**Analytics with Azure Machine Learning (Microsoft Data Camp)**

**Hands On Lab Document Url:** <https://aka.ms/azuremlhol>

This is an interactive workshop. To get the most from your session, please don't forget to bring your laptop and power supply. In addition:

* Activate your Azure account and bring your Microsoft account credentials. Don't have a Microsoft account? [Sign up now](https://signup.live.com/newuser.aspx?cr_cc=200744396&wt.mc_id=usdx_evan_events_reg_dev_0_iottour_0_0)
* Activate a [free 30-day trial Microsoft Azure account](https://azureinfo.microsoft.com/us-freetrial.html?cr_cc=200744395&wt.mc_id=usdx_evan_events_reg_dev_0_iottour_0_0)
* If you subscribe to MSDN, activate your free Azure [MSDN subscriber benefits](https://azure.microsoft.com/en-us/pricing/member-offers/msdn-benefits/?cr_cc=200744394&wt.mc_id=usdx_evan_events_reg_dev_0_iottour_0_0)
* Windows 10 must be [installed](https://www.microsoft.com/en-us/windows/windows-10-upgrade?cr_cc=200744397&wt.mc_id=usdx_evan_events_reg_dev_0_iottour_0_0)on your laptop before you arrive – we won’t be able to assist with the install at the event.
* Visual Studio 2015 Community, Pro, or Enterprise must be [installed](https://www.visualstudio.com/downloads/download-visual-studio-vs?cr_cc=200744398&wt.mc_id=usdx_evan_events_reg_dev_0_iottour_0_0); we won’t be able to assist with the install at the event.
* Office 2013 or later. We won't be able to assist with the install at the event.
* Create Azure ML workspace for free by [signing up here](https://studio.azureml.net/).

**Business Problem: Predict Income >50K (Level 100) [Estimated: 1 hour]**

This experiment demonstrates how we can build a binary classification model to predict income levels of adult individuals.

Url: <https://gallery.cortanaintelligence.com/Experiment/Sample-5-Train-Test-Evaluate-for-Binary-Classification-Adult-Dataset-6>

You can copy the final experiment into your account by clicking on **Open in Studio**. Then, you can use the completed solution as a reference into building your experiment.

**Business Problem: Flight delay prediction (Level 300) [Estimated: 3 hours]**

In this experiment, we predict whether scheduled passenger flight is delayed or not using a Binary-classifier.

The problem statement is describing how to build this experiment from scratch. Follow these steps to build this binary classifier experiment using Azure ML Studio.

Url: <https://gallery.cortanaintelligence.com/Experiment/Binary-Classification-Flight-delay-prediction-3>

You can copy the final experiment into your account by clicking on **Open in Studio**. Then, you can use the completed solution as a reference into building your experiment.

Understanding Flight Delays Data Set

A flight is counted as "on time" if it operated less than **15 minutes** later the scheduled time shown in the carriers' Computerized Reservations Systems (CRS). Arrival performance is based on arrival at the gate. Departure performance is based on departure from the gate. (Ref. [link](http://www.rita.dot.gov/bts/help/aviation/index.html))

Some important notes about the dataset & AzureML experiment:

* 1. Departure or Arrival Date is in HHMM format. So to round any the closest hour, we will divide it by 100 and then round the result (using floor() math operation) to get the integer number of a given number. For example: 1135. We will apply divide op.: (1135/100) = 11.35, then round op. using Floor(11.35) = 11.
  2. Categorical Variables ([Wiki](https://en.wikipedia.org/wiki/Categorical_variable)): are those that represent a fixed number of possible values, rather than a continuous number. In this experiment, the columns Carrier, OriginAirportID, and DestAirportID represent categorical attributes.
  3. Binning or grouping data (some time called quantization) is an important tool in preparing numeric data for machine learning. In this experiment, we are using group data into bins using Quantiles binning mode. Quantiles is also known as equal hight binning. This applies to all numeric values in selected flight delays dataset.
  4. In this experiment, we used sweep parameters (Tune Model Hyperparameter): *Performs a parameter sweep on a model to determine the optimum parameter settings.*
  5. In this experiment, we used **AUC** as a metric for measuring performance for classification. Read more about other metrics ([here](https://msdn.microsoft.com/library/azure/038d91b6-c2f2-42a1-9215-1f2c20ed1b40)).
  6. **Conclusion**: Since the experiment is using two different algorithms: **Two-Class Boosted Decision Tree & Two-Class Logistic Regression**, you will notice that boosted decision has AUC of **0.64** on the validation set which is better than logistic regression model with AUC of **0.628.**

**The evaluation here:** [**https://aka.ms/mddatacamp**](https://aka.ms/mddatacamp)