

Integration of Python in Power BI Desktop



Cognitive Convergence

<http://www.cognitiveconvergence.com>

+1 4242530744

shahzad@cognitiveconvergence.com

Cognitive Convergence is Subject Matter Expert in Office 365, Dynamics 365, SharePoint, Project Server, Power Platform: Power Apps-Power BI-Power Automate-Power Virtual Agents.

We offer Power BI consulting services covering solution architecture refinement, customization, integration, transformation, visualization and analytics to uncover insights hidden within data and enhance data exploration.

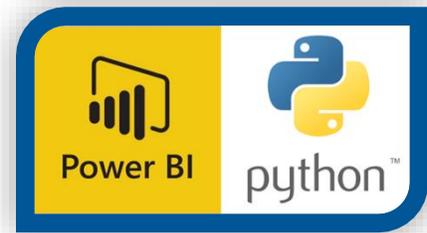
Contents

Objectives.....	2
Background	2
<i>Python</i>	2
<i>Python scripts in Power BI Desktop</i>	2
Python integration with power BI Desktop.....	3
Enabling Python in Power BI.....	3
<i>Pandas</i>	3
<i>Matplotlib</i>	3
Configure Power BI to use Python	4
Run Python scripts.....	5
Create Custom visualization using Python.....	7
<i>Create a scatter plot</i>	9
Type of visualization	10
Security of Python Script	11
Limitation	12
Conclusion	12



OBJECTIVES

This paper will discuss about Python and its integration with power BI Desktop. It development patterns and implementation details along with some examples and creation of visuals using python script in Power BI Desktop.



BACKGROUND

The Power BI has added a feature to the run Python scripts easily in Power BI Desktop. For creating data model and reports visualization, integrate Python, perform data cleaning, perform advanced data shaping and analytics in datasets. After that, find out the missing data. Now, use Python to create new visuals on report. Python visuals is updated with data refreshes and cross-filtering also. The addition of Python integration in Power BI is one of the greatest things that Microsoft could have done. This gives the user the ability to utilize amazing visual libraries such as Seaborn. Beyond visualization, can also utilize some of the machine learning packages.

Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.



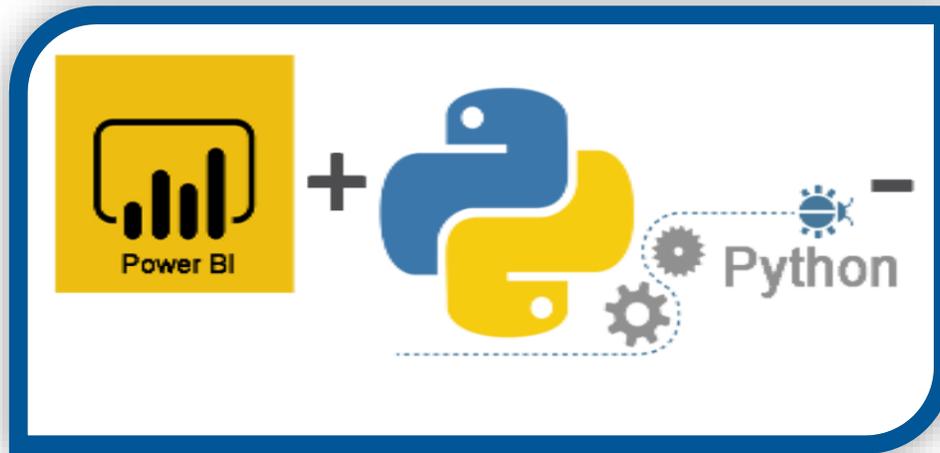
Python scripts in Power BI Desktop

Python scripts can directly run in Power BI Desktop and import the resulting datasets into a Power BI Desktop data model with Power BI Desktop, Python can be used to visualize data. Python, a programming language can be used widely by statisticians, data scientists, and data analysts, in the Power BI Desktop Query Editor. This integration of Python into Query Editor lets perform data cleansing using Python, and perform advanced data shaping and analytics in datasets, including completion of missing data, predictions, and clustering, just to name a few. Python is a powerful language, and can be used in Query Editor to prepare data model and create reports.



PYTHON INTEGRATION WITH POWER BI DESKTOP

Python integration has been generally available in Power BI and can now use Python for importing data, data transformation and data visualization. Firstly needed to prepare a Python environment, enable Python in Power BI, import data and apply clustering to the data and create custom visuals in Power BI using Python.



ENABLING PYTHON IN POWER BI

To run Python scripts in Power BI Desktop, need to install Python on local machine. The current Python scripting release supports Unicode characters and spaces in the installation path. The Power BI Python integration requires the installation of two Python packages:

Pandas

A software library for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series. Imported data must be in a pandas data frame. A data frame is a two-dimensional data structure. For example, data is aligned in a tabular fashion in rows and columns.

```
pip install pandas
```

Matplotlib

A plotting library for Python and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits, such as Tkinter, wxPython, Qt, or GTK+.

In a console or shell, use the pip command-line tool to install the two packages. The pip tool is packaged with more recent Python versions.

We understand the integrating capabilities of Power BI and Python going hands-on to perform certain task.

Cognitive Convergence

<http://www.cognitiveconvergence.com>

+1 4242530744

shahzad@cognitiveconvergence.com



```
pip install matplotlib
```

```
C:\Users\inspiron\Python>pip install pandas
Requirement already satisfied: pandas in c:\python38\lib\site-packages (1.0.5)
Requirement already satisfied: numpy>=1.13.3 in c:\python38\lib\site-packages (from pandas) (1.19.0)
Requirement already satisfied: pytz>=2017.2 in c:\python38\lib\site-packages (from pandas) (2020.1)
Requirement already satisfied: python-dateutil>=2.6.1 in c:\python38\lib\site-packages (from pandas) (2.8.1)
Requirement already satisfied: six>=1.5 in c:\python38\lib\site-packages (from python-dateutil>=2.6.1->pandas) (1.15.0)

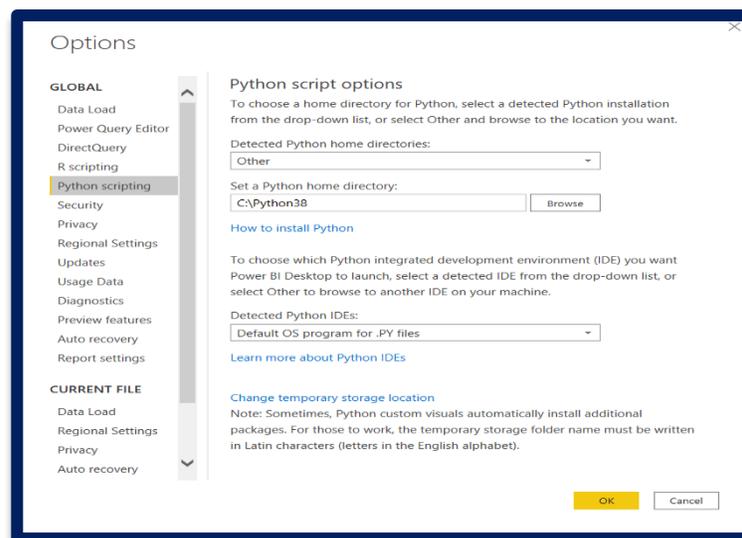
C:\Users\inspiron\Python>pip install matplotlib
Collecting matplotlib
  Downloading matplotlib-3.2.2-cp38-cp38-win_amd64.whl (9.2 MB)
    |#####| 9.2 MB 67 kB/s
Requirement already satisfied: python-dateutil>=2.1 in c:\python38\lib\site-packages (from matplotlib) (2.8.1)
Collecting kiwisolver>=1.0.1
  Downloading kiwisolver-1.2.0-cp38-none-win_amd64.whl (58 kB)
    |#####| 58 kB 83 kB/s
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in c:\python38\lib\site-packages (from matplotlib) (2.4.7)
Requirement already satisfied: numpy>=1.11 in c:\python38\lib\site-packages (from matplotlib) (1.19.0)
Collecting cycler>=0.10
  Using cached cycler-0.10.0-py2.py3-none-any.whl (6.5 kB)
Requirement already satisfied: six>=1.5 in c:\python38\lib\site-packages (from python-dateutil>=2.1->matplotlib) (1.15.0)
Installing collected packages: kiwisolver, cycler, matplotlib
Successfully installed cycler-0.10.0 kiwisolver-1.2.0 matplotlib-3.2.2

C:\Users\inspiron\Python>
```

CONFIGURE POWER BI TO USE PYTHON

To enable Python scripting:

- In Power BI Desktop, select File > Options and settings > Options > Python scripting. The Python script options page appears.



- If necessary, specify local Python installation path in Detected Python home directories.

Once specify Python installation, ready to begin running Python scripts in Power BI Desktop.

RUN PYTHON SCRIPTS

In just a few steps, run Python scripts and create a data model. From this model, create reports and share them on the Power BI service.

- First, create a script in local Python development environment and make sure it runs successfully. For example, here's a simple Python script that imports pandas and uses a data frame:

```
import pandas as pd
data = [['Alex',10],['Bob',12],['Clarke',13]]
df = pd.DataFrame(data,columns=['Name','Age'],dtype=float)
print (df)
```

- Which return the following output

```
Name Age
0 Alex 10.0
1 Bob 12.0
2 Clarke 13.0
```

- To run Python Script in Power BI Desktop. In the Home ribbon, select Get Data > Other. Select Other > Python script as shown in the following image:

We use Python's statistical within Power BI to create a correlation matrix in it.

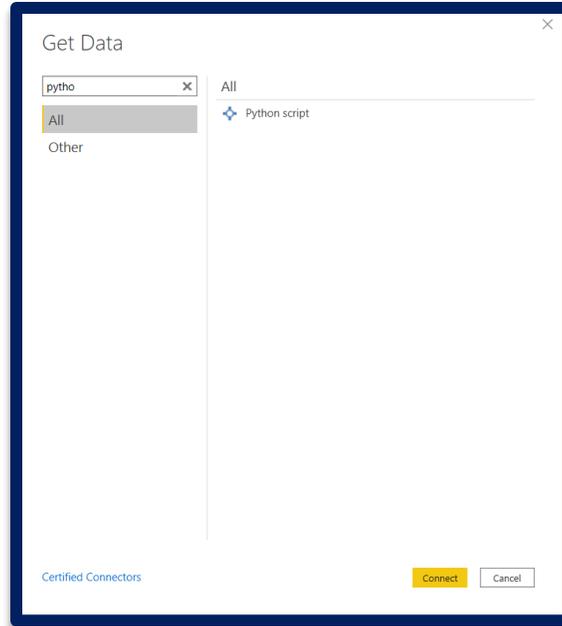
Cognitive Convergence

<http://www.cognitiveconvergence.com>

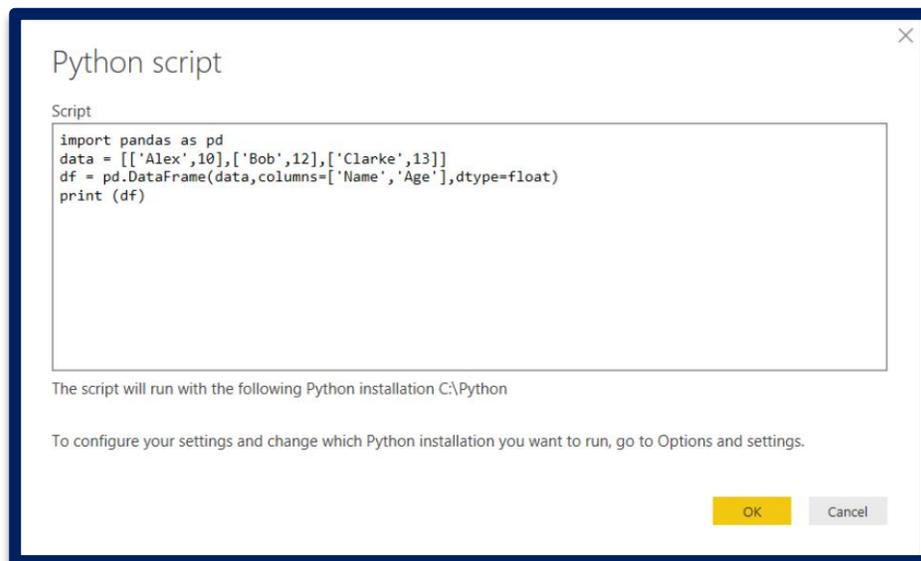
+1 4242530744

shahzad@cognitiveconvergence.com



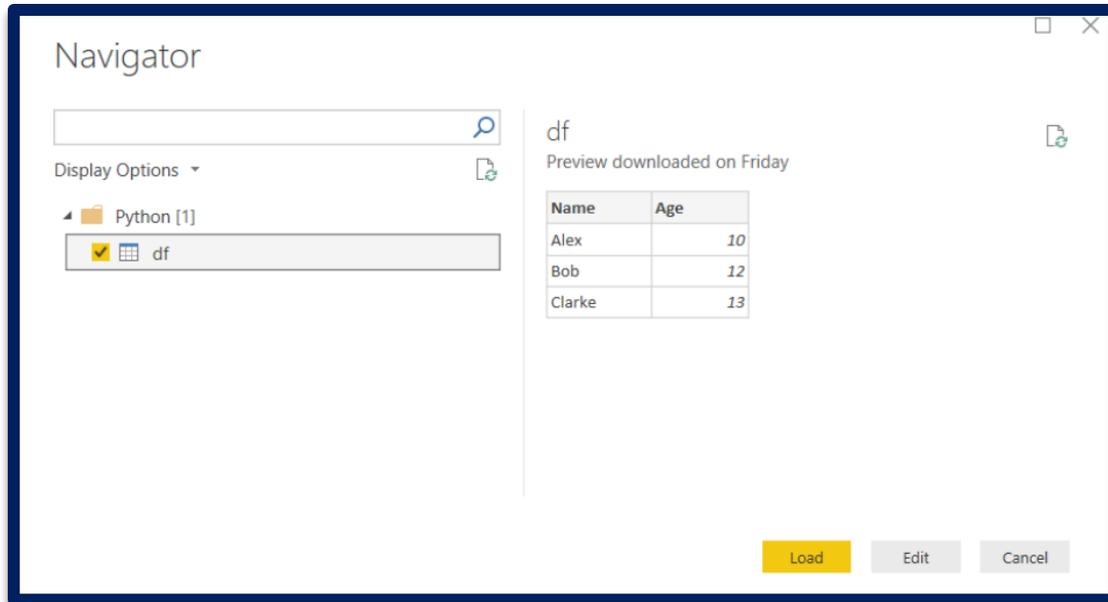


- Select Connect. Local latest installed Python version is selected as Python engine. Copy script into the Python script dialog box that appears. Here, we enter the simple Python script shown before.



- Select OK. If the script runs successfully, the Navigator appears and can load the data and use it. For the example, select df, as shown in the image, then Load.

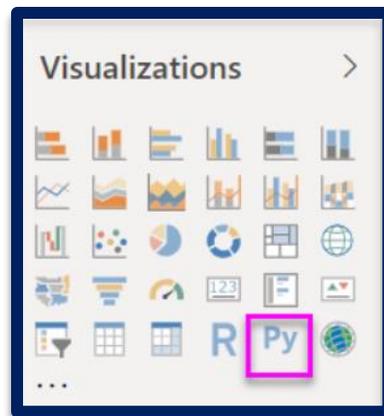




CREATE CUSTOM VISUALIZATION USING PYTHON

With Power BI Desktop, use Python to visualize data.

1. Select the Python visual icon in the Visualizations pane.



2. In the Enable script visuals dialog box that appears, select Enable. When add a Python visual to a report, Power BI Desktop takes the following actions:
 - A placeholder Python visual image appears on the report canvas.
 - The Python script editor appears along the bottom of the center pane.

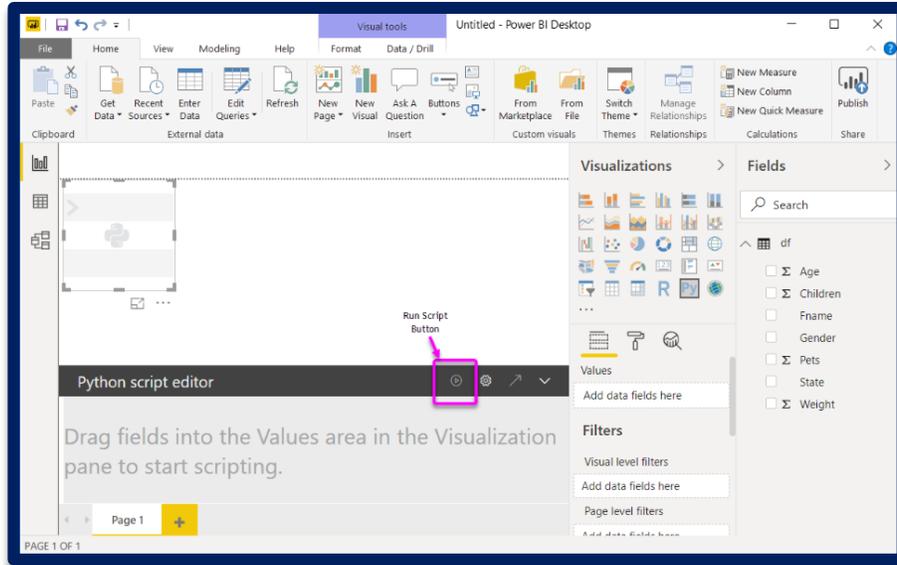
Our consulting services of Integrating Python in Power BI will get you the best of both worlds.

Cognitive Convergence

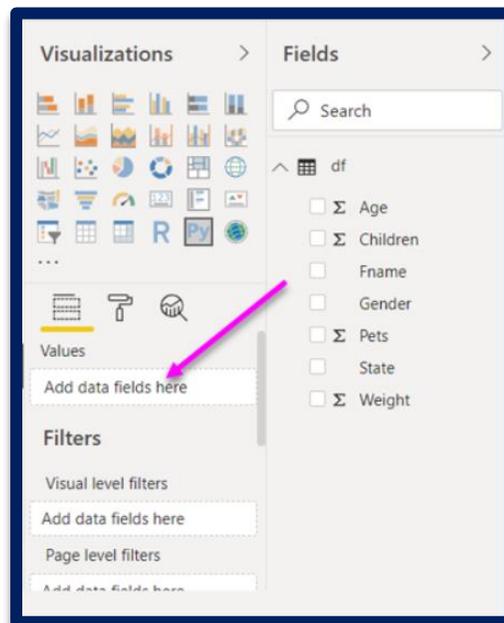
<http://www.cognitiveconvergence.com>

+1 4242530744

shahzad@cognitiveconvergence.com

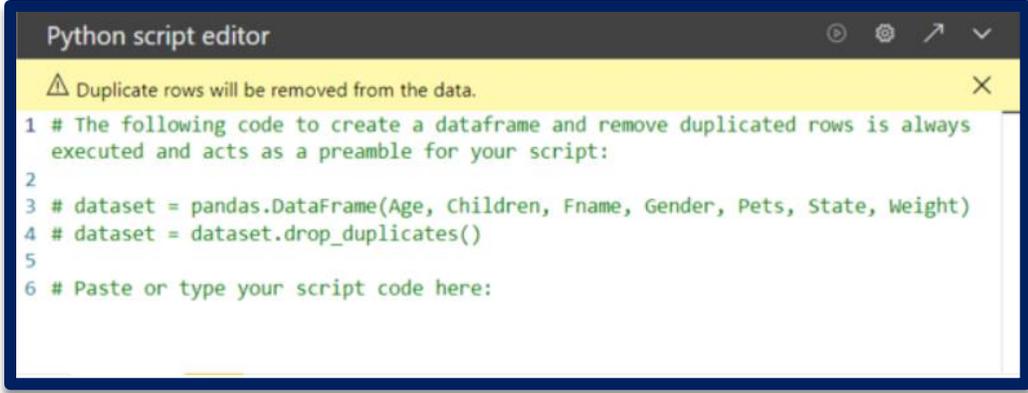


- Next, drag the Age, Children, Fname, Gender, Pets, State, and Weight fields to the Values section where it says Add data fields here.



- Python script can only use fields added to the **Values** section. Add or remove fields from the **Values** section while working on Python script. Power BI Desktop automatically detects field changes.
- Now use the data selected to create a plot. As select or remove fields, supporting code in the Python script editor is automatically generated or removed. Based on selections, the Python script editor generates the following binding code. The editor created a dataset dataframe, with the fields added. The default aggregation is: do not summarize. Similar to table visuals, fields are grouped and duplicate rows appear only once.



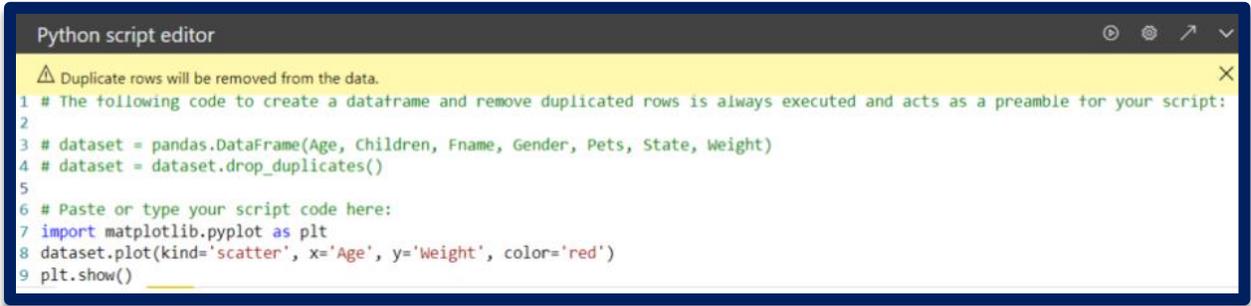


```
Python script editor
Duplicate rows will be removed from the data.
1 # The following code to create a dataframe and remove duplicated rows is always
  executed and acts as a preamble for your script:
2
3 # dataset = pandas.DataFrame(Age, Children, Fname, Gender, Pets, State, Weight)
4 # dataset = dataset.drop_duplicates()
5
6 # Paste or type your script code here:
```

Create a scatter plot

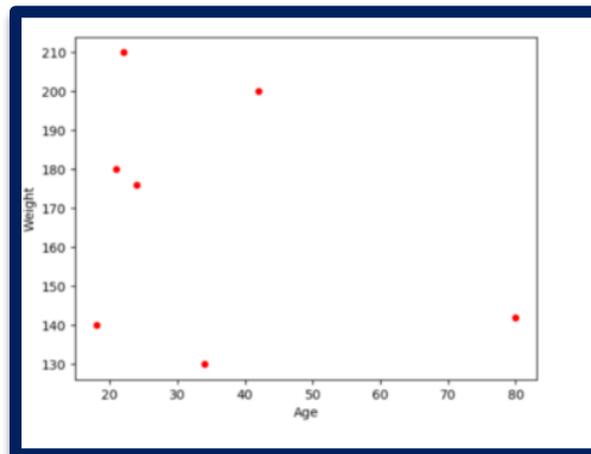
Let's create a scatter plot to see if there's a correlation between age and weight. Under Paste or type script code here, enter this code.

```
import matplotlib.pyplot as plt
dataset.plot(kind='scatter', x='Age', y='Weight', color='red')
plt.show()
```



```
Python script editor
Duplicate rows will be removed from the data.
1 # The following code to create a dataframe and remove duplicated rows is always
  executed and acts as a preamble for your script:
2
3 # dataset = pandas.DataFrame(Age, Children, Fname, Gender, Pets, State, Weight)
4 # dataset = dataset.drop_duplicates()
5
6 # Paste or type your script code here:
7 import matplotlib.pyplot as plt
8 dataset.plot(kind='scatter', x='Age', y='Weight', color='red')
9 plt.show()
```

The matplotlib library is imported to plot and create our visuals. When select the Run script button, the following scatter plot generates in the placeholder Python visual image.

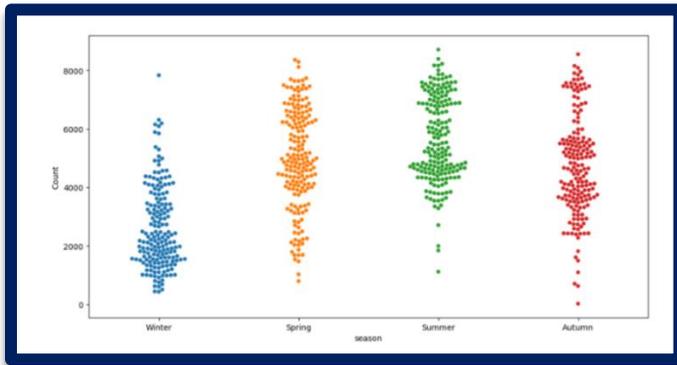


TYPE OF VISUALIZATION

Before choosing the right visualizations, the construction of some exploratory dataset is important. After knowing the exact data, the data types and other useful information selecting the appropriate images will be cooperatively easier. An enormous number of visualization packages are available in Python. Matplotlib can work various operating systems and graphic background. It possesses high quality images with the finest quality and thereby giving accurate results. It can be effectively used in Python scripts and all other web application servers.

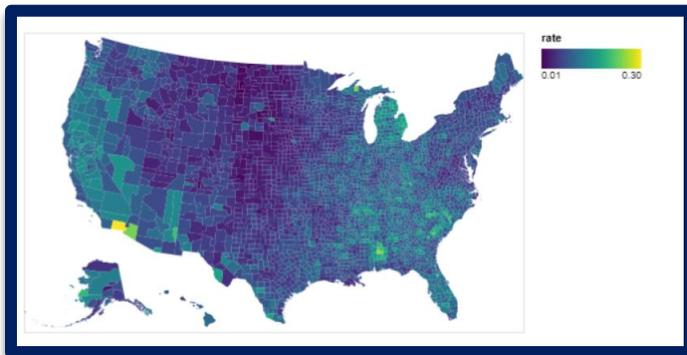
Seaborn

Built on top of the default matplotlib plotting library, seaborn offers a great extension to the matplotlib library allowing to generate more complex plots quickly.



Altair

A declarative library to generate plots unlike other libraries which require to create the legend, axis and labels. Altair aim is to alleviate some of these pain points and focus on the plot itself rather than specify every single element of the chart. By default, the axis/legend are generated based on the data pass to the plotting function.



We use the reporting capabilities of Power BI along with the analytical capabilities of Python to build analytical reports.

Cognitive Convergence

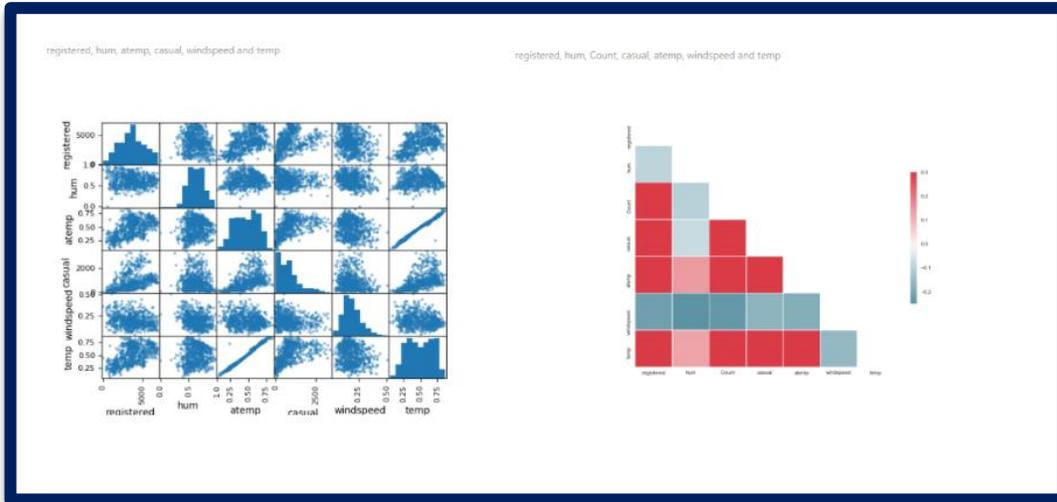
<http://www.cognitiveconvergence.com>

+1 4242530744

shahzad@cognitiveconvergence.com

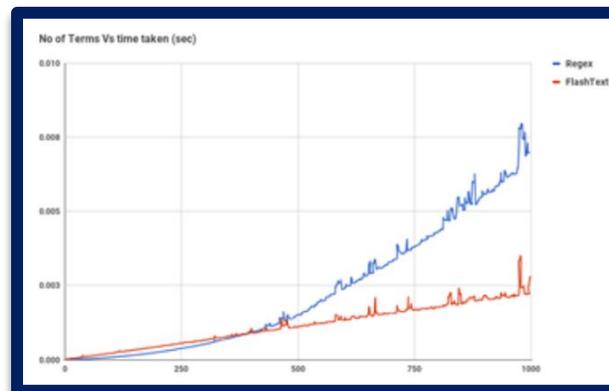
Scikit

Power BI, now use the many python libraries to create own machine learning models and easily operationalize them within your Power BI reports. One of the things want to do first is generate a matrix chart showing the correlations/histogram between each pairwise variables.



FlashText

A performant library for searching and replacing words within a text. Finding keywords or replacing values can now be done in minutes rather than hours.



SECURITY OF PYTHON SCRIPT

The execution of Python scripts in the reports is done by the Power BI system. This is to guarantee that the collected data and the administration of the Power BI are not likely to be susceptible to any assault. Python script execution on the Power BI has got size restriction and the execution time is restricted to 300 seconds.

LIMITATION

When preparing and running a Python script in Power BI Desktop, there are a few limitations:

- Only pandas data frames are imported, so make sure the data import to Power BI is represented in a data frame
- When setting the working directory within the Python script, must define a full path to the working directory, rather than a relative path
- Nested tables are currently not supported.
- Data size limitations. Data used by the Python 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.
- Relationships. As with other Power BI Desktop visuals, if data fields from different tables with no defined relationship between them are selected, an error occurs.
- Python visuals are refreshed upon data updates, filtering, and highlighting. However, the image itself isn't interactive and can't be the source of cross-filtering.
- Python visuals respond to highlighting other visuals, but can't click on elements in the Python visual to cross filter other elements.
- Only plots that are plotted to the Python default display device are displayed correctly on the canvas. Avoid explicitly using a different Python display device.
- Python visuals do not support renaming input columns. Columns will be referred to by their original name during script execution.

CONCLUSION

This paper will discuss about Python and its integration with Power BI Desktop. Their development patterns and implementation details along with some examples are also discussed. This paper prepares a Python environment, enables Python in Power BI, import data and apply clustering to the data and create custom visuals in Power BI using Python.

Cognitive Convergence will provide consulting services that help in designing, deploying, managing, enhancing or troubleshooting on-premises, cloud-based or hybrid Power BI environment. Cognitive Convergence will provide help to start fresh with Power BI to modernize current business analytics solution or revamp existing Power BI deployment by incorporating new data sources or adding new services.

Contact Us

Cognitive Convergence

<http://www.cognitiveconvergence.com>

+1 4242530744

shahzad@cognitiveconvergence.com