Soccer Match Analysis
Reddit Threads

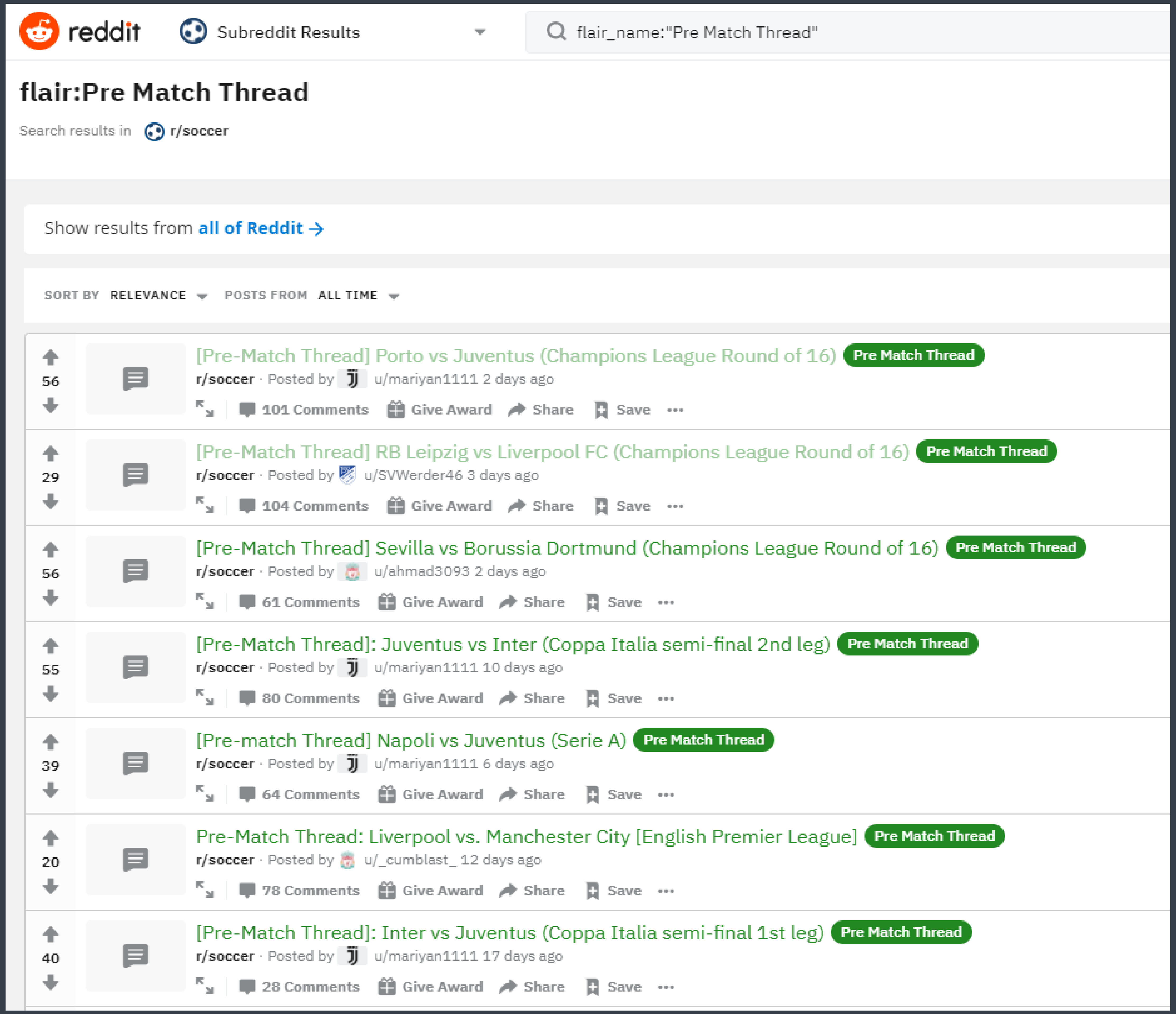
Richard Gao - 260729805
Anukriti Yadav - 260986353
Siddharth Singhal - 260986354
Bogdan Tanasie - 260747949

Problem Statement

- ✓ Can we predict the outcome of games from pre-match threads?
- ✓ What are the popular sentiments among the fans of popular teams before a soccer match?
- ✓ What are the popular topics and key words discussed by the fans of the top soccer teams around the globe?
- ✓ Comparing teams from different leagues and know how are they perceived?
- ✓ Can we calculate these things on the fly and display it on a dashboard?

Data Source- Reddit

- ✓ Pre-Match Threads- Scrapped all the pre-match threads from one of the most popular social, web content and rating aggregator website Reddit.
- ✓ Other sources can also be useful. Ex. http://www.footballforums.net/threads/uefa-champions-league-2020-2021.276873/
- ✓ Reddit API was used to pull the threads from Reddit soccer forums.
- ✓ Our Dataset- 98 Unique Pre-Match threads, ~18327 Comments/Posts



flair:Pre Match Thread

Search results in r/soccer

Show results from all of Reddit →

SORT BY RELEVANCE POSTS FROM ALL TIME

56 [Pre-Match Thread] Porto vs Juventus (Champions League Round of 16) Pre Match Thread
r/soccer · Posted by u/mariyan1111 2 days ago
101 Comments Give Award Share Save ···

29 [Pre-Match Thread] RB Leipzig vs Liverpool FC (Champions League Round of 16) Pre Match Thread
r/soccer · Posted by u/SVWerder46 3 days ago
104 Comments Give Award Share Save ···

56 [Pre-Match Thread] Sevilla vs Borussia Dortmund (Champions League Round of 16) Pre Match Thread
r/soccer · Posted by u/ahmad3093 2 days ago
61 Comments Give Award Share Save ···

55 [Pre-Match Thread]: Juventus vs Inter (Coppa Italia semi-final 2nd leg) Pre Match Thread
r/soccer · Posted by u/mariyan1111 10 days ago
80 Comments Give Award Share Save ···

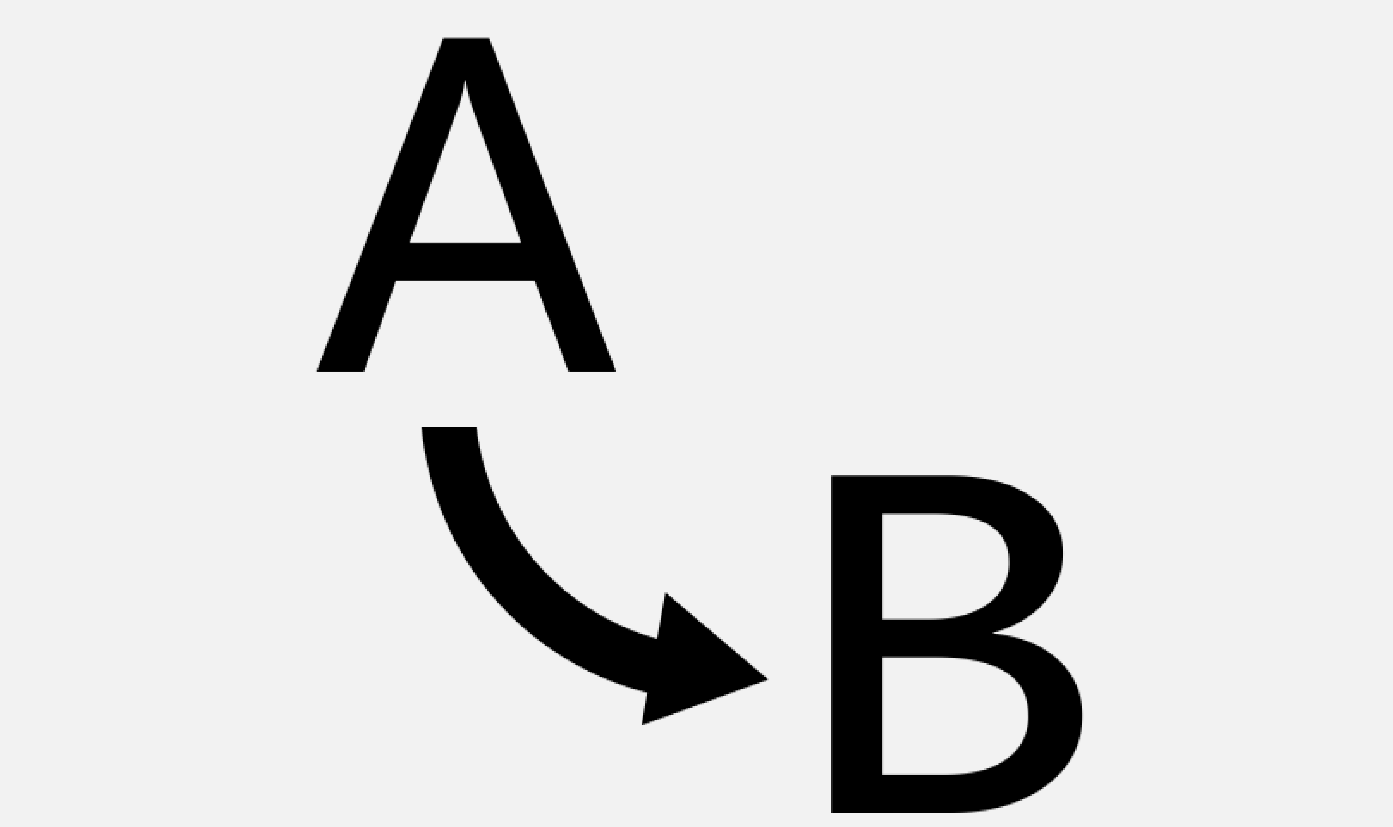
39 [Pre-match Thread] Napoli vs Juventus (Serie A) Pre Match Thread
r/soccer · Posted by u/mariyan1111 6 days ago
64 Comments Give Award Share Save ···

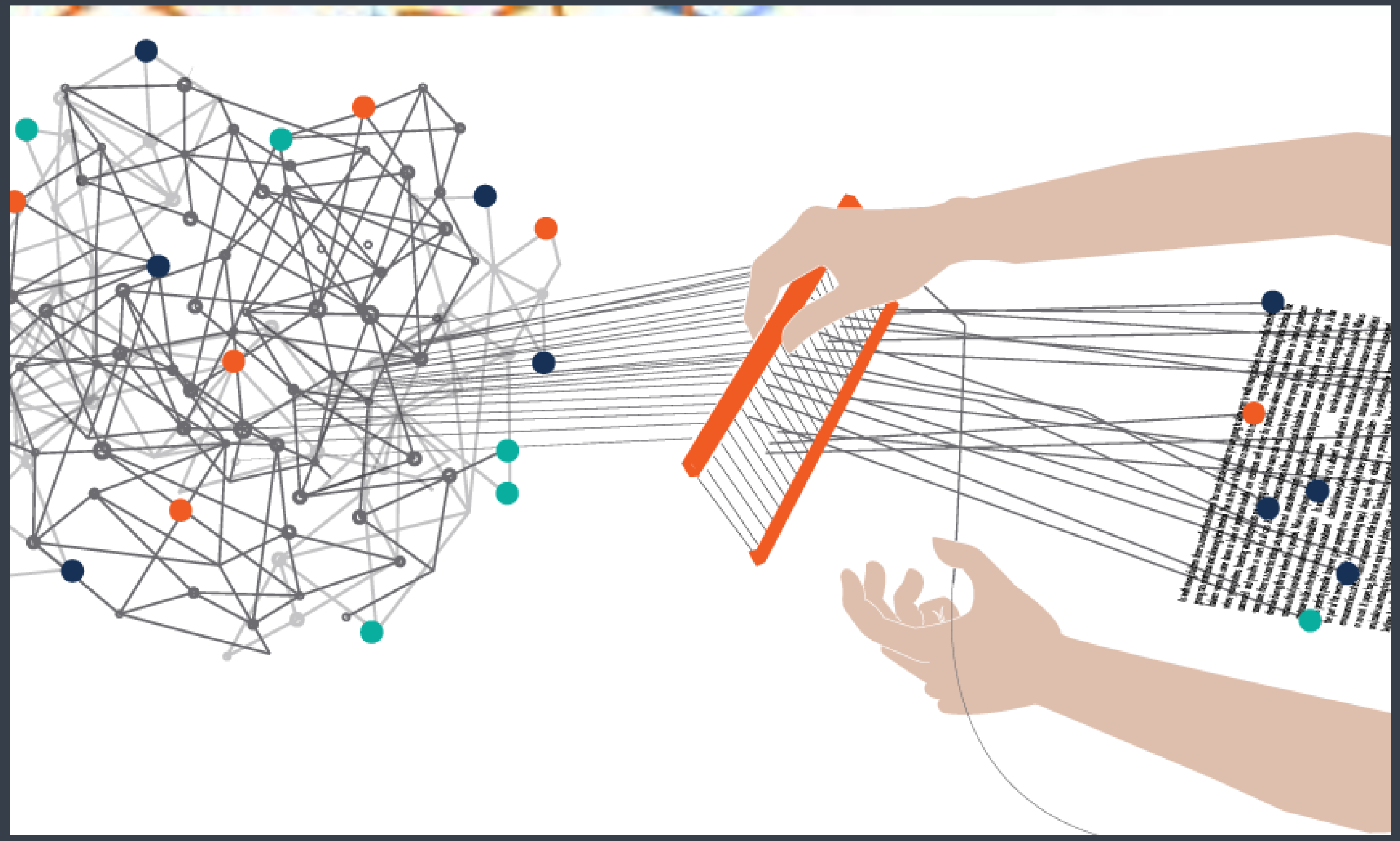
20 Pre-Match Thread: Liverpool vs. Manchester City [English Premier League] Pre Match Thread
r/soccer · Posted by u/_cumblast_ 12 days ago
78 Comments Give Award Share Save ···

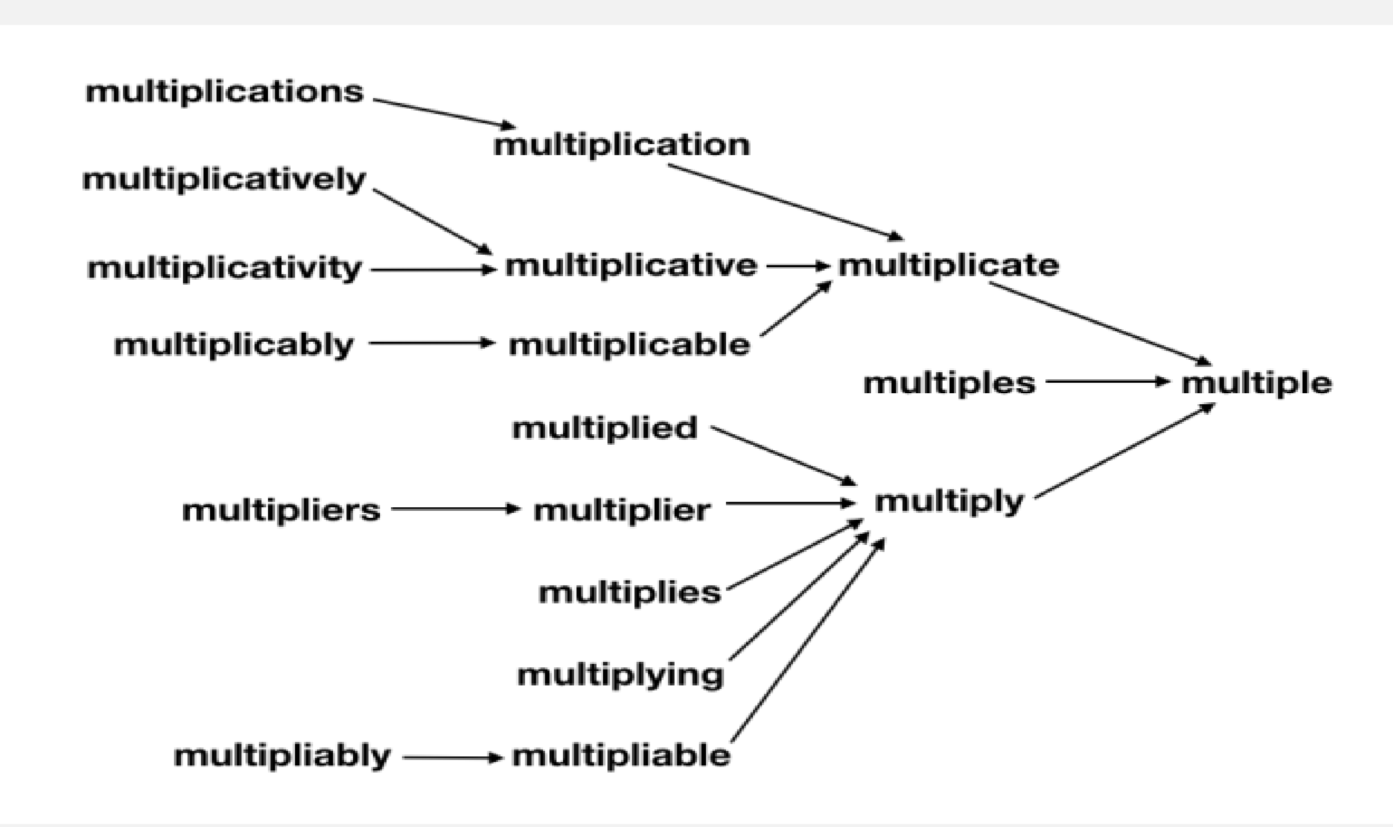
40 [Pre-Match Thread]: Inter vs Juventus (Coppa Italia semi-final 1st leg) Pre Match Thread
r/soccer · Posted by u/mariyan1111 17 days ago
28 Comments Give Award Share Save ···

Data Pre-Processing

“Data is like crude. It’s valuable, but if unrefined it cannot really be used.”







Replacement

Find & Replaced

Team names
Emotions

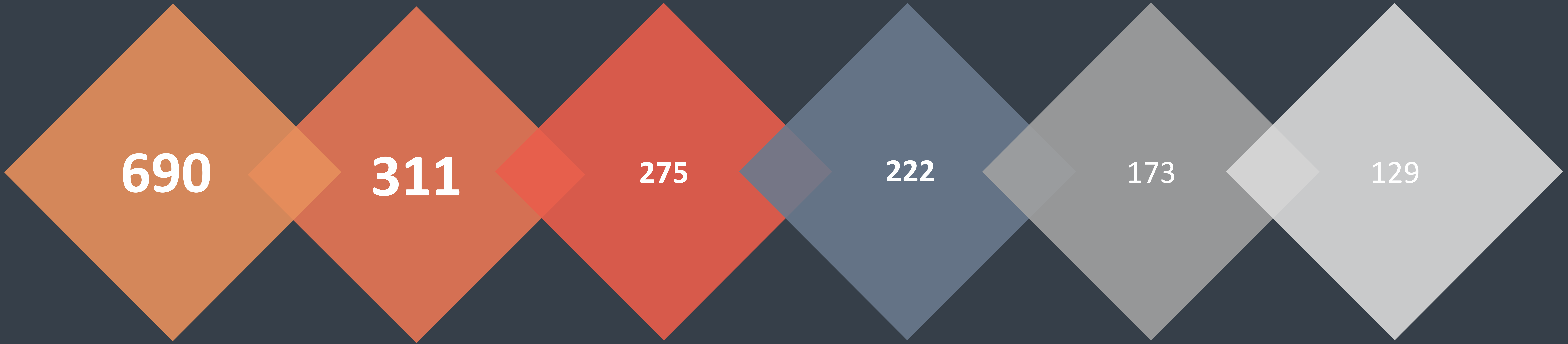
Tokenization

Converted a sentence into a list of words
Remove Punctuations
Tokenize
Remove stop words
Remove alpha-alphabetic tokens

Lemmatization

Transformed any form of a word to its root word

Frequency of important words



690

311

275

222

173

129

Good

Fans have used a lot of positive words. This can be related to the performance of teams.

Angry

Yes, fans have right to get angry with their favorite team when the team underperforms or when a fan of the other team boo's

Lose

Before the match kicks-off, soccer enthusiasts start predicting the final score by looking at the line-ups and the forms of the teams

Bad

Fans tend to be more brutal towards the certain teams because of the past revelry. Ex. We have El-Classico in La Liga and Manchester Derby in EPL.

Injury

Injury is one of the important topic of discussion amongst the fans. Makes sense as injury of an important player may cost a match to the team

Defense

Fans are focused more on game defense rather than attack. This shows the importance of a team with a good defense tactics and players.

Dashboard

Insights (From Lift Scores):

- People **anticipate** Liverpool matches the most.

- People seem to relate Chelsea and Man United to **losing**.

- People talk a lot about Man United's and PSG's **defense**.

- People seem to be least **angry** at Ajax.

- People mention **good** alongside Ajax.

- People talk a lot about Bayern's **offense**.

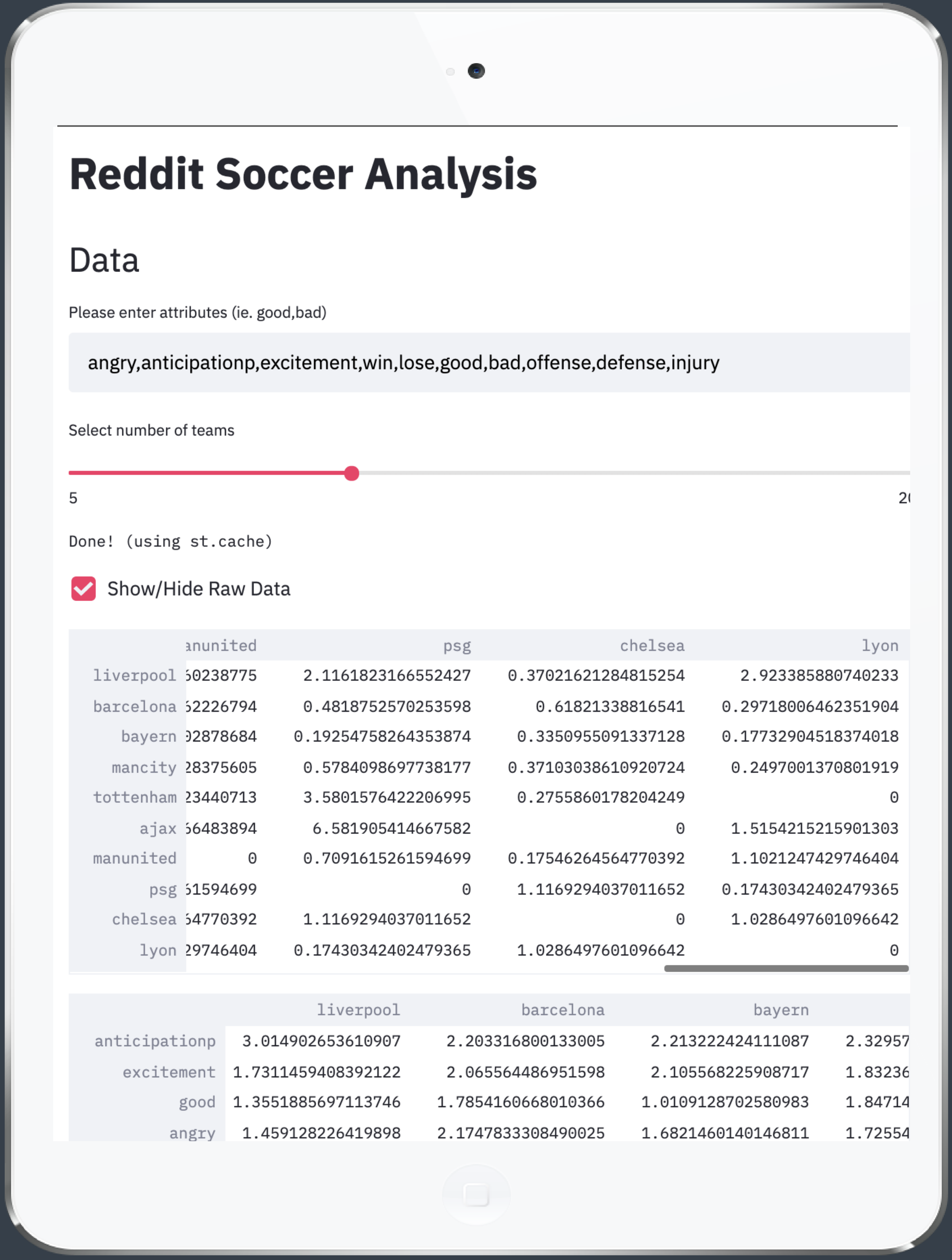
- **Injuries** are talked about a lot with Tottenham, Ajax and Porto.

- Barcelona, Man City, and PSG seem attract the word **bad**.

- **Excitement** is highest for Juventus.

Insights MDS:

- Liverpool, Arsenal, Man United and Ajax separate from the rest of the teams: They tend to be at the top of their respective leagues.



Reddit Soccer Analysis

Data

Please enter attributes (ie. good,bad)

angry,anticipationp,excitement,win,lose,good,bad,offense,defense,injury

Select number of teams

```
Done! (using st.cache)
```

☑ Show/Hide Raw Data

	anunited	psg	chelsea	lyon
liverpool	0238775	2.1161823166552427	0.37021621284815254	2.923385880740233
barcelona	2226794	0.4818752570253598	0.61821338816541	0.29718006462351904
bayern	2878684	0.19254758264353874	0.3350955091337128	0.17732904518374018
mancity	28375605	0.5784098697738177	0.37103038610920724	0.2497001370801919
tottenham	23440713	3.5801576422206995	0.2755860178204249	0
ajax	6483894	6.581905414667582	0	1.5154215215901303
manunited	0	0.7091615261594699	0.17546264564770392	1.1021247429746404
psg	1594699	0	1.1169294037011652	0.17430342402479365
chelsea	4770392	1.1169294037011652	0	1.0286497601096642
lyon	29746404	0.17430342402479365	1.0286497601096642	0

	liverpool	barcelona	bayern	
anticipationp	3.014902653610907	2.203316800133005	2.213222424111087	2.32957
excitement	1.7311459408392122	2.065564486951598	2.105568225908717	1.83236
good	1.3551885697113746	1.7854160668010366	1.0109128702580983	1.84714
angry	1.459128226419898	2.1747833308490025	1.6821460140146811	1.72554

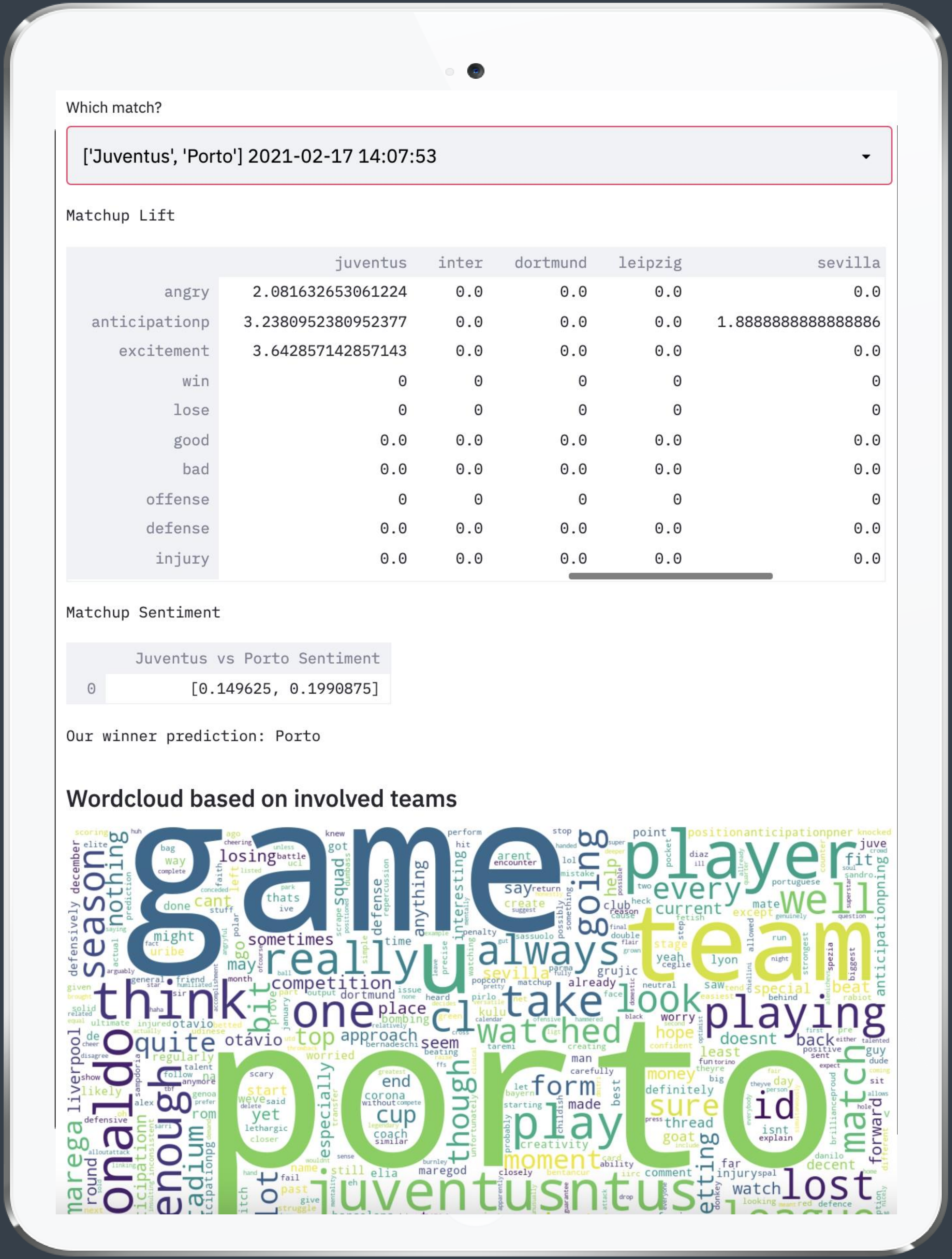
Dashboard

Insights (Per Match Sentiment Analysis):

- Example: Sentiment for Porto is higher than Juventus
- ~40 % accuracy (including draws)

Insights Topic Modelling + Word Cloud (Per Match):

- Star players such as Messi, Ronaldo always stand out
- We see much more directed anger towards Juventus
- We see more positive words and topics such as good involved with Porto.



Which match?

['Juventus', 'Porto'] 2021-02-17 14:07:53

Matchup Lift

	juventus	inter	dortmund	leipzig	sevilla
angry	2.081632653061224	0.0	0.0	0.0	0.0
anticipationp	3.2380952380952377	0.0	0.0	0.0	1.8888888888888886
excitement	3.642857142857143	0.0	0.0	0.0	0.0
win	0	0	0	0	0
lose	0	0	0	0	0
good	0.0	0.0	0.0	0.0	0.0
bad	0.0	0.0	0.0	0.0	0.0
offense	0	0	0	0	0
defense	0.0	0.0	0.0	0.0	0.0
injury	0.0	0.0	0.0	0.0	0.0

Matchup Sentiment

	Juventus vs Porto Sentiment
0	[0.149625, 0.1990875]

Our winner prediction: Porto

Wordcloud based on involved teams

Learnings and Future Work

Difficult to predict

- Reddit data is hard to deal with and can be biased

- Sentiment & Lift values are not enough for accurate predictions

Player level insights

- We are only looking at a team level

- Would be interesting to use similar measure on players or even managers to gather insights

Use more data

- Only looking at about 50 games from top teams

- Include upvotes

- Need more historical data and potentially some contextual data such as team ranking to improve predictions

Automate process

- Replacement for teams should be automated

- Checking results is manual and takes a lot of time

Thank You!