1. Main page: http://cortanaanalytics.com
2. Before you begin this module, you should have:
    1. Basic Math and Stats skills
    2. Business and Domain Awareness
    3. A general Computing Background
    4. A general Azure Awareness
    5. An understanding of and interpretation of the problem Statement for a Customer Scenario

Cortana Analytics Scenario - Churn
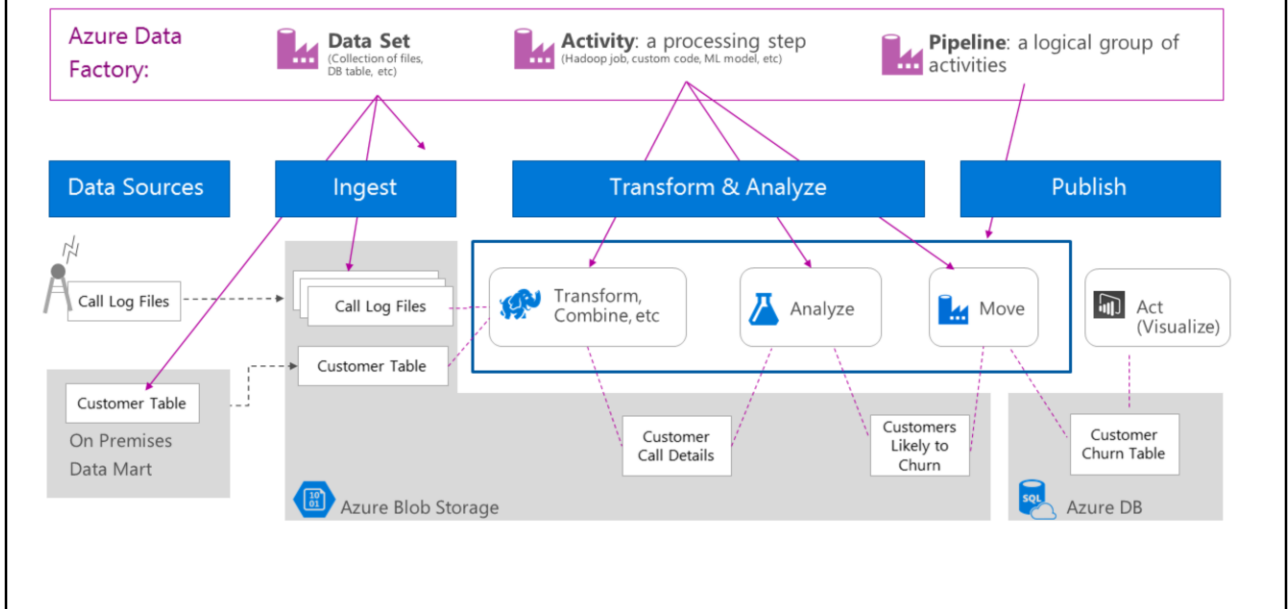
1. Full explanation of this example: https://azure.microsoft.com/en-us/blog/getting-started-with-azure-data-factory-and-azure-machine-learning-4/

Cortana Analytics

Cortana	Cortana
Power BI	Power BI
	Azure Stream Analytics
	Azure HDInsight
Azure Machine Learning	
	Azure SQL DB, Data Warehouse, DocumentDB
	Azure Data Lake
	Azure Event Hubs
	Azure Data Catalog
	Azure Data Factory
	Microsoft Azure

1. Platform: Microsoft Azure – http://microsoftazure.com
2. Storage: https://azure.microsoft.com/en-us/documentation/services/storage/ (Blobs)
3. Azure Data Lake: http://azure.microsoft.com/en-us/campaigns/data-lake/ (from Bing)
4. Azure Data Factory: http://azure.microsoft.com/en-us/services/data-factory/ (Orchestration and Movement)
5. Azure Event Hubs: http://azure.microsoft.com/en-us/services/event-hubs/ (Ingest to streaming (IoT))
6. Azure Data Catalog: http://azure.microsoft.com/en-us/services/data-catalog (Social networking for data)
7. Azure DocumentDB: https://azure.microsoft.com/en-us/services/documentdb/?WT.srch=1&WT.mc_ID=SEM_JQ3fO8dU
8. Azure SQL Data Warehouse: http://azure.microsoft.com/en-us/services/sql-data-warehouse/
9. Azure Stream Analytics: http://azure.microsoft.com/en-us/services/stream-analytics/ (Consume from EH, Output to Alerting (Telco Example))
10. Azure HDInsight: http://azure.microsoft.com/en-us/services/hdinsight/ (Platform on a Platform)
11. Azure Machine Learning: http://azure.microsoft.com/en-us/services/machine-learning/ (Prediction and Prescription - to API)
12. Revolution R Enterprise: http://www.revolutionanalytics.com/revolution-r-enterprise (ML)
13. Power BI: https://powerbi.microsoft.com/
14. Cortana: http://blogs.windows.com/buildingapps/2014/09/23/cortana-integration-and-speech-recognition-new-code-samples/ and https://blogs.windows.com/buildingapps/2015/08/25/using-cortana-to-interact-with-your-customers-10-by-10/

Learning objectives

1. Understand the role of Azure Machine Learning in the Azure data analysis workflow
2. Understand feature selection options in Azure ML
3. Use Azure ML to identify high-value features

1. When you finish this Module, you will be able to:
   1. Understand the role of Azure Machine Learning in the Azure data analysis workflow
   2. Understand feature selection options in Azure ML
   3. Use Azure ML to identify high-value features

Azure Machine Learning

- Import and clean data
- Build, compare, and share models
- Deploy as an API
- Use APIs from the market

1. Home: https://azure.microsoft.com/en-us/services/machine-learning/
2. Gallery of Experiments: https://gallery.cortanaanalytics.com/experiments
3. Currently-available APIs: https://gallery.cortanaanalytics.com/machineLearningAPIs

1. Home: https://azure.microsoft.com/en-us/services/machine-learning/
2. Gallery of Experiments: https://gallery.cortanaanalytics.com/experiments
3. Currently-available APIs: https://gallery.cortanaanalytics.com/machineLearningAPIs

se-Cases

Sales and marketing	Finance and risk	Customer and channel	Operations and workforce
Sales forecasting	Fraud detection	User segmentation	Agent allocat
Demand forecasting	Credit risk management	Personalized offers	Warehouse e
Sales lead scoring		Product recommendation	Smart buildin
Marketing mix optimization			Predictive maintenance
			Supply chain optimization

1. Collections: http://gallery.cortanaanalytics.com/collections
2. Real-time recommendation: https://gallery.azureml.net/MachineLearningAPI/3574432384684cac9cc766e57729ea4c
3. Customer churn forecasting: https://gallery.azureml.net/MachineLearningAPI/7d86b89faf2e4cbcabb84a02179da99e
4. Fraud detection: https://gallery.cortanaanalytics.com/Collection/Online-Fraud-Detection-Template-1
5. Predictive maintenance: https://gallery.cortanaanalytics.com/Collection/Predictive-Maintenance-Template-3
6. Perceptual Intelligence - See - https://gallery.azureml.net/MachineLearningAPI/b0b2598aa46c4f44a08af8891e415cc7 and https://gallery.azureml.net/MachineLearningAPI/02ce55bbc0ab4fea9422fe019995c02f and Hear: https://gallery.azureml.net/MachineLearningAPI/89d229231a72471ebf7280fb5bd3e18c abd Read: https://gallery.azureml.net/MachineLearningAPI/6948e0a54fe44e6fb70cbcc143b31298
7. Personal Assistance - Learning, human interaction, proactive - https://gallery.cortanaanalytics.com/browse/?categories=["Collection"]
8. Example Video: https://blogs.microsoft.com/business-matters/2015/07/13/dartmouth-hitchcock-ushers-in-a-new-age-of-proactive-personalized-healthcare-using-cortana-analytics-suite/
9. Example of HowOld.net - https://how-old.net/#
10. Mechanics: http://blogs.technet.com/b/machinelearning/archive/2015/05/04/fun-with-ml-stream-analytics-and-powerbi-observing-virality-in-real-time.aspx?

Types of Machine Learning Problems:

Classification: Assign a category to each item (Chinese | French | Indian | Italian | Japanese restaurant).

Regression: Predict a real value for each item (stock/currency value, temperature).

Ranking: Order items according to some criterion (web search results relevant to a user query).

Clustering: Partition items into homogeneous groups (clustering twitter posts by topic).

Dimensionality reduction: Transform an initial representation of items into a lower-dimensional representation while preserving some properties (preprocessing of digital images).

- Provided APIs: http://gallery.cortanaanalytics.com/machineLearningAPIs
- http://openclassroom.stanford.edu/MainFolder/courses/MachineLearning/exercises/ex8materials/ex8b_10.png
- http://spectrum.ieee.org/img/RecommendNEWf1-1348253703748.jpg
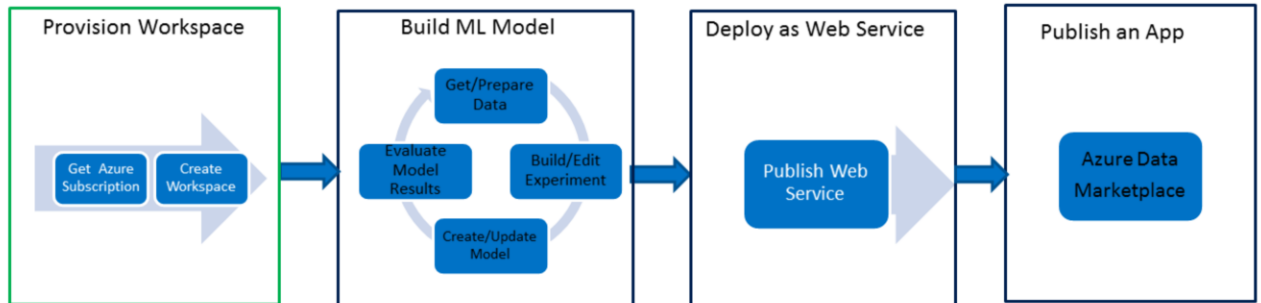
1. Simple overview of this process: https://azure.microsoft.com/en-us/documentation/articles/machine-learning-create-experiment/

1. Process to learn: https://azure.microsoft.com/en-us/documentation/learning-paths/machine-learning-self-guided-predictive-analytics-training/
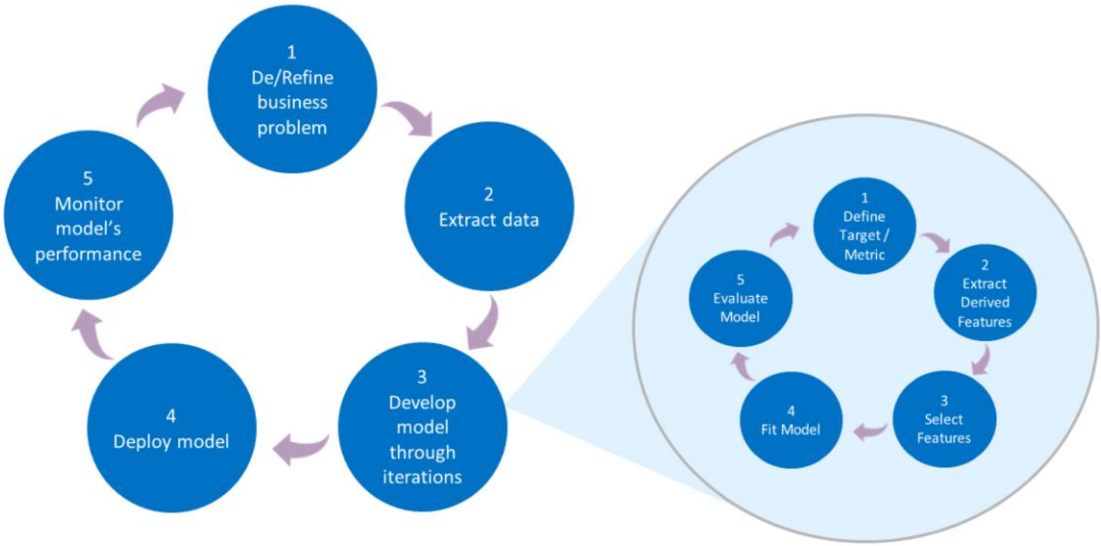
Azure Machine Learning
Data Ingress

- Read data from:
  - Web URL
  - Hive query
  - Azure SQL database
  - Azure table
  - Azure blob storage

- Load a collection of images from blob storage, for use in image classification tasks

Enter Data
1

Reader
1

1. https://azure.microsoft.com/en-us/documentation/articles/machine-learning-import-data/
2. https://azure.microsoft.com/en-us/documentation/articles/machine-learning-data-science-preprocess-steps/
3. https://msdn.microsoft.com/en-us/library/azure/dn905948.aspx
4. https://msdn.microsoft.com/library/azure/dn905997.aspx
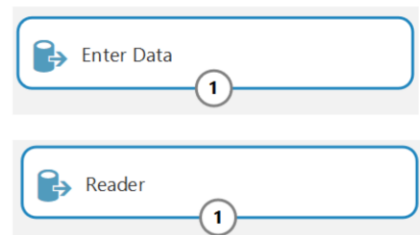
Azure Machine Learning
Data Egress

- Write data out to:
  - An Azure table
  - Azure blob storage
  - Azure SQL database
  - Hive query

Writer

1. https://msdn.microsoft.com/en-us/library/azure/dn905984.aspx
2. http://blogs.technet.com/b/machinelearning/archive/2014/11/25/azureml-web-service-parameters.aspx
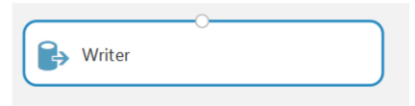
Azure Machine Learning
Feature Selection & Extraction

- Select predictive features
- Create robust new features

1. https://azure.microsoft.com/en-us/documentation/articles/machine-learning-algorithm-choice/
2. https://msdn.microsoft.com/library/azure/818b356b-045c-412b-aa12-94a1d2dad90f
3. https://msdn.microsoft.com/library/azure/8be18eb5-ddd8-4d12-8573-7ae10d5f72fb
4. https://msdn.microsoft.com/library/azure/dcaab0b2-59ca-4bec-bb66-79fd23540080
5. https://msdn.microsoft.com/library/azure/2e010ee4-714e-44e9-933e-62d8c41818a9

Azure Machine Learning

ML/Statistics Algorithms

- Regression
- Classification
- Clustering
- Text Analysis
- Image Analysis
- Model validation and comparison

But first: let's select features

1. Model evaluation: https://azure.microsoft.com/en-us/documentation/articles/machine-learning-evaluate-model-performance/

Why Use Feature Selection?

- Reduce model complexity
- Reduce overfitting
- Speed up data processing and model building

1.   An Introduction to Feature Selection: http://machinelearningmastery.com/an-introduction-to-feature-selection/

Feature Selection in Azure ML

- Azure ML
  - Filter-based
  - Permutation Feature Importance
  - Fisher Linear Discriminant Analysis
  - Principal Components Analysis (for dimensionality reduction)
- R (caret, Boruta, etc)
- Python (scikit-learn, etc)

1. Filter-based: https://msdn.microsoft.com/library/azure/818b356b-045c-412b-aa12-94a1d2dad90f
2. Permutation Importance: https://msdn.microsoft.com/library/azure/2e010ee4-714e-44e9-933e-62d8c41818a9
3. Fisher LDA: https://msdn.microsoft.com/library/azure/dcaab0b2-59ca-4bec-bb66-79fd23540080
4. PCA: https://msdn.microsoft.com/library/azure/8be18eb5-ddd8-4d12-8573-7ae10d5f72fb
5. Comparsion of LDA and PCA: http://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_vs_lda.html
6. caret: http://topepo.github.io/caret/varimp.html
7. Boruta: http://www.r-bloggers.com/feature-selection-all-relevant-selection-with-the-boruta-package/
8. scikit-learn overview: https://www.quora.com/How-do-I-perform-feature-selection

1. Labs are real-time for this module – follow along with the instructor in the following slides

Start Azure ML Studio

- Go to: http://studio.azureml.net/
- Click **Get started** ⊕
- If you have an Azure account, sign in; otherwise click Guest Access
- In the bottom left corner, click **➕ NEW**
- Click "Blank Experiment"

Our Dataset

- Telecom data for 4,667 fictitious customers
  - Account tenure
  - Number of calls by time of day (day, evening, night)
  - Length of calls by time of day
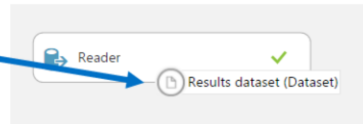 - Number of customer service interactions

1. https://aka.ms/AzureMLChurnTutorial

Read the Data into Azure ML

Data URL:
https://telecomchurndemo.blob.core.windows.net/churnfeatures/sourcedata.csv

1. Paste from link here: https://telecomchurndemo.blob.core.windows.net/churnfeatures/sourcedata.csv
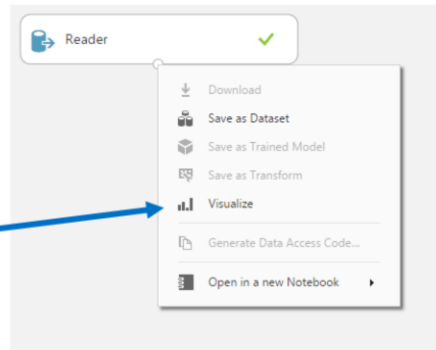
Examine the Imported Data

Hover over the output port

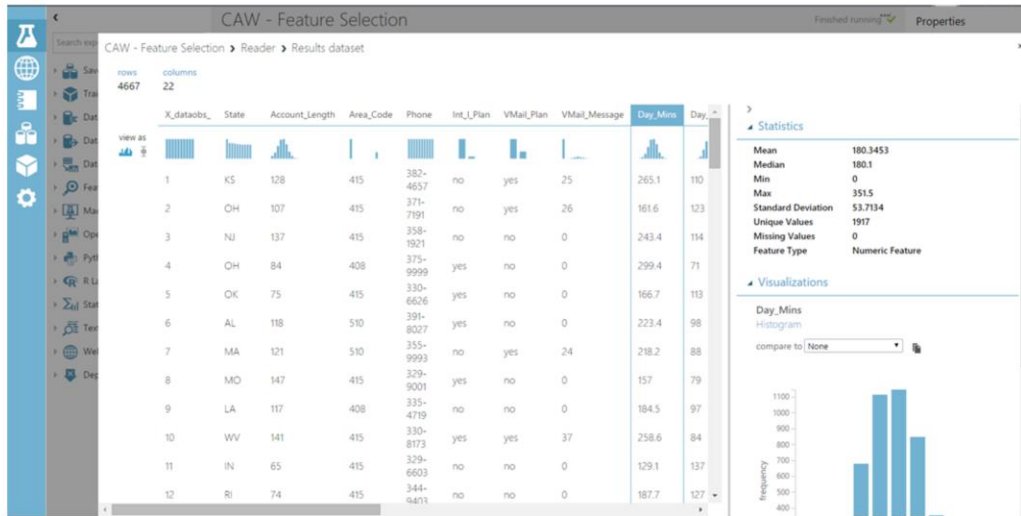Click to see options

Click "Visualize"
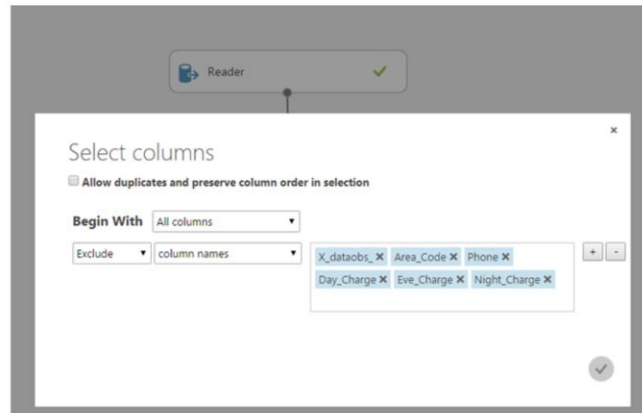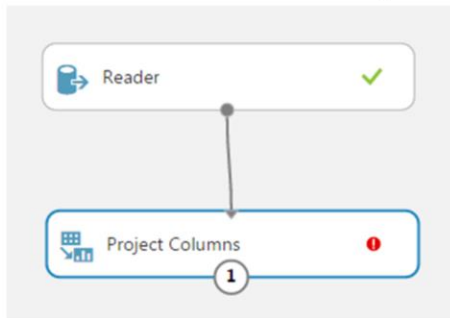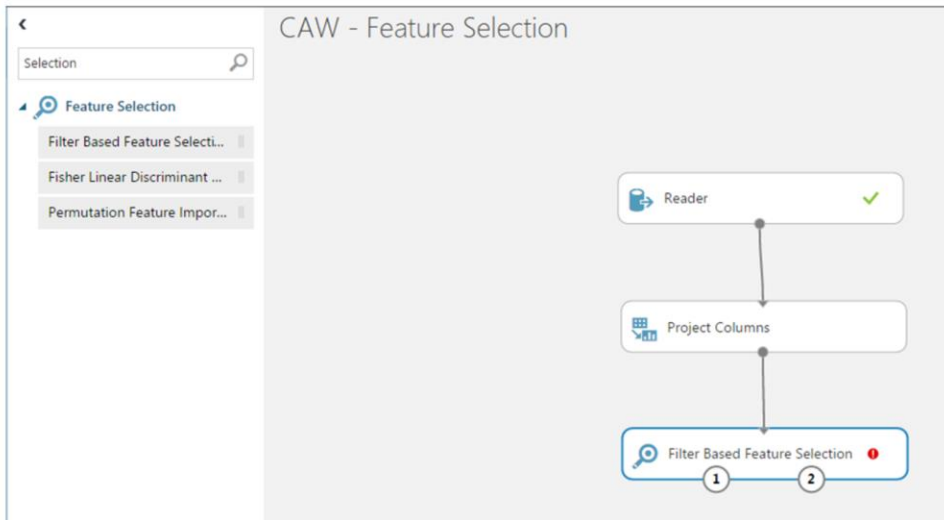
Examine the Imported Data

Drop Columns

Use the Project Columns module to drop
- Non-predictors (keys, IDs, row numbers, etc)
- Highly-correlated variables (depending on your feature selection method)

Reader ✓

Project Columns ❗ (1)

Reader ✓

Select columns
☐ Allow duplicates and preserve column order in selection

Begin With All columns ▼

Exclude ▼ column names ▼   X_dataobs_ ✕ Area_Code ✕ Phone ✕
Day_Charge ✕ Eve_Charge ✕ Night_Charge ✕   [+] [-]

- Other Data Transformation Options: https://msdn.microsoft.com/en-us/library/azure/dn905863
- R and Python scripts can also be used for transformations, missing-data management, etc.
  - R: https://msdn.microsoft.com/en-us/library/azure/dn905920.aspx
  - Python: https://msdn.microsoft.com/en-us/library/azure/dn927167.aspx

1. Filter-based: https://msdn.microsoft.com/library/azure/818b356b-045c-412b-aa12-94a1d2dad90f
2. Fisher LDA: https://msdn.microsoft.com/library/azure/dcaab0b2-59ca-4bec-bb66-79fd23540080
3. Permutation Importance: https://msdn.microsoft.com/library/azure/2e010ee4-714e-44e9-933e-62d8c41818a9
4. Principal Components Analysis (not shown): https://msdn.microsoft.com/library/azure/8be18eb5-ddd8-4d12-8573-7ae10d5f72fb

Select Features

In draft

Draft saved at 1:54:40 PM

Reader

Project Columns

Filter Based Feature Selection
1 2

Properties ›

▲ Filter Based Feature Selection

Feature scoring method

Pearson Correlation ▼

☑ Operate on feature co... ≡
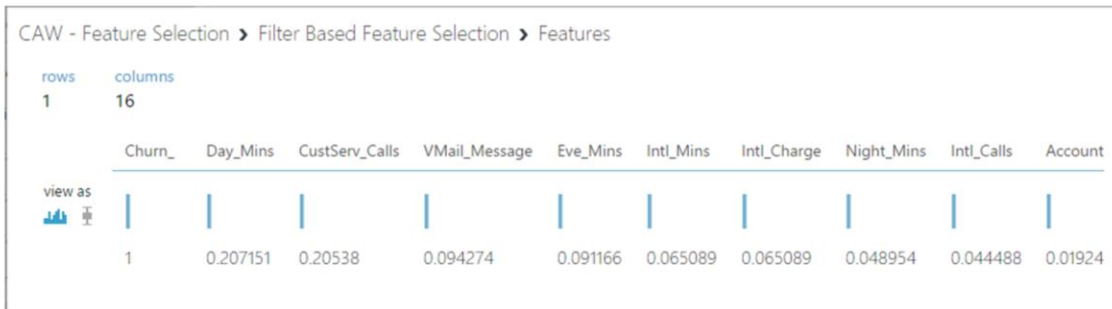
Target column

Selected columns:
Column names: Churn_

Launch column selector

Number of desired features ≡

5

1. We'll continue this in the next Module – make sure you save your experiment.

View Filtered Dataset

- Output Port #1

- The original dataset, now with only the top five most correlated variables

- Ready for model-building

CAW - Feature Selection › Filter Based Feature Selection › Filtered dataset

rows 4667   columns 6

Churn_	Day_Mins	CustServ_Calls	VMail_Message	Eve_Mins	Intl_Mins
false	265.1	1	25	197.4	10
false	161.6	1	26	195.5	13.7
false	243.4	0	0	121.2	12.2
false	299.4	2	0	61.9	6.6
false	166.7	3	0	148.3	10.1
false	223.4	0	0	220.6	6.3
false	218.2	3	24	348.5	7.5
false	157	0	0	103.1	7.1
false	184.5	1	0	351.6	8.7
false	258.6	0	37	222	11.2
true	129.1	4	0	228.5	12.7
false	187.7	0	0	163.4	9.1

Next: build a churn model, or save the output data for later use

1. Use this for Q/A time