TourneyMachine Scrape Details

First Process – Scrape Homepage and clean data

1. Scrape homepage and get tournament data
   1. Insert into [stg].[tourneymachine\_mainpage\_data]
2. Separate out dates and clean
   1. Insert cleaned and “new” dates into [data].[tm\_tournament\_dates]
3. Clean tournament data and insert new events into [data].[tm\_mainpage\_data]
   1. All “new” rows have a 1 for “is\_active”

Second Process – Take data from first process, follow to endpoints, and get game data

1. Within the tourneymachine scraping script, there is a file named TMachineExtractor.py
   1. This file will take all “is\_actives” from [data].[tm\_mainpage\_data] WHERE is\_active = 1 and end\_date < GETDATE()
2. The script follows the endpoints and inserts game data into [clubstats].[stg].[tourneymachine\_data]
3. Data needs to be ‘cleaned’ from there…