Paradigma

WEBINAR - noviembre, 25 2020

ML - Power BI: Visualizando nuestros modelos.

ponente: Marco Russo

Quién soy.

- **Consultor en Data** en Paradigma Digital, con más de 8 años como docente para importantes escuelas de negocios y profesor colaborador en la UOC.
- **Especializado** en data mining, optimización de modelos y machine learning en área del Marketing, Retail y Banca-Finanzas entre otras. Además de especialista en analítica digital, SEO y PPC en digital marketing y visualización de datos - BI.
- **Apasionado** de IoT, datos y robótica, dedico el tiempo con mi familia y a mi deporte favorito, bici de carretera.

Marco Russo (aka marcusRB)

@rb_marcus

github.com/marcusRB

marcusRB

Qué vamos a ver.

1. Configuración de los entornos:
    a. Power BI
    b. R studio
    c. Python venv
    d. Azure free trial
    e. Databricks evaluation
    f. Knime
2. Get Data (SQL database, dataset, opendata)
3. Modeling
    a. Jupyter Notebooks
    b. Azure Machine Learning
    c. Azure autoML
    d. Power BI Machine Learning
4. Visualization (elementos nativos, librerías)
5. Conclusiones

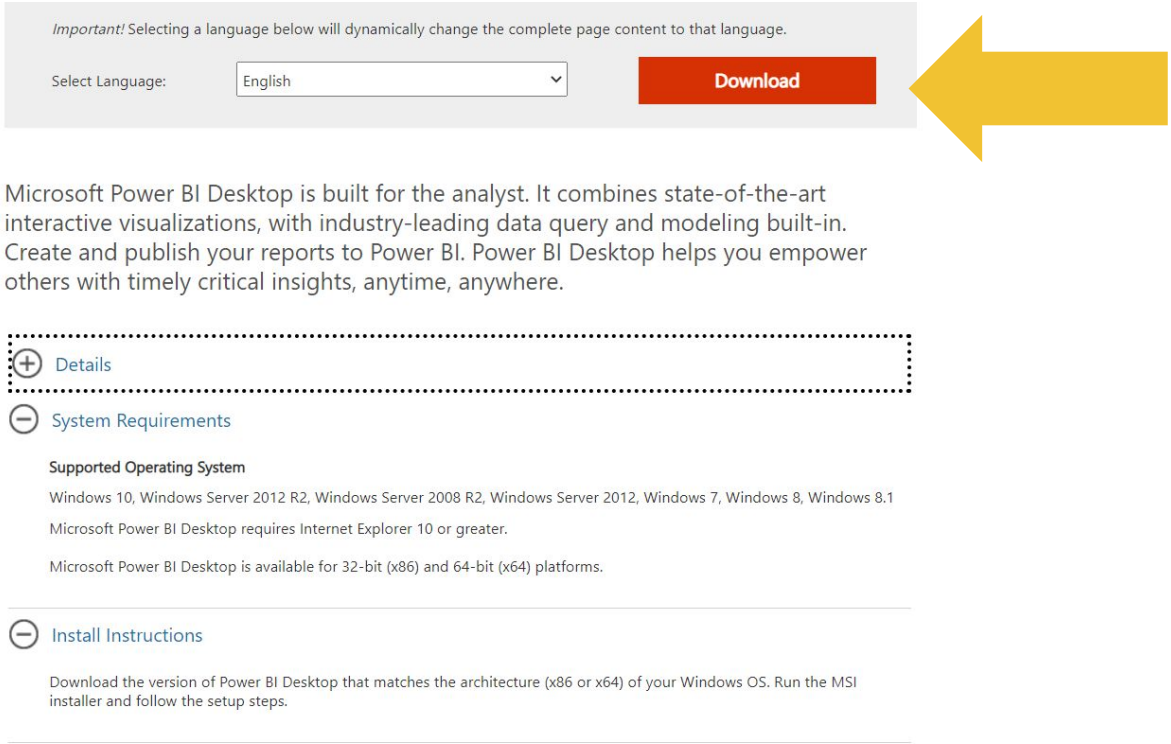
01.01

Introducción

Configuración entornos.

Power BI Desktop: entorno Windows

Power BI



Power BI Desktop: entorno virtual

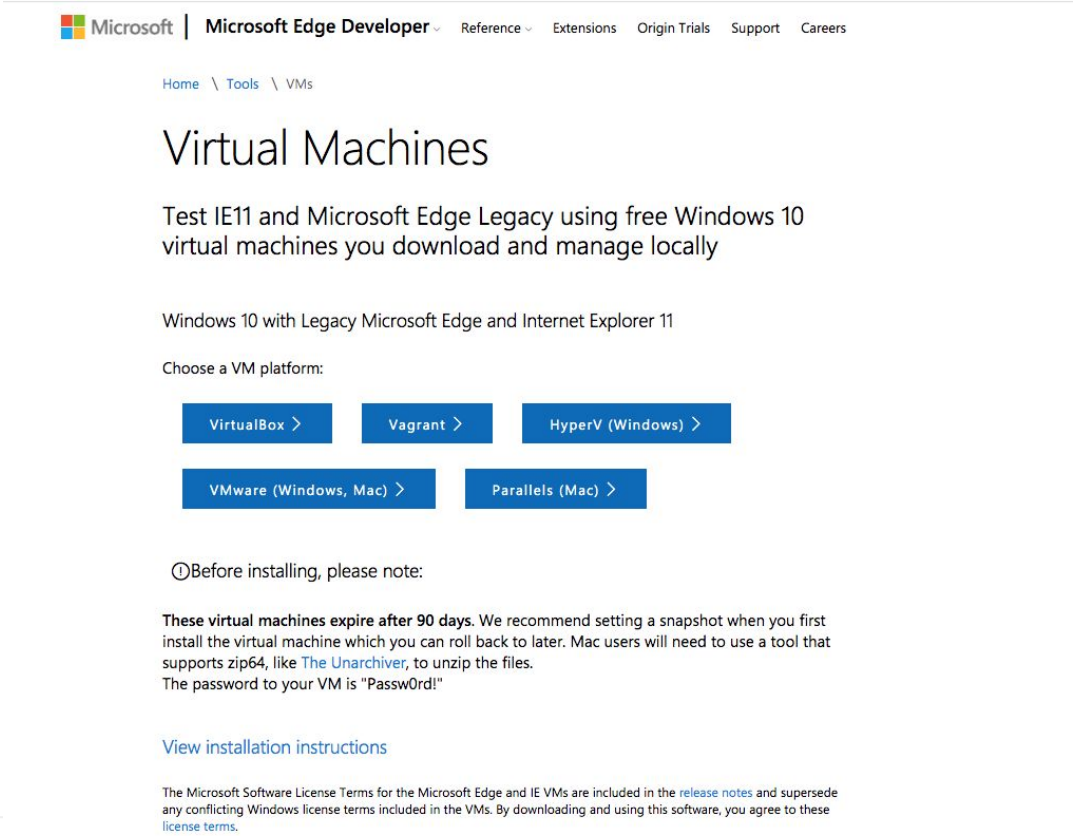
Power BI

Una máquina virtual replica un entorno Windows dentro de otro sistema operativo (sea Mac OS o Linux, e incluso otro Windows) y permite realizar test o pruebas como en este caso, de tener nuestra aplicación de Power BI .

Actualmente podemos instalar esta versión de Windows10 para las aplicaciones:

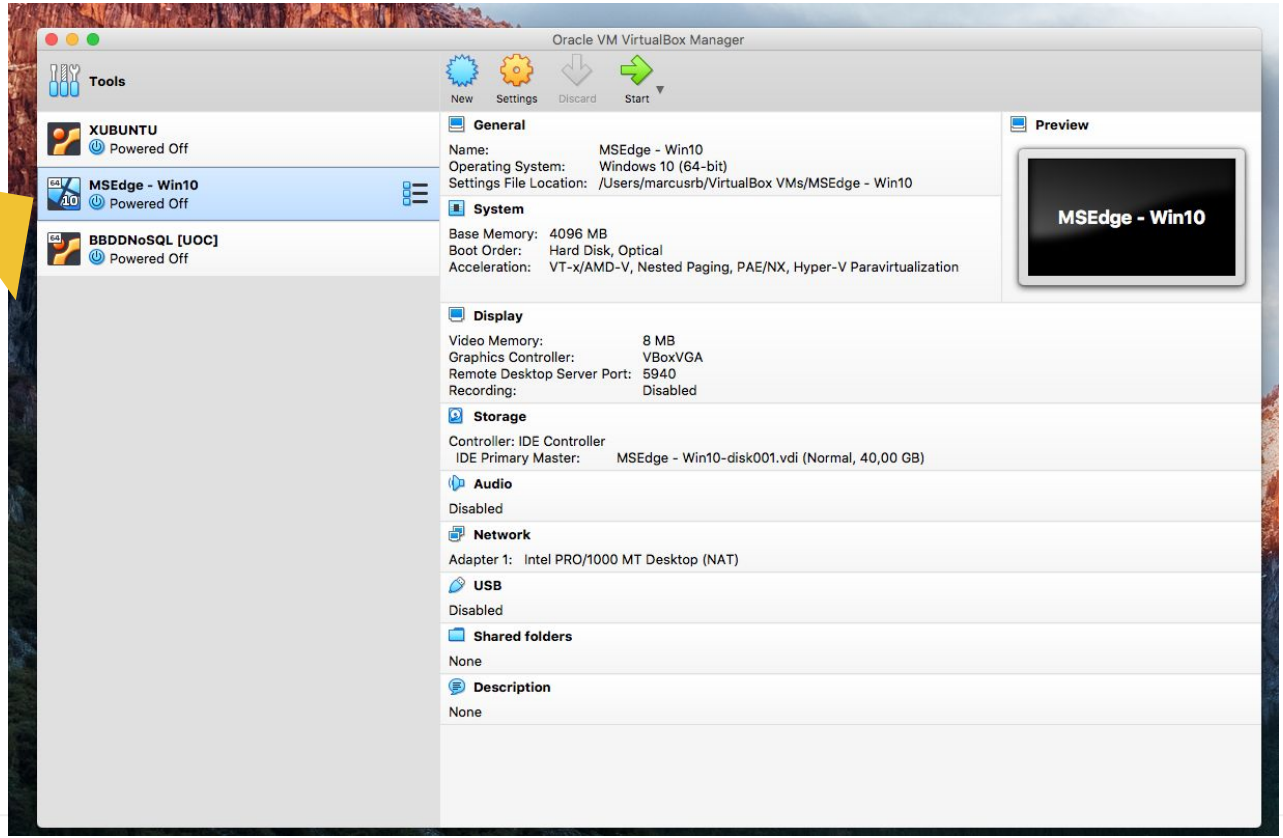
- Virtualbox
- VMware
- Parallels
- Vagrants
- HyperV

Y disponible a este enlace:
https://developer.microsoft.com/en-us/microsoft-edge/tools/vms/



Power BI Desktop: entorno virtual Virtualbox

Power BI



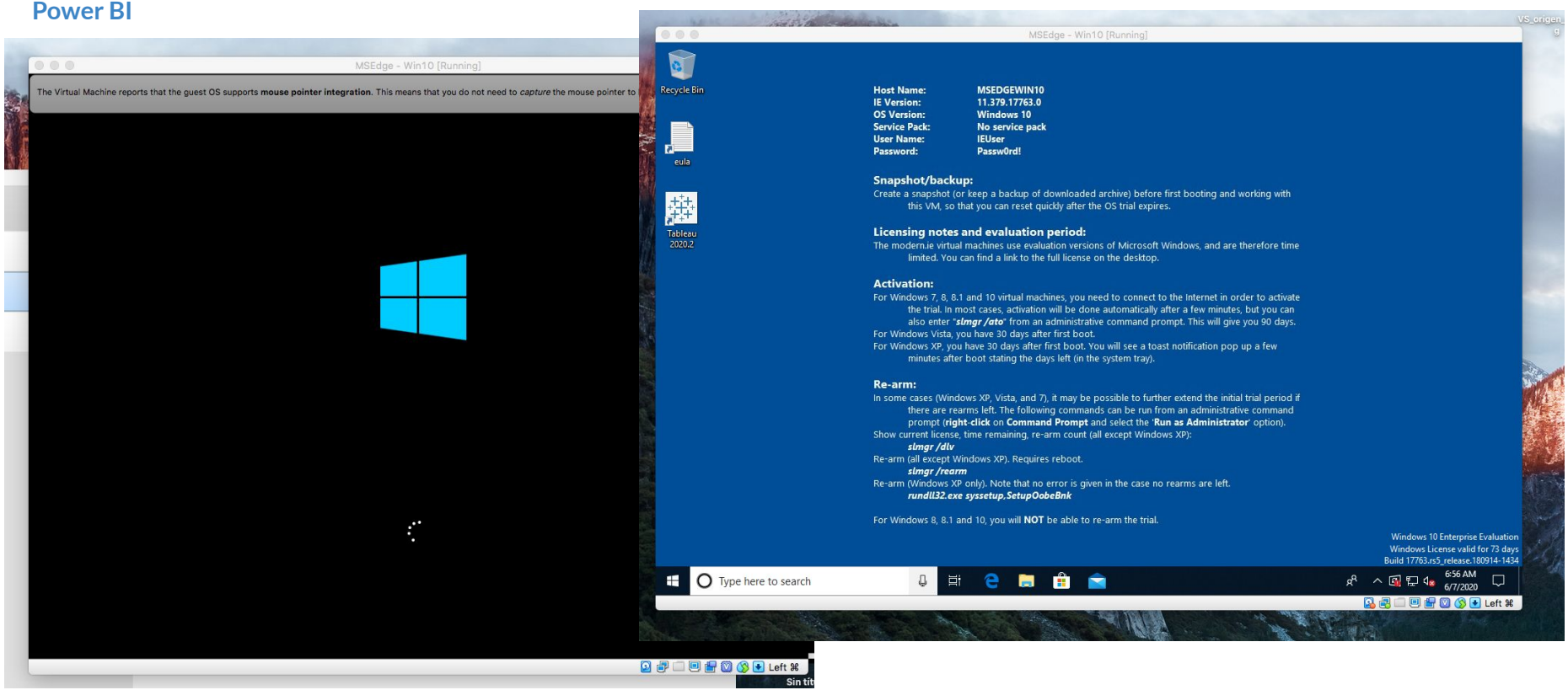
En caso de Virtualbox asignamos un mínimo de:

- 4gb de memoria RAM
- 2 CPU

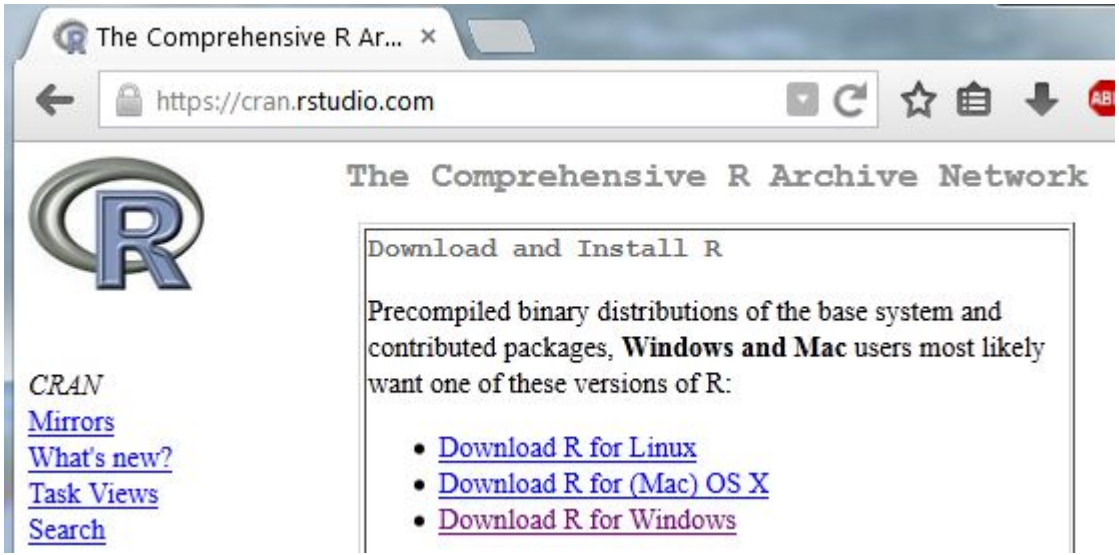
El resto de opciones vienen por defecto siendo el espacio disponible en disco duro de un mínimo de 50 gb

Power BI Desktop: entorno virtual Virtualbox

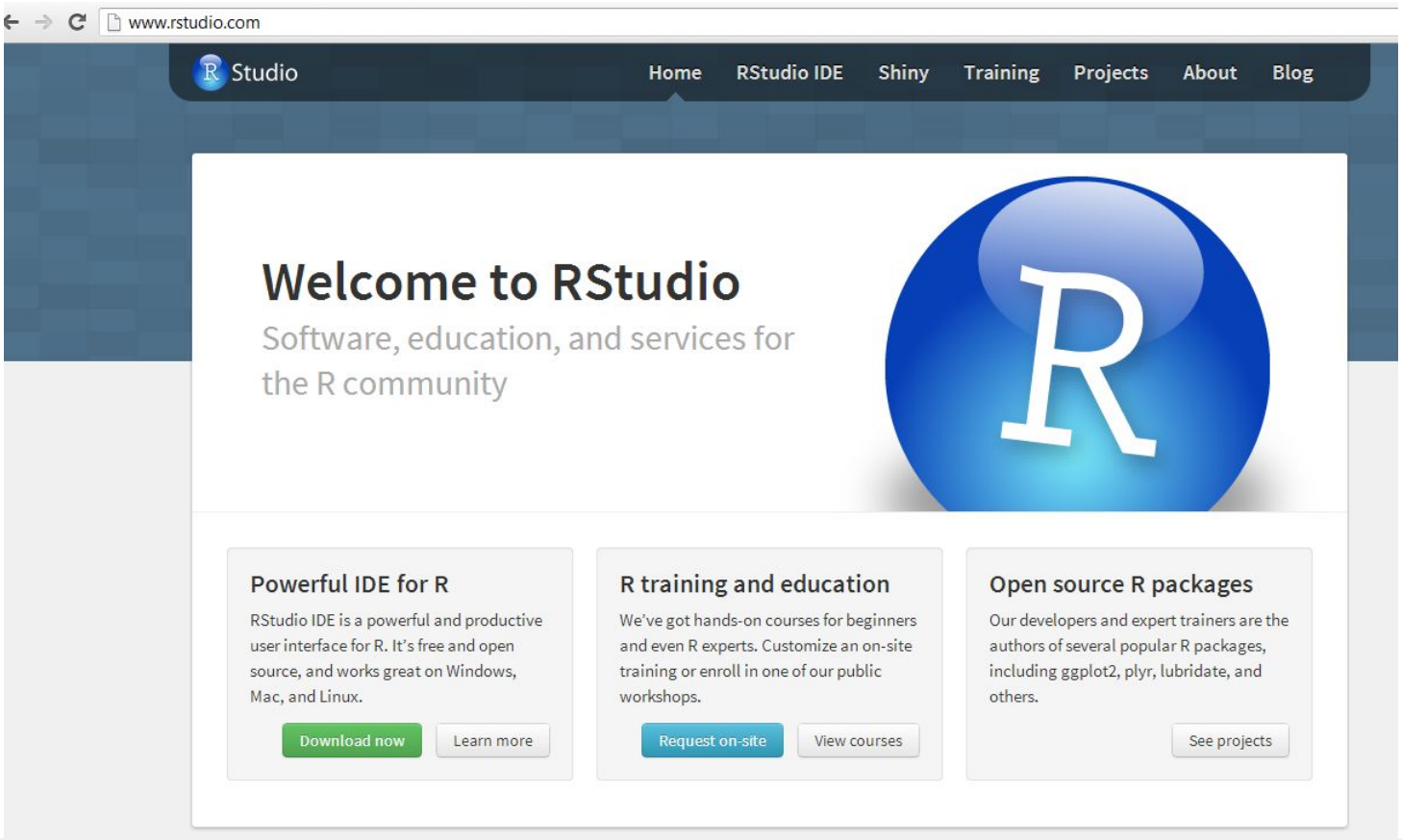
Power BI



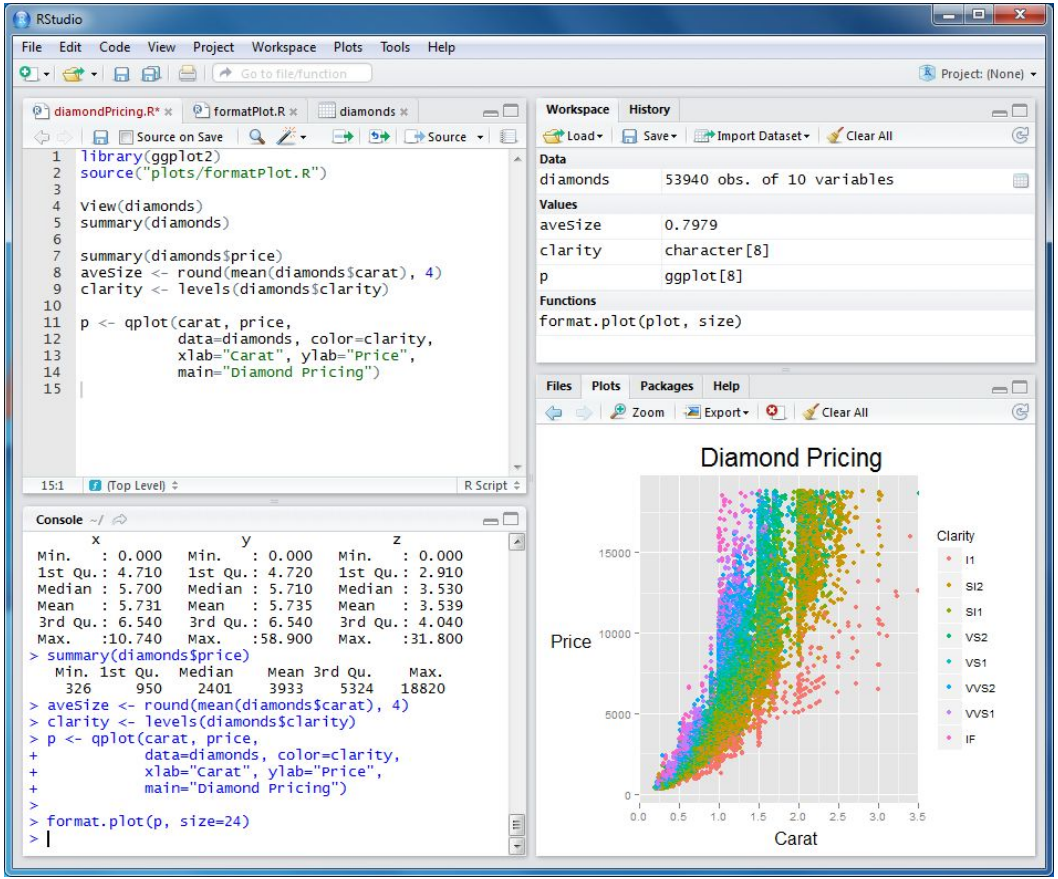
Configuración R commander.



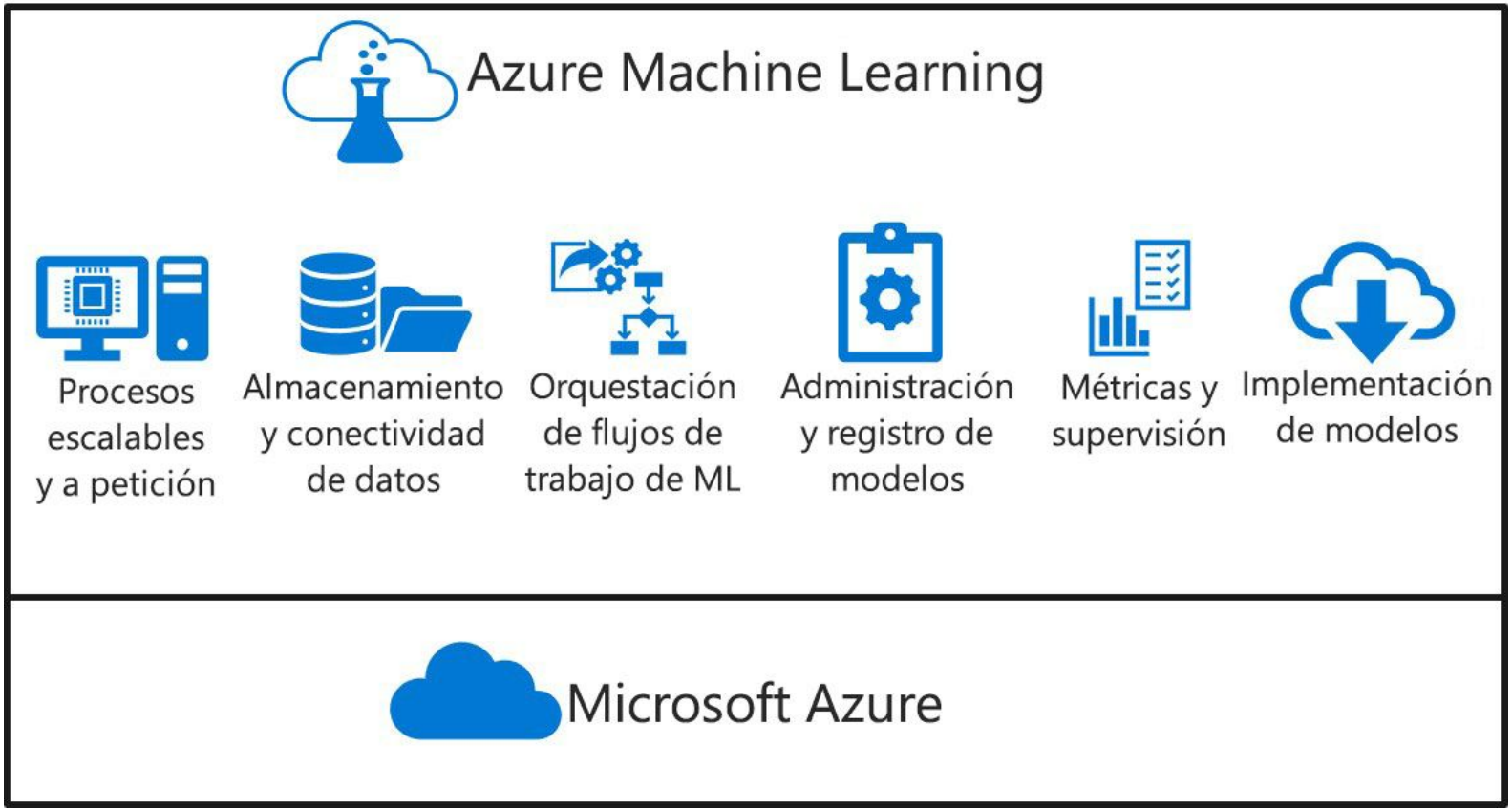
Configuración Rstudio.



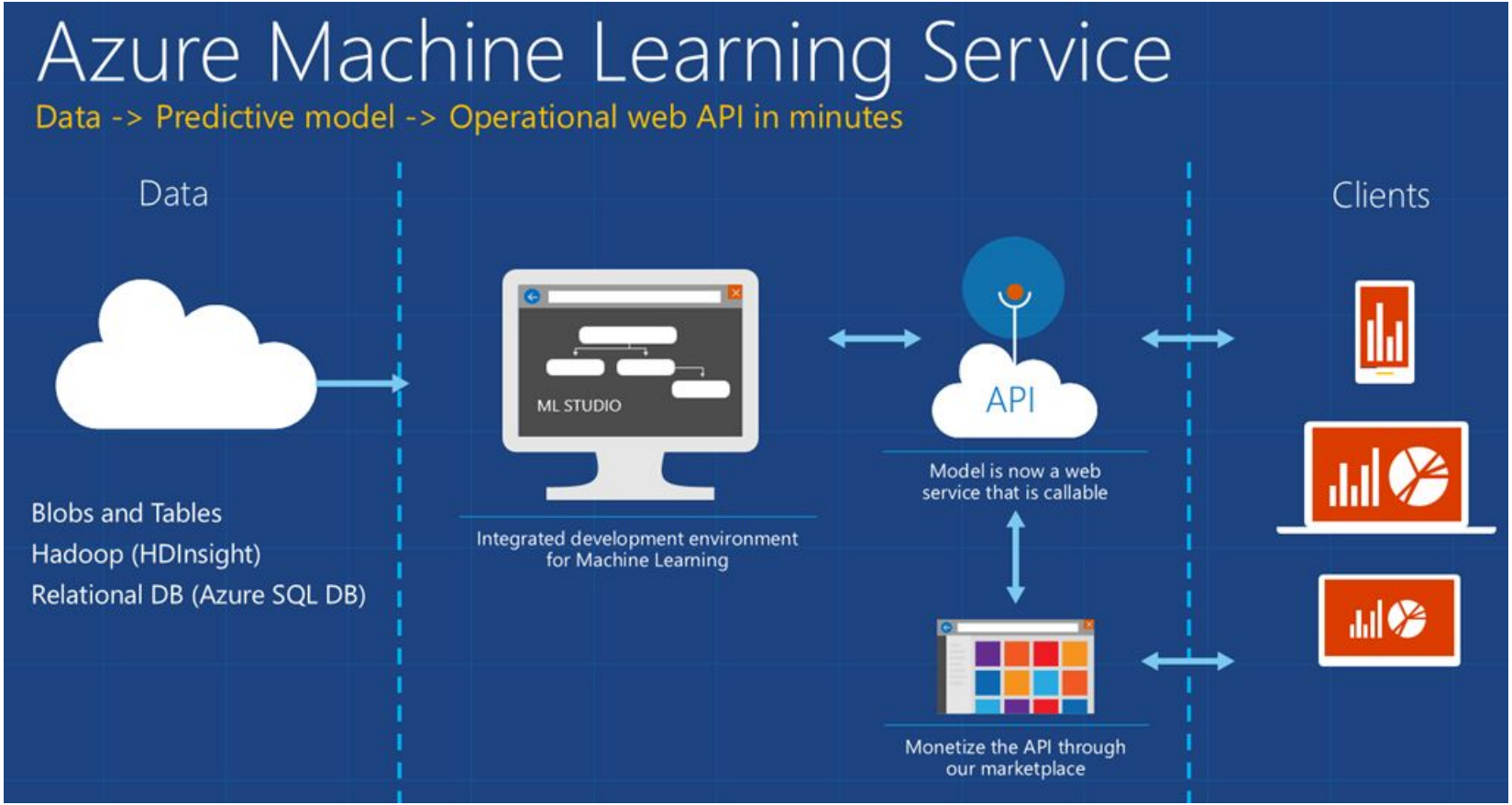
Comprobación de R Studio.



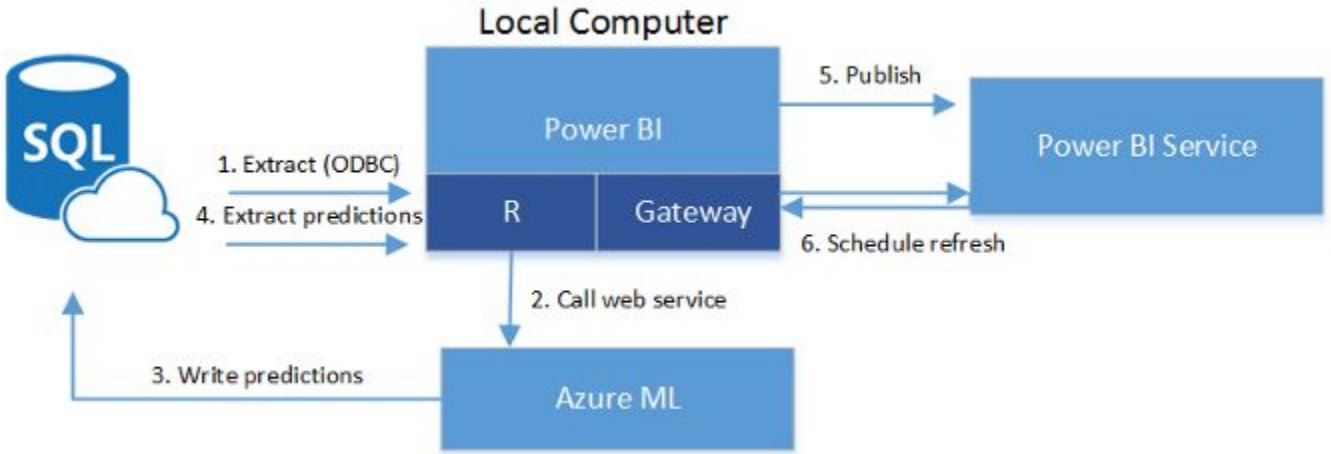
Azure services.



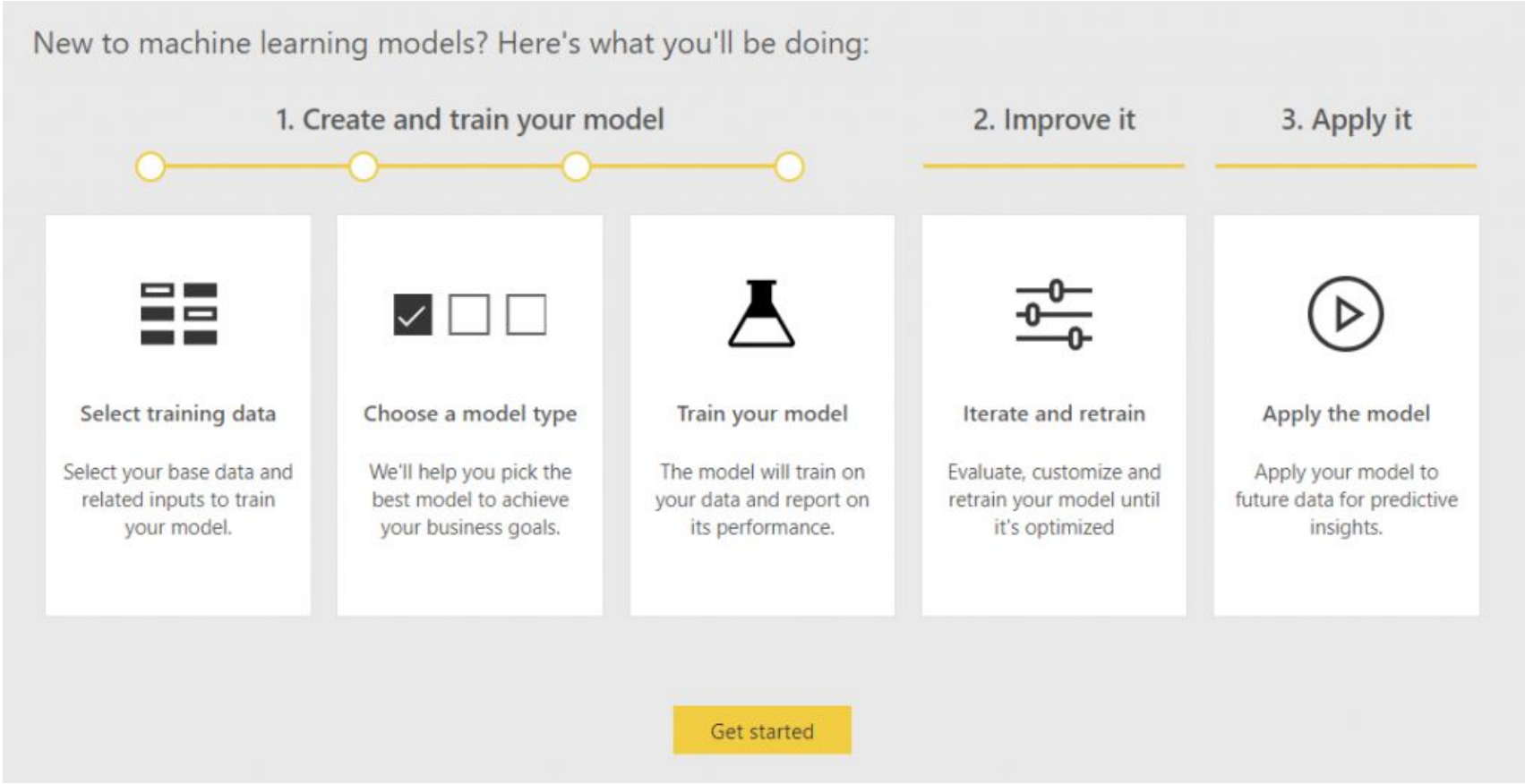
Azure ML.


Azure Machine Learning Service
Data -> Predictive model -> Operational web API in minutes
Data
Blobs and Tables
Hadoop (HDInsight)
Relational DB (Azure SQL DB)
ML STUDIO
Integrated development environment for Machine Learning
API
Model is now a web service that is callable
Monetize the API through our marketplace
Clients

Azure services.



Power BI.


New to machine learning models? Here's what you'll be doing:
1. Create and train your model
2. Improve it
3. Apply it
Select training data
Select your base data and related inputs to train your model.
Choose a model type
We'll help you pick the best model to achieve your business goals.
Train your model
The model will train on your data and report on its performance.
Iterate and retrain
Evaluate, customize and retrain your model until it's optimized
Apply the model
Apply your model to future data for predictive insights.
Get started

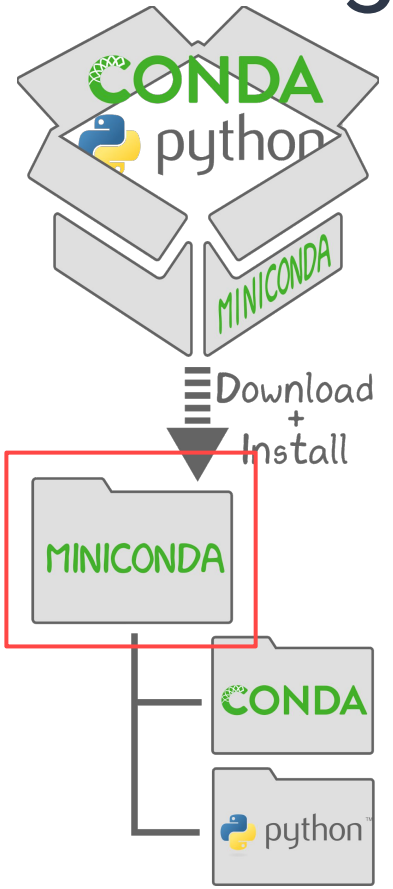
Power BI

Demo 1.

Configuración entorno Python.

https://www.microsoft.com/en-us/p/python-38/

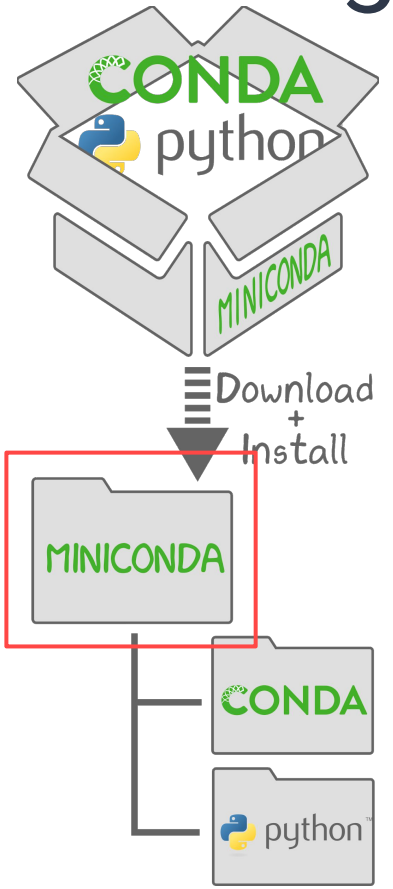
Configuración Miniconda for Win10.

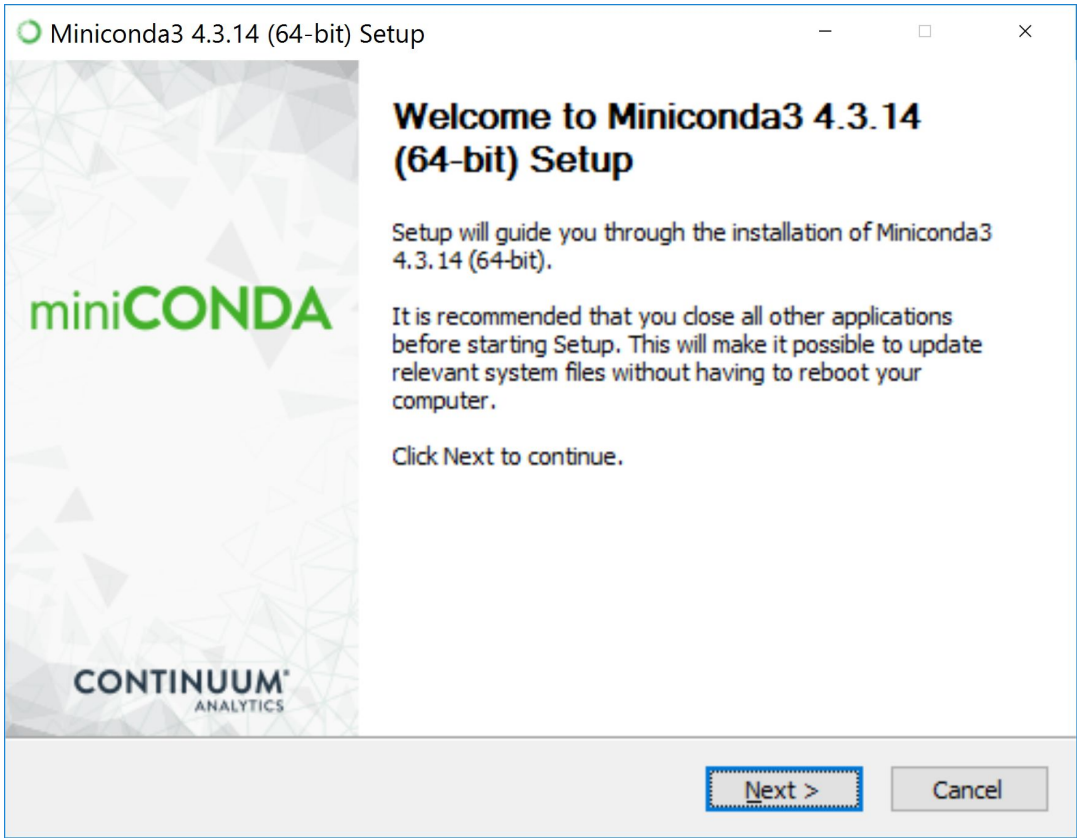


System requirements

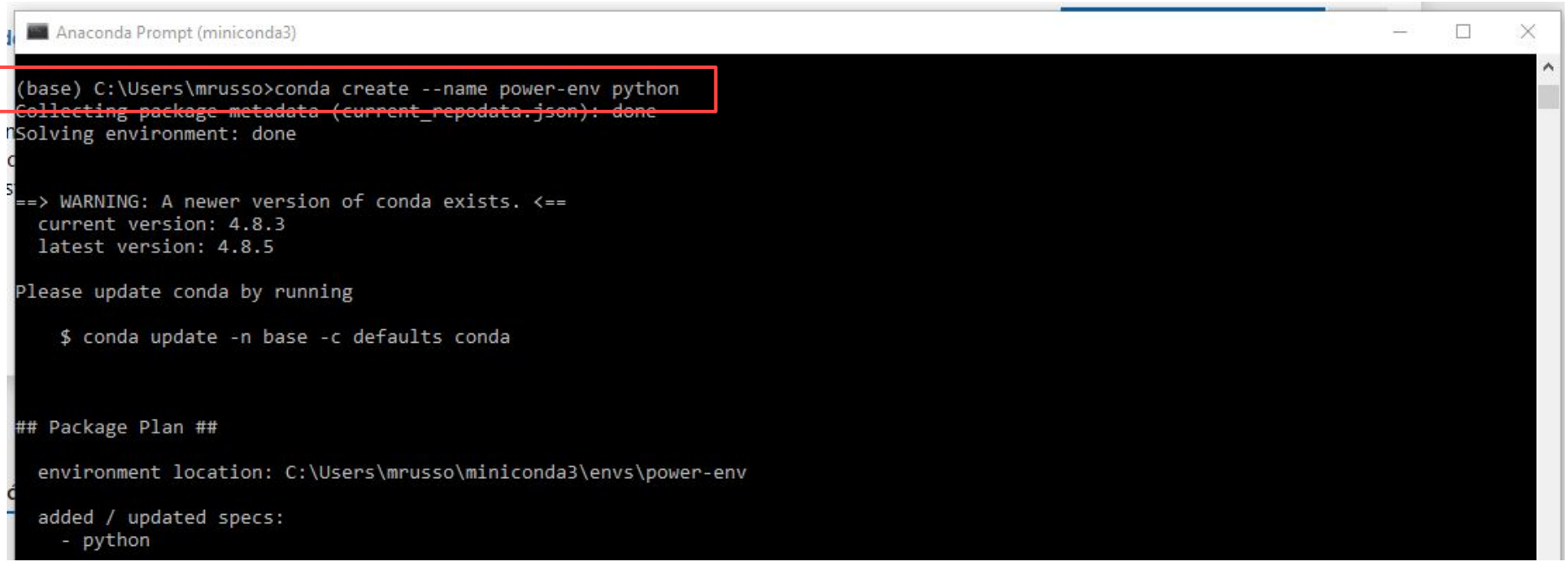
- 32- or 64-bit computer.
- For Miniconda---400 MB disk space.
- For Anaconda---Minimum 3 GB disk space to download and install.
- Windows, macOS, or Linux.

Configuración Miniconda for Win10.





Configuración Miniconda for Win10.



Configuración Miniconda for Win10.

```
Proceed ([y]/n)? y


Downloading and Extracting Packages
python-3.8.5          | 15.7 MB   | ########################################################################## | 100%
sqlite-3.33.0         | 809 KB    | ########################################################################## | 100%
certifi-2020.6.20     | 157 KB    | ########################################################################## | 100%
openssl-1.1.1h        | 4.8 MB    | ########################################################################## | 100%
vs2015_runtime-14.16  | 1.2 MB    | ########################################################################## | 100%
wheel-0.35.1          | 37 KB     | ########################################################################## | 100%
setuptools-50.3.0     | 741 KB    | ########################################################################## | 100%
ca-certificates-2020  | 122 KB    | ########################################################################## | 100%
pip-20.2.3            | 1.8 MB    | ########################################################################## | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate power-env
#
# To deactivate an active environment, use
#
#     $ conda deactivate


(base) C:\Users\mrusso>conda activate power-env
```

https://conda.io/projects/conda/en/latest/user-guide/tasks/**manage-environments.html**

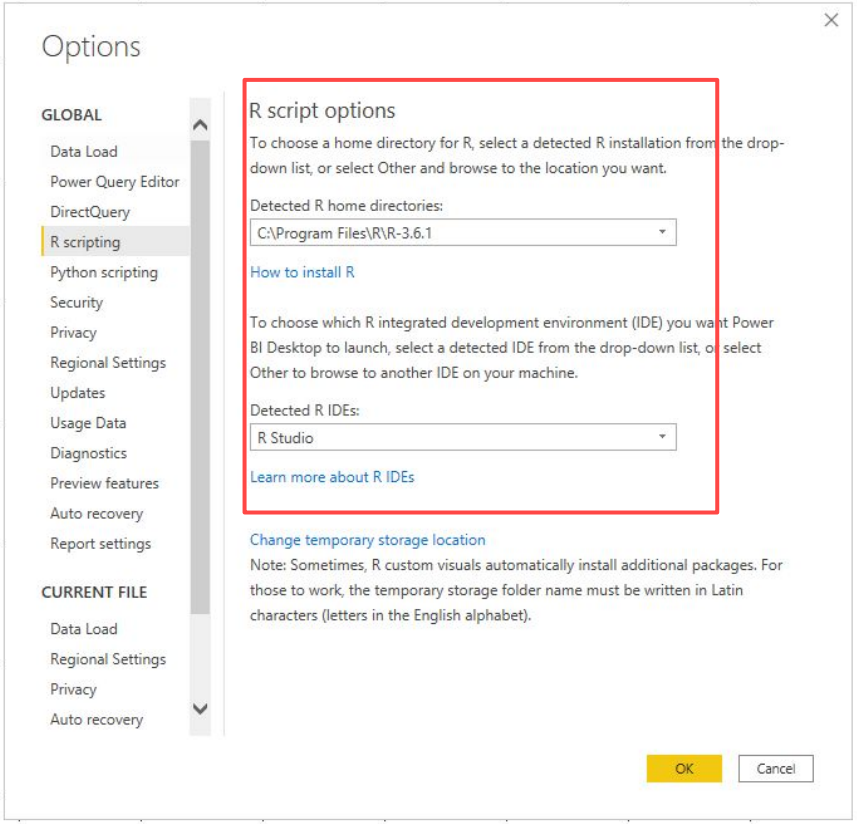
Requirements.txt.

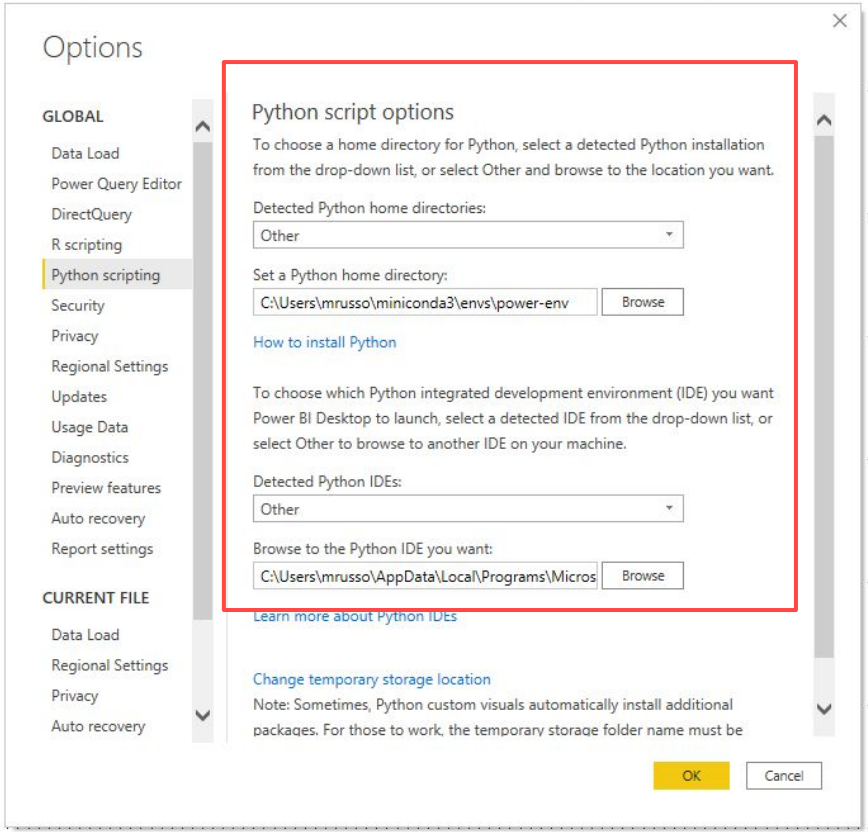
pandas==1.1.3
matplotlib==3.3.2
seaborn==0.11.0
numpy==1.18.5
scipy==1.5.3
scikit-learn==0.23.2
pycaret==2.0

Power BI

Demo 2.

Configuración en Power BI.





Power BI

Demo 3.

02.01

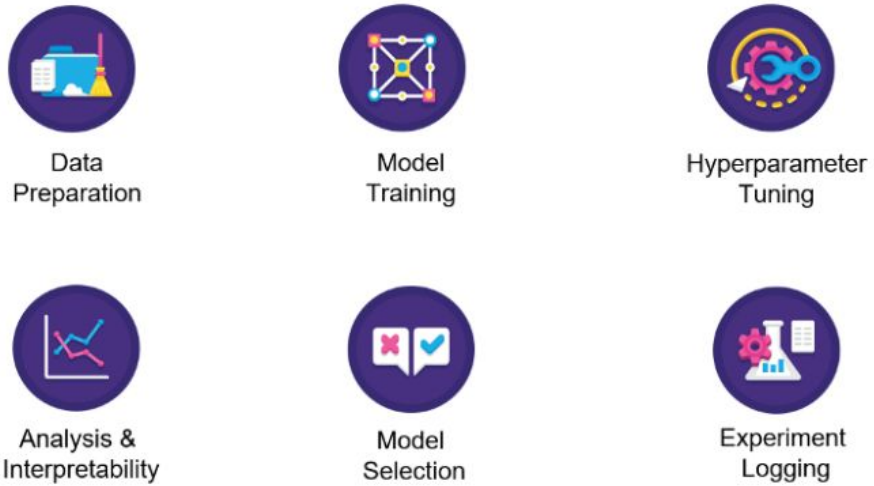
Power BI

Pycaret autoML en Power BI.

¿Qué es Pycaret?.

PyCaret is seamlessly integrated with BI

PyCaret and its Machine Learning capabilities are seamlessly integrated with environments supporting Python such as Microsoft Power BI, Tableau, Alteryx and KNIME to name a few. This gives immense power to users of these BI platforms who can now integrate PyCaret into their existing workflows and add a layer of Machine Learning with ease.



PyCaret is ideal for

- Experienced Data Scientists who want to increase productivity.
- Citizen Data Scientists who prefer a low code machine learning solution.
- Students of Data Science.
- Data Scientists and Consultants involved in building Proof of Concept projects.

Power BI

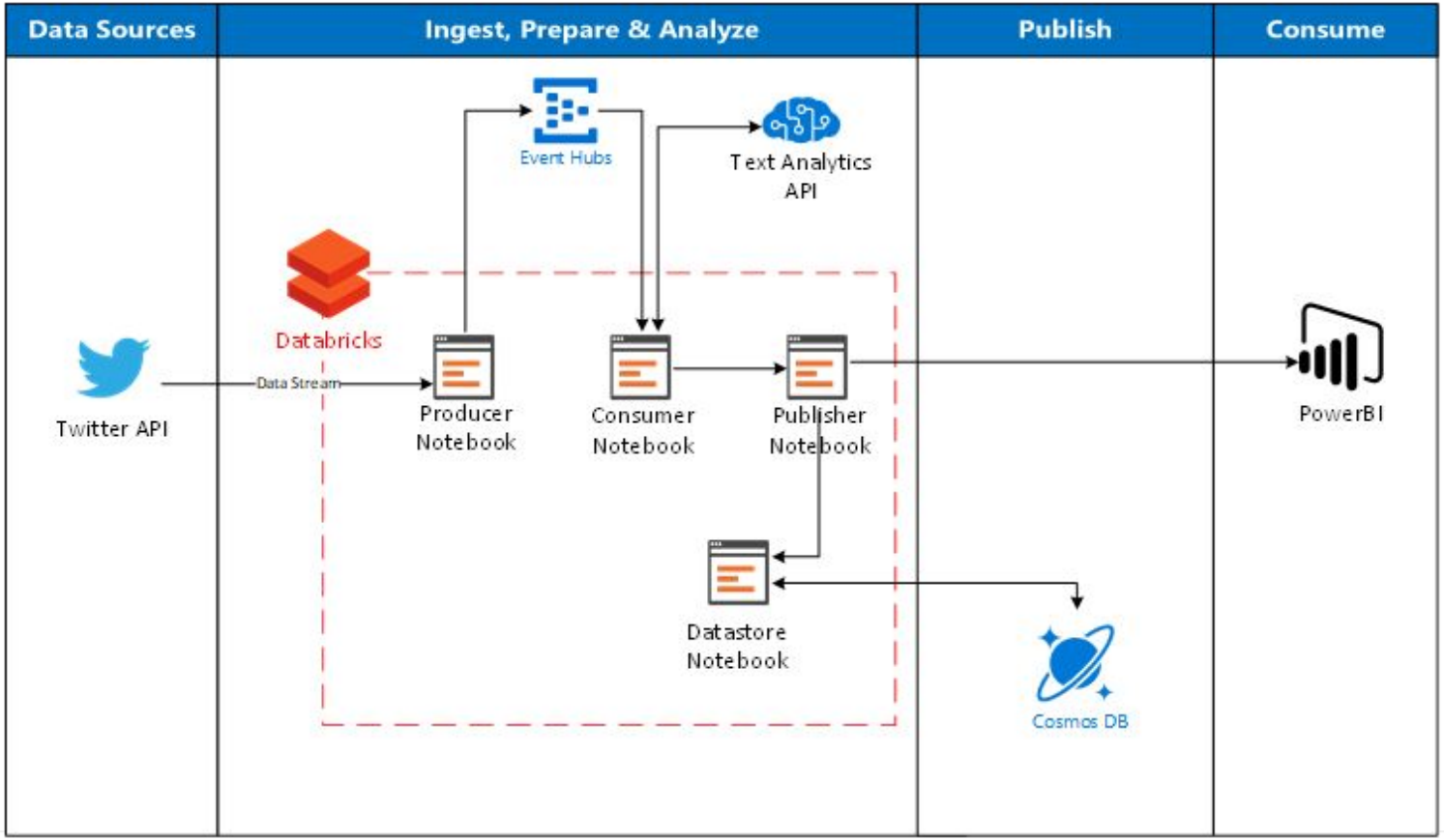
Demo 4.

02.02

Power BI

Get Data con Databricks.

Databricks - Azure Databricks.



Power BI

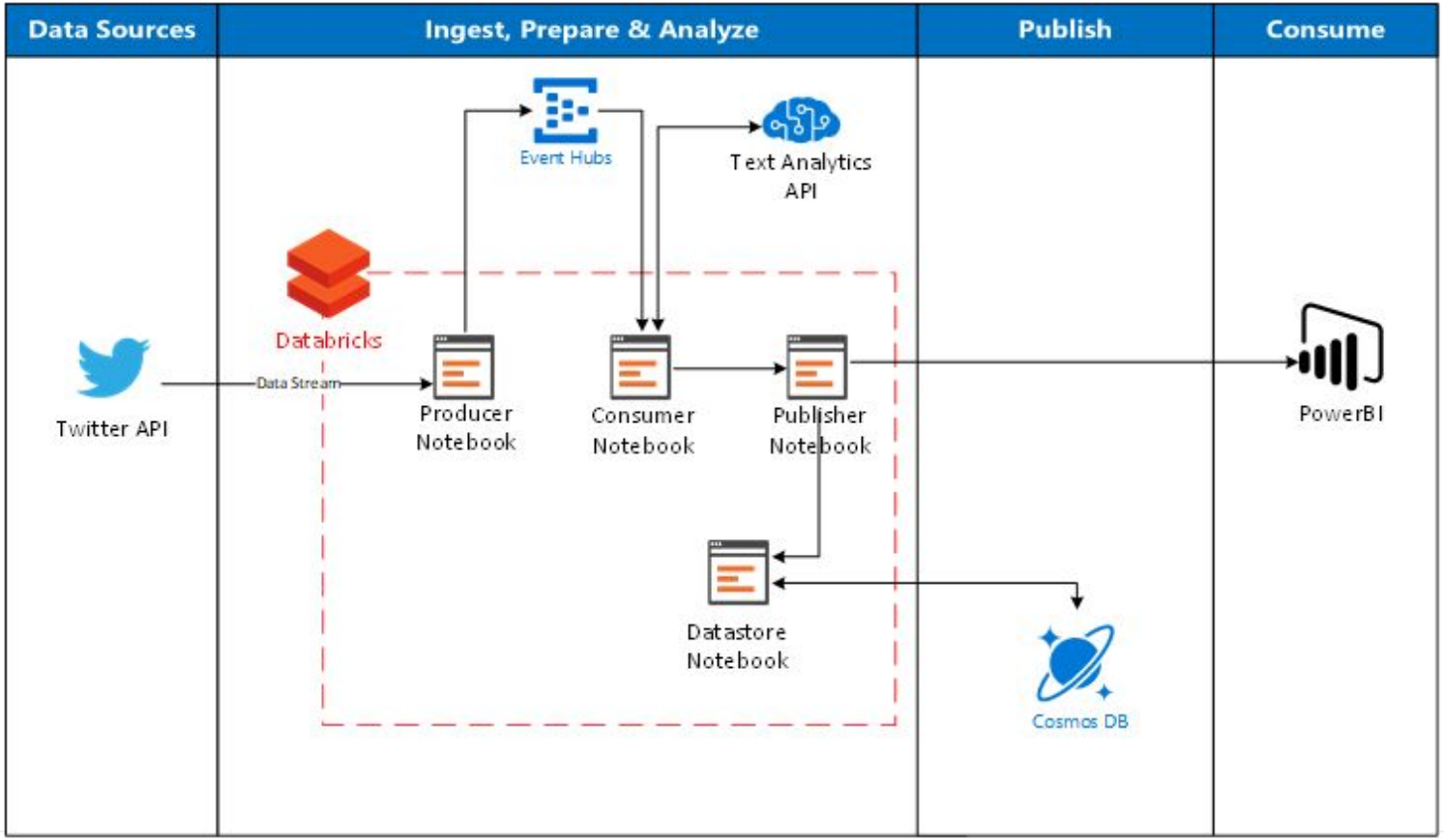
Demo 5.

02.03

Power BI

Get Data con Databricks.

Databricks - Azure Databricks.



Power BI

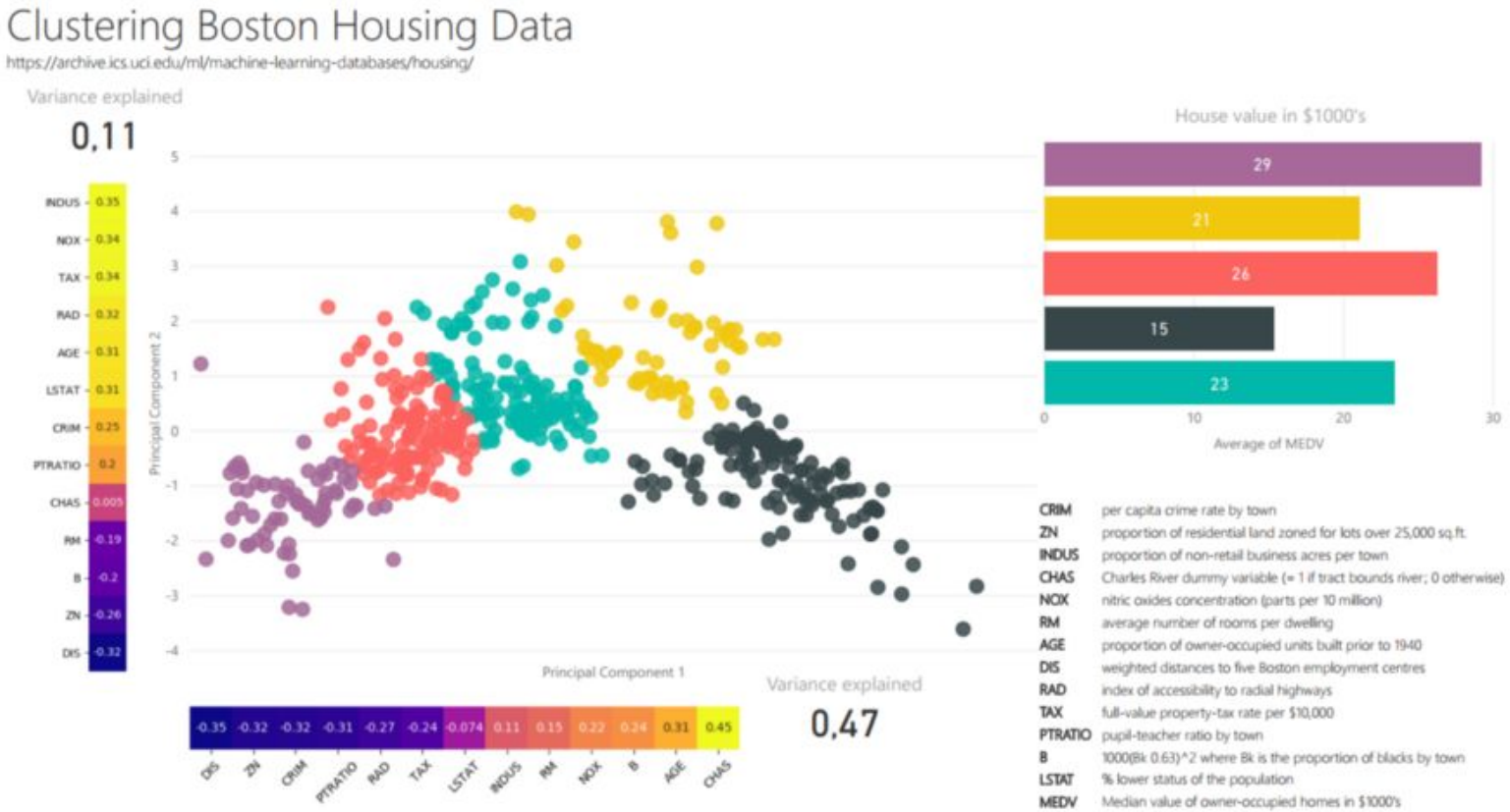
Demo 5.

04.01

Power BI

Visualization.

Visualización con R y Python.



05.01

Power BI

Limitation.

Limitaciones con Python.

The Power BI service also imposes other limits on Python script execution.

CONSTRAINT	LIMIT
Dataframe size	150,000 rows 250MB when serialized in CSV format
Execution time	60 seconds
Memory consumption	1 GB
Disk I/O	1 GB/sec
Image size for a visual in PNG format	2 MB

These limits ensure that user-provided scripts are performant, and moderate in the consumption of resources.

Limitaciones con R.

Known Limitations

R visuals in the Power BI service have a few limitations:

- R visuals support is limited to the packages identified in Learn which R packages are supported. There currently is no support for custom packages.
- Data size limitations – data used by the R visual for plotting is limited to 150,000 rows. If more than 150,000 rows are selected, only the top 150,000 rows are used and a message is displayed on the image. Additionally, the input data has a limit of 250 MB.
- Resolution - all R visuals are displayed at 72 DPI.
- Plotting device - only plotting to the default device is supported.
- Calculation time limitation – if an R visual calculation exceeds 60 seconds the script times out, resulting in an error.
- R visuals are refreshed upon data updates, filtering, and highlighting. However, the image itself is not interactive and does not support tool tips.
- R visuals respond to highlighting other visuals, but you cannot click on elements in the R visual in order to cross filter other elements.
- R visuals are currently not supported for the *Time* data type. Please use Date/Time instead.
- R visuals do not display when using **Publish to web**.
- R visuals do not support renaming input columns. Columns will be referred to by their original name during script execution.
- R visuals currently do not print with dashboard and reports printing
- R visuals are currently not supported in the DirectQuery mode of Analysis Services

uencers

https://docs.microsoft.com/en-us/power-bi/visuals/service-r-visuals#known-limitations

¡Muchas Gracias!

Recursos del webinar.

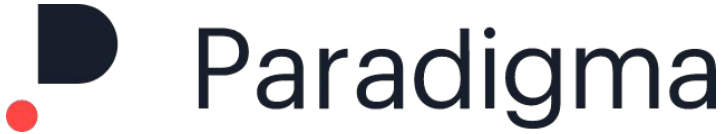
https://github.com/marcusRB/WEBINAR_ML_PowerBI

https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-r-visuals

https://powerbi.microsoft.com/es-es/blog/python-visualizations-in-power-bi-service/

https://powerbi.microsoft.com/en-us/blog/pythonblogepisode1/

https://powerbi.microsoft.com/fr-fr/blog/data-cleansing-with-r-in-power-bi/



- https://youtu.be/AFTanu_iBD4

- https://www.paradigmadigital.com/dev/5-tips-power-bi-sencillos-utiles/
- https://www.paradigmadigital.com/dev/bigquery-y-power-bi-emplatado-datos/
- https://www.paradigmadigital.com/techbiz/cocinando-datos-bigquery-powerbi-primeros-pasos/