Research on Hattrick's Youth Academy Scouting Pools

To gain good young players you can start using the Youth Academy. You can hire (up to three) scouts looking for talents / your future stars. For each scout you can choose in which region and for what kind of player type he should look for. Each region has at least one player pool, people assume that there is a pool for each type of player (goalkeeper, defender, winger, midfielder, forwarder) and region, in which the scouts are looking for talents. You can also search for any type of player - a former research has shown, that the chance of getting a player of a specific type is (slightly) higher when search for this type. Since players got pulled from the pool into the youth squads Hattrick has to refill them regularly.

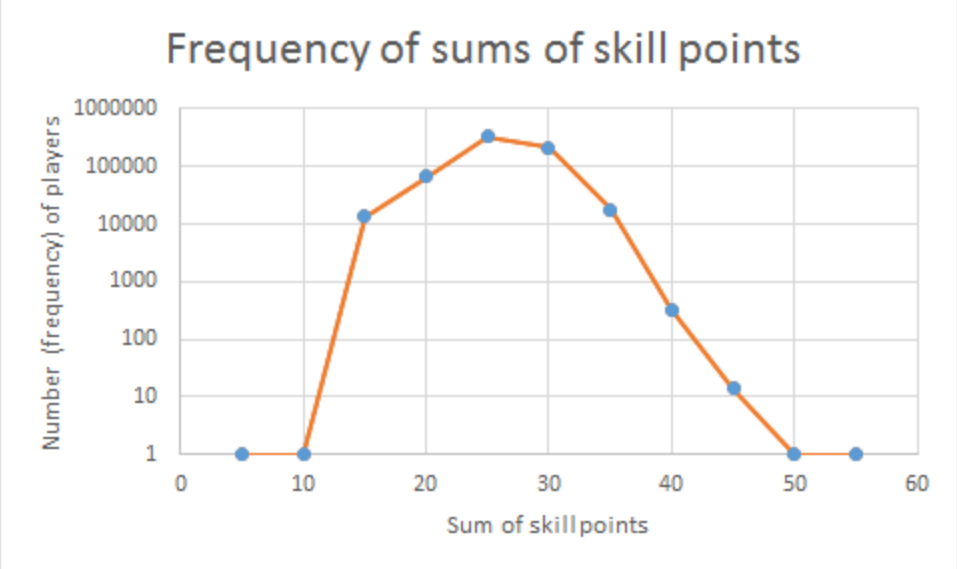
HT-Tools with Dantekavala and hattrick youthclub with Mackshot have analyzed **645,129** youth players to identify trends, distributions and frequencies of talented players. We divided the research in three parts. In the first one, we studied the talent of youth players in general through the last years. In the second part, we examined if the point of time of scouting a youth player has an important impact on the talent and finally, we searched for trends in order to see if pool refills are important on scouting.

We would like to note that this is not a guide that one should follow with closed eyes. Since it is a statistical analysis based on historical data it is likely that if many Hattrick managers change their behavior (due to this research) the results of a repetition of this study would be different.

Frequency of talented Youth Players

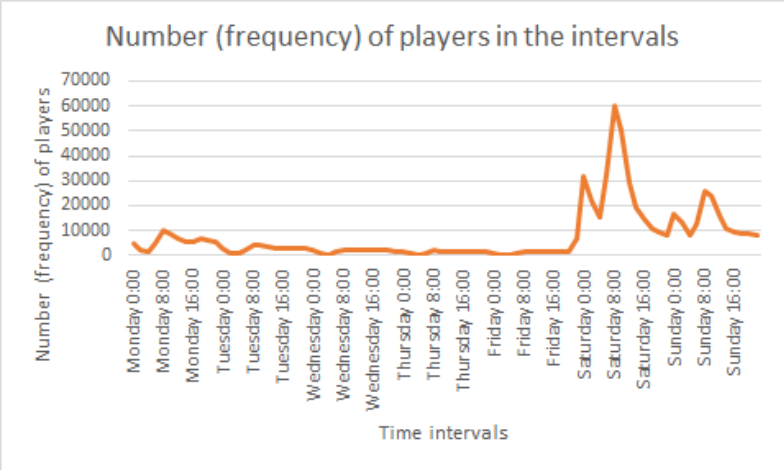
We grouped the players by their sum of skill points (keeper, defending, playmaking, winger, passing, scoring and set pieces)when the player got promoted from youth to senior squad. For example, a player with as sum of 17 skill points will be in the group of 16-20.
The graph fits to a Right-Skewed Normal distribution which seems logical in order to have more bad and average than super talent players in the database. Additionally we have to take into account that not all managers can train all of their youth players to perfection.

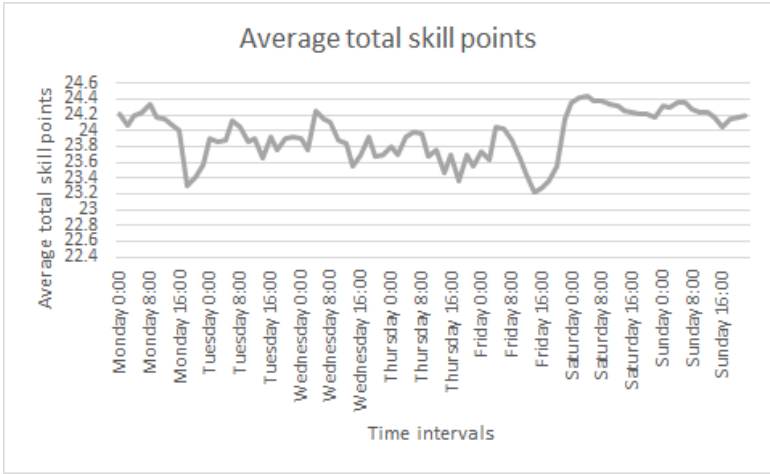
Groups	Frequency
0-5	0
6-10	0
11-15	13478
16-20	68124
21-25	331514
26-30	213722
31-35	17954
36-40	329
41-45	14
46-50	0
51-55	0

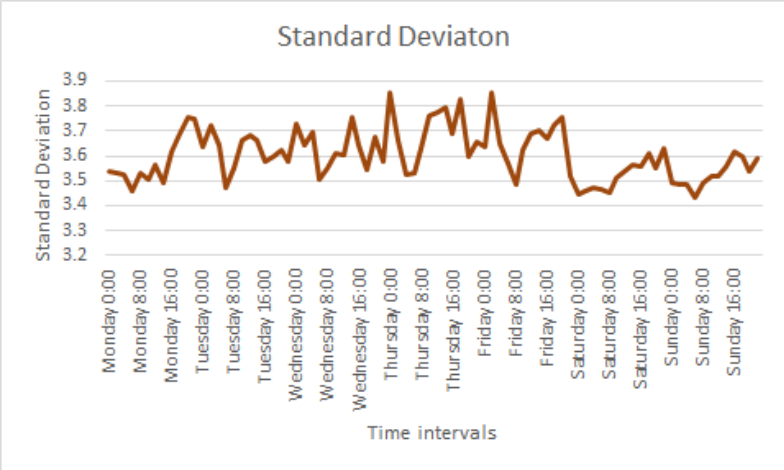


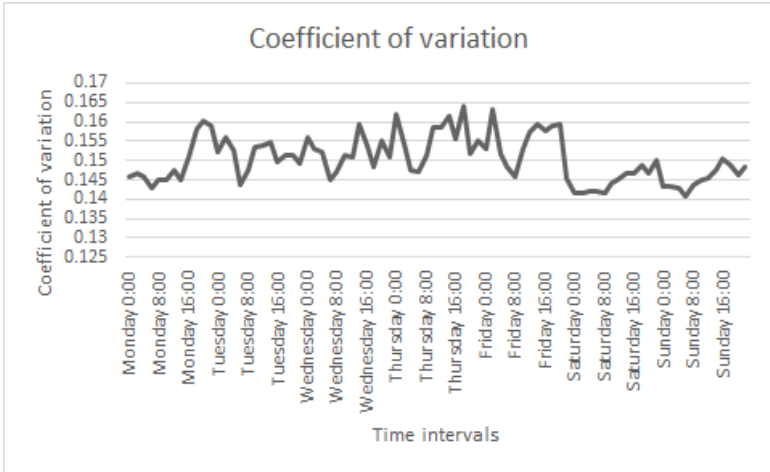
Importance of the point of time for scouting

Moreover, we have studied the importance of the point of time of scouting players / calling the scouts. We split the dataset into groups with an interval of 2 hours grouped by weekday (Monday to Sunday). We calculated the sum, the frequency and finally the mean and standard deviation of each group. One can easily see that managers call their scouts more often on weekends than on weekdays, which is pretty reasonable since senior matches take place on the weekend and thus most users are logged-in anyway. Additionally, managers can call their scouts only once a week and this blockage is reset by Hattrick shortly before the weekend starts. However, the gap in the sum of skill points on weekends is not higher than 0 to 1.0 skill point from weekdays. Examining also the standard deviation and the coefficient of variation it seems that the lowest numbers are at 6:00 to 8:00 every day. This means that during this period, all youth players are relatively similar with regards to their skill points sum. There aren't a lot of bad, average and talent players at this period. In specific, you will only find either a lot of bad or a lot of talented players. Combining this with the graph of average sums of skill points, we observe a peak every day that time.
This means that (based on this dataset and data analysis) you have a slightly higher probability to scout a talented player between 6:00 and 8:00 in the morning.





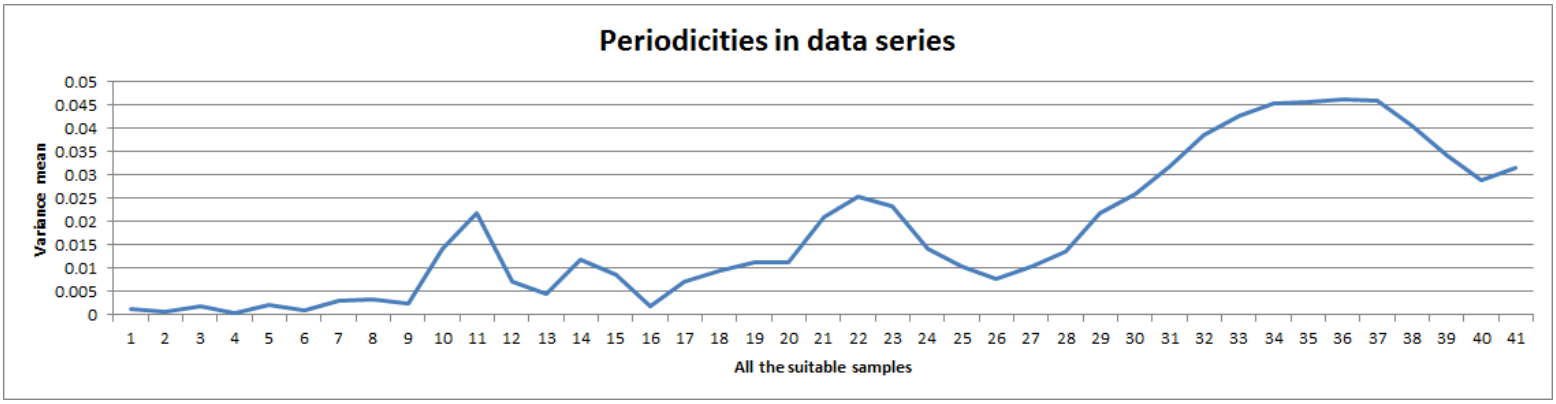




Looking for trends

Using an algorithm for searching for periodicities in data series based on systematic sampling procedures, we examined if there is any important trend of talented youth players after youth player pools have been refilled. The algorithm split the dataset in all possible sublists, calculate the mean and standard deviation in every sublist and check if there is any case that the data seems to be duplicated. If there is a part of the

graph that is repeated, we should have find a trend in the time serie. However, there isn't anything that could make us conclude in a trend for the time intervals, thus we believe that the pool refills are dynamic and they depend from the current available size.



Conclusion and Future Research

In general, there seems to be no certain point of time when all pool get refilled. Thus we form the thesis that Hattrick refills a pool when the filling level has fallen below a certain threshold (this check may be implemented as post-scouting hook or is called, for example, every hour). We want to note, that to confirm or reject this thesis it would be helpful to analyze if the player quality differs for any point of time for a single region or country.