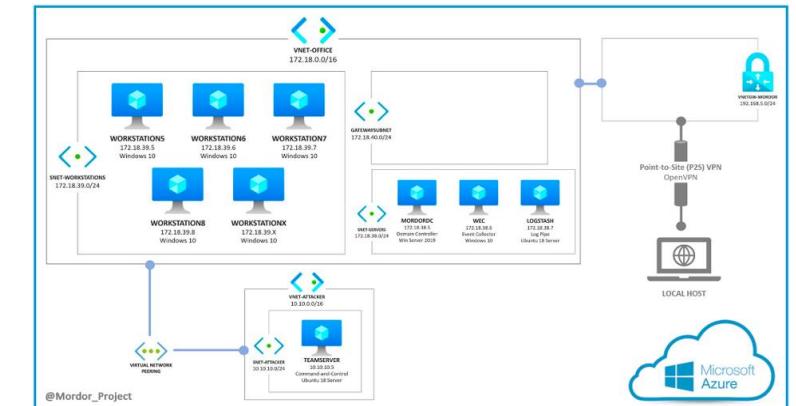
Necesitamos
Data?...

MORDOR



@Mordor_Project

- Eventos de seguridad pre-grabados, generados a través de la simulación de técnicas usadas por adversarios en
- Formato JavaScript Object Notation (**JSON**)
- Sets de datos categorizados por plataformas, grupos de adversarios, tácticas y técnicas definidas por MITRE - ATT&CK
- Datasets pequeñas y grandes



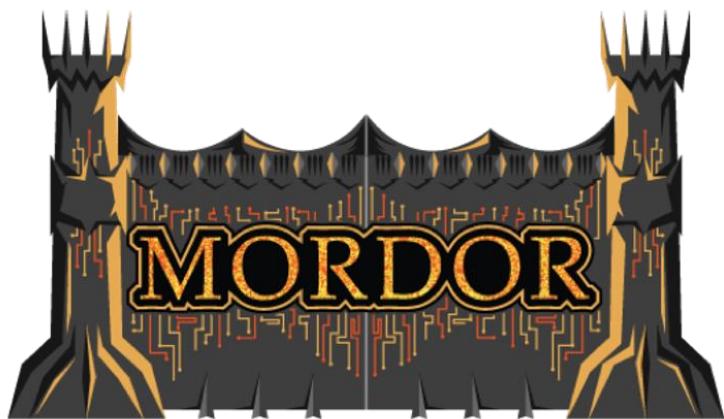
@Mordor_Project

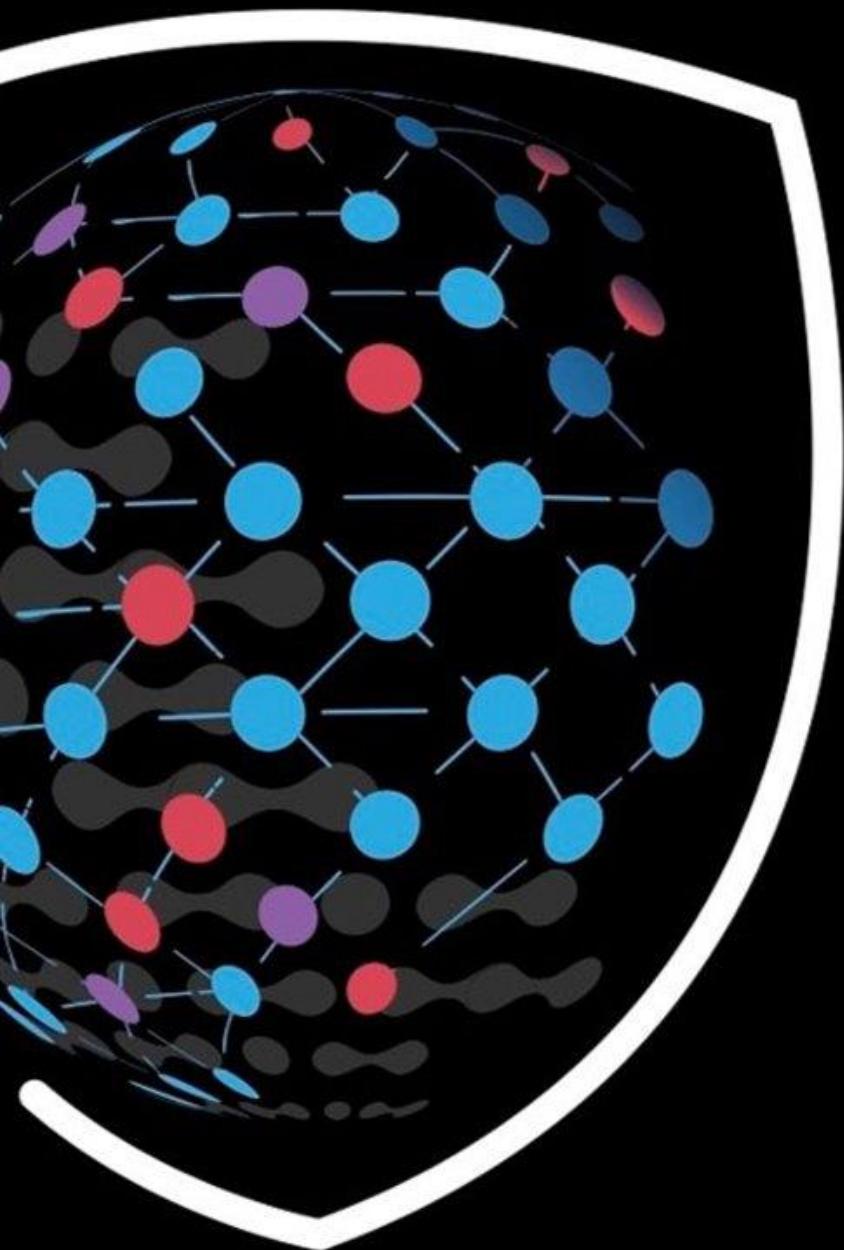
The screenshot shows a web browser window with the following details:

- Address Bar:** https://mordordatasets.com/notebooks/small/windows/08_lateral_movement/SDWIN-190518210652.html
- Left Sidebar (Lateral Movement):**
 - Empire Invoke SMBExec
 - Covenant PowerShell Remoting
 - Command
 - DCOM RegisterXLL
 - Covenant SharpWMI Exec
 - Covenant SharpSC Start
 - Covenant Remote File Copy
 - Empire Invoke PSRemoting
 - Covenant SharpSC Create
 - Empire Over-Pass-The-Hash
 - Covenant SharpSC Stop Service
 - Empire Invoke DCOM ShellWindows
 - Covenant Remote WMI Eventing
 - ActiveScriptEventConsumers
 - Empire Invoke PsExec**
 - Covenant SharpSC Query
 - DCOM ExecuteExcel4macro
 - Covenant SC.exe Utility Query
 - Empire Invoke Execute MSBuild
- Main Content Area:**
 - Dataset Description:** This dataset represents adversaries remotely creating and starting a service via RPC methods over TCP.
 - Datasets Downloads:** A table showing download links:

Type	Link
Host	https://raw.githubusercontent.com/OTRF/mordor/master/datasets/small/windows/lateral_movement/host/empire_psexec_dcerpc_tcp_svcctl.zip
Network	https://raw.githubusercontent.com/OTRF/mordor/master/datasets/small/windows/lateral_movement/network/empire_psexec_dcerpc_tcp_svcctl.zip

Ejemplo 1: Importando sets de datos





OPEN THREAT RESEARCH

EMPOWERING THE INFOSEC COMMUNITY

Quieres Seguir Los Demos? (Notebook #2)

https://otrf.github.io/workshop-ekoparty-bluespace-2020/conceptos-basicos/2_dataframe_desde_Set_datos_Mordor.html

Importando sets de datos desde Mordor

The screenshot shows a Jupyter Notebook interface with the title "jupyter dataframe_desde_Set_datos_Mordor". The notebook contains the following content:

Dataframe desde un Set de Datos de Mordor

- Autor: Jose Rodriguez (@Cyb3rPandah)
- Proyecto: Infosec Jupyter Book
- Organización Pública: [Open Threat Research](#)
- Licencia: [Creative Commons Attribution-ShareAlike 4.0 International](#)
- Referencia: https://mordordatasets.com/notebooks/small/windows/02_execution/SDWIN-200806115603.html

Obteniendo el archivo JSON

a) Descargando el archivo Zip

Usaremos el comando `wget` y la opción `-O` (output document file) para guardar el archivo Zip en la carpeta `sets_datos`.

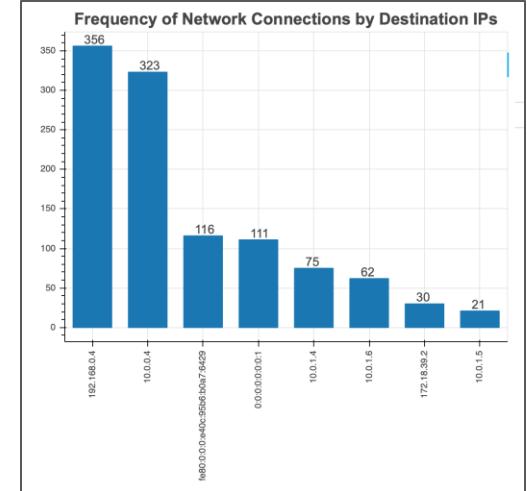
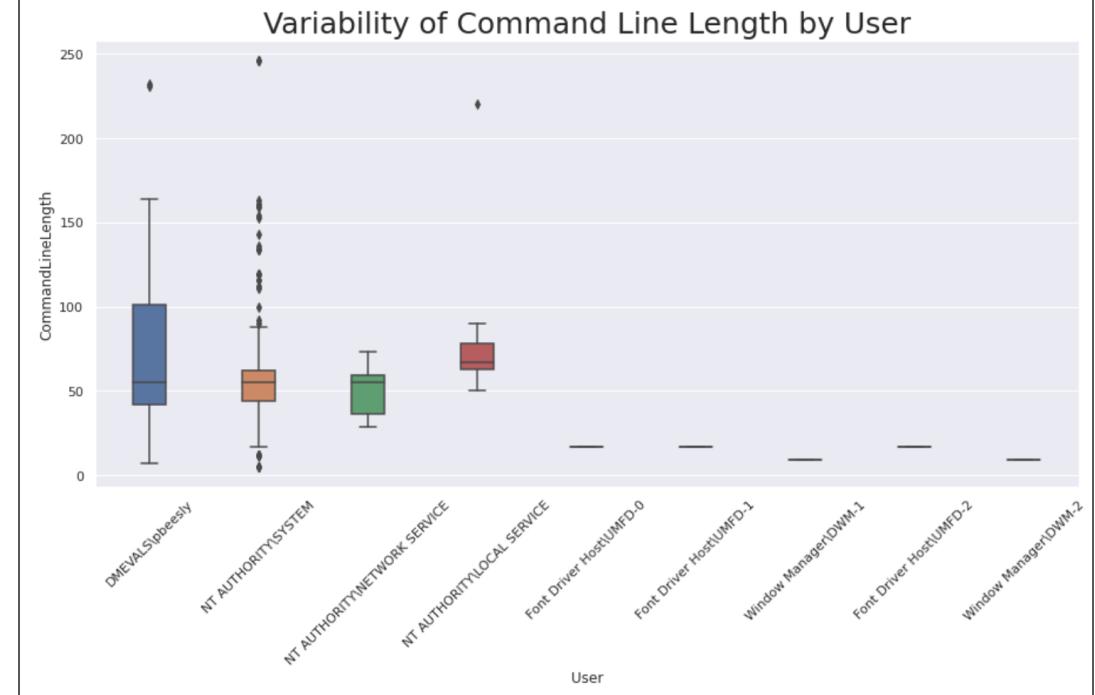
```
In [1]: ! wget https://raw.githubusercontent.com/OTRF/mordor/master/datasets/small/windows/lateral_movement/host/covenant_psremoting_command.zip
```

--2020-09-26 10:43:39-- https://raw.githubusercontent.com/OTRF/mordor/master/datasets/small/windows/lateral_movement/host/covenant_psremoting_command.zip
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 151.101.200.133
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|151.101.200.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 359550 (351K) [application/zip]
Saving to: 'sets_datos/covenant_psremoting_command.zip'

sets_datos/covenant 100%[=====] 351.12K ---KB/s in 0.1s
2020-09-26 10:43:40 (2.32 MB/s) - 'sets_datos/covenant_psremoting_command.zip' saved [359550/359550]

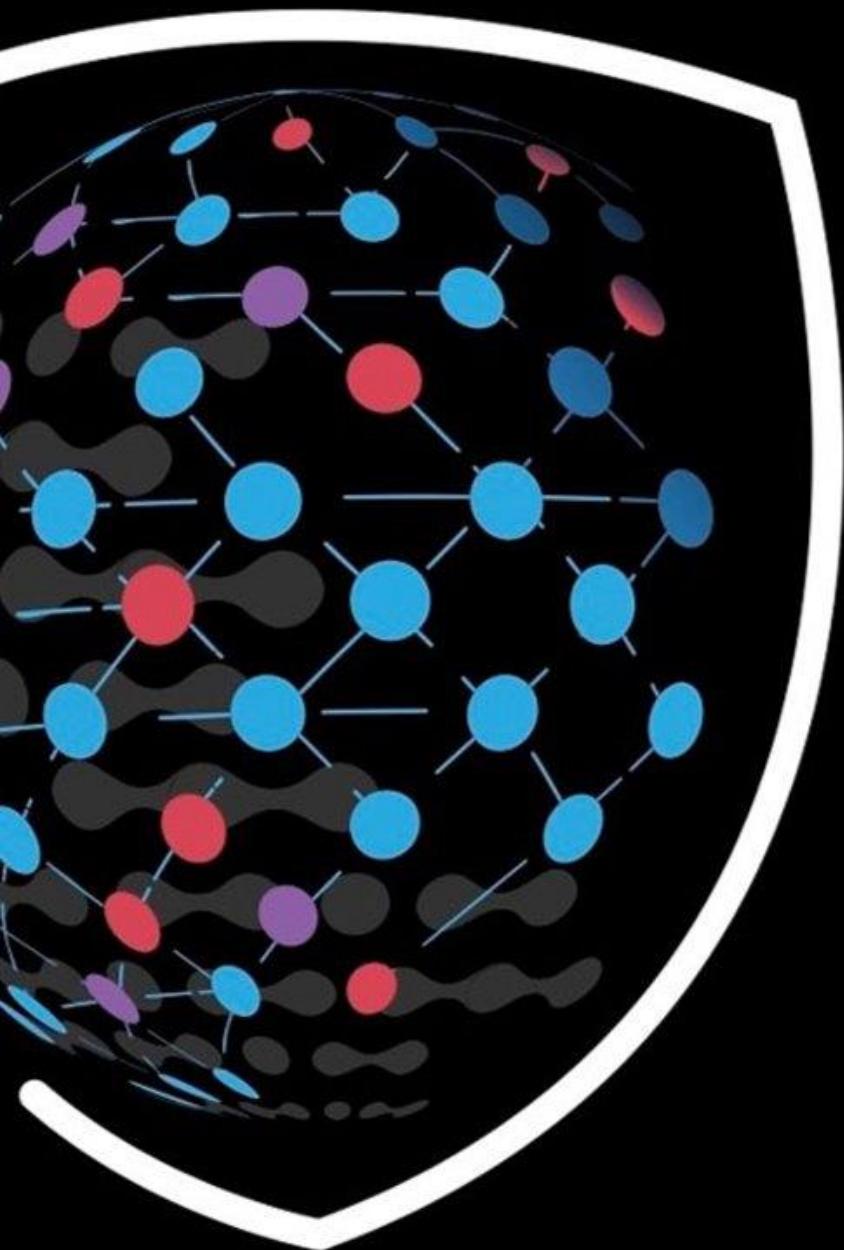
Técnicas para Análisis de Datos

GrantedAccess	Count
0x1000	463
0x3000	83
0x40	4
0xfffffff	2
0x1400	2
0x1410	2
0x1478	2
0xf3ffff	1
0x100000	1
0x101541	1



Algunas técnicas basicas son:

- Resumir data
- Filtrar columnas y/o filas
- Transformar data
- Correlacionar data
- Visualizar data



OPEN THREAT RESEARCH

EMPOWERING THE INFOSEC COMMUNITY

Quieres Seguir Los Demos? (Notebook #3)

https://otrf.github.io/workshop-ekoparty-bluespace-2020/conceptos-basicos/3_Analisis_Datos_Pandas_Filtrando_Resumiendo_Datos.html

Resumiendo data con Pandas

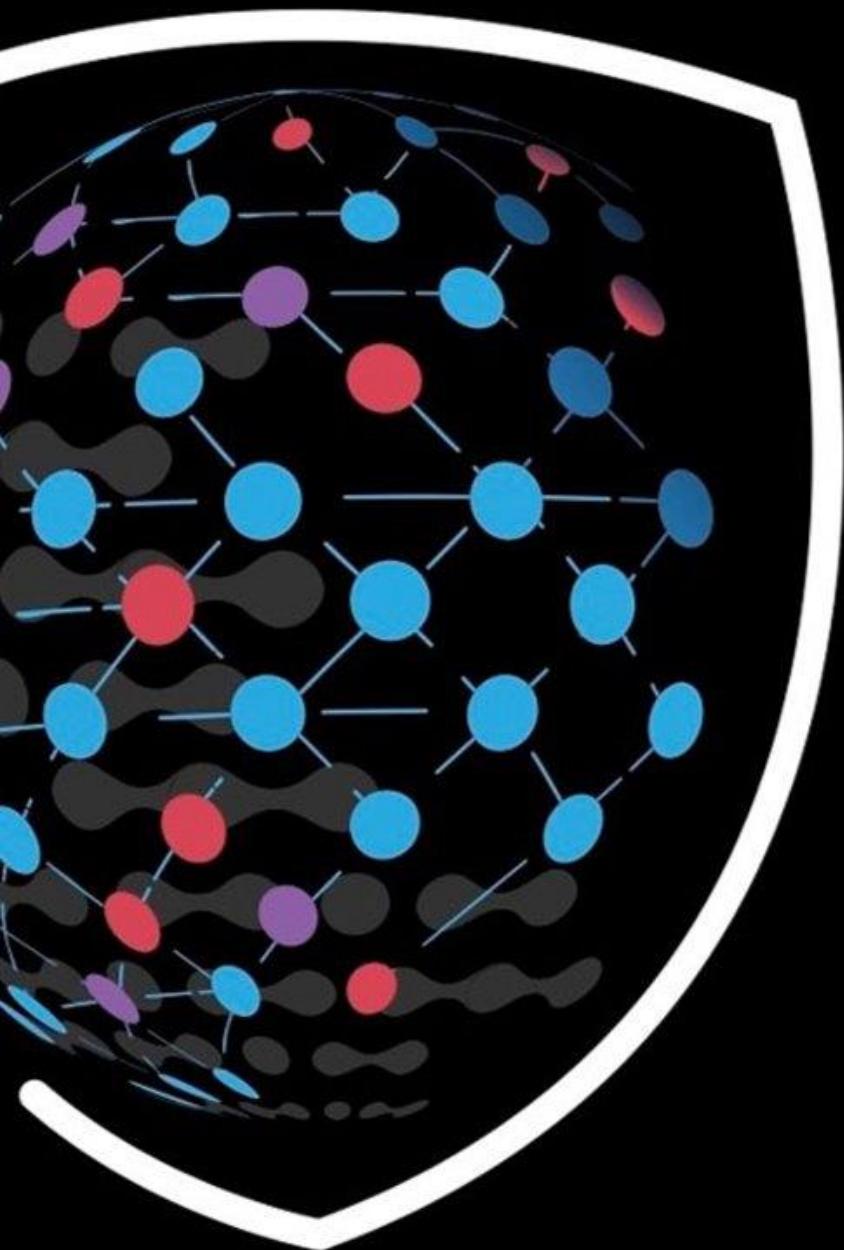
The screenshot shows a Jupyter Notebook interface with the title "jupyter Analisis_Datos_Pandas_Filtrando_Resumiendo_Datos Last Checkpoint: 3 hours ago (autosaved)" and a "Logout" button. The toolbar includes File, Edit, View, Insert, Cell, Kernel, Help, and various cell type icons. The kernel is set to "PySpark_Python3".

In [51]:

```
(  
df[['@timestamp','Hostname','Channel','ParentImage','Image','EventID']]  
  
[(df['Image'].str.contains('wsmprovhost.exe',case = False, na = False, regex = False)) |  
 (df['ParentImage'].str.contains('wsmprovhost.exe',case = False, na = False, regex = False))]  
  
.head(5)  
)
```

Out[51]:

	@timestamp	Hostname	Channel	ParentImage	Image	EventID
656	2020-08-06T15:56:24.416Z	WORKSTATION6.theshire.local	Microsoft-Windows-Sysmon/Operational	C:\Windows\System32\svchost.exe	C:\Windows\System32\wsmprovhost.exe	1
666	2020-08-06T15:56:24.419Z	WORKSTATION6.theshire.local	Microsoft-Windows-Sysmon/Operational	NaN	C:\Windows\System32\wsmprovhost.exe	7
669	2020-08-06T15:56:24.420Z	WORKSTATION6.theshire.local	Microsoft-Windows-Sysmon/Operational	NaN	C:\Windows\System32\wsmprovhost.exe	7
673	2020-08-06T15:56:24.423Z	WORKSTATION6.theshire.local	Microsoft-Windows-Sysmon/Operational	NaN	C:\Windows\System32\wsmprovhost.exe	7
675	2020-08-06T15:56:24.424Z	WORKSTATION6.theshire.local	Microsoft-Windows-Sysmon/Operational	NaN	C:\Windows\System32\wsmprovhost.exe	7



OPEN THREAT RESEARCH

EMPOWERING THE INFOSEC COMMUNITY

Quieres Seguir Los Demos? (Notebook #4)

https://otrf.github.io/workshop-ekoparty-bluespace-2020/conceptos-basicos/4_Analisis_Datos_Pandas_Transformando_Datos.html

Transformando Datos

Calculando la Longitud del CommandLine

Usaremos el método `assign` para agregar una columna nueva a nuestro dataframe. Esta nueva columna mostrará el calculo de la longitud del command line que el proceso utilizó.

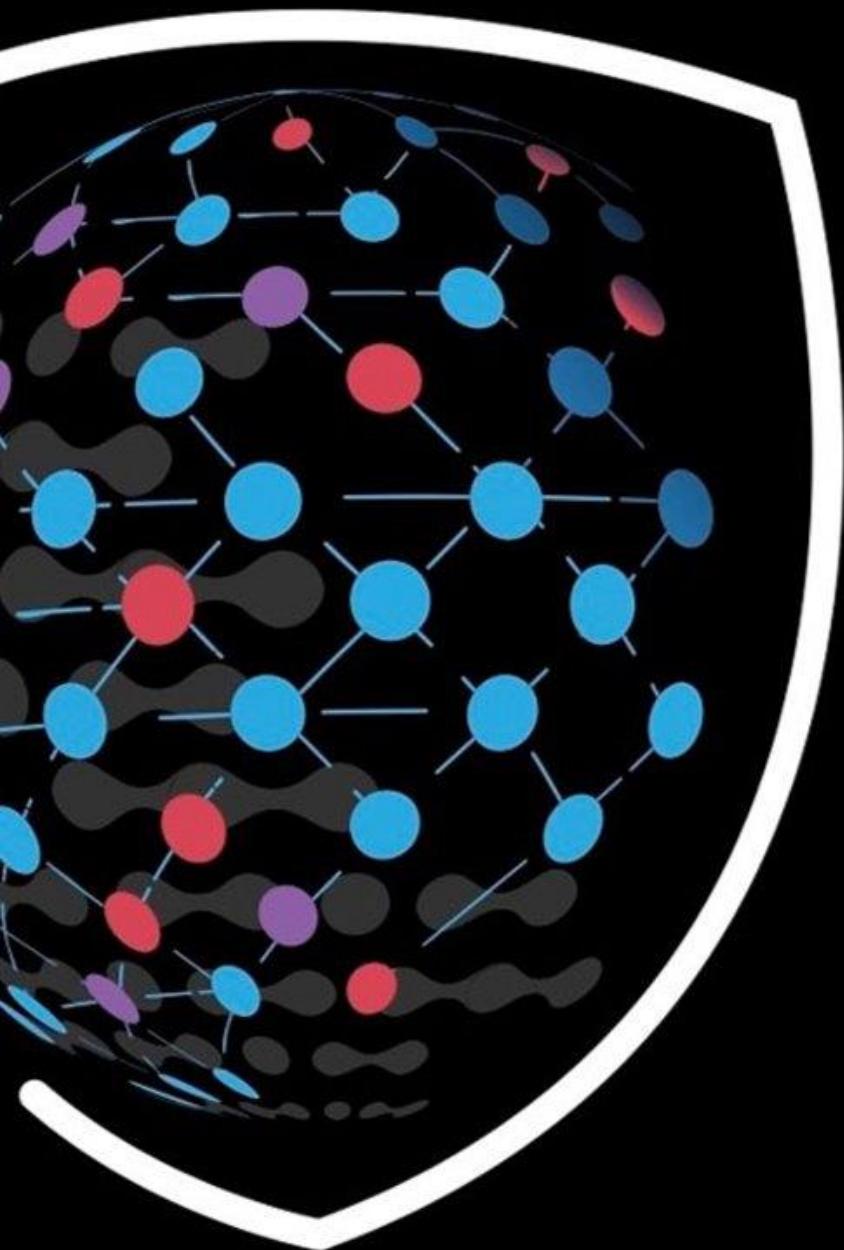
Referencia: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.assign.html>

```
In [6]:  
(  
df[['@timestamp', 'Image', 'CommandLine']]  
  
[(df['EventID'] == 1) & (df['Channel'].str.contains('sysmon', case = False, na = False, regex = False))]  
  
.assign(Command_Length = df['CommandLine'].str.len())  
)
```

Out[6]:

		@timestamp	Image	CommandLine	Command_Length
372	2020-05-02T02:55:57.730Z	C:\ProgramData\victim\cod.3aka3.scr	"C:\ProgramData\victim\cod.3aka3.scr" /S		43.0
621	2020-05-02T02:56:05.822Z	C:\Windows\System32\conhost.exe	\?\C:\windows\system32\conhost.exe --headless...		99.0
649	2020-05-02T02:56:05.830Z	C:\Windows\System32\cmd.exe	"C:\windows\system32\cmd.exe"		29.0
827	2020-05-02T02:56:15.884Z	C:\Windows\System32\WindowsPowerShell\v1.0\pow...		powershell	10.0
3620	2020-05-02T02:57:02.831Z	C:\Windows\System32\SearchProtocolHost.exe	"C:\windows\system32\SearchProtocolHost.exe" G...		308.0
...
193780	2020-05-02T03:25:33.847Z	C:\Windows\System32\UsoClient.exe	C:\windows\system32\usoclient.exe StartScan		43.0
193812	2020-05-02T03:25:33.858Z	C:\Windows\System32\usocoreworker.exe	C:\Windows\System32\usocoreworker.exe -Embedding		48.0
194036	2020-05-02T03:25:50.013Z	C:\Windows\System32\taskhostw.exe		taskhostw.exe Logon	19.0
194568	2020-05-02T03:26:12.287Z	C:\Windows\System32\UsoClient.exe	C:\windows\system32\usoclient.exe StartScan		43.0
194596	2020-05-02T03:26:12.298Z	C:\Windows\System32\usocoreworker.exe	C:\Windows\System32\usocoreworker.exe -Embedding		48.0

447 rows × 4 columns



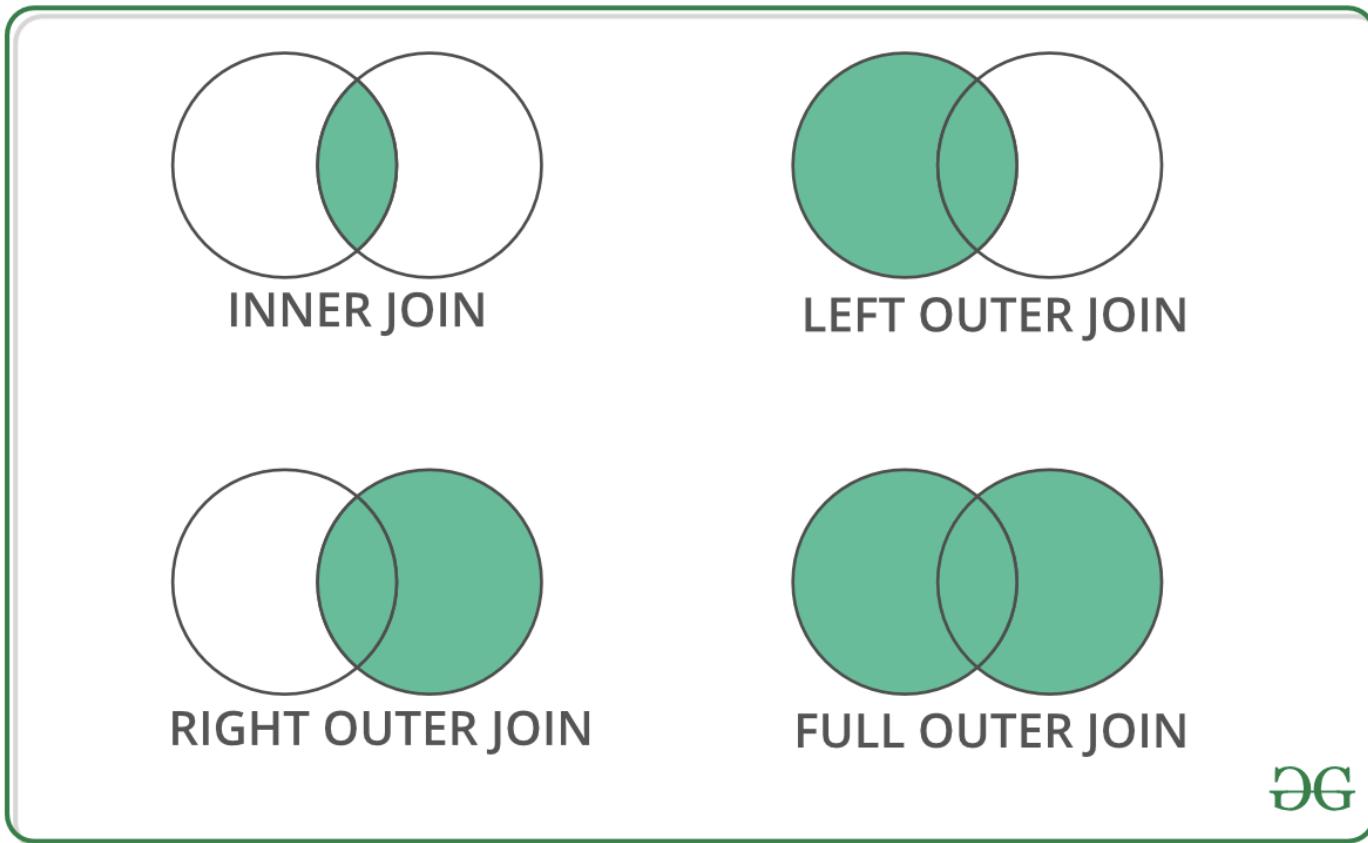
OPEN THREAT RESEARCH

EMPOWERING THE INFOSEC COMMUNITY

Quieres Seguir Los Demos? (Notebook #5)

https://otrf.github.io/workshop-ekoparty-bluespace-2020/conceptos-basicos/5_Analisis_Datos_Pandas_Correlacionando_Datos.html

Correlacionando Data con Pandas (JOINs)



JOIN

- Allows you to combine rows from the same or different tables
 - JOINs can be performed with **join()** or **merge()** with the following options (**LEFT, RIGHT, INNER, FULL**) and the columns to join on (**column names or indices**).

Preparando dataframes para JOIN

```
# Creating a dataframe with information of Security event 4624: An account was successfully logged on
Security4624 = (
    apt29[(apt29['Channel'].str.lower() == 'security') & (apt29['EventID'] == 4624)].dropna(axis = 1, how = 'all')
)
Security4624.shape

(297, 55)
```

```
# Creating a dataframe with information of Security event 4688: A new process has been created
Security4688 = (
    apt29[(apt29['Channel'].str.lower() == 'security') & (apt29['EventID'] == 4688)].dropna(axis = 1, how = 'all')
)
Security4688.shape

(460, 42)
```

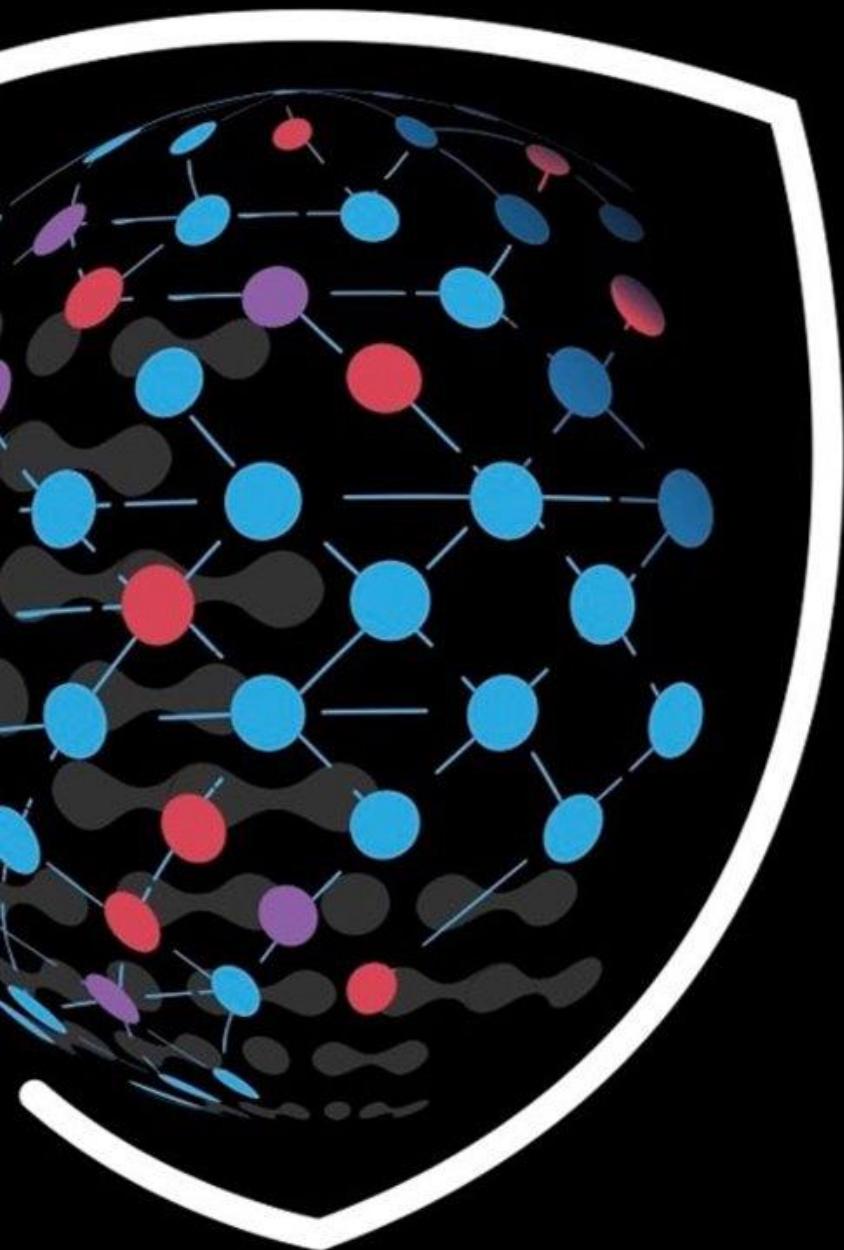
```
# Creating a dataframe with information of Security event 4697: A service was installed in the system
Security4697 = apt29[
    (apt29['Channel'].str.lower() == 'security') & (apt29['EventID'] == 4697)
].dropna(axis = 1, how = 'all')
Security4697.shape

(23, 37)
```

Procesos Siendo Ejecutados en el contexto de un logon de network (Tipo 3)

```
# merge performs an INNER JOIN by default
(
pd.merge(Security4688, Security4624[Security4624['LogonType'] == 3],
         on = 'TargetLogonId', how = 'inner')
[[ 'NewProcessId', 'NewProcessName', 'ProcessId_x', 'ParentProcessName', 'TargetUserName_y', 'IpAddress' ]]
)
```

	NewProcessId	NewProcessName	ProcessId_x	ParentProcessName	TargetUserName_y	IpAddress
0	0x1e28	C:\Windows\System32\wsmprovhost.exe	0x374	C:\Windows\System32\svchost.exe	pbeesly	-



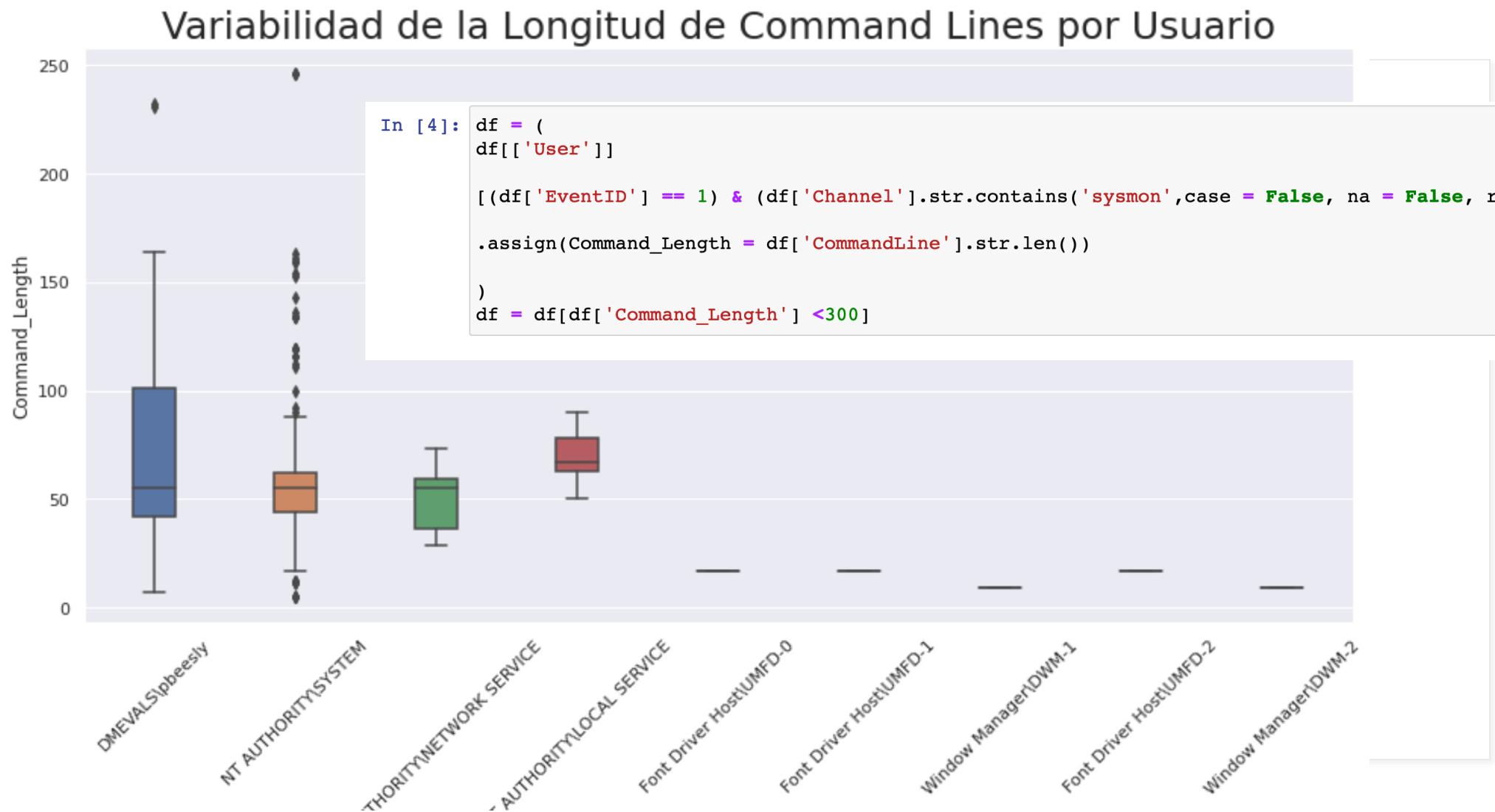
OPEN THREAT RESEARCH

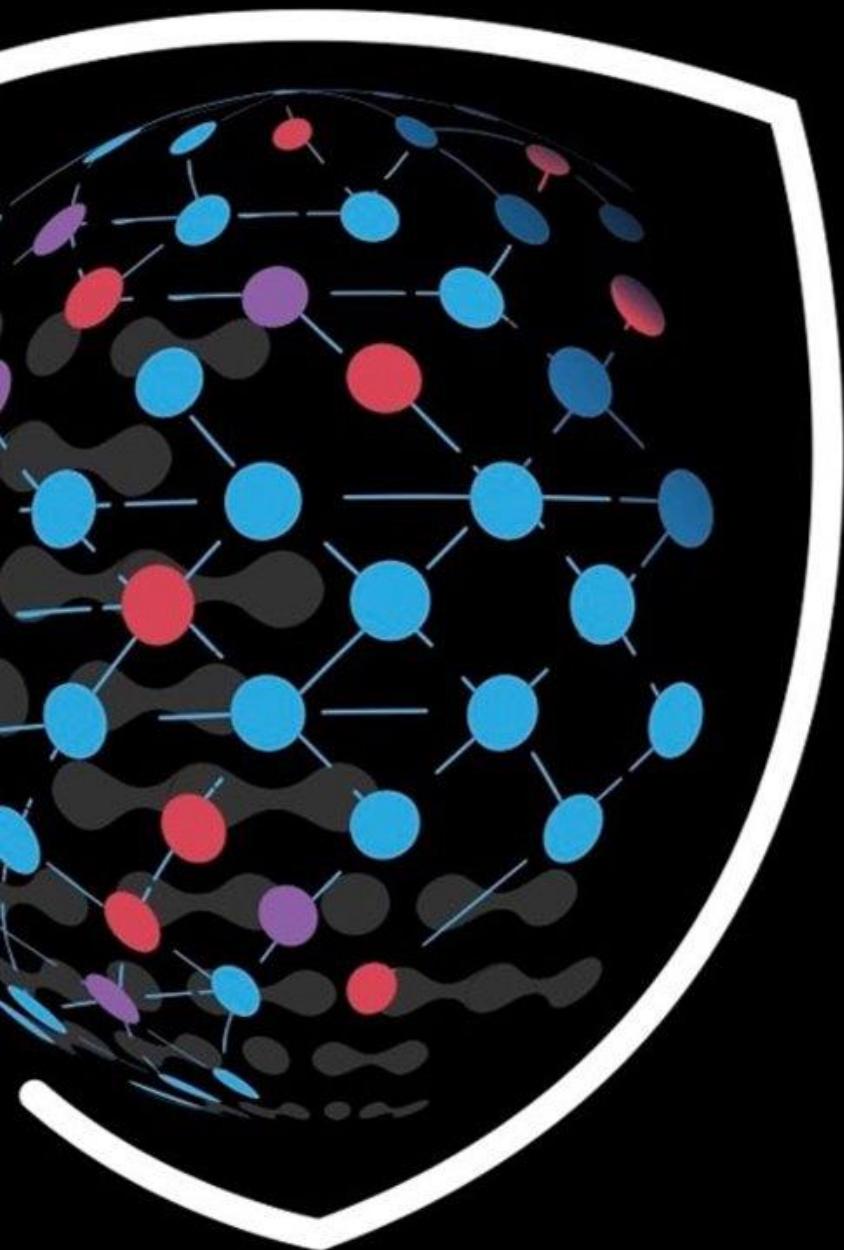
EMPOWERING THE INFOSEC COMMUNITY

Quieres Seguir Los Demos? (Notebook #6)

https://otrf.github.io/workshop-ekoparty-bluespace-2020/conceptos-basicos/6_Analisis_Datos_Pandas_Visualizando_Datos.html

Visualizando Datos





OPEN THREAT RESEARCH

EMPOWERING THE INFOSEC COMMUNITY