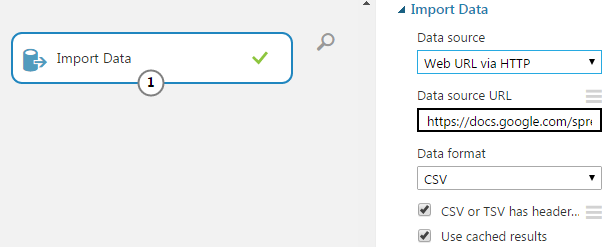
Companion Document for Week 5

**Detailed steps for the Machine Learning in Azure segments**

1. Import the data:

Drag the **Import Data** box to the experiment. In the **Data source** line, select **Web URL vis HTTP**. In the **Data source URL** line, enter the following url: https://docs.google.com/spreadsheets/d/1SNfa8yWRDQ9aU5cVQMdUdx8D64gYq06iTh0D6QlHq44/pub?gid=104186993&single=true&output=csv. In the **Data format** line, select **CSV**. Check the box in front of **CSV or TSV has header row** and **Use cached results**. Click **Run**.



1. Select necessary data to be use

Drag the **Select Columns in Dataset** box and connect it to the **Import Data** box. Click **Launch column selector** option. Select the columns with name DEFAULT, LIMIT\_BAL, AGE, MARITAL, and PAY\_1.



1. Split the dataset into training set and test set

Drag the **Split Data** box and connect it to the **Select Columns in Dataset** box. Set the value of **Fraction of rows in the first output dataset** to be 0.7; this means that 70% of the data will be used as the training set and the remaining 30% as the test set.



1. Drag the **Two-class Decision Forest** box. Set the value of **Number of decision trees** to be 1. Set the **Maximum depth of the decision trees** to be 4.



1. Drag the **Train Model** box. Connect its left inlet with the **Two-class Decision Forest** box. Connect its right inlet to outlet 1 of the **Split Data** box.



1. Drag the **Score Model** box. Connect its left inlet with the **Train Model** box. Connect its right inlet to outlet 2 of the **Split Data** box.

Drag the Evaluate Model. Connect its left inlet with the **Score Model** box.



1. You can find the trained model from the outlet of the **Train Model** box. You can evaluate the quality of the model by looking at the output of the **Evaluate Model** box.