OVERVIEW

This lab covers the steps for creating a click-through rate (CTR) prediction pipeline. You will work with the [Criteo Labs](http://labs.criteo.com/) dataset that was used for a recent [Kaggle competition](https://www.kaggle.com/c/criteo-display-ad-challenge" \t "[object Object]).

The lab is due July 25, 2015 at 07:00 UTC. There is a three day grace period for late submissions until July 28, 2015 at 07:00 UTC. Submissions after that time will lose 20 points.

NOTEBOOK DOWNLOAD

The notebook can be found on GitHub at the following links:

* [Lab4 CTR ipynb for VM download](https://raw.githubusercontent.com/spark-mooc/mooc-setup/master/ML_lab4_ctr_student.ipynb)
* [Lab4 CTR dbc for Databricks download](https://github.com/spark-mooc/mooc-setup-dbc/blob/master/ML_lab4_ctr_student.dbc?raw=true)

When working on these notebooks, out of respect for current (and future) students taking this course, please**do not store your solutions in publicly visible repositories** such as GitHub.  Similarly, please**do not post code snippets on Piazza**.

VIEW NOTEBOOK

GitHub automatically renders IPython notebooks.  To view the notebook follow this GitHub link: [Lab4 CTR web preview](https://github.com/spark-mooc/mooc-setup/blob/master/ML_lab4_ctr_student.ipynb).  Note: this link should not be used to download the notebook.  Use the raw link in the Notebook Download section when downloading.

DETAILED INSTRUCTIONS FOR RUNNING NOTEBOOKS IN VM

* Download the appropriate raw Lab4 file for your environment:
  + [Lab4 CTR ipynb for VM download](https://raw.githubusercontent.com/spark-mooc/mooc-setup/master/ML_lab4_ctr_student.ipynb) - an IPython notebook.  **Make sure that the file extension is .ipynb**.  If the download adds an extension (e.g. ".txt"), rename the file so that the extension is just .ipynb.
  + [Lab4 CTR dbc for Databricks download](https://github.com/spark-mooc/mooc-setup-dbc/blob/master/ML_lab4_ctr_student.dbc?raw=true) - a Databricks notebook.  The extension should be **.dbc**, for importing into the Databricks account that you got via email if you were randomly selected for the Databricks option.
* Upload or import the raw Lab4 file.  This process was explained during "Setting up the Course Software Environment" in the Week 0 courseware.
* See our [Databricks FAQ](https://courses.edx.org/wiki/BerkeleyX.CS190.1x.1T2015/databricks-faq/" \t "[object Object]) for details on importing to Databricks.  If you are using the VM:
  1. Start the VM - To start the VM, from a DOS prompt (Windows) or Terminal (Mac/Linux), issue the command "vagrant up", while in the custom directory created for this course (you should have created this directory as part of the Week 0 segment on "Downloading and Installing the VM Image").
  2. Once the VM is running, access the Jupyter web UI for running IPython notebooks by navigating your web browser to "<http://localhost:8001/>" (or "<http://127.0.0.1:8001/>").
  3. Shut down any notebooks you have running, as only **ONE** notebook should run at a time.  Running notebooks have a green icon to the left of the notebook name and green text to the right of the screen that says "Running".  Shutdown running notebooks by clicking the checkbox next to the notebook and then clicking the orange "Shutdown" button.
* For the **Lab4** notebook, please follow the instructions in the notebook and replace <FILL IN> sections with your solutions, and submit to the autograder, following the same [guidelines](https://courses.edx.org/courses/BerkeleyX/CS190.1x/1T2015/wiki/BerkeleyX.CS190.1x.1T2015/autograder/) as in previous weeks.
* When you have submitted successfully, if you are using the VM you can shut it down by issuing the command "vagrant halt".