**CSCI-620: Introduction to Big Data**

**Assignment 3: Data Transformation Techniques**

**Due** on Friday, 2/23/2018 @ 11:00pm

**The Goal:** Students must show mastery level of applying data transformation techniques.

**Requirements:**

1. Search and download a dataset of any data-type and topic of your choice.
2. Your dataset **MUST** contain at least 100,000 observations and 20 attributes.
3. Apply whatever data transformation techniques you believe appropriate to your dataset.

**Deliverables**:

You have to submit the following:

50 points

40 points

10 points

Total:

100 points

* Assignment **report** that explains in details how did you explore your dataset. You should include in your report the utilized R/Rattle commands, screenshots of the results, and sufficient discussions to emphasize the most important features of your dataset!!
* **One text file** that contains all executed commands on R and Rattle. The file must have two main titles to clearly distinguish R-commands from Rattle-commands.
* The **dataset**.

**Individual work:** This is an individual work assignment. Although cooperation between students is highly encouraged and desirable, every student must submit her/his **original work!!**

\*\*\*\*\*\* GOOD LUCK \*\*\*\*\*\*