I am a visual person and drawing out this post is helpful. Rather than copy ESPN’s model for sport categorization, I dreamed up the categorization below.

Sports

College

High School

Middle School

Non-Competitive

City

Intermural

States

Regional

Professional

Amateur

Cities

Each category above delineates specific text in an effort to enhance the quality of data. The first categories are between Amateur and Professional and are child tags of sports. Money is the biggest delineating factor between these two categories. While both categories encompass vast amounts of money, a professional sport directs its money to individuals and for-profit institutions. The predictive analytical capabilities around this text include mainstream techniques. Furthermore, the regional, states, and cities, tags focus the query to specific geographical locations, which enriches the sentiment analysis as the generation of tags become younger. An example of the predictive capability of this categorization is an investment strategy. One could analyze specific sports by city and delineate the financial growth and predict which specific teams to buy ‘shares’ from in the form of tickets to re-sell.

Amateur and the subsequent child tags capture a completely different concept as well as terms. The primary concept behind amateur is non-financial in its nature. One could mine specific team keywords in Sports🡪Amateur🡪College to find team updates on an alma mater. The most intense sports recruiting occur at the High School and Middle School level, which has its own domain of keywords. One could predict potential athletic talent based on mining specific texts within this category. Non-Competitive and the child tags City and Intermural capture the domain of mass participant involvement. Schedules, leagues, costs, and age ranges are subjective to this category but yield considerable information. Local governments could mine text in this domain to anticipate employee staffing needs and field assignments.