**ETL Project Report**

*Jacob Klug, Nico Boyajian, Louis McCullagh*

Extract:  
 Our original datasets came a variety of sources openly available on the internet. Our valuations data came from Forbes and showed us the 2018 valuations of all 30 MLB teams in USD. Our standings data and our attendance data both originated from Baseball References online repository. Each of our three datasets came in the form of .CSV files which was useful for standardized, however, some key values throughout differed from dataset to dataset.

Transform:   
 In order to provide precise standardization throughout our three datasets, we cleaned our team designations in order to have congruent three letter acronyms. In addition, we deleted frivolous data I the form of unnecessary columns. After this initial cleaning, using the Pandas library we loaded each set into a data frame and then proceeded to merge all three on the pivot point of team name.   
Load:   
 In order to perform the final part of our project we spun up a local instance of a SQL database and renamed ETL\_Baseball. Finally after configuring the engine path, we joined the aforementioned data frame to a newly created table within our SQL database.