Vince Laird

SPORTS BETTING LINE MOVEMENT: DATA WRANGLING

DATA SCIENCE - CAREER TRACK

Data wrangling steps

The first step in wrangling the data for the sports betting line movement project was determining what to wrangle. The four major sports in America are:

MLB baseball (162 game season, 30 teams, 2017 season starts April 2)

NHL hockey (82 game season, 30 teams, season already underway: October 12, 2016 to April 9, 2017)

NFL football (16 game season, 32 teams, 2017 season starts September 7)

NBA basketball (82 game season, 30 teams, season already underway: October 25, 2016 – April 12, 2017)

Based on today's date (March 15, 2017), MLB baseball seems to be the most ideal sport to analyze and start with for this project; NHL/NBA seasons are both winding down and NFL football, the most popular sport, doesn't start for several months.

One advantage of looking at baseball is that there are so many games, there will be a good number of data points to look at every day. The disadvantage of baseball is that the scores tend to be pretty low (average of 8.33 runs per game during the 2013 season) so there may not much line movement prior to each contest, but there should be something worth looking at.

Data: past and present

With baseball in mind, determining where to get the data was the next step. For last year's data, vegasinsider.com provides historical line movements. Below is a screenshot of the url that can be manipulated to see each game's line movements, and the table we are interested in is the one following the heading "VI CONSENSUS LINE MOVEMENTS" - this table can be copied to the clipboard for each game, and pasted into a pandas dataframe to be saved as a .csv file.


www.vegasinsider.com/mlb/odds/las-vegas/line-movement/mariners-@-yankees.cfm/date/4-15-16/

VI CONSENSUS LINE MOVEMENTS

		Money Line		Run Line		Total		1st Half (5 Inn.)		2nd Half	
Date	Time	Fav	Dog	Fav	Dog	Over	Under	Fav	Dog	Fav	Dog
04/14	1:55pm	NYY-147	SEA+137			8 -120	8 +100				
04/14	4:05pm	NYY-145	SEA+135			8 -120	8 +100				
04/14	4:10pm	NYY-145	SEA+135	SEA-165	NYY+145	8.5 +100	8.5 -120				
04/14	4:30pm	NYY-145	SEA+135	SEA-160	NYY+140	8.5 +100	8.5 -120				
04/14	6:20pm	NYY-150	SEA+130	SEA-160	NYY+140	8.5 +100	8.5 -120				
04/14	6:25pm	NYY-145	SEA+135	SEA-160	NYY+140	8.5 +100	8.5 -120				
04/14	7:20pm	NYY-145	SEA+135	SEA-165	NYY+145	8.5 +100	8.5 -120				
04/14	9:00pm	NYY-130	SEA+120	SEA-165	NYY+145	8.5 +100	8.5 -120				
04/14	9:10pm	NYY-130	SEA+120	SEA-190	NYY+165	8.5 +100	8.5 -120				
04/14	9:20pm	NYY-130	SEA+120	SEA-165	NYY+145	8.5 +100	8.5 -120				
04/14	9:25pm	NYY-130	SEA+120	SEA-190	NYY+165	8.5 +100	8.5 -120				
04/14	9:40pm	NYY-130	SEA+120	SEA-185	NYY+165	8.5 +100	8.5 -120				
04/15	2:00am	NYY-132	SEA+122	SEA-185	NYY+165	8.5 +100	8.5 -120				
04/15	3:55am	NYY-132	SEA+122	SEA-185	NYY+165	8 -110	8 -110				
04/15	4:50am	NYY-133	SEA+123	SEA-185	NYY+165	8 -110	8 -110				
04/15	4:55am	NYY-133	SEA+123	SEA-185	NYY+165	8.5 +100	8.5 -120				
04/15	5:05am	NYY-130	SEA+120	SEA-185	NYY+165	8.5 +100	8.5 -120				
04/15	7:15am	NYY-125	SEA+115	SEA-185	NYY+165	8.5 +100	8.5 -120				
04/15	7:20am	NYY-125	SEA+115	SEA-185	NYY+165	8.5 +100	8.5 -120	NYY-130	SEA+110		
04/15	8:30am	NYY-126	SEA+116	SEA-185	NYY+165	8.5 +100	8.5 -120	NYY-130	SEA+110		
04/15	8:40am	NYY-127	SEA+117	SEA-185	NYY+165	8.5 +100	8.5 -120	NYY-130	SEA+110		
04/15	9:10am	NYY-127	SEA+117	SEA-185	NYY+165	8 -110	8 -110	NYY-130	SEA+110		
04/15	9:10am	NYY-127	SEA+117	SEA-185	NYY+165	8 -110	8 -110	NYY-130	SEA+110	SEA-125	NYY-105
04/15	9:20am	NYY-130	SEA+120	SEA-185	NYY+165	8 -110	8 -110	NYY-130	SEA+110	SEA-125	NYY-105

For present data, jsonodds.com provides an API with up-to-date odds. These odds can be stored in a mongoDB in json format - simply calling the API every minute is where we'll start, though frequency of these calls will be reduced once the data is examined and it's determined when we need more frequent calls (so as to run out of space in our database).

A web app has been created and deployed to Heroku to do the data collection; code for this app can be found here:

https://github.com/vincelaird/ds-mlb