Wigan Squad Analysis

1 Introduction

After being invited to interview I decided to go through the motions of building a framework to analyse a given team's squad and see how their stats differ to "target teams". These deltas between a given squad and their target team's stats were used to identify weaknesses in the squad and as a starting point for targeting replacements.

In Wigan's case I decided that the target of promotion back into the Championship and a finish in the top half of the table, as to avoid a fifth demotion into League One, felt like a lofty but achievable goal. Once the goal had been established I had to decide which stats I would use when comparing Wigan's squad to the target teams. To do this I analysed the relationship between each stat and goals scored (GS) and each stat and goals against (GA). This was done to understand which stats are important for scoring goals and which stats are important for conceding fewer. After some testing I landed on KNN permutation feature importance as the method for extracting the relationship between statistics and GS/GA.

Using the most important stats for scoring and conceding fewer goals I analysed the weaknesses in the current Wigan squad and targeted signings from leagues where club wage structures track similarly to Wigan's to fix said weaknesses.

2 Feature Selection

The get the results for the KNN permutation feature importance I ran the process five times and averaged the feature importance provided on each run, these results can be seen in table 1. These models were run with the target variable being goals scored and goals conceded for offensive and defensive feature importance respectively. These top features shown are the features with the highest scores before there is a significant drop off, rather than a set number of features for each model, the full scores can be seen by running stat_selection.py. It must be noted I used data from the top 5 leagues and not Championship and League One due to ease of access. This will breed some inaccuracies but as this is a proof of concept I felt it acceptable.

The biggest surprise in defensive importance is the high score of crosses into the penalty area, this might suggest that teams that play a brand of football where they cross the ball into the box a lot concede fewer goals. However, as this analysis is based in not only data but my personal analysis I decided to drop it from the defensive squad analysis. The defensive importance scores shows that number of blocks is the most important factor in determining the number of goals conceded. This shows that the ability for a team to make itself compact and block shots is important for defensive solidity. However, it also demonstrates that a high number of blocks might be indicative of a team that is under pressure for extended periods and might be liable to concede more goals.

As for offensive importance this analysis highlights the outstanding importance of shots on target, teams that generate a large volume of shots on target are more likely to score. Furthermore, players that generate large numbers of shots on target and have a strong goals per shot ratio are extremely valuable. In later analysis I will split the offensive importance into the top and bottom two relating to scoring and creativity analysis respectively.

3 Setting the Goal

Over the past 10 seasons Wigan have been demoted from the Championship four times and re-promoted a corresponding three times. Given that they were in the Premier League from 2005-2013 spending significant time in League One is a serious disappointment. A significant amount of this downturn in performance is due to their financial issues, having had points deducted on numerous occasions due to lack of access to funding. However, this write-up is focused on the football side, if the financial issues and points deductions had not occurred top half of the Championship table would almost certainly have been within grasp. Regardless of

Defensive Importance Scores	
Blocks	7.72
AerWon	6.48
Crosses into PA	5.29
Progressive Passes	4.19
Recoveries	3.98

Offensive Importance Scores	
Shots on Target	34.32
Goals per Shot	23.14
Passes into PA	14.41
Progressive Passes	7.90

Table 1: Top Feature Importance Scores

the realities of their current situation, this is where Wigan must strive to be if they want to achieve financial sustainability.

The team's that I targeted were recently promoted from League One and have since remained in the championship or been promoted to the Premier League, I took data from seasons where the target team finished in the top half of the table. These teams and corresponding seasons were; Sunderland (22/23), Luton (21/22, 22/23), Blackburn (21/22, 22/23), Coventry (21/22, 22/23).

These team's stats were averaged to create a set of stats that a team in the top half of the championship is aiming to achieve. Before comparing the stats I normalised the target team stats and Wigan's stats from last season in the Championship to make it easier to understand where the biggest delta in stats lay as they are on different scales. These deltas between the goal stats and Wigan's stats can be seen in table 2 below. These deltas show that Wigan is a long way off a top half Championship finish, this however could be seen from the outset. These deltas outline that Wigan's biggest issue seems to be creating and scoring goals as goals scored and stats associated with goals scored take up four of the top five top spots. For this reason in the next section where I target potential transfers I will focus on players with strong attacking numbers.

Stat	Delta
Goals Scored	-0.68
Shots on Target	-0.59
Goals Against	0.53
Key Passes[1]	-0.48
Goals/Shot	-0.48
Recoveries	-0.46
Progressive Passes	-0.44
Passes into PA	-0.32
Blocks	-0.052
Aerials Won	0.11

Table 2: Wigan vs Top Half Team Deltas

4 Transfer Targeting

It is important to only target relevant transfers, Wigan would never be able to entice or afford a Premier League player. For this reason I decided to target players in leagues where clubs have similar wage structures. I have not included the Championship as their club's wage structures are too dissimilar. Most of my analysis will focus on players that fit Wigan's current wage structure, which is solidly that of a League one team with a weekly wage bill of ∼£120,000. For this reason any player that came up in my analysis with a wage over Wigan's top paid player (Charlie Wyke £12k p.w) would be ignored. It should also be noted I unfortunately don't have access to in-depth League One stats as FBref.com does not host them, so I cannot target players from League One.

To target attacking players I split the attacking stats into two subsections, creativity and scoring. A player's creativity score is calculated using their Assists, Progressive Passes, Key Passes and Passes into the Penalty Area. A player's scoring score on the other hand is calculated using their Goals Scored, Shots on Target and Goals per Shot. To actually quantify the score I rank each player on each of these stats relative to the league they're playing in and sum these ranks. If in their league a player is 2nd in goals, 10th in G/Sh and 14th in Shots on Target their scoring score will be 26. When presented with the players that had the highest scoring score or creativity score I would then investigate them individually and find the most viable candidates before making a recommendation.

I will be creating two sets of recommendations, the first will be a tentative recommendation based on a player's performance so far this season, as seen in table 3. This is to catch players breaking out into a team and finding form or young players just finding their feet. These recommendations are more tentative as they are based on data from an extremely small sample size of ~18 games. The second set of reccomendations will be based off of data from last season, these will have a full season's worth of numbers so may be more reliable.

Player	Age	Wages	YoC²	Player Type	Creativity	Scoring
Baris Atik	29	7,500	NA	Creative	10	288
Juan Brunetta	26	7,100	4 yrs	Creative	10.5	118.5
Paul Nebel	21	7,800	1 yr	Creative	44.5	484
Jann Fiete-Arp	24	6,300	6 mth	Finisher	734	83

Table 3: Current Season Transfer Targets

Player	Age	Wages	YoC	Player Type	Creativity	Scoring
Marvin Wanitzek	30	7,900	3 yrs	Creative	5.5	244.5
Antonio Palumbo	27	2,500	3 yrs	Creative	25.5	466
Juan Brunetta	26	7,100	4 yrs	Creative	10.5	118.5
Robert Leipertz	30	6,700	6 mth	Finisher	617.5	75.5

Table 4: 22/23 Transfer Reccomendations

These results show a selection of the players with the highest creative of scoring rank who are of an acceptable age and on affordable wages. A majority of the players who ranked highly on last season's scoring have secured themselves transfers to bigger teams, or are the best paid players on their team and out of budget. The full list of said players can be seen by running analysis.ipynb. A notable exception to this is Juan Brunetta, having performed extremely well last season and already attaining 19 goal involvements in the first 16 games for mid-table Santos Laguna this season, he seems to be an exciting prospect on affordable wages. The only downside of this target is the he only moved to Santos last season. Marvin Wanitzek seems to be another standout from this list, whilst his age is a slight problem, he has averaged 16 goal involvements in his last four seasons with lower mid-table Karlsruher. These results also show an interesting trend, almost all of the top finishers were somewhat unsurprisingly far out of budget. Similar to Charlie Wyke being the top paid player at Wigan, most teams prize their top goalscorers and therefore they are almost always out of budget. All of these players, whilst having different levels of promise, would almost certainly bolster Wigan's current attack and seem to be within budget.

5 Conclusion

I'm glad I decided on a whim to do this quick project, it certainly falls into the proof of concept category so there are a whole raft of improvements that need to be made, such as.

- Pulling data from Transfermarkt and salarysport to automate sifting out players on long contracts and/or high wages.

- Aggregating data across many seasons to create a clearer picture of a player's profile.

- Having access to in-depth data for League One.

- A better personal understanding of the needs of the club being analysed

- Having data for League One and the Championship to use for creating the stat to GS/GA correlations

- Spend more time on individual player identification from candidates presented by the framework

That all being said I feel gratified in completing this project and am happy with the very simple framework I've created. The weaknesses in the framework are clear but if anything they drive me to improve the product and keep working on it into the future. This area of data science is my passion and this project was thoroughly interesting for me personally, more importantly, I hope the process outlined in this write-up illuminates how I would approach a problem like this.