Project Report

E(xtract):

* We gathered our data from Kaggle for March Madness data
* Our data had season results and tournament results data from 2003 – 2015(16 \*tourney)
* The data was formatted in csv
* We used data that was an expanded box score and not the basic score data set it included

T(ransform):

* For our transformations we changed the names of the columns to be understandable
* We also combined some columns to become either percentages and a total
* We dropped columns we did not see as valuable moving forward in information we wanted
* We changed the names of the winning and losing team rows to know the team names
* Then we saved our transformation data in new csv’s

L(oad):

* To load the data we created a database called MarchMadness2.0
* Uploaded our transformed results and tourney data into tables in Postgres
* Unioned the tables to combine season and tourney data in one table to show a teams whole season record
* We created two final tables to contain 2015 data for both Duke and Wisconsin season records since both team reached the final