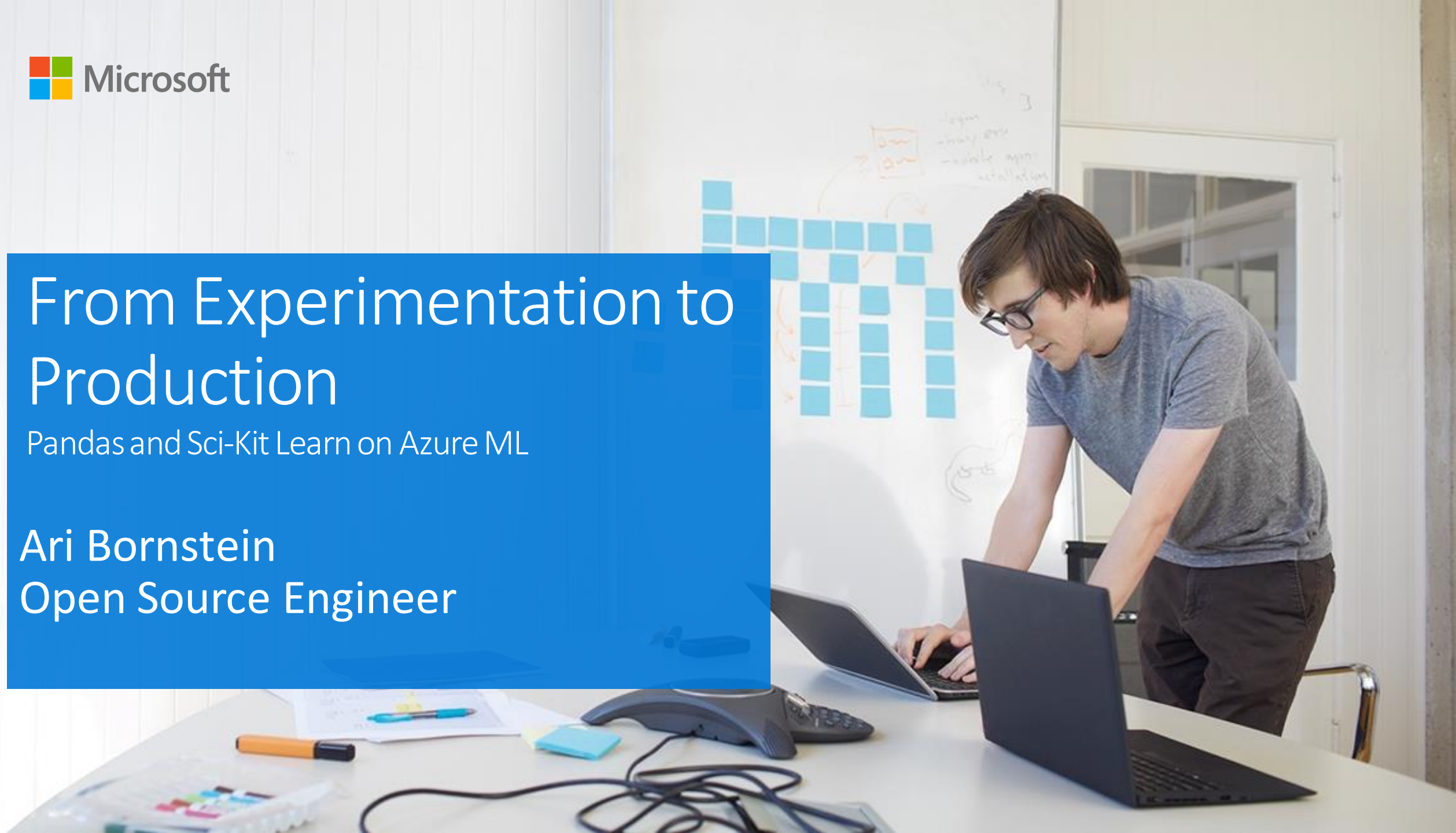




# From Experimentation to Production

Pandas and Sci-Kit Learn on Azure ML

Ari Bornstein  
Open Source Engineer



# About Me

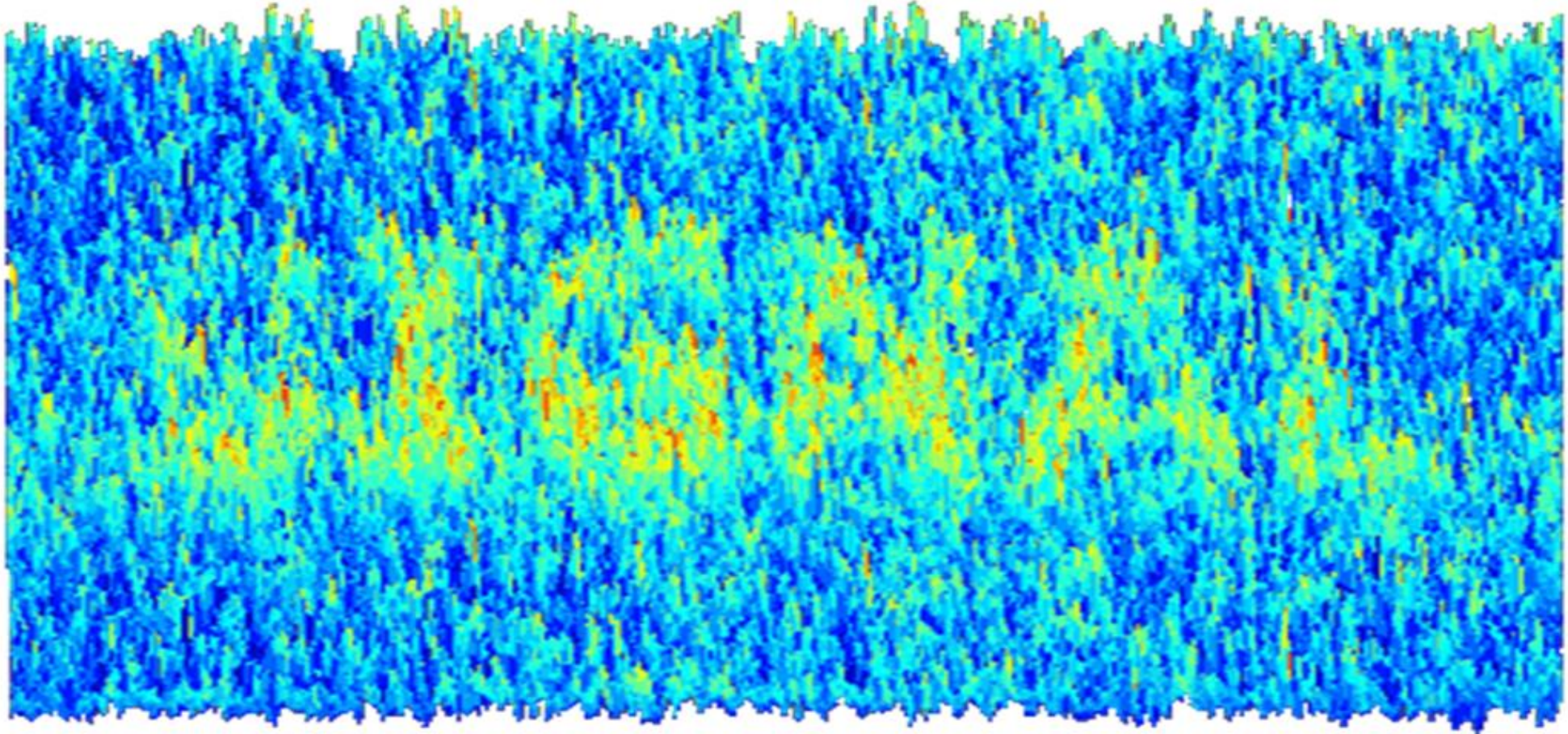


- Website: <http://thepythiccoder.com>
- Twitter: @pythiccoder
- Github: [aribornstein](#)

# Part 1: Experimentation



# Signal and Noise



*The illustrations are from an article on cameras that can be found [Cambridge In Colour](#)*

# Signal and Noise



<http://www.criticalcommons.org/Members/ccManager/clips/alphabet-soup-is-used-as-an-interface-for-the-bat>

What does this look like?

**350,000,000**







# The amount of new data is overwhelming



# For the average doctor.

NCBI Resources How To Sign in to NCBI

PubMed.gov  
US National Library of Medicine  
National Institutes of Health

PubMed  Search

Create RSS Create alert Advanced Help

Article types  
Clinical Trial  
Review  
Customize ...

Text availability  
Abstract  
Free full text  
Full text

PubMed Commons  
Reader comments  
Trending articles

Publication dates  
5 years  
10 years  
Custom range...

Species  
Humans  
Other Animals

Clear all  
[Show additional filters](#)

Summary 20 per page Sort by Most Recent Send to: Filters: [Manage Filters](#)

See 2303 articles about [BRCA1 gene function](#)  
See also: [BRCA1 breast cancer 1](#) in the Gene database  
[brca1](#) in [Homo sapiens](#) [Mus musculus](#) [Rattus norvegicus](#) [All 186 Gene records](#)  
See also: [191 tests](#) for [BRCA1](#) in the Genetic Testing Registry

**Search results**  
Items: 1 to 20 of 12435

<< First < Prev Page 1 of 622 Next > Last >>

☐ [Resveratrol inhibits cell cycle progression by targeting Aurora kinase A and Polo-like kinase 1 in breast cancer cells.](#)  
Medina-Aguilar R, Marchat LA, Arechaga Ocampo E, Gariglio P, García Mena J, Villegas Sepúlveda N, Martínez Castillo M, López-Camarillo C.  
Oncol Rep. 2016 Apr 4. doi: 10.3892/or.2016.4728. [Epub ahead of print]  
PMID: 27109433  
[Similar articles](#)

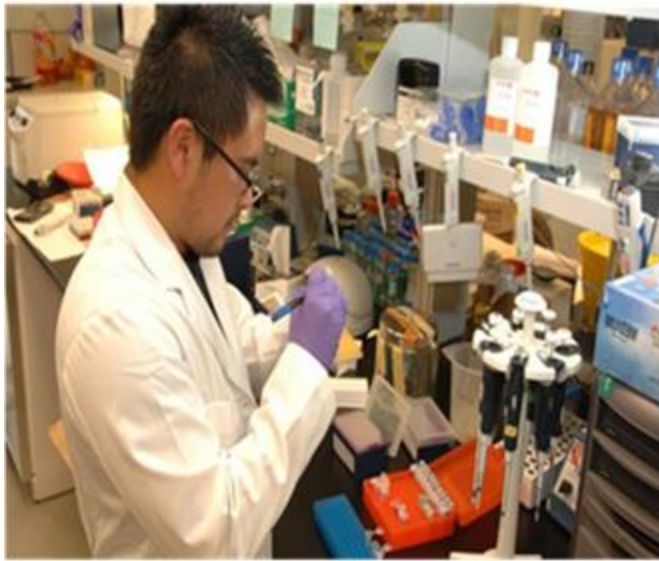
☐ [Resveratrol Ameliorates Cardiac Hypertrophy by Down-regulation of miR-155 Through Activation of Breast Cancer Type 1 Susceptibility Protein.](#)  
Fan Y, Liu L, Fang K, Huang T, Wan L, Liu Y, Zhang S, Yan D, Li G, Gao Y, Lv Y, Chen Y, Tu Y.  
J Am Heart Assoc. 2016 Apr 22;5(4). pii: e002648. doi: 10.1161/JAHA.115.002648.  
PMID: 27107135 **Free Article**

**New feature**  
Try the new Display Settings option - [Sort by Relevance](#)

**Results by year**  
2015: 959  
Download CSV

**Related searches**  
[brca1 2](#)  
[brca1 breast cancer](#)  
[cancer brca1 brca2](#)  
[brca1 gene](#)  
[brca1 mutations](#)

# Medical Research Types



Basic



Clinical



Translational

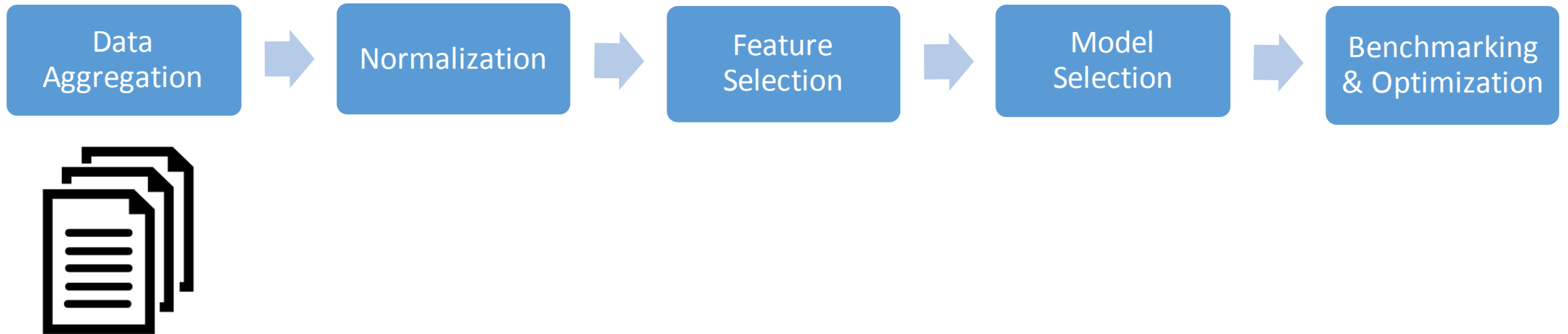




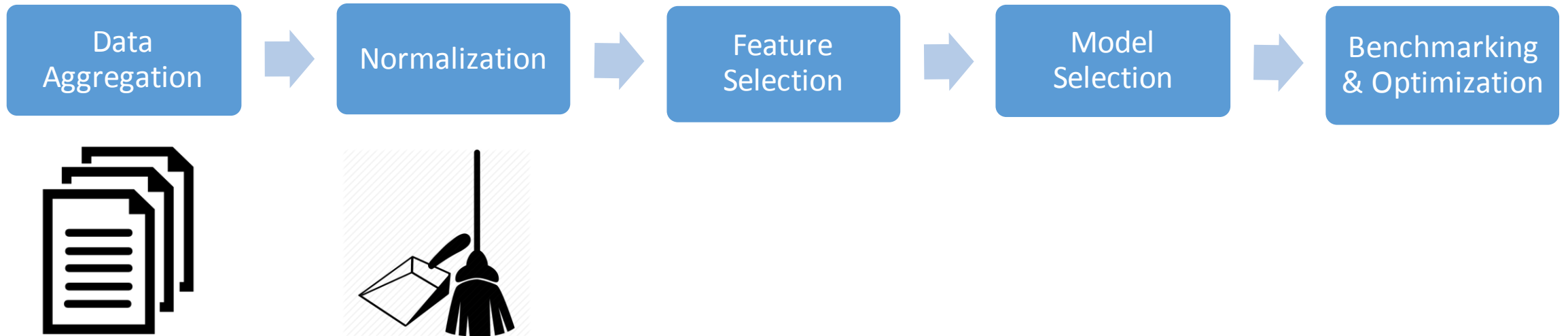
# Document Classification Process



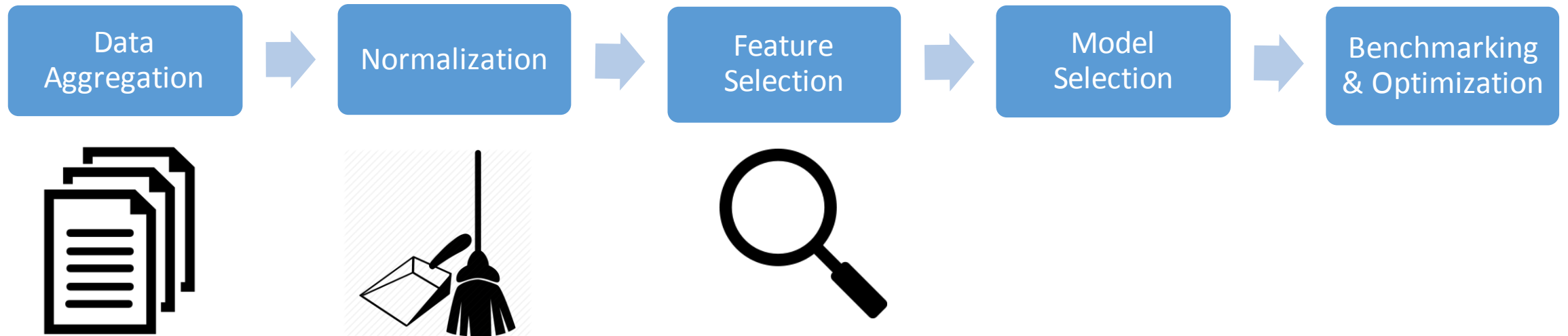
# Document Classification Process



# Document Classification Process

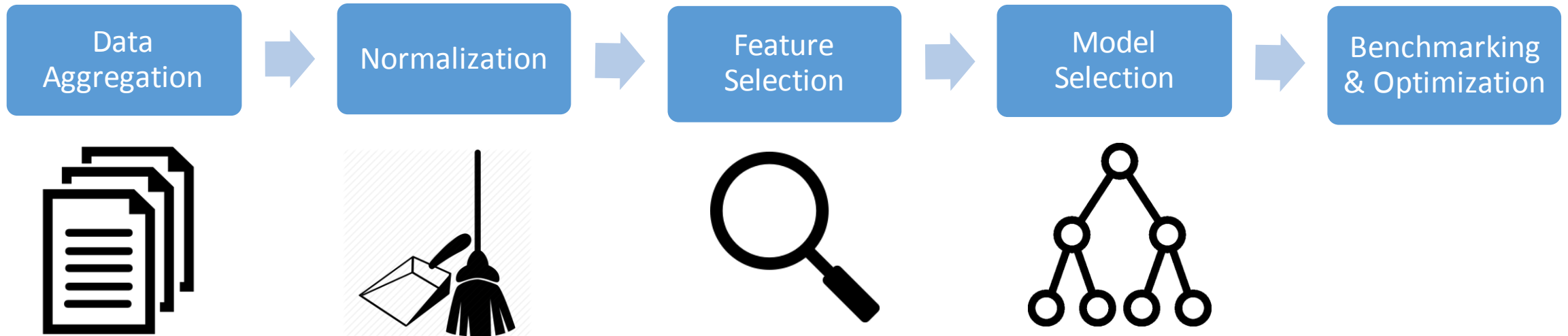


# Document Classification Process

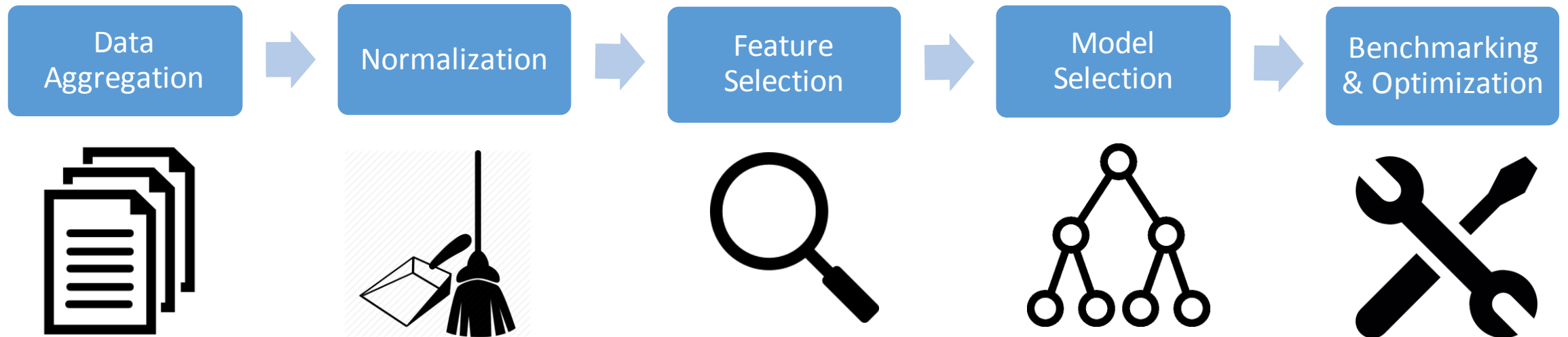




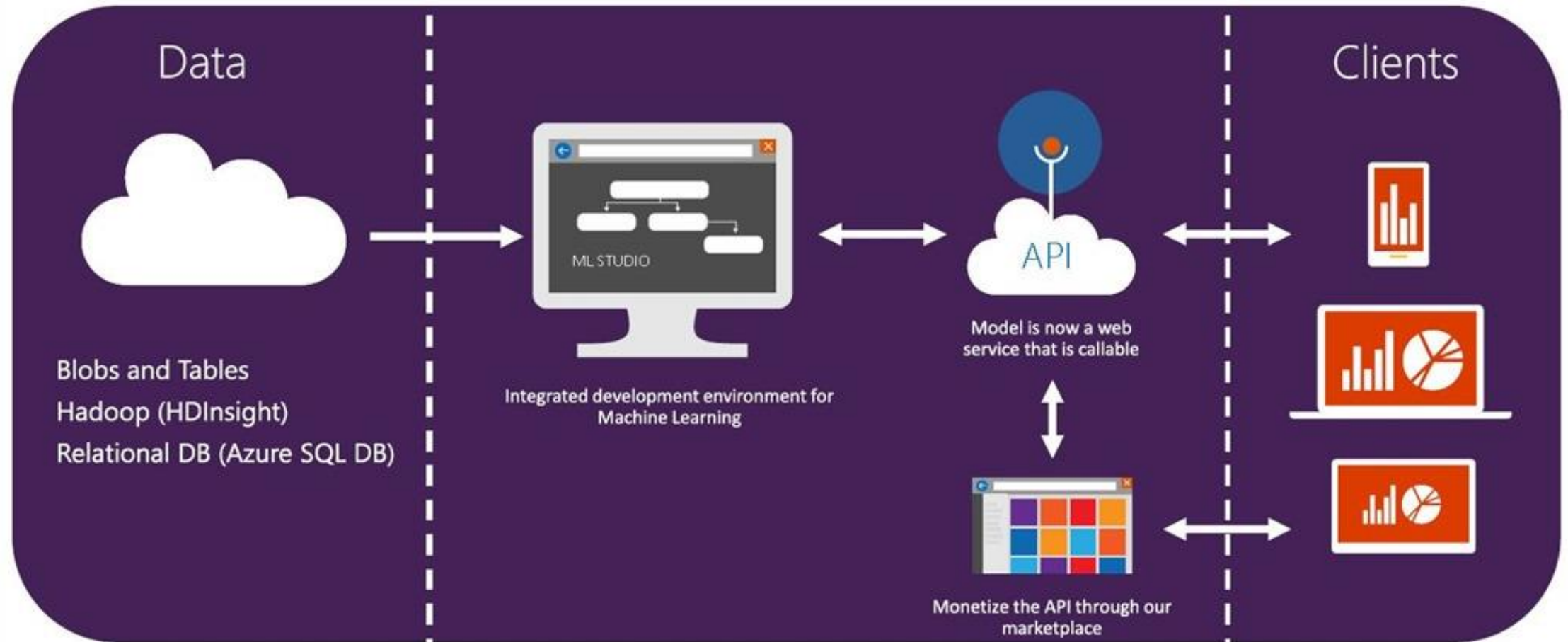
# Document Classification Process



# Document Classification Process



# Azure ML



# Using Experiments to Augment and Operationalize my code

The screenshot displays the Microsoft Azure Machine Learning (ML) interface. The central workspace shows a workflow titled "Bethesda Experiment" in draft status. The workflow consists of three main components: "Project Bethesda DataS...", "Apply SQL Transformation", and "Execute Python Script". The "Execute Python Script" node is highlighted with a blue border and numbered 1 and 2. A "Stopwords.csv" file is also shown as an input to the script. The left sidebar contains a navigation pane with categories like "Saved Datasets", "Data Input and Output", "Data Transformation", "Feature Selection", "Machine Learning", "OpenCV Library Modules", "Python Language Modules", "R Language Modules", "Statistical Functions", "Text Analytics", "Web Service", and "Deprecated". The right sidebar shows the "Properties" panel for the "Execute Python Script" node, displaying a Python script for text processing. The script includes comments and code for creating a normalized abstract entry, removing punctuation, stand-alone numbers, and stop words. The bottom status bar includes icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SETUP WEB SERVICE", and "PUBLISH TO GALLERY".

```
5 # Param(dataframe1): a pandas.DataFrame
6 # Param(dataframe2): a pandas.DataFrame
7 def azureml_main(dataframe1 = None, dataframe2 = None):
8     import re
9
10    # Create normalized abstract entry
11    dataframe1["Normalized"] = dataframe1["TextInput"]
12
13    numWords = ["zero", "one", "two", "three", "four", "five", "six", "seven", "eight", "nine",
14               "thousand", "million", "billion", "trillion"]
15
16    for i in range(len(dataframe1["TextInput"])):
17        Abstract = dataframe1["TextInput"][i]
18        if(Abstract != None):
19            #to lower
20            abstract = Abstract.lower()
21            #remove punctuation
22            abstract = re.findall(r'[^\W]', abstract, flags = re.UNICODE | re.LOCALE)
23            #remove stand alone numbers
24            abstract = " ".join([x for x in abstract if not x.isdigit()])
25            #remove stand alone numbers and number words
26            abstract = " ".join(filter(lambda w: not w in numWords, abstract.split(" ")))
27
28            #remove stop words
29            NormAbs = " ".join(filter(lambda w: not w in set(dataframe2["StopWords"]), abstract.split(" ")))
30
31        else:
32            NormAbs = ''
33            abstract = ''
34
35    dataframe1["Normalized"] = [NormAbs for i in range(len(dataframe1["TextInput"]))]
```



# Azure ML Workbook Demo

# Part 2: Production

Great so now I have a SciKit Learn model how can we push it to production?

# Production Process today





# Pushing my code to production in python with Azure ML

```
@services.publish('workspace_id', 'workspace_token')  
@services.types(abstract = str)  
@services.returns(list)  
def classify(abstract):  
    return clf.predict(abstract)
```

# That's easy but what's in it for me?

- Scalable Restful and Swagger Endpoints
- Auto generated documentation
- Sample Code and Excel integration
- Monetization through the azure marketplace

# Azure ML Studio Demo

# Recap

# Recap

- Pandas and SciKit learn provide valuable tools for building machine learning solutions

# Recap

- Pandas and SciKit learn provide valuable tools for building machine learning solutions
- Azure ML provides mechanisms to augment these tools and operationalize your code



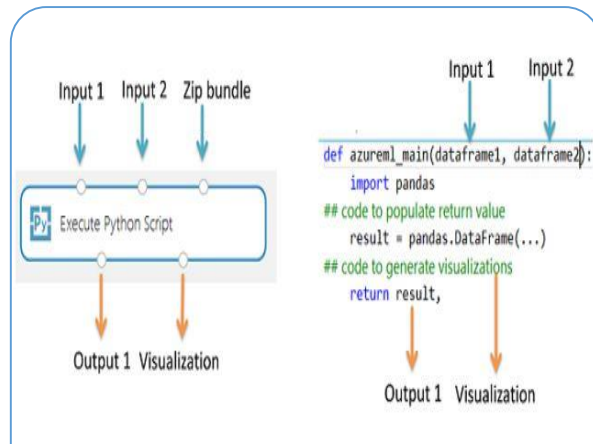
# Questions

# Resources

- [MVA Getting Started with Microsoft Azure Machine Learning](#)
- [Azure Machine Learning \(FAQ\) Types](#)
- [Blog: TechNet Machine Learning Blog](#)
- [Module Descriptions: Machine Learning Module Descriptions](#)
- [Project Bethesda: Research Classification Model](#)

# Using Azure ML as a Workbook Service

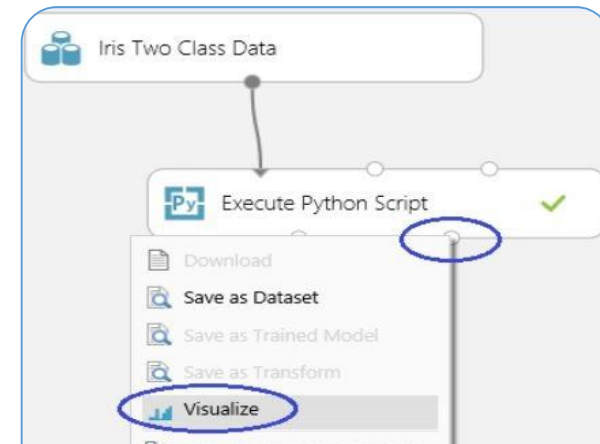
# Using Experiments to Augment and Operationalize my code



Attach Python  
Scripts to  
workflow



Import Custom  
Dependencies



Visualize