

Cognos Analytics Tech Talk



Introduction to Jupyter Notebooks and Cognos Analytics

Matthew Robinson
Peter D'Haeyer



Agenda

- What are Jupyter Notebooks?
Examples
- Benefits of Combining Jupyter Notebooks with Cognos Analytics
- Demonstration of Notebooks in Cognos Analytics
- Options for Cloud Implementations

What are Jupyter Notebooks?

- An open-source web application that supports interactive data science and scientific computing across all programming languages
Jupyter Notebooks are open-source web applications that let you create and share documents that contain live code, equations, visualizations and narrative text.
- Rebranded from IPython (interactive Python) as it supports multiple languages, including Julia, Python, R and many others
JU-PY-(TE)R
- Great for prototyping as cells are independent



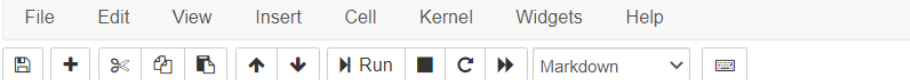
Demonstrations

Analysing text

Unstructured data
i.e. block of text to understand if there are patterns in the text. Start with most frequently used words after some exclusions

Market Basket Analysis

As a retailer, I want to understand which products are bought together. This might allow me to consider marketing campaigns or store layouts such as “beer and nappies”



Exploratory Analysis

To begin this exploratory analysis, first import libraries and define functions for plotting to be made.

```
In [1]: # Loading relevant libraries into notebook
from mpl_toolkits.mplot3d import Axes3D
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt # plotting
import numpy as np # linear algebra
import os # accessing directory structure
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
```

There is 1 csv file in the current version of the dataset:

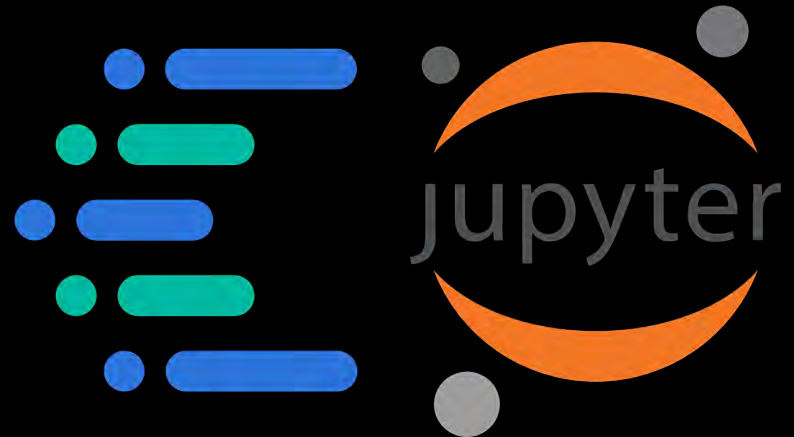
```
In [2]: # This cell is not required for the coffee shop MBA
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

The next hidden code cells define functions for plotting data. Click on the "Code" button

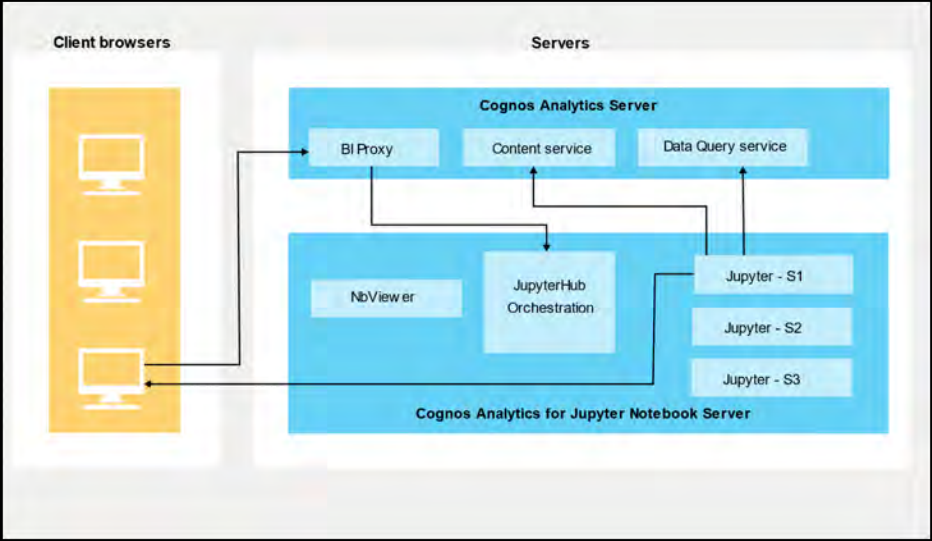
```
In [3]: # Distribution graphs (histogram/bar graph) of column data
def plotPerColumnDistribution(df, nGraphShown, nGraphPerRow):
    nunique = df.nunique()
    df = df[[col for col in df if nunique[col] > 1 and nunique[col] <= 5]]
    nRow, nCol = df.shape
    columnNames = list(df)
    nGraphRow = (nCol + nGraphPerRow - 1) // nGraphPerRow
    plt.figure(num = None, figsize = (6 * nGraphPerRow, 8 * nGraphPerRow),
              dpi = 80, facecolor = 'white', edgecolor = 'k')
    for i in range(min(nCol, nGraphShown)):
        plt.subplot(nGraphRow, nGraphPerRow, i + 1)
        columnDf = df.iloc[:, i].copy()
        if not np.isnan(columnDf.values).any():
            columnDf.hist()
            plt.ylabel('counts')
            plt.xticks(rotation = 30)
            plt.title(f'{columnNames[i]} (count)')
```

Benefits of Combining Cognos Analytics with Jupyter Notebooks

- Enhance the capabilities of CA with advanced analytics and visualisations
- Incorporate the power of Open Source quickly and easily into CA
 - Share work done by others
 - Trusted sources
- Address a key issue with Data Science
 - Deployment



Cognos Analytics and Jupyter Notebooks



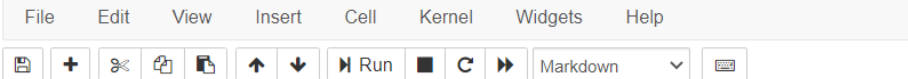
Demonstrations

Analysing text

Unstructured data
i.e. block of text to
understand if there
are patterns in the
text. Start with
most frequently
used words after
some exclusions

Market Basket Analysis

As a retailer, I want to
understand which
products are bought
together. This might
allow me to consider
marketing campaigns
or store layouts such
as “beer and nappies”



Exploratory Analysis

To begin this exploratory analysis, first import libraries and define functions for plotting to be made.

```
In [1]: # Loading relevant libraries into notebook
from mpl_toolkits.mplot3d import Axes3D
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt # plotting
import numpy as np # linear algebra
import os # accessing directory structure
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
```

There is 1 csv file in the current version of the dataset:

```
In [2]: # This cell is not required for the coffee shop MBA
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

The next hidden code cells define functions for plotting data. Click on the "Code" button

```
In [3]: # Distribution graphs (histogram/bar graph) of column data
def plotPerColumnDistribution(df, nGraphShown, nGraphPerRow):
    nunique = df.nunique()
    df = df[[col for col in df if nunique[col] > 1 and nunique[col] <= 5]]
    nRow, nCol = df.shape
    columnNames = list(df)
    nGraphRow = (nCol + nGraphPerRow - 1) // nGraphPerRow
    plt.figure(num = None, figsize = (6 * nGraphPerRow, 8 * nGraphRow),
              dpi = 80, facecolor = 'white', edgecolor = 'k')
    for i in range(min(nCol, nGraphShown)):
        plt.subplot(nGraphRow, nGraphPerRow, i + 1)
        columnDf = df[columnNames[i]]
        columnDf = columnDf.dropna()
        columnDf = columnDf.astype(float)
        columnDf.hist()
        plt.ylabel('counts')
        plt.xticks(rotation = 30)
        plt.title('%s\nDistribution: %s counts' % (columnNames[i], nunique[columnNames[i]]))
```

Options for Cloud Deployments

Different approach

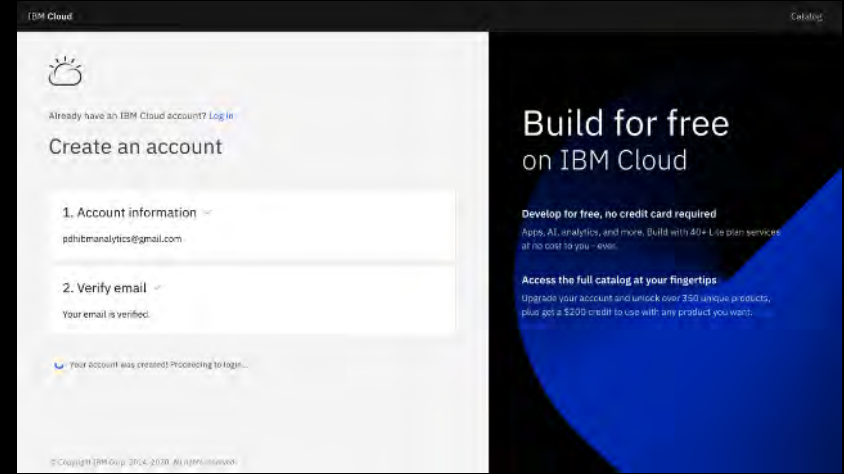
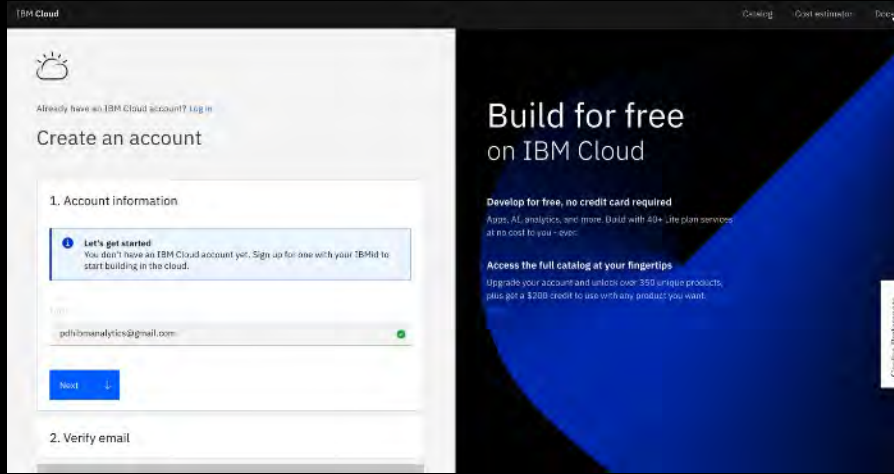
Notebooks are not supported within Ca on Cloud environment

- *Would not make sense to install for each CA Cloud environment a separate notebook server*
- *... while we have a similar offering already running in IBM Cloud*
- *Watson Studio supports Notebooks*
- *... and can leverage CADATAConnector*
- *However, notebooks output cannot be displayed within CA Dashboards and Reports*
- *Best practice:*
 - *Leverage Capabilities within WS (AutoAI)*
 - *Write Notebook result data back to CA data sources*
 - *Leverage CA visualizations to display insights*

Setup within Watson Studio

1. **Sign up to IBM Cloud**
2. **Provision Watson Studio service (free plan or other)**
3. **Create a new project**
4. **Define the assets you need in the project**
 - **Data Connection**
 - **Notebook**
 - **Other as required**

1) Sign up to IBM Cloud: <https://cloud.ibm.com/>

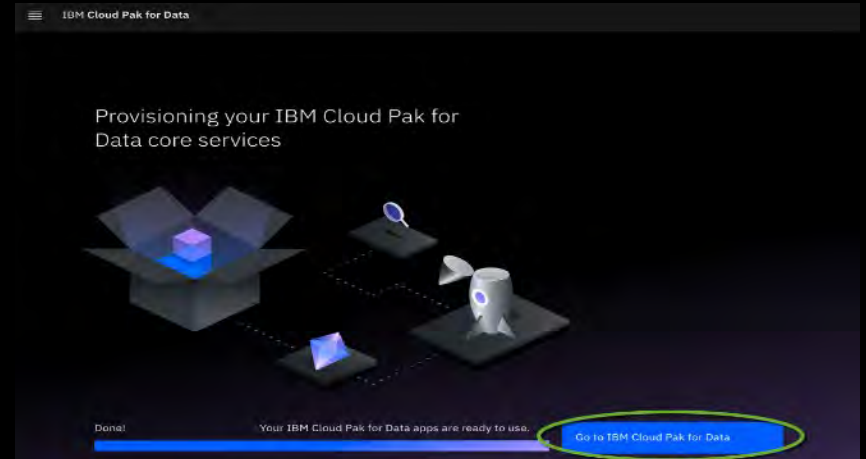
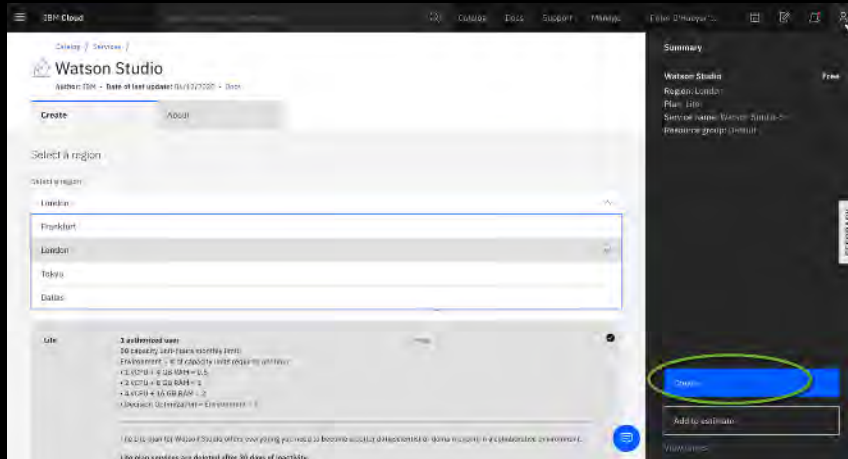
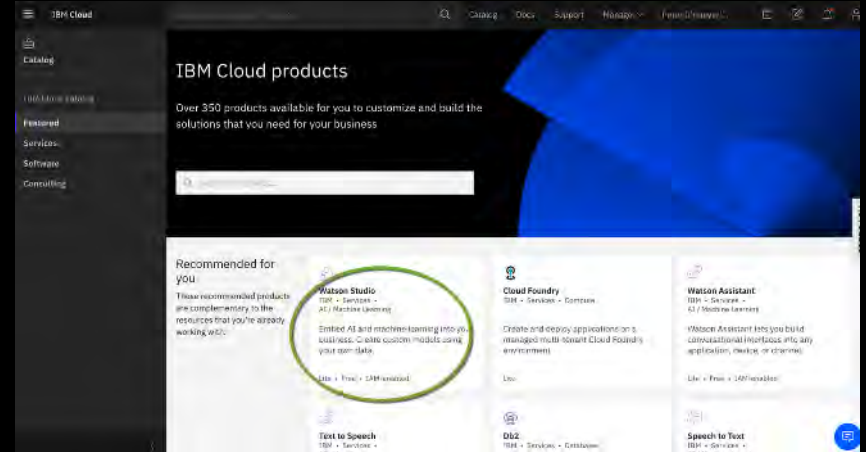
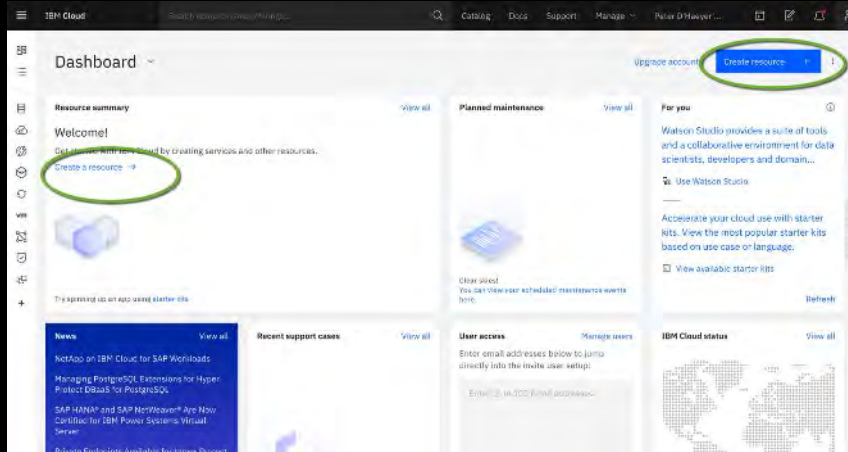


Use your IBM Id to sign up to IBM Cloud

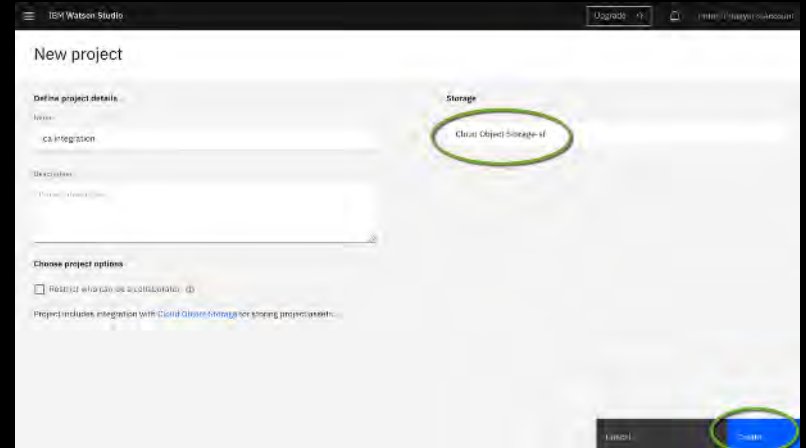
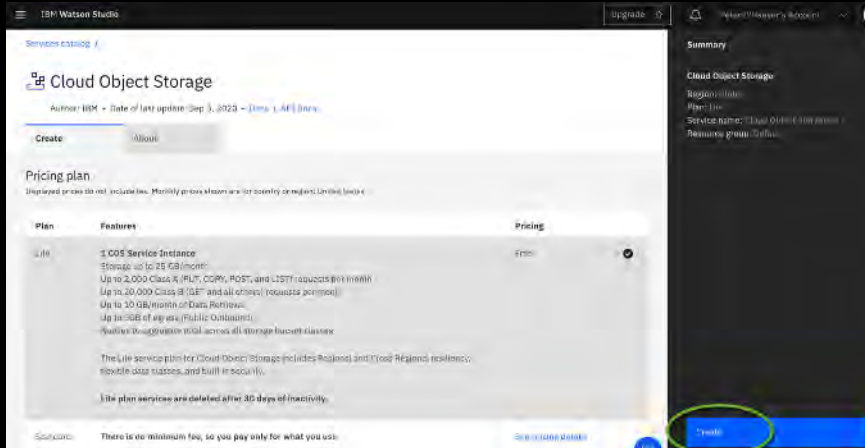
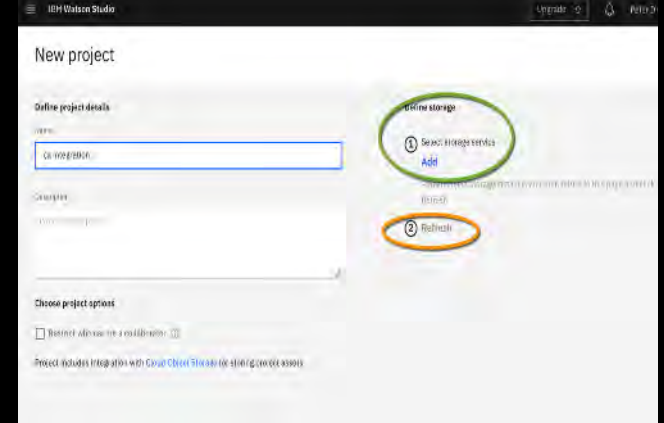
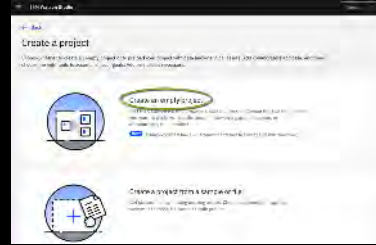
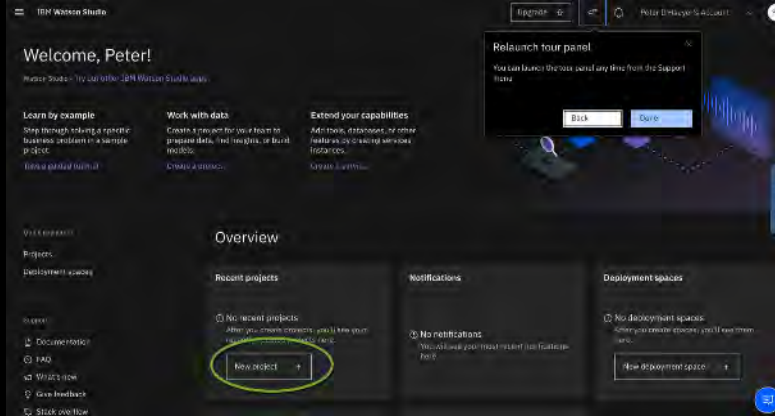
An email will be sent for verification of your account

Enter the code to setup your IBM Cloud environment

2) Provision Watson Studio service



3) Create your project



Demo

Connect to CA data for retrieving a training data set for risk scoring

Watson studio notebook will score loan records to be risky or not

Read a new sample of data from CA and write result data set back into CA

Use dashboards to show results

- **While the algorithm in WS was able to score data about being at risk or not**
- **We can still leverage CA build in predictive capability to find driver based patterns on whether loans or at risk or not**

More use cases for CA notebooks will be explained in a separate tech talk session

Next steps: Join our Community!

- Access our new client webinar series
- Join discussion forums with peers
- Stay up to date on new product features
- Help steer product roadmaps

ibm.biz/bizanalyticscommunity



Next Week

Topic

Cognos Analytics provides unrivaled insight into your data and Cognos onDemand (Software as a Service) provides organisations with a fast onboarding experience, low cost and managed system.

What if your data is not in another cloud? What if it's behind firewalls and you want direct, realtime access to the data without uploading spreadsheets? IBM Secure Gateway (part of CA onDemand) is the answer.

In this session you will understand what IBM Secure Gateway is and how it works. You will see a live demonstration of how to configure it and use Cognos onDemand to access on premise data, behind a firewall, without exposing your data to the rest of the world.

Speaker : Paul Prieto

Date and Time : Wednesday 4th November @ 14:00 GMT

Register at <https://ibm.biz/catechtalks>