INSY 336 Final Project

Premier League Performance and Google Engagement

Samuel Ferreira, Sebastian Andersen, Shane Silversmith

Table of contents

1

Introduction

Our initial thoughts and predictions

Introduction

- Our initial motivations were to explore the link between a Premier League teams' social media engagement on matchdays and their game performance
- Some relationships we wanted to explore were:
  - The correlation between social media engagement and game performance
  - The correlation between the competitiveness of matches and social media engagement
  - The difference between lower performing teams' and higher performing teams social media engagement

What's the relationship between Premier League teams' performance and online engagement?

2

Data Collection

Our process and difficulties we faced

Data Collection

1. Scraped tables from 2021/22-2023/24 season from fbref.com
2. Created a script that made links with Google trends' standardized link format
3. Scraped the daily data from each Premier League teams' data for a season
4. Files were automatically renamed to their team names and placed in folders corresponding to season
5. Each CSV was combined by Season and added with other CSV's for data on match results etc
6. All data was added to a CSV appropriately titled "the_holy_grail.csv"

Data Collection

Online Engagement

Daily Google Trends Data

Daily Google Trends data is standardized and ranges from 0-100. All of the data is relative to the 100th value, the day with the largest internet traffic in the given time period.

Weekly Google Search Traffic

Weekly Google Search Traffic shows actual traffic for Google internet searches for a team in a given week, making it suitable for comparing data across teams.

Performance

Matchday Performance

Matchday performance data looked at the performance of each individual team on days that they played a Premier League match; this could result in a win, tie or loss.

Weekly Ranking

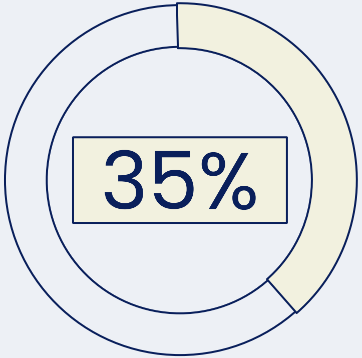
Weekly ranking data involved the team's ranking on the table at the moment of a game in that season.

3

Results

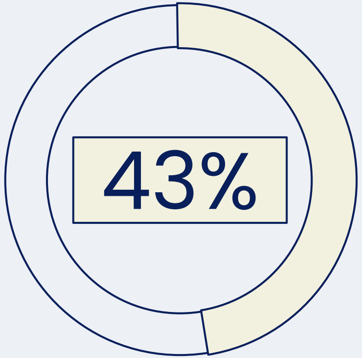
Our findings and analysis

Daily Google Trends Data and Match Outcome



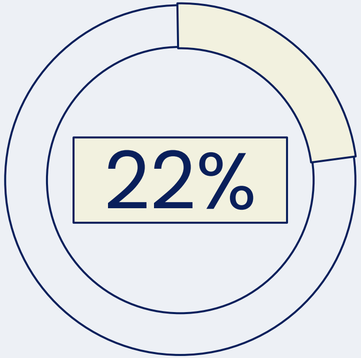
Win

When having a higher normalized daily traffic index than the opponent



Loss

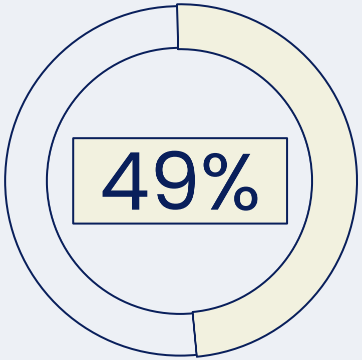
When having a higher normalized daily traffic index than the opponent



Draw

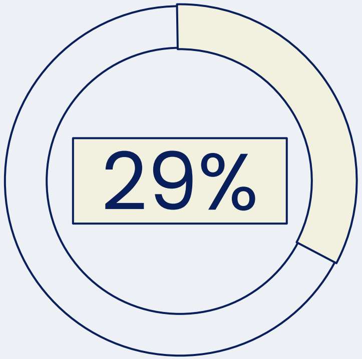
When having a higher normalized daily traffic index than the opponent

Weekly Google Search Traffic and Match Outcome



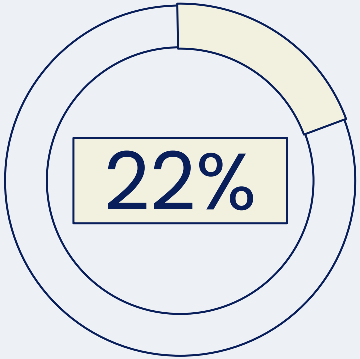
Win

When average weekly engagement is higher than the opponents



Loss

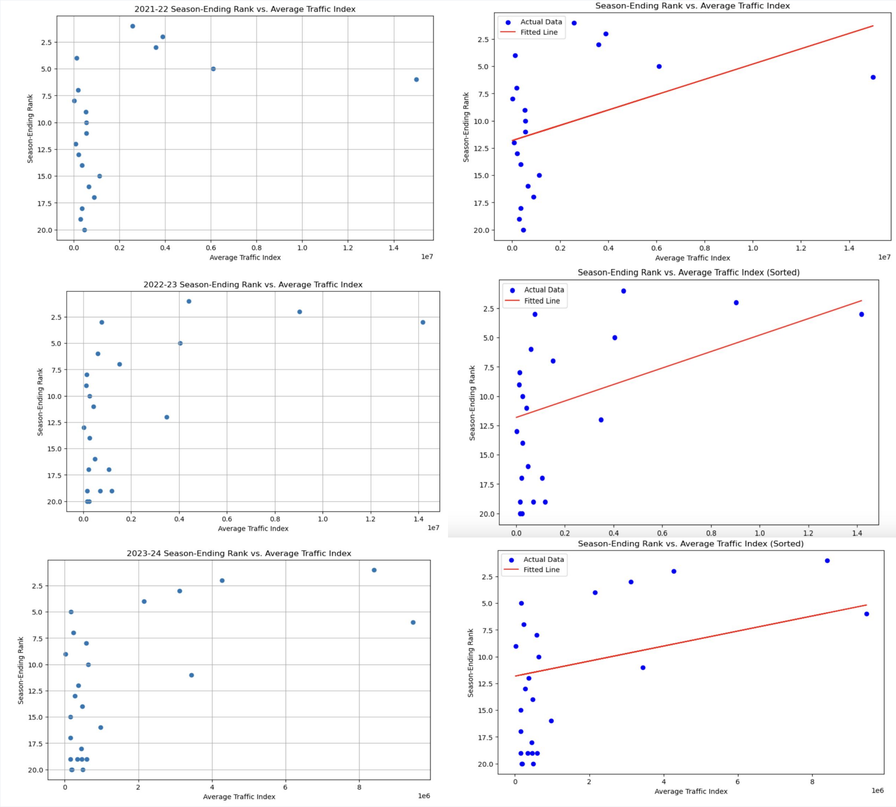
When average weekly engagement is higher than the opponents



Draw

When average weekly engagement is higher than the opponents

End of Season Rank and Weekly Google Search Traffic



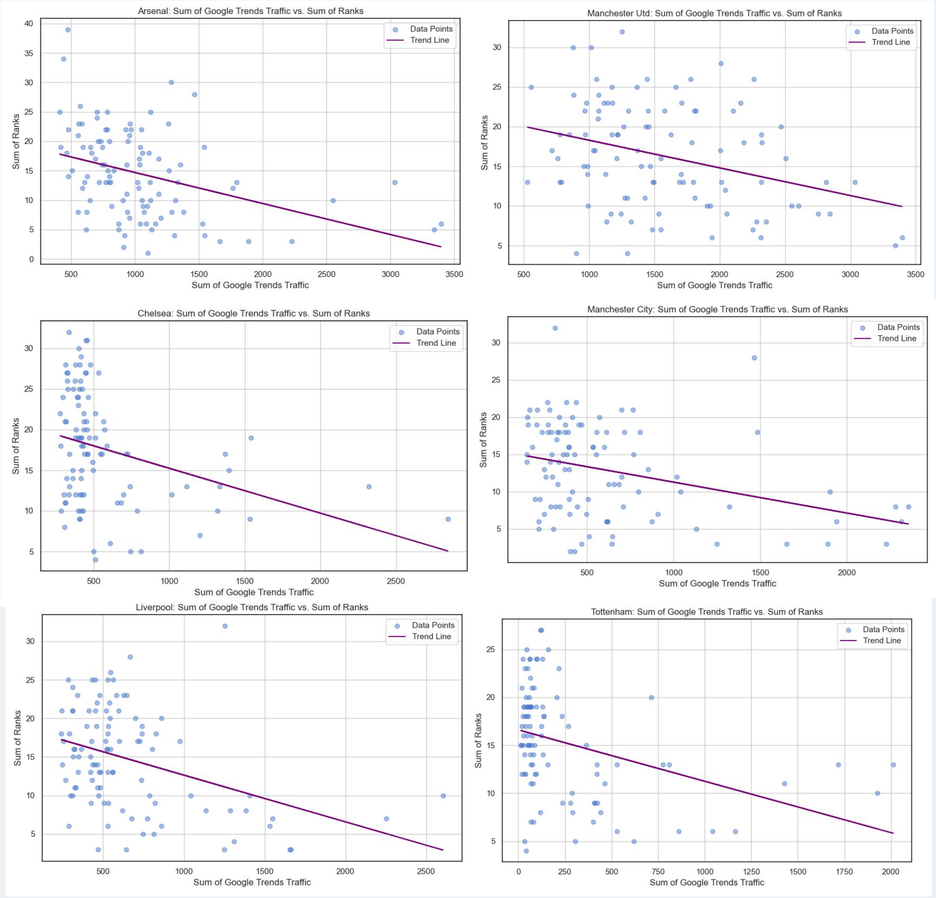
-Positive correlation between positioning and weekly Google Search Traffic for all 3 seasons

-The teams who are positioned better in the standings tend to have a higher average Google Search Traffic

-A positive coefficient and a p-value for the 'Average Traffic' coefficient of 0.02, which shows that having high engagement has a high association with being a higher ranked team

Match Competitiveness and Google Traffic Engagement

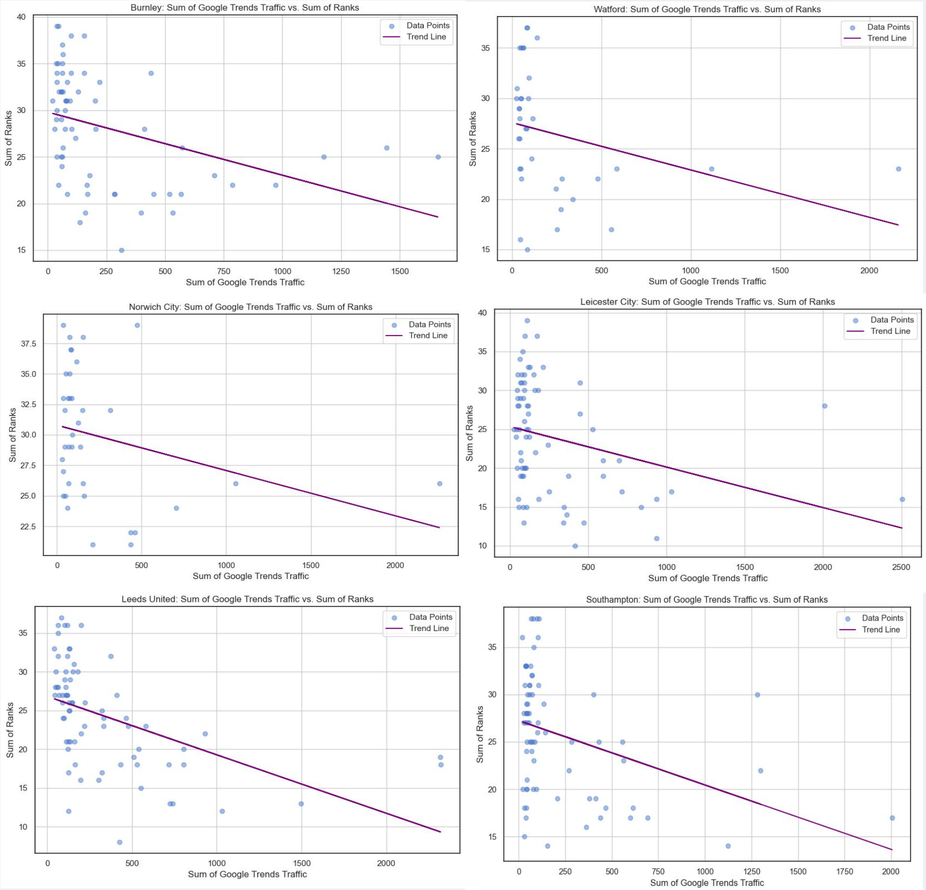
The ‘Big 6’



- Across all teams there was positive correlation between ‘match competitiveness’ and Google search traffic
- ‘Match Competitiveness’ was defined by the sum of the rank of the two teams at the time of the game: a match between first and second represented a Sum of Ranks of 3, while a match between last and second last represented a sum of 39
- While the relationship was significant across all teams (p-value<0.05), the R^2 was consistently low, generally ranging around 0.1-0.2
    - The model (Google traffic) did a poor job explaining the variability of the ranks of the teams

Match Competitiveness and Google Traffic Engagement

The 6 relegated teams in 22 and 23 seasons

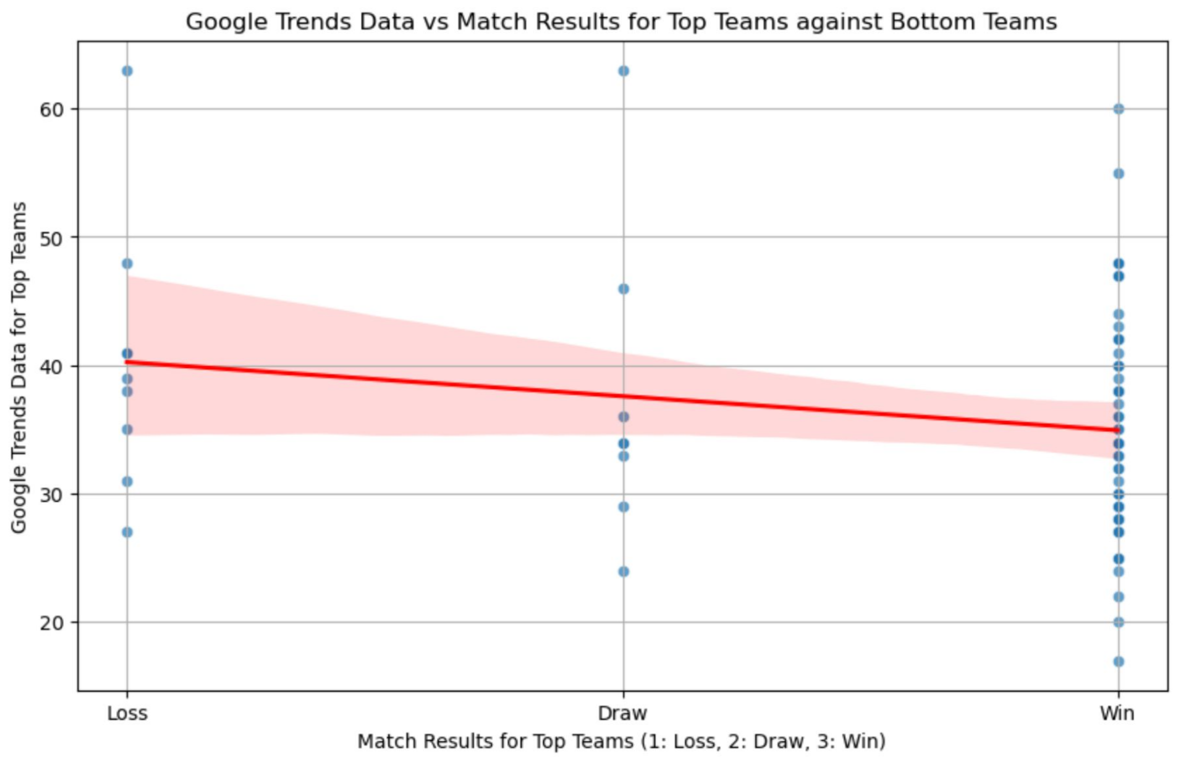



```
OLS Regression Results for Arsenal:
                            OLS Regression Results
==============================================================================
Dep. Variable:           Sum of Ranks   R-squared:                       0.147
Model:                            OLS   Adj. R-squared:                  0.138
Method:                 Least Squares   F-statistic:                     17.05
Date:                Sun, 07 Apr 2024   Prob (F-statistic):           7.61e-05
Time:                        21:11:08   Log-Likelihood:                -338.11
No. Observations:                 101   AIC:                             680.2
Df Residuals:                      99   BIC:                             685.5
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
const                        19.9808      1.500     13.322      0.000      17.005      22.957
Sum of Google Trends Traffic -0.0053      0.001     -4.129      0.000      -0.008      -0.003
==============================================================================
Omnibus:                        3.451   Durbin-Watson:                   1.629
Prob(Omnibus):                  0.178   Jarque-Bera (JB):                3.089
Skew:                           0.427   Prob(JB):                        0.213
Kurtosis:                       3.061   Cond. No.                     2.55e+03
==============================================================================
```

```
OLS Regression Results for Manchester City:
                            OLS Regression Results
==============================================================================
Dep. Variable:           Sum of Ranks   R-squared:                       0.114
Model:                            OLS   Adj. R-squared:                  0.105
Method:                 Least Squares   F-statistic:                     12.77
Date:                Sun, 07 Apr 2024   Prob (F-statistic):           0.000547
Time:                        21:06:17   Log-Likelihood:                -322.51
No. Observations:                 101   AIC:                             649.0
Df Residuals:                      99   BIC:                             654.2
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
const                        15.4405      0.932     16.572      0.000      13.592      17.289
Sum of Google Trends Traffic -0.0041      0.001     -3.573      0.001      -0.006      -0.002
==============================================================================
Omnibus:                        1.978   Durbin-Watson:                   1.564
Prob(Omnibus):                  0.372   Jarque-Bera (JB):                1.457
Skew:                           0.270   Prob(JB):                        0.483
Kurtosis:                       3.234   Cond. No.                     1.26e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.26e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Traditional top 4 teams against bottom 5



OLS Regression Results

Dep. Variable:	Google Trends Data	**R-squared:**	0.064
Model:	OLS	**Adj. R-squared:**	0.053
Method:	Least Squares	**F-statistic:**	5.767
Date:	Sun, 07 Apr 2024	**Prob (F-statistic):**	0.0185
Time:	20:34:22	**Log-Likelihood:**	-310.46
No. Observations:	86	**AIC:**	624.9
Df Residuals:	84	**BIC:**	629.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	40.6916	3.173	12.825	0.000	34.382	47.001
Result	-2.9831	1.242	-2.401	0.019	-5.453	-0.513

Omnibus:	4.544	**Durbin-Watson:**	1.504
Prob(Omnibus):	0.103	**Jarque-Bera (JB):**	3.871
Skew:	0.498	**Prob(JB):**	0.144
Kurtosis:	3.299	**Cond. No.**	9.47

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

4

Conclusion

Conclusions reached and limitations

Conclusion

1. Winning games and higher rankings have a significant relationship with Google Search Traffic
2. Matches with high Google Search Traffic showed a significant correlation with better ranked teams
3. There was little to no relationship between online engagement of top performing teams to their performance against low performing teams

Limitations in our approach

Keywords used posed challenges

- For Google search data, many teams had standard names; searches under their name were not necessarily related to the team. Example: Wolverhampton Wanderers and Wolves

Not enough data in general

- While our data frames included thousands of data points, there is high variability within each season. More seasons are needed to account for seasonal differences

Weekly and daily data did not account for external factors

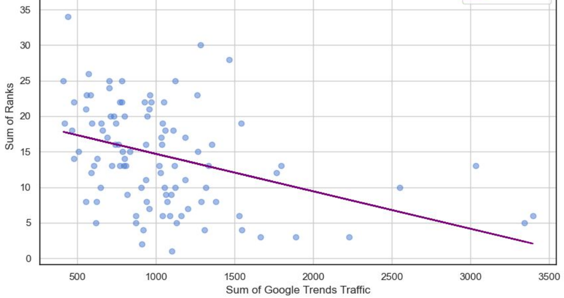
- Teams play matches in other competitions outside the Premier League that drive online engagement
- Club news also drives online engagement. Example: Cristiano Ronaldo’s move to Manchester United in August 2021

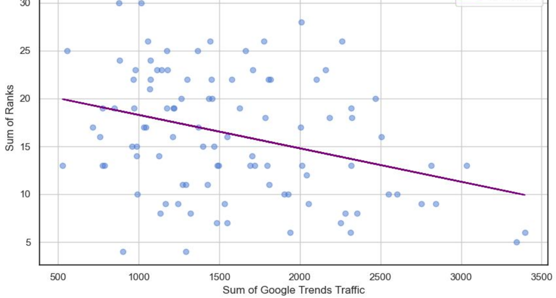
Google search traffic and trends may not be the best measure for online engagement

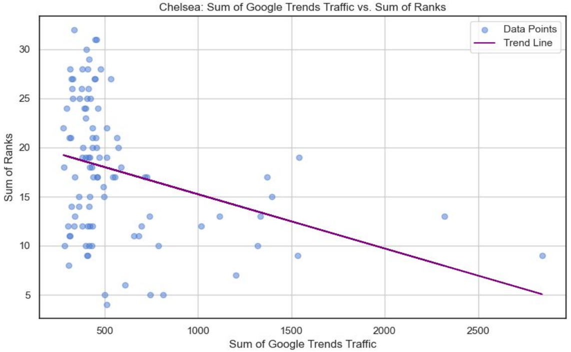
- We believe that other measures of social media engagement, like proactive involvement on social media apps would lead to stronger results

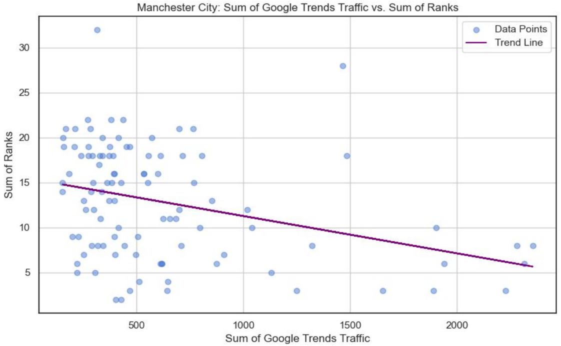
Thanks!

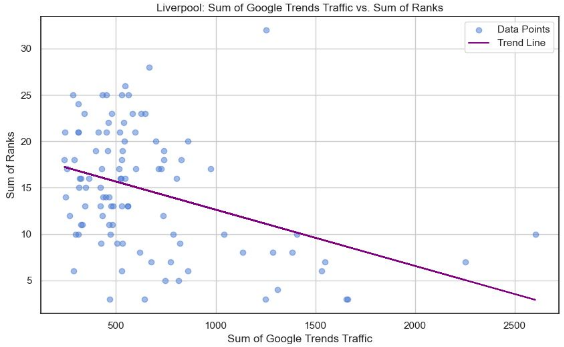
Questions?

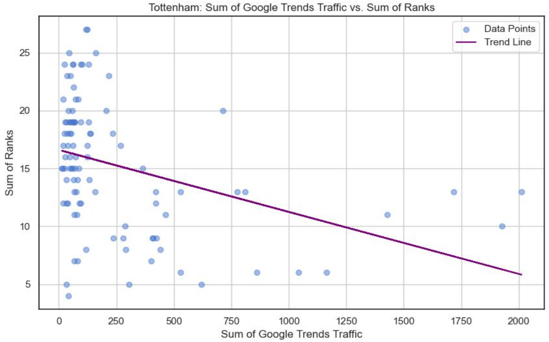

Sum of Ranks
Sum of Google Trends Traffic


Sum of Ranks
Sum of Google Trends Traffic


Chelsea: Sum of Google Trends Traffic vs. Sum of Ranks
Data Points
Trend Line
Sum of Ranks
Sum of Google Trends Traffic


Manchester City: Sum of Google Trends Traffic vs. Sum of Ranks
Data Points
Trend Line
Sum of Ranks
Sum of Google Trends Traffic


Liverpool: Sum of Google Trends Traffic vs. Sum of Ranks
Data Points
Trend Line
Sum of Ranks
Sum of Google Trends Traffic


Tottenham: Sum of Google Trends Traffic vs. Sum of Ranks
Data Points
Trend Line
Sum of Ranks
Sum of Google Trends Traffic