Cole Barcia

Jhuelian Gasmena

Jean Medina

Joel Woodward

ETL Project

NFL Combine Data of 99-2015

Extraction: Our extraction process began with searching through Kaggle for relatively clear csv files concerning NFL scouting combines. Our goal within this section was to find accurate and relative datasets amongst those who participate in the various drills. We also took into considerations certain body measurements to see if there would be any correlation with regards to ability to perform. As a group, we wanted to make sure we could reflect with significant confidence a relationship between given NFL draft grade and where they actually ended up being drafted.

Transformation: The transformation aspect of our project started with reviewing the csv’s we grabbed from Kaggle and cleaning through the dataset. Our method of paring the dataset started with dropping any unnecessary columns. For example, we had separate first and last name columns accompanied with a full name column. Another key piece of our transformation process was to clear away any Nan values with regards to integer inputs. We didn’t want Nan’s to interfere with the graphing visualizations.

Loading: The loading phase in our project surmounted to building data visuals so that we may draw significant relationships from what we had plotted. We chose to compare the times of combine participants bench press results to their total body weight. We measured the linear correlation between the two measures and found a moderate correlation. Correlation is measured from (-1) to 1 with 0 representing no correlation between the variables. Our results showed a correlation score of 0.63. We chose to compare the times of combine participants' NFL Grade to their Draft Pick number. We measured the linear correlation between the two measures and found a moderate negative correlation (High NFL grades are correlated to an earlier pick). Correlation is measured from (-1) to 1 with 0 representing no correlation between the variables. Our results showed a correlation score of (-0.62). We chose to compare the times of combine participants 40 Yard Dash to their total body weight. We measured the linear correlation between the two measures and found a strong correlation. Correlation is measured from (-1) to 1 with 0 representing no correlation between the variables. Our results showed a correlation score of 0.88 which was the highest of any two measures recorded at the NFL combine.