In this assignment you will practice using Alteryx to create a workflow to prepare loan data for analysis.

**Instructions**

1. **(1 point)** Create a step to read in the loan\_application.csv file.
2. **(2 points)** Convert the data types of the following columns to the data type listed after the hyphen:
   * Columns starting with “CNT” – INT32
   * Columns starting with “AMT” – Float
3. **(1 point)** Remove rows with NAs or NULL in the AMT\_INCOME\_TOTAL
4. **(1 point)** Filter data to keep only the rows where AMT\_CREDIT is at least 300,000
5. **(2 points)** Join this data with credit\_bureau dataset (read it first) using the common field “SK\_ID\_CURR”
6. **(2 points)** Keep only the following columns from the joint dataset: SK\_ID\_CURR, TARGET, AMT\_INCOME\_TOTAL, and AMT\_CREDIT\_SUM. Make sure you set their data types correctly. Also remove any duplicate observations from this dataset, using the ***unique*** tool.
7. **(1 point)** Finally, create a step to save the datafile to your desktop as netid\_HW4\_data.csv. For example, if your email address is jdoe@illinois.edu, then save the data as jdoe\_HW4\_data.csv.

**Submission\***

1. You are given two data files: credit\_bureau.csv and loan\_application.csv.
2. Use Alteryx to complete the required ETL steps.
3. Save your Alteryx workflow file (.yxmd format) as netid\_HW4.yxmd
4. Compress the jdoe\_HW4.yxmd file as a .zip or .rar file.
5. Submit the .zip or .rar file on Compass2g before the deadline.