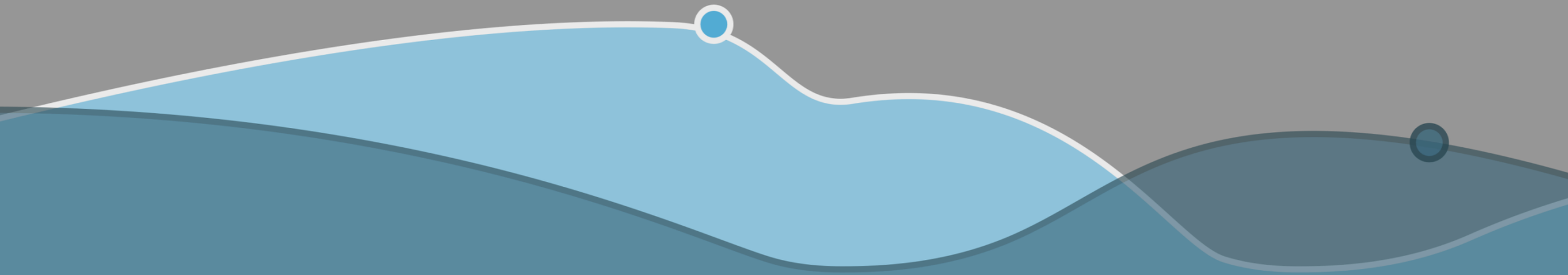
Microsoft

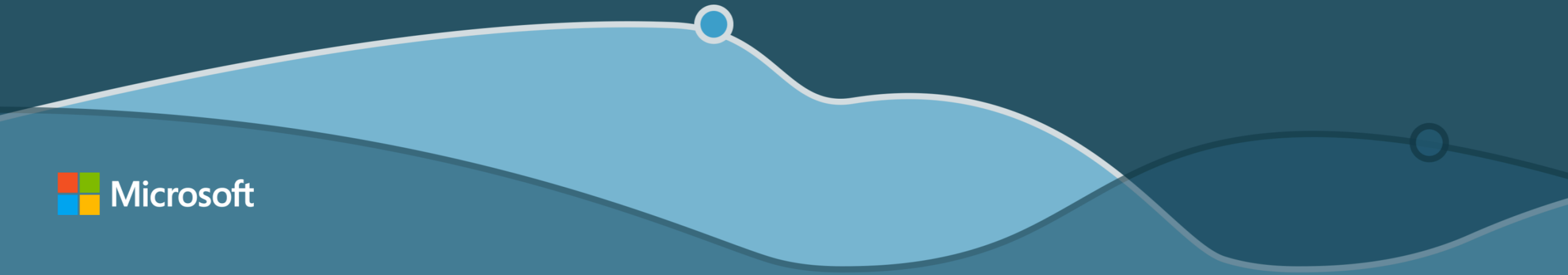
Cortana Analytics Workshop

Sept 10 – 11, 2015 • MSCC



Introduction to Data Science with Cortana: Microsoft Azure Machine Learning

Stephen F. Elston
Principle Consultant, Quantia Analytics, LLC



Microsoft

Overview

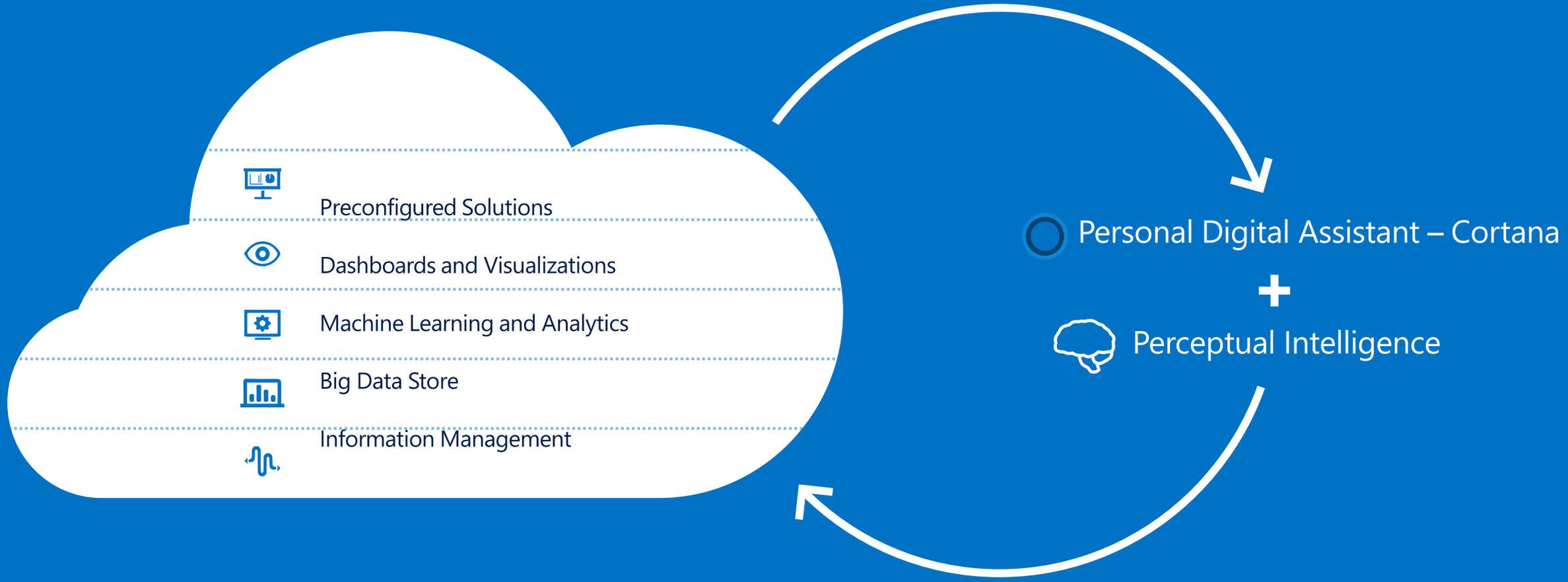
- Introduction to Azure Machine Learning
- Tour of Azure ML Studio
- Building a first Azure ML experiment
- A forecasting example with Azure ML and R
- Publishing a web service

Why Azure ML?

- Quickly deploy production solutions as web services
- Models run in a highly scalable secure cloud environment
- Powerful, efficient built-in algorithms
- Extensible with, SQL, Python, and R
- Integrated in Cortana stack

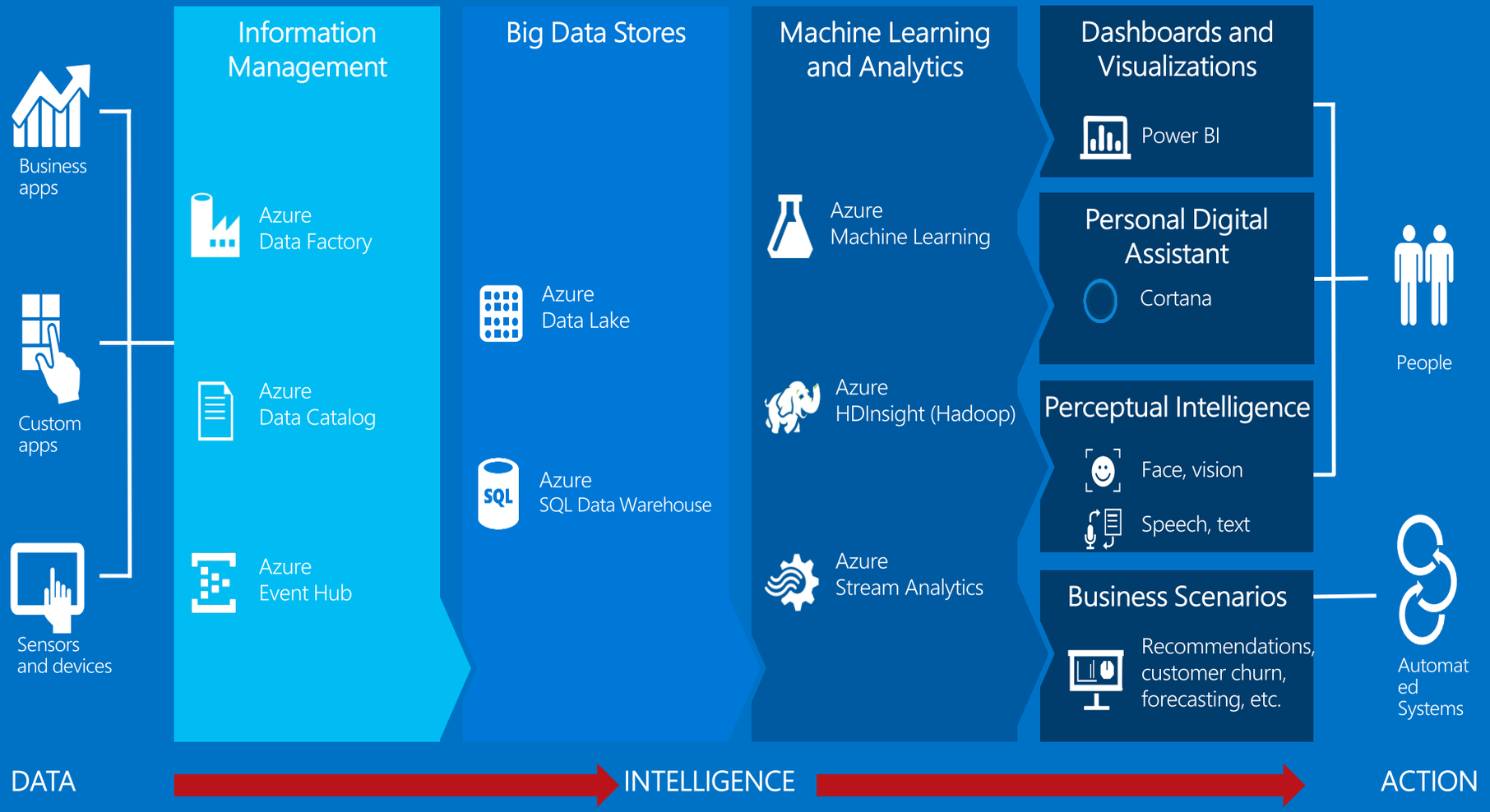
Cortana Analytics Suite:

Transform data into intelligent action



The "Distro" for Intelligence

http://microsoft.com/cortanaanalytics



Why Open-Source Tools?

- R and Python widely used in data science
- Highly interactive
- Good visualization
- Vast packages (libraries) of utilities and machine learning algorithms
- Excellent development environments

Azure ML Free Tier Account

- Free Tier Account
  - http://bit.ly/azureml_login
- Unlimited time, with restricted priority
- Paid account provides full performance

Quick Start Guide to Azure Machine Learning

Azure ML Studio

- Experiments contain workflow
- Experiments constructed of modules
- Experiments in sharable workspace
- Modules transform data, compute models, score models, and evaluate models
- Create custom modules with SQL, R and Python
- Deploy solutions as web services

Azure ML Documentation Resources

Azure ML tutorials and resources:

http://azure.microsoft.com/en-us/documentation/services/machine-learning/

Azure ML Gallery:

http://azure.microsoft.com/en-us/documentation/services/machine-learning/

Documentation and examples for each module

Sample Experiments tab in studio

Azure ML Learning Resources

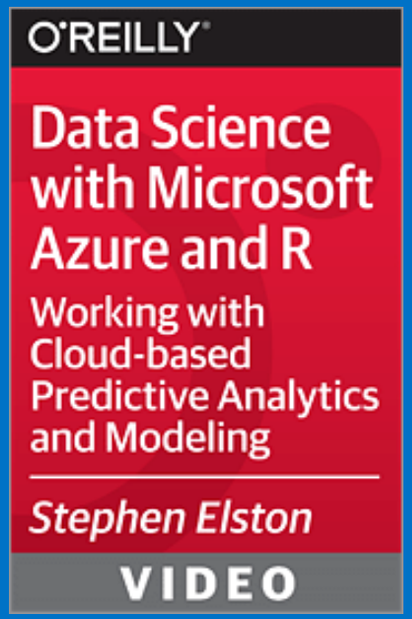
Book, Microsoft Azure Essentials: Azure Machine Learning

http://www.microsoftvirtualacademy.com/ebooks#9780735698178

Data Science in the Cloud Microsoft Azure Machine Learning and R, O'Reilly Media

http://www.oreilly.com/data/free/data-science-in-the-cloud.csp

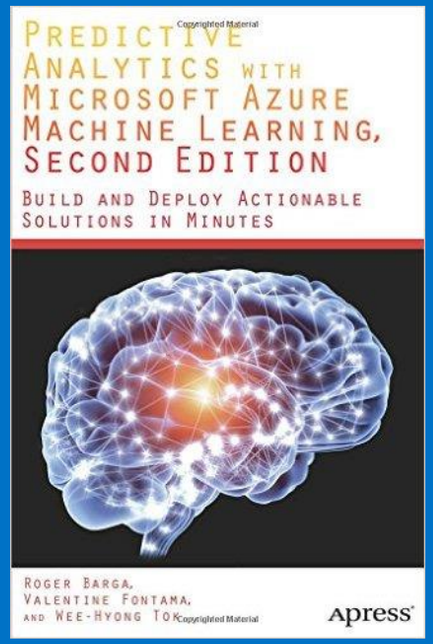
Azure ML Learning Resources



http://shop.oreilly.com/product/0636920040255.do

Data Science and Machine Learning Essentials
With
Stephen Elston and
Cynthia Rudin

https://www.edx.org/course/data-science-machine-learning-essentials-microsoft-dat203x



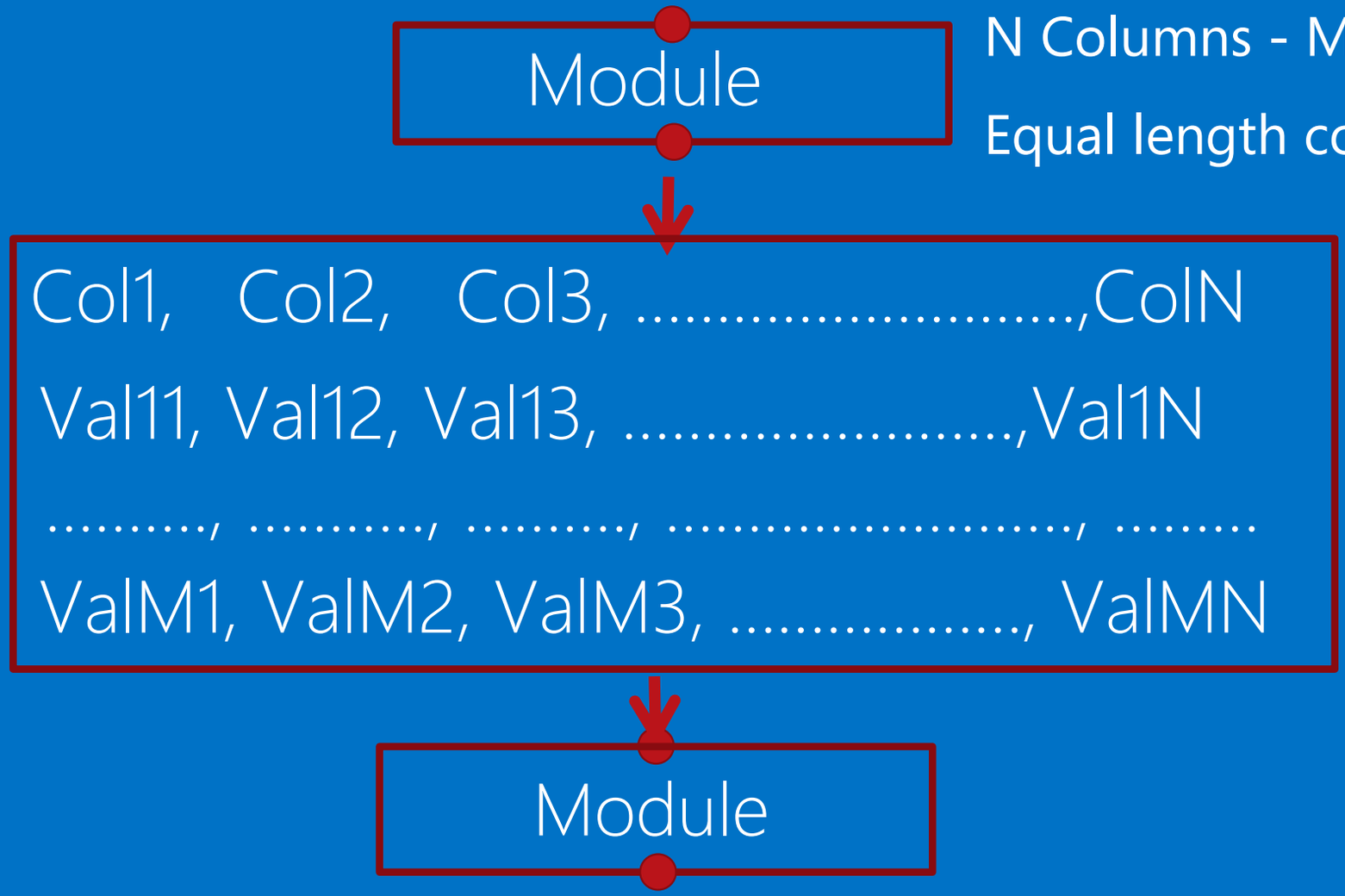
http://www.amazon.com/Predictive-Analytics-Microsoft-Machine-Learning/dp/1484212010/ref=la_B00NBELJJI_1_1?s=books&ie=UTF8&qid=1441060294&sr=1-1

Data Passed from Module to Module in Azure ML Tables

Rectangular table

N Columns - M Rows

Equal length columns



Azure ML Table Data Types

- Numeric: Floating Point
- Numeric: Integer
- Boolean
- String
- Categorical
- Date-time
- Time-Span
- Image

Building a First Model

Building machine learning models

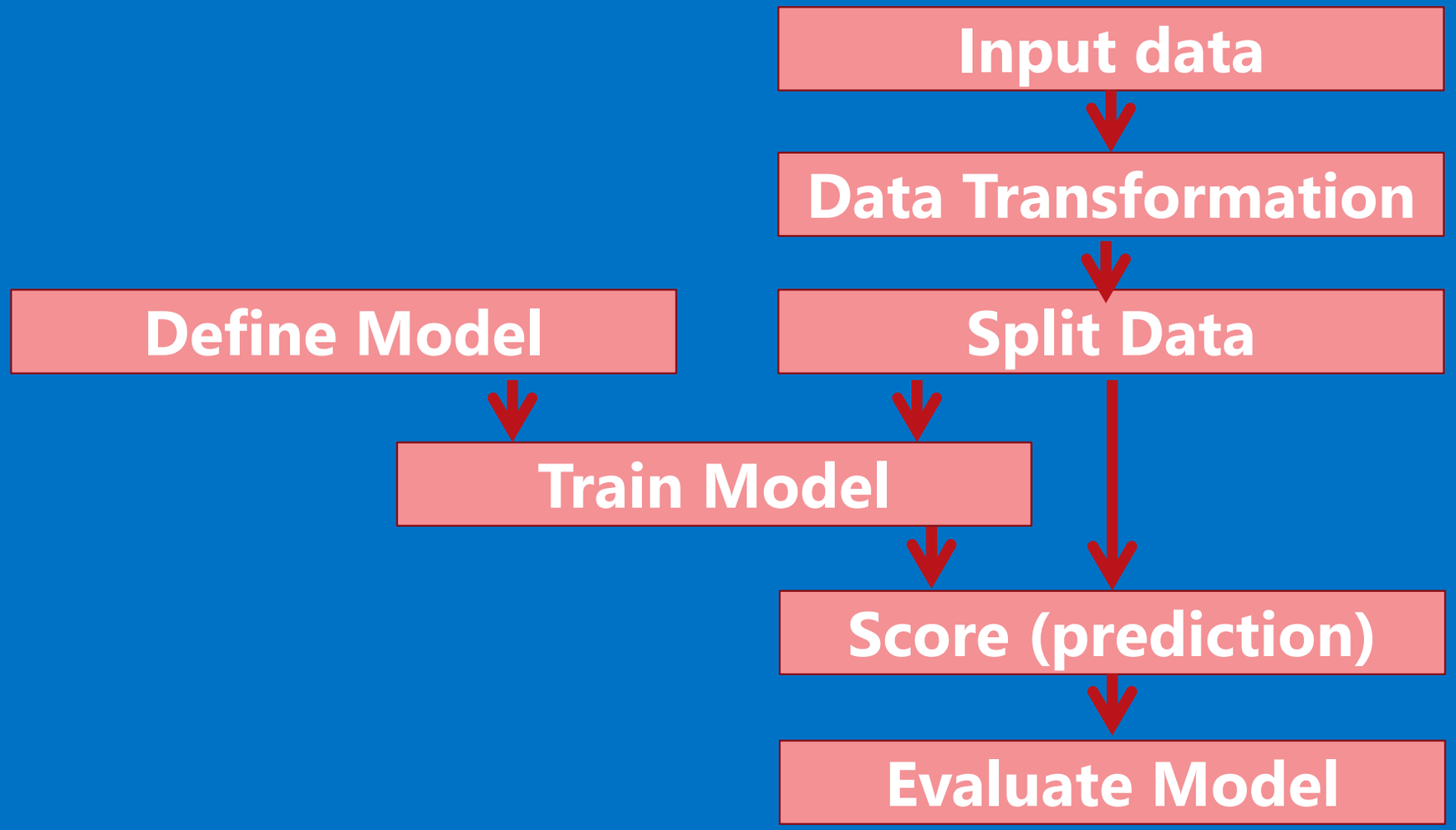
- Define business problem
- Understand data relationships
- Prepare data
- Construct models
- Evaluate models
- Improve models
- (Cross) validate model
- Publish model

Classification

“Science is the systematic classification of experience.”
George Henry Lewes

- Two class and multi-class
- Examples:
  - Species
  - Move genre
  - Fraud detection

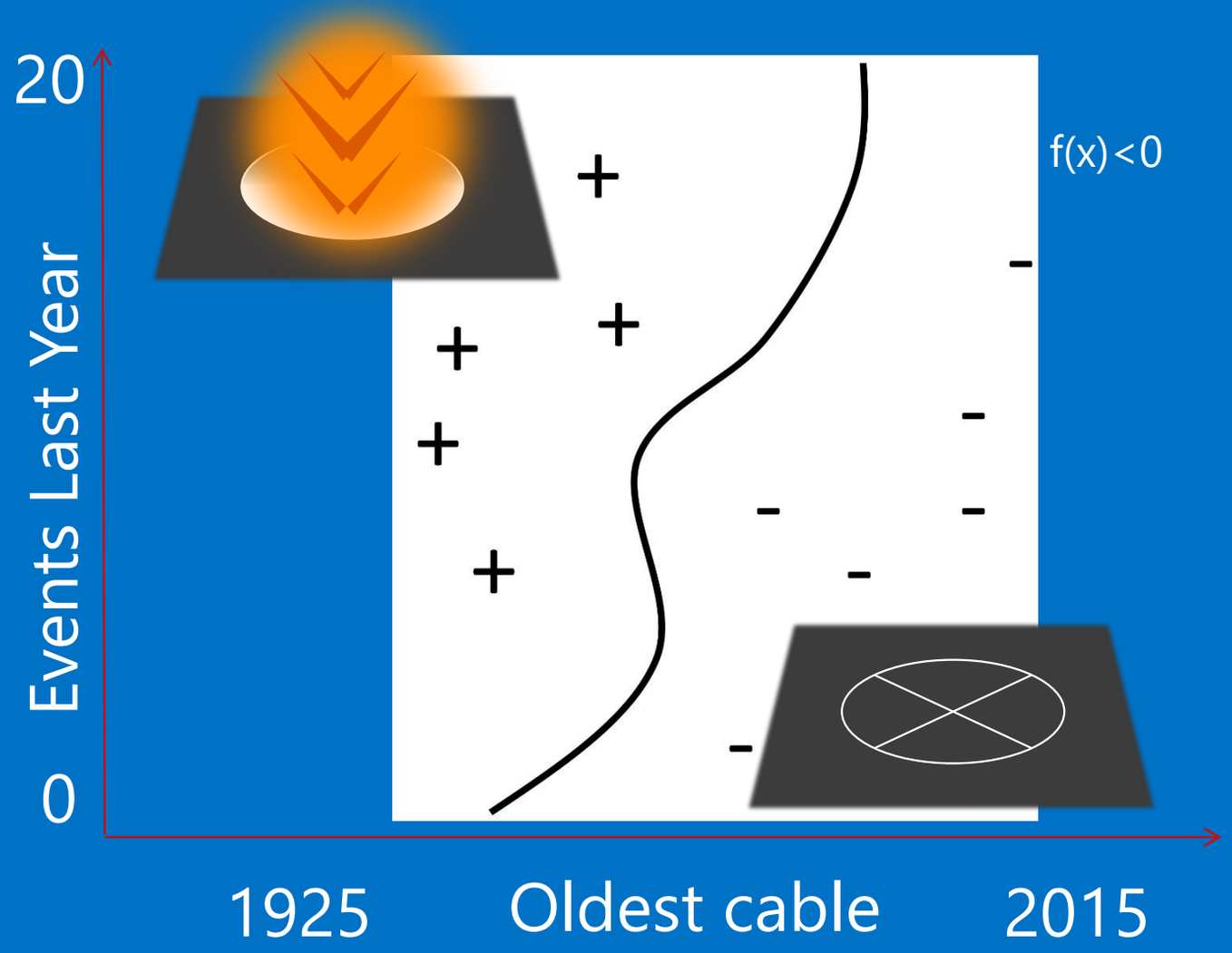
Machine learning workflow



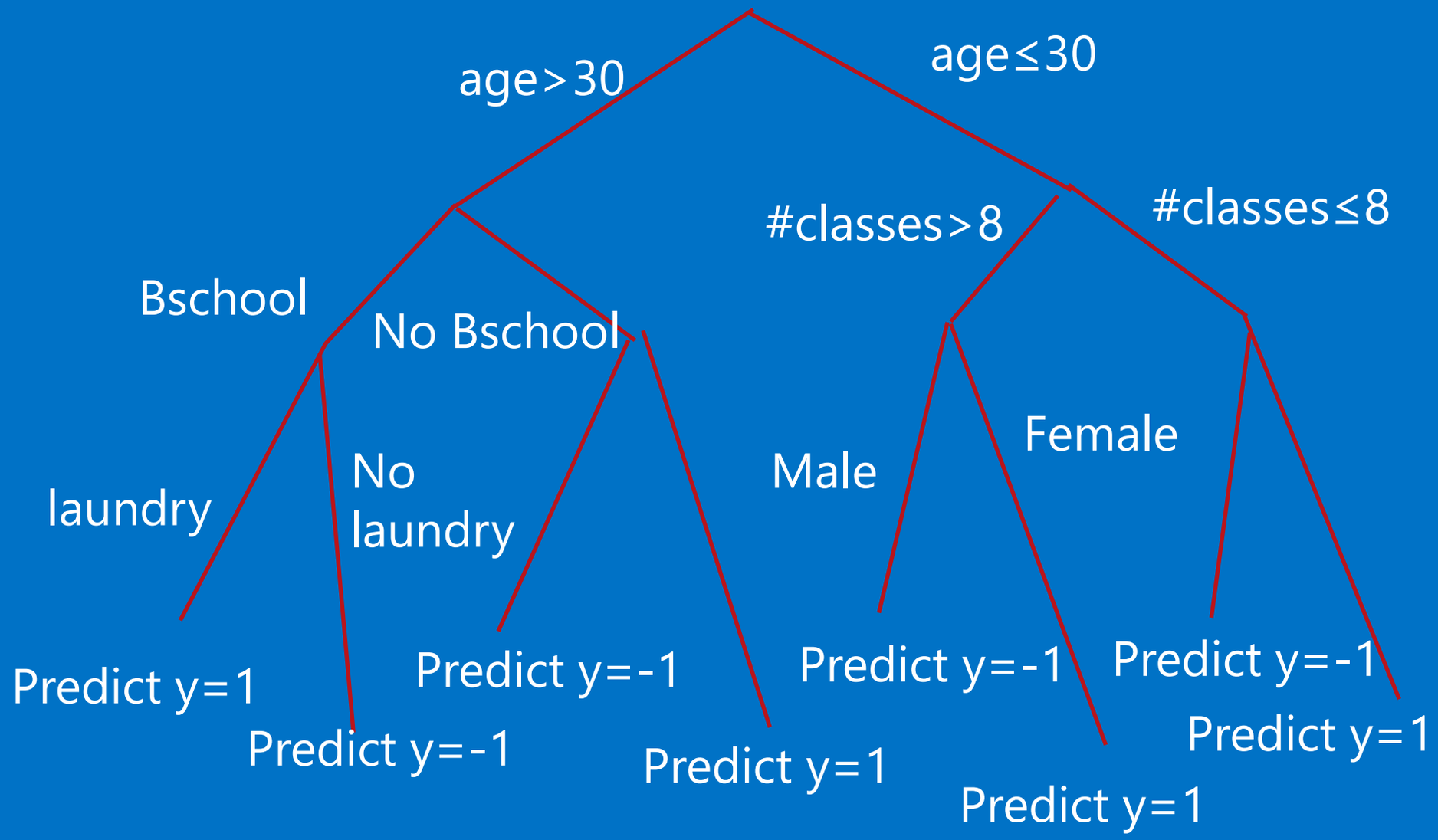
Classification

Formally, given training set (x_i, y_i) for i=1…n, we want to create a classification model f that can predict label y for a new x.

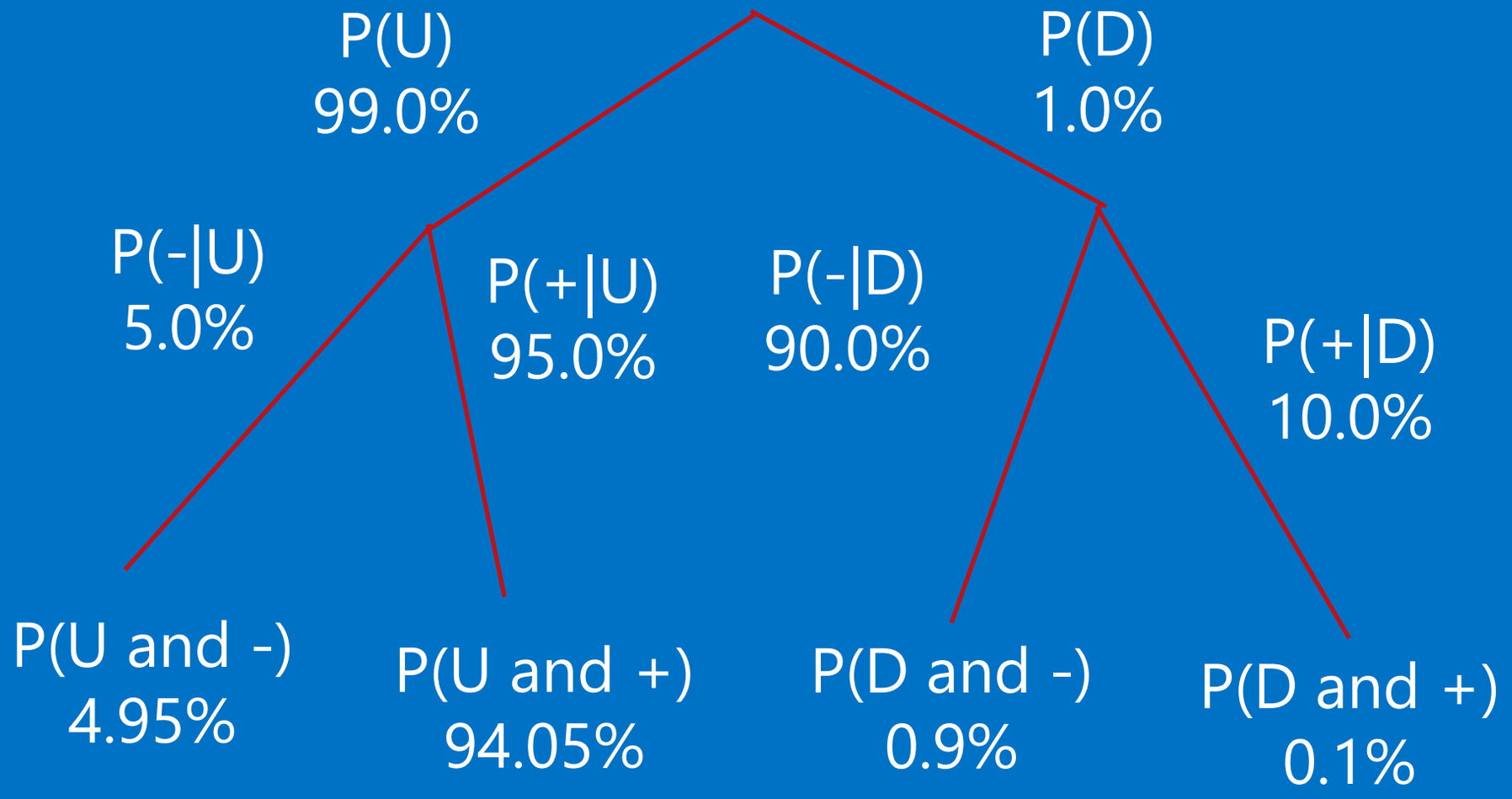
f(x) = function(Events Last Year, Oldest Cable)



Decision Trees For Classification



Empirical Bayes Model



Metrics for Classification

Confusion matrix

	Predicted Positive	Predicted Negative
Actual Positive	**TP**	**FN**
Actual Negative	**FP**	**TN**

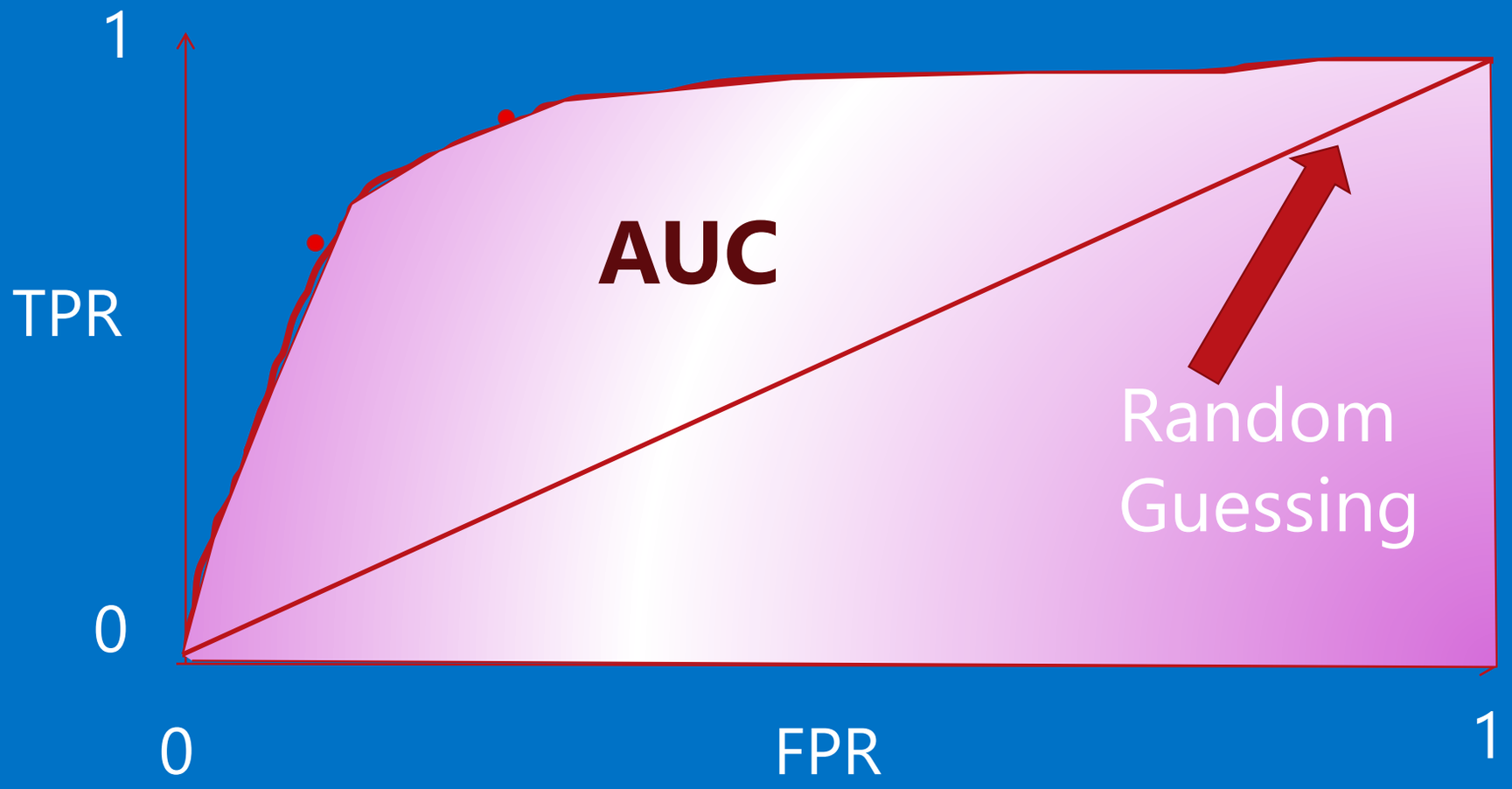
Metrics for Classification

- Accuracy = TP + TN / (TP + TN + FP + FN)
- Precision or positive predictive value = TP/(TP + FP)
- Recall = TP/(TP + FN)
- F1 = Precision * Recall / (Precision + Recall)

	Predicted Positive	Predicted Negative
Actual Positive	**TP**	**FN**
Actual Negative	**FP**	**TN**

ROC Curves

For a particular False Positive Rate (FPR), what is the True Positive Rate (TPR)?



Demo: Classification Example

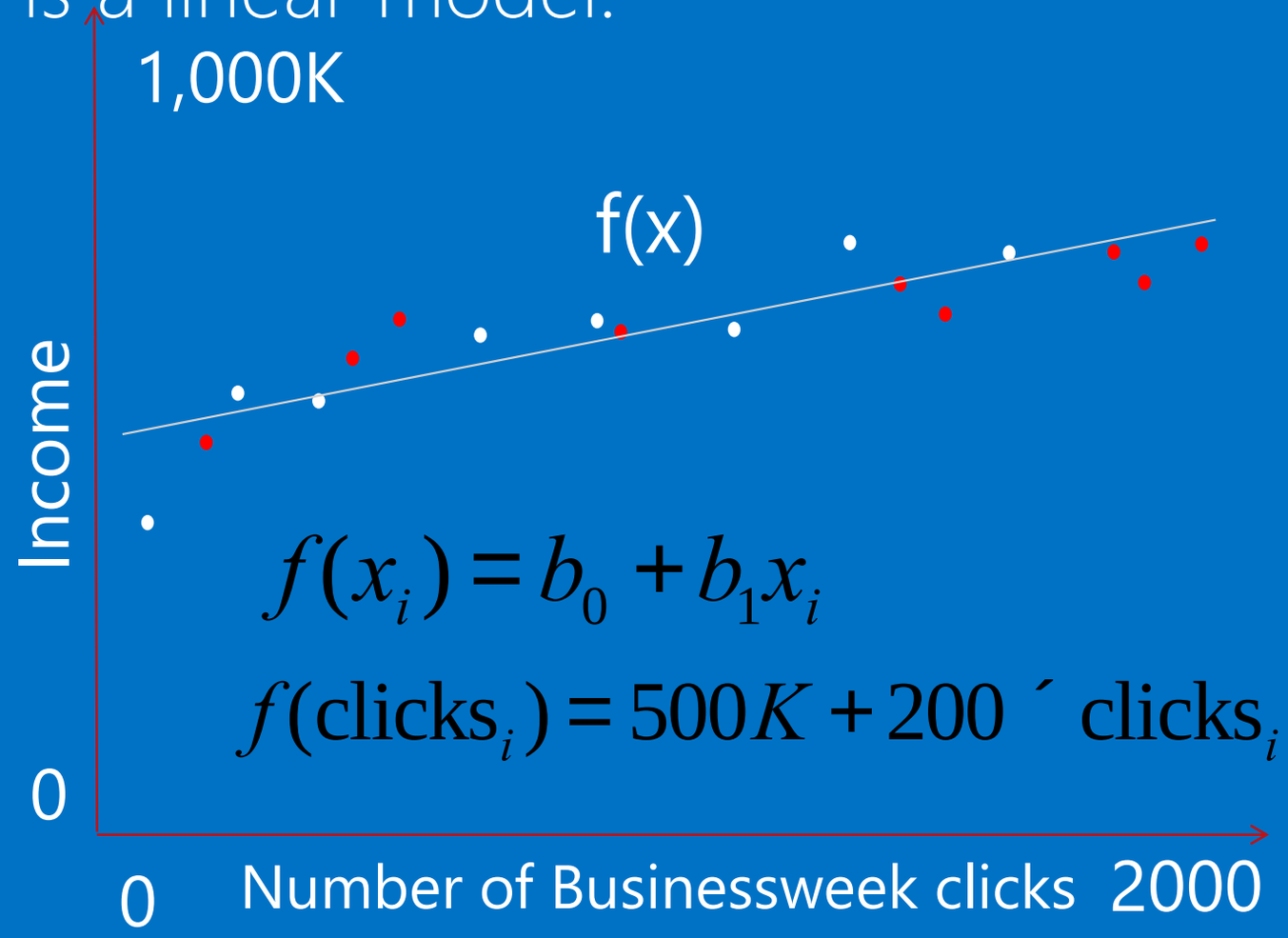
Regression Model Example with R or Python in Azure ML

Forecasting

- A forecast is a prediction of a future value
- Examples:
  - Inventory levels
  - Utility demand
  - Service requirements

Linear regression

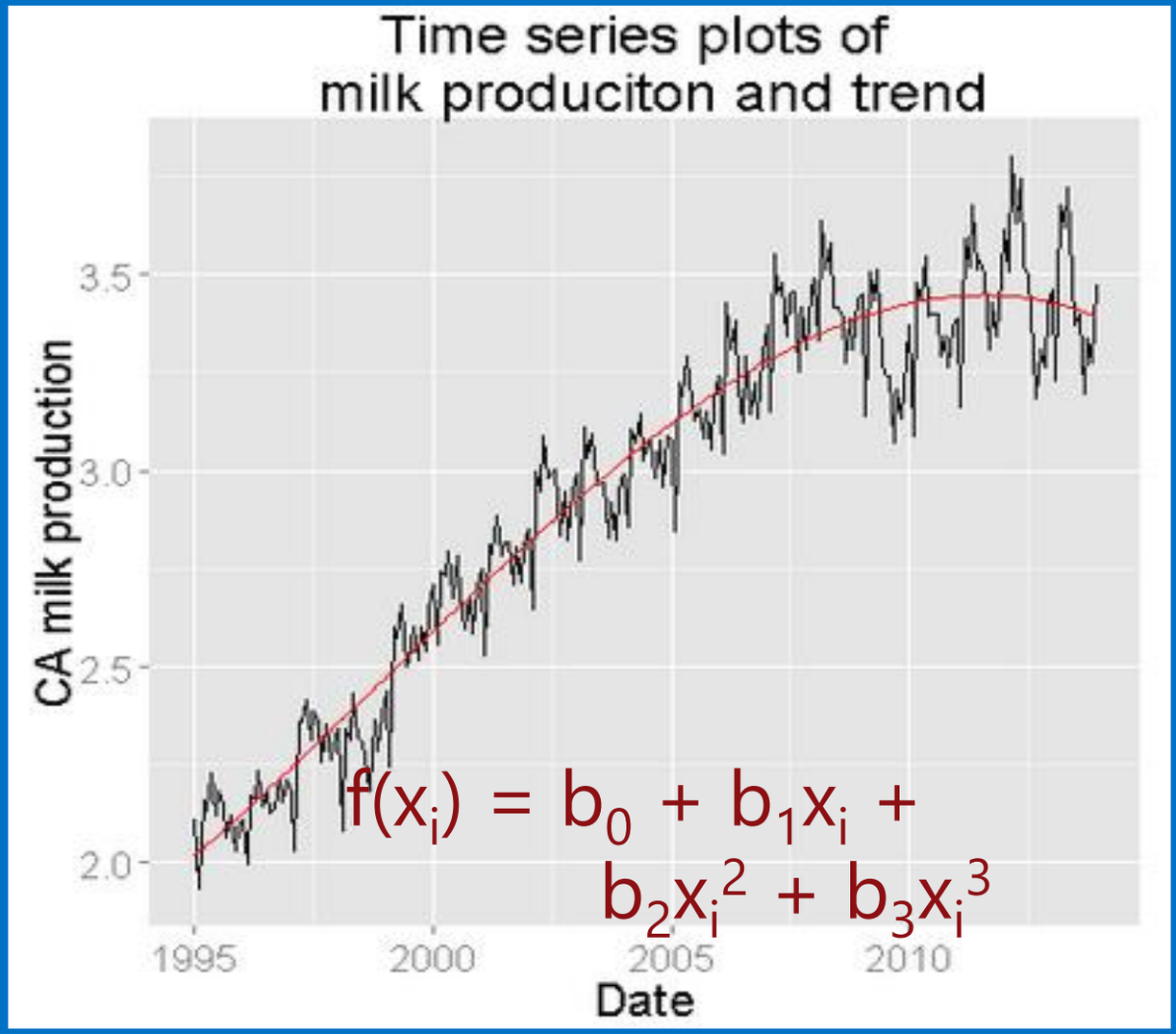
- Need a function that estimates y for a new x.
- The simplest is a linear model.



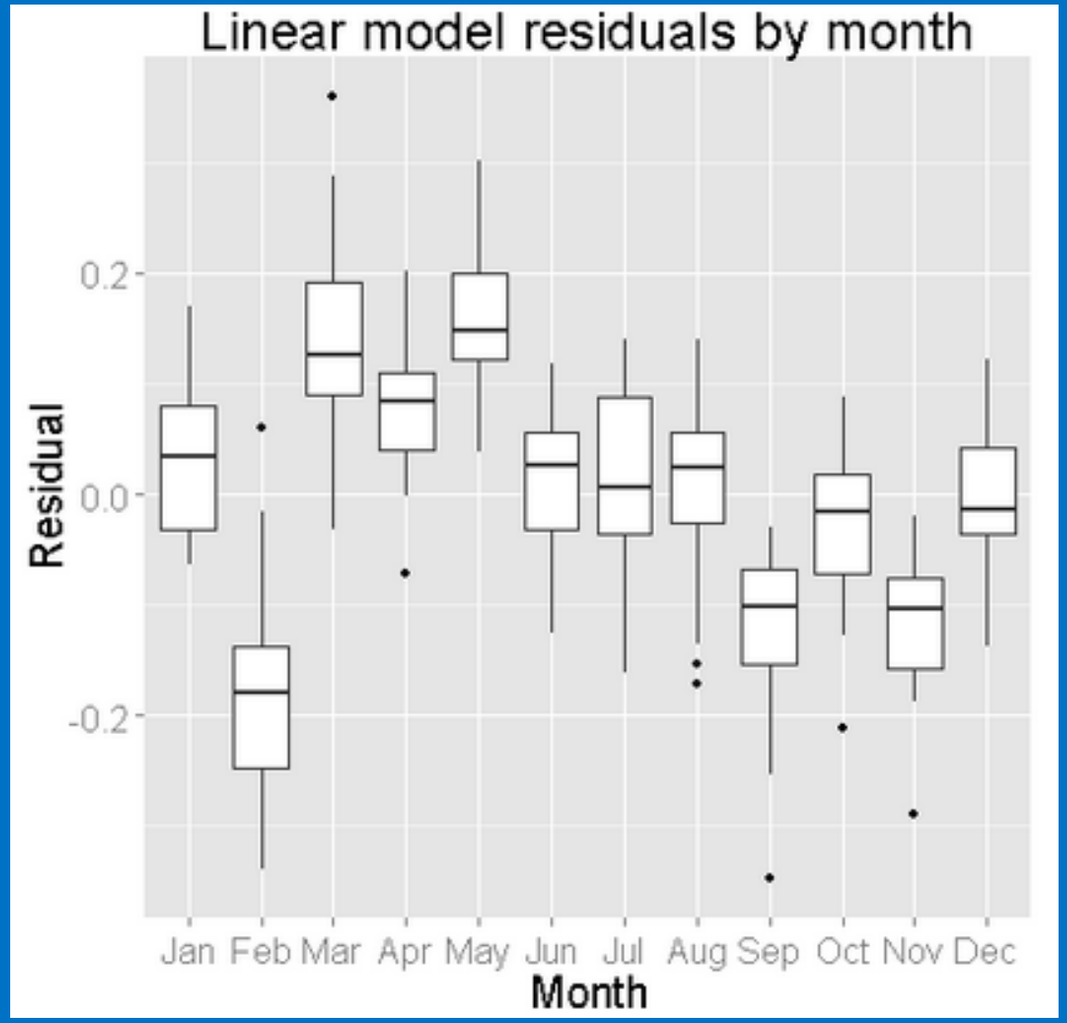
$$f(x_i) = b_0 + b_1 x_i$$

$$f(\text{clicks}_i) = 500K + 200 \text{ ´ } \text{clicks}_i$$

Time series trend and seasonal variation



Time series trend and seasonal variation



R or Python?

- R and Python are widely used in data science
- Powerful open-source data science tools
- Python tends to be more systematic and faster
- R contains wider range of packages and analytics capabilities
- R support currently deeper in Azure ML

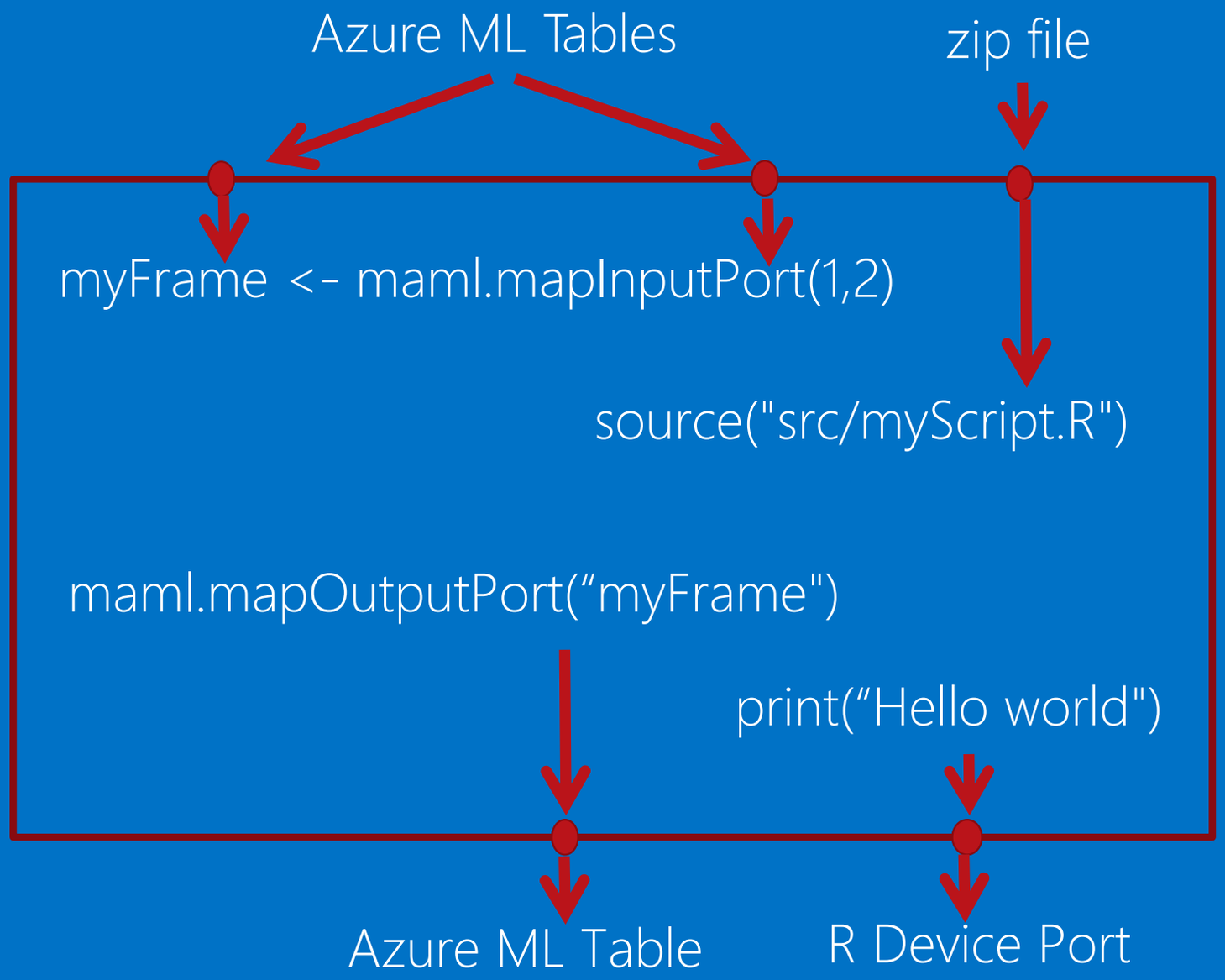
Developing and testing R and Python

- Azure ML is a production environment
- Interactively develop and test in IDE
- Subset data as needed – download as .csv
- IDE has powerful editor and debugger
- Cut and paste code into Execute R/Python Script module to test in Azure ML
- Jupyter notebooks in preview (Python only, R coming)

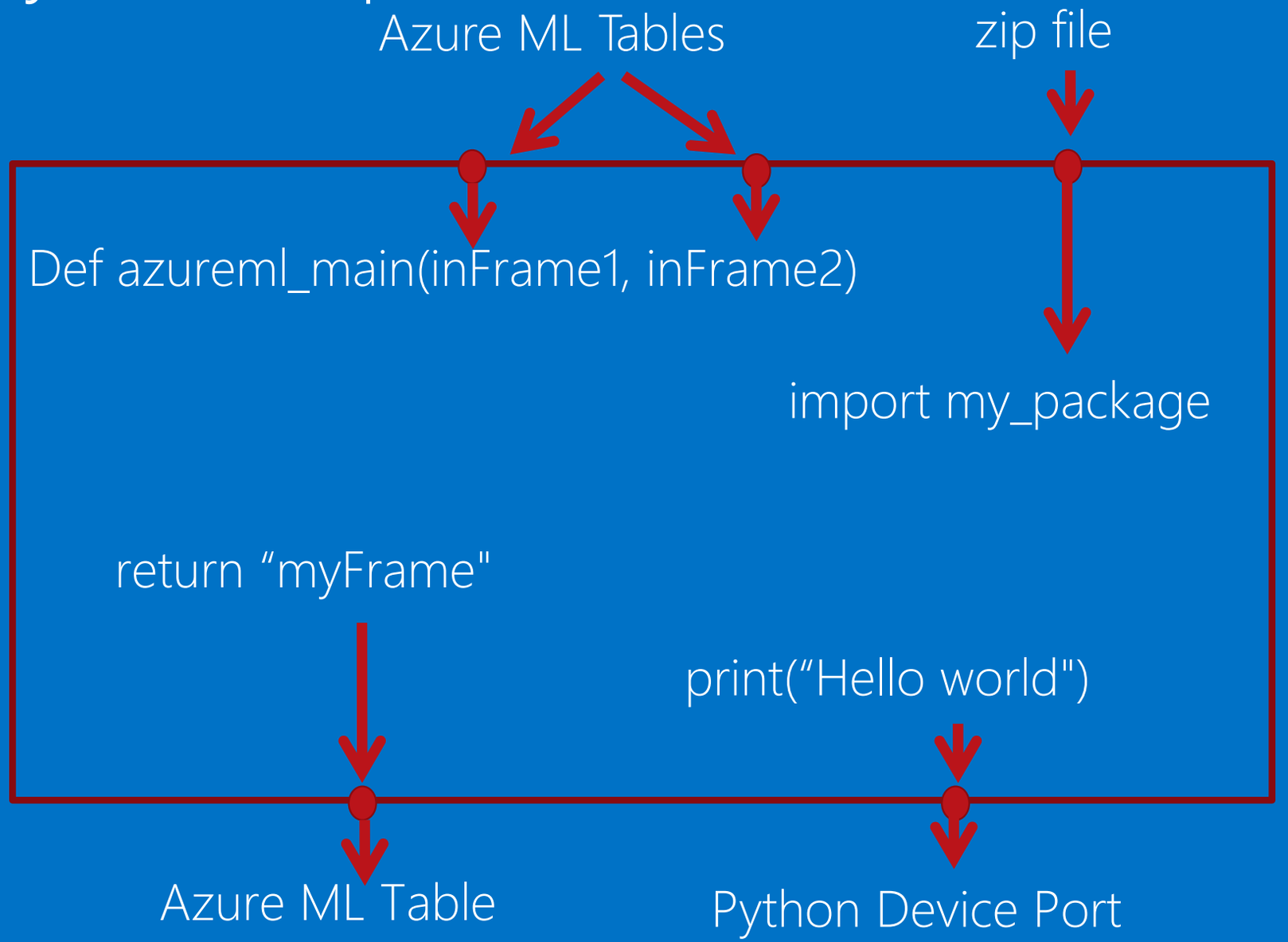
Debugging R and Python in Azure ML

- Code tested in IDE should run in Azure ML, but……
- If error occurs look at the error.log or output.log
- From R use print() function
- From Python use sys.stderr.write() from sys

Execute R Script



Execute Python Script



Python Plotting in Azure ML

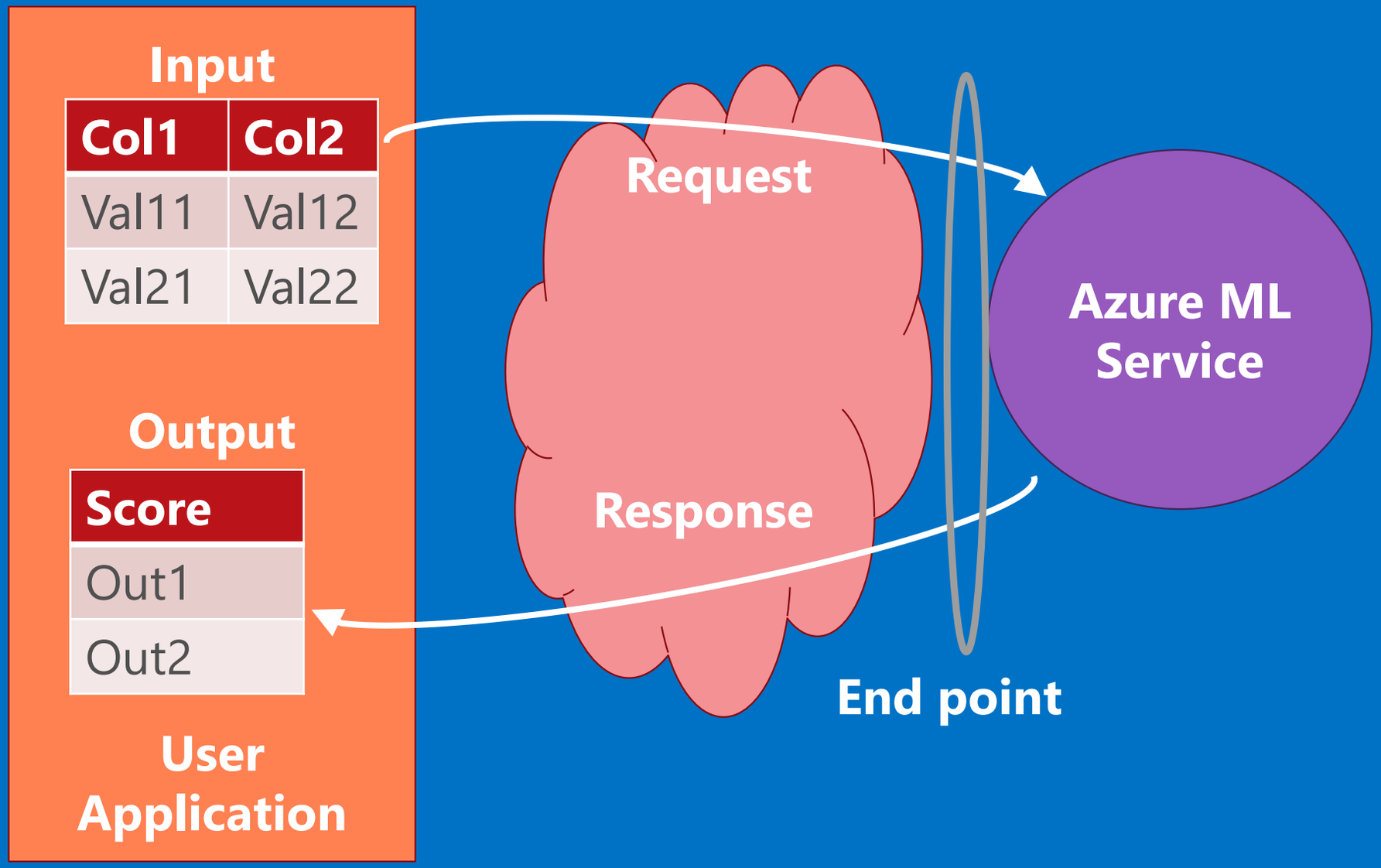
```
def azureml_main(frame1):
# Set graphics backend: Do this first!
  import matplotlib
  matplotlib.use('agg')

  ## Code to create plots

  ## Save figure in a file for output
  fig.savefig('scatter2.png')

  return frame1
```

Forecasting Regression Example

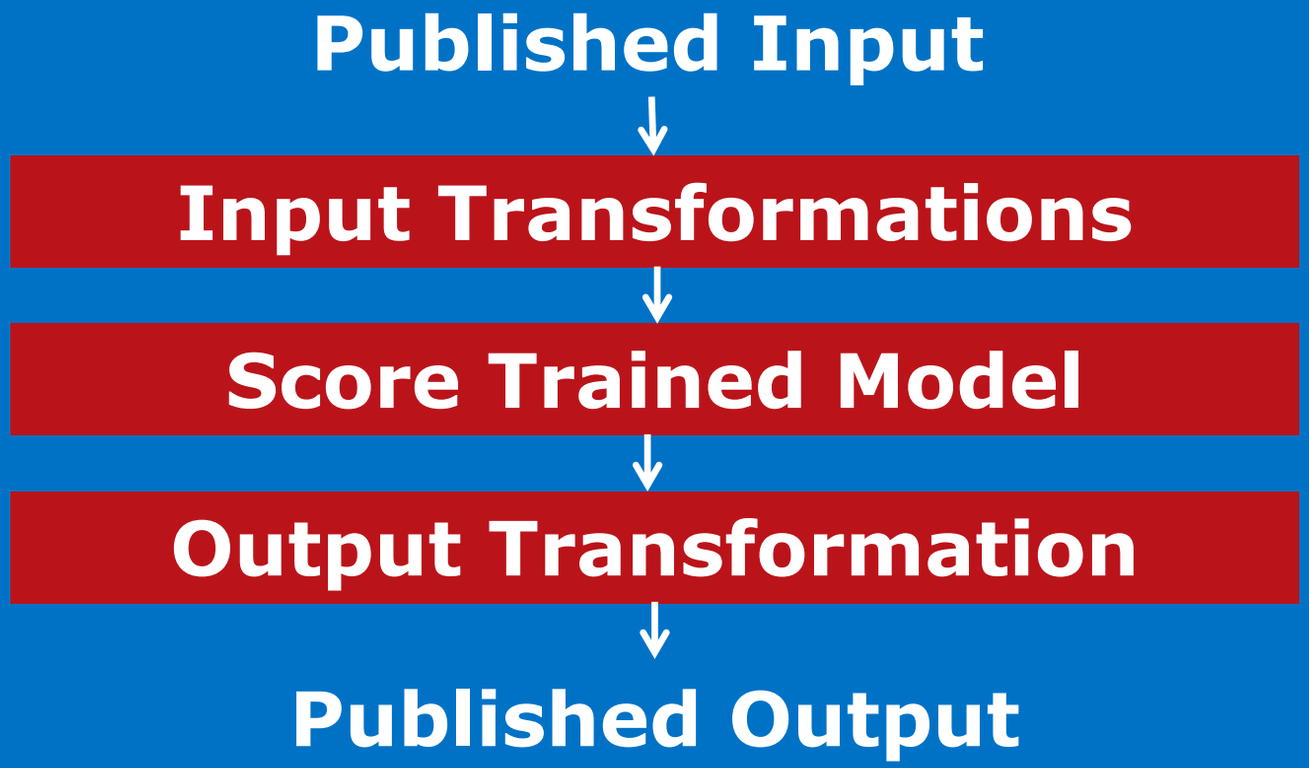
- Time series regression forecast milk production in the State of California
- In Gallery (R and Python)
- https://gallery.azureml.net/Experiment/e616740e68c647ba9bbefa663d037df5
- https://gallery.azureml.net/Experiment/c8c8fe15c4ee470685cc91d5e19c77dc
- On Github (R and Python)
- https://github.com/Quantia-Analytics/Contana-Data-Science-Example-R
- https://github.com/Quantia-Analytics/Cortana-Data-Science-Example-Python

Demo: Regression Example

Metrics for Classification



Azure ML Web Services Data Flow



Demo: Create Web Service

Building machine learning models

- Define business problem
- Understand data relationships
- Prepare data
- Construct models
- Evaluate models
- Improve models
- (Cross) validate model
- Publish model

Why Azure ML?

- Quickly deploy production solutions as web services
- Models run in a highly scalable secure cloud environment
- Powerful, efficient built-in algorithms
- Extensible with, SQL, Python, and R
- Integrated in Cortana stack

Microsoft