



# FIFA20

**FIFA®**

OFFICIAL  
LICENSED  
PRODUCT

C H A M P I O N S E D I T I O N

- Question: How do teams fare in the next game after playing the dominant team?
- Are teams with better records more likely to beat the dominant team? Are they more likely to win the next game?
- Has playing the dominant team changed their odds in the next game?

#### Data & Methodology

- **Kaggle Database of 11 European Soccer Leagues along with their EA Sports' FIFA players' ratings and results of 25k matches from 2008 to 2016.**
- **Tables include Country, Leagues, Matches, Players, Player Attributes, Team Attributes**
- **Focused on 2015-16 season, which was most recent in database. Could easily have looked at earlier seasons.**

## By the Numbers for the 2016 season we used

- 6,652 matches
- 188 teams
- 11 leagues
- 332 matches against dominant team
- 332 “next games”
- 1 league had 28 games with the exact same team following the dominant team game

Kaggle



database.sqlite

Jupyter  
Notebook



```
import sqlite3  
conn = sqlite3.connect('database.sqlite')
```

DataFrame  
Object



```
COUNTRY_df = pd.read_sql_query("select *  
from COUNTRY;", conn)
```

Kaggle



database.sqlite

Jupyter  
Notebook



```
import sqlite3  
conn = sqlite3.connect('database.sqlite')
```

DataFrame  
Object



```
COUNTRY_df = pd.read_sql_query("select *  
from COUNTRY;", conn)
```

## COUNTRY

## TEAM

## LEAGUE

id	name
1	Belgium
1729	England
4769	France
7809	Germany
10257	Italy
13274	Netherlands
15722	Poland
17642	Portugal
19694	Scotland
21518	Spain
24558	Switzerland

team_api_id	team_long_name	team_short_name
9987	KRC Genk	GEN
9993	Beerschot AC	BAC
10000	SV Zulte-Waregem	ZUL
9994	Sporting Lokeren	LOK
9984	KSV Cercle Brugge	CEB
8635	RSC Anderlecht	AND
9991	KAA Gent	GEN
9998	RAEC Mons	MON
7947	FCV Dender EH	DEN
9985	Standard de Liège	STL
8203	KV Mechelen	MEC

id	country_id	name
1	1	Belgium Jupiler League
1729	1729	England Premier League
4769	4769	France Ligue 1
7809	7809	Germany 1. Bundesliga
10257	10257	Italy Serie A
13274	13274	Netherlands Eredivisie
15722	15722	Poland Ekstraklasa
17642	17642	Portugal Liga ZON Sagres
19694	19694	Scotland Premier League
21518	21518	Spain LIGA BBVA
24558	24558	Switzerland Super League

# MATCH

date	league_id	home_team_api_id	away_team_api_id	home_team_goal	away_team_goal
2015-07-24 00:00:00	1	9997	8342	2	1
2015-07-25 00:00:00	1	8571	9985	2	1
2015-07-25 00:00:00	1	9987	1773	3	1
2015-07-25 00:00:00	1	8573	8203	3	1
2015-07-25 00:00:00	1	10000	9994	3	1
2015-07-26 00:00:00	1	10001	9991	1	1
2015-07-26 00:00:00	1	8635	8475	3	2
2015-07-26 00:00:00	1	9986	274581	2	1
2015-10-04 00:00:00	1	8635	8203	1	1
2015-10-04 00:00:00	1	9991	8342	4	1
2015-10-04 00:00:00	1	9987	9985	3	1

For every game  
against Dominant  
team ...

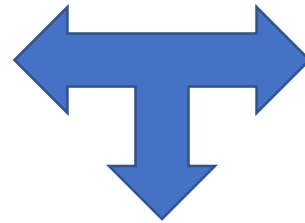
Team's  
game 1



Dominant  
Team

... what happened  
in the next game?

Team's  
game 2

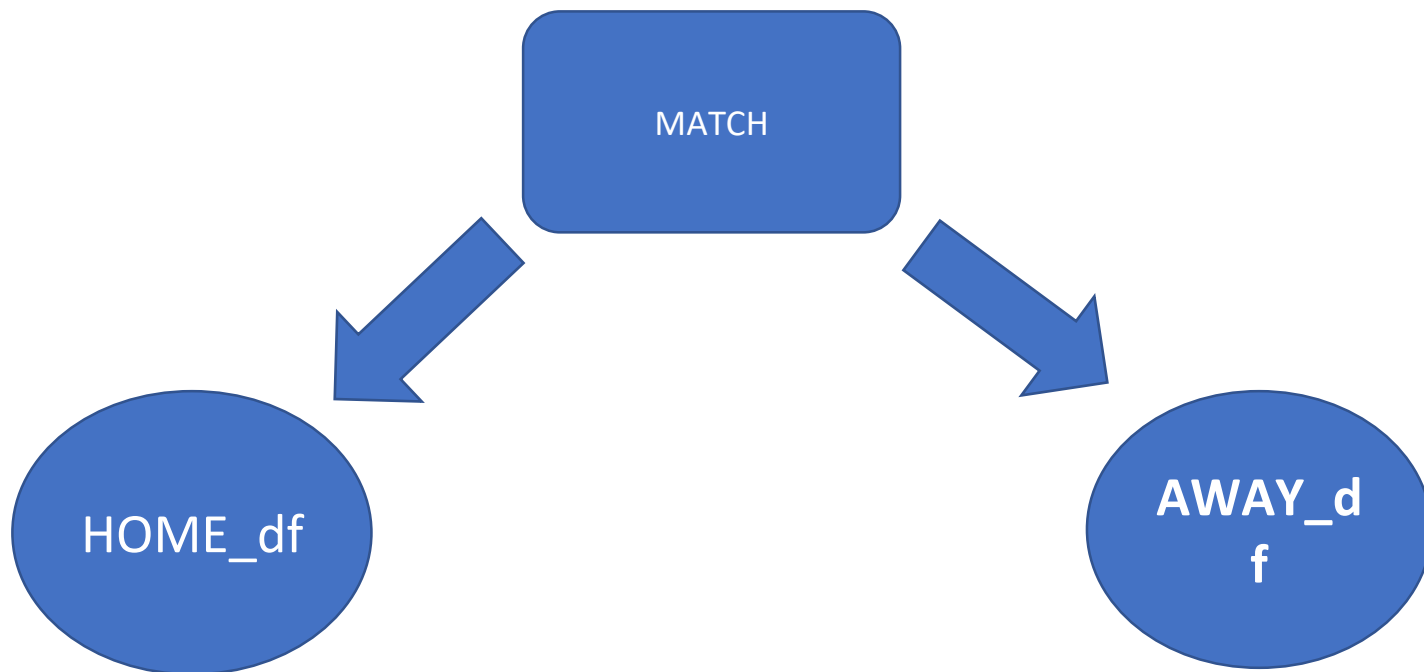


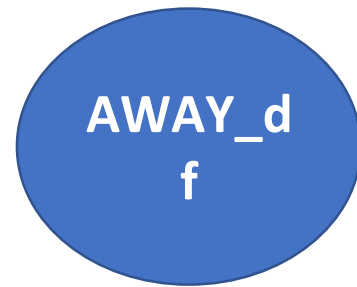
Some  
other team

Win Loss or Tie? By how much?

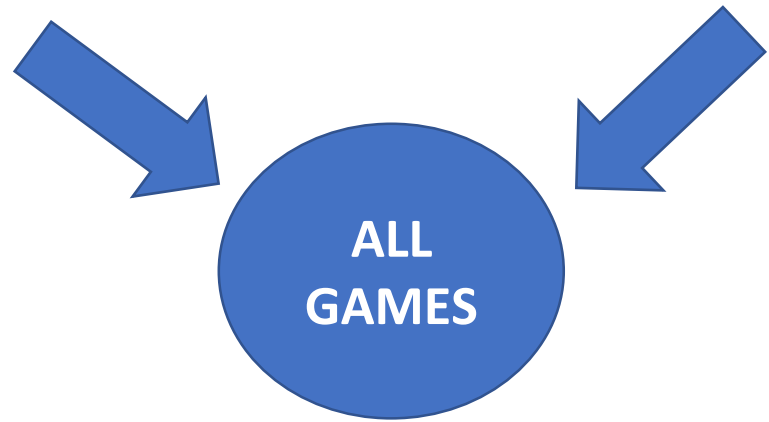


# Split out Home And Away Games





home\_team\_goal - away\_team\_goal **DIFF** away\_team\_goal - home\_team\_goal



```
frames = [HOME, AWAY]  
pd.concat(frames)
```

# The tricky part - getting the next game:

Step 1 – Copy and sort the data

## Do not use:

- `next_game = all_games` ? Would make a new “nickname” for all\_game  
It does not make a new dataframe  
Changes to next\_game would apply to all\_game

## Use:

- `next_game = all_games.copy()` ? Makes a new copy of next\_game  
Can change each dataframe  
independently

## Step 2 – Shift the data to see two consecutive games in one row

all\_games

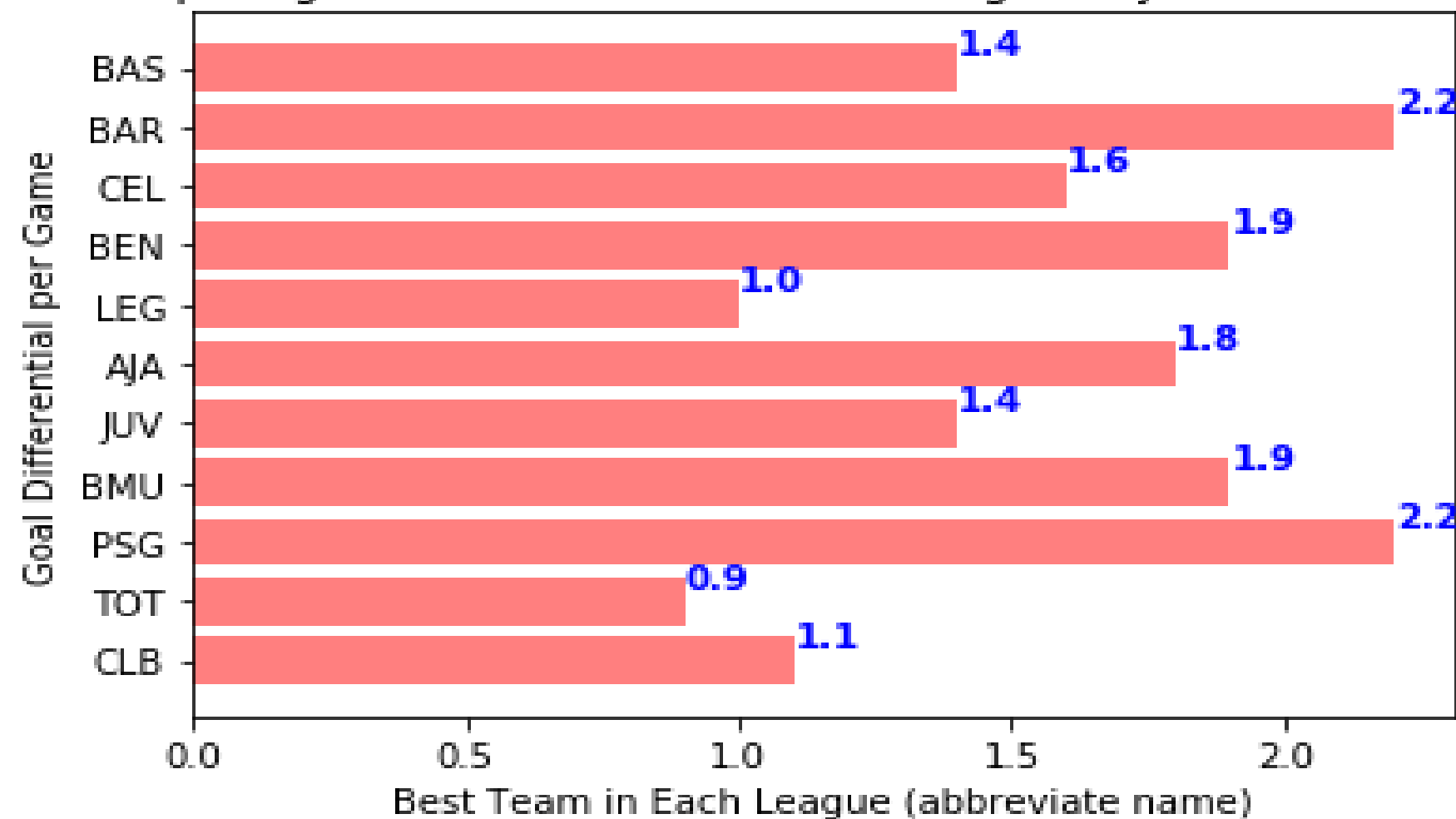
team_api_id	DIFF	league_id	opponent	W	L	T
1601	1	15722	8245	1	0	0
1601	-1	15722	1957	0	1	0
1601	-2	15722	8021	0	1	0
1601	-1	15722	2186	0	1	0
1601	0	15722	2182	0	0	1

next\_game

team_api_id	DIFF	league_id	opponent	W	L	T
1601	1	15722	8245	1	0	0
1601	-1	15722	1957	0	1	0
1601	-2	15722	8021	0	1	0
1601	-1	15722	2186	0	1	0
1601	0	15722	2182	0	0	1

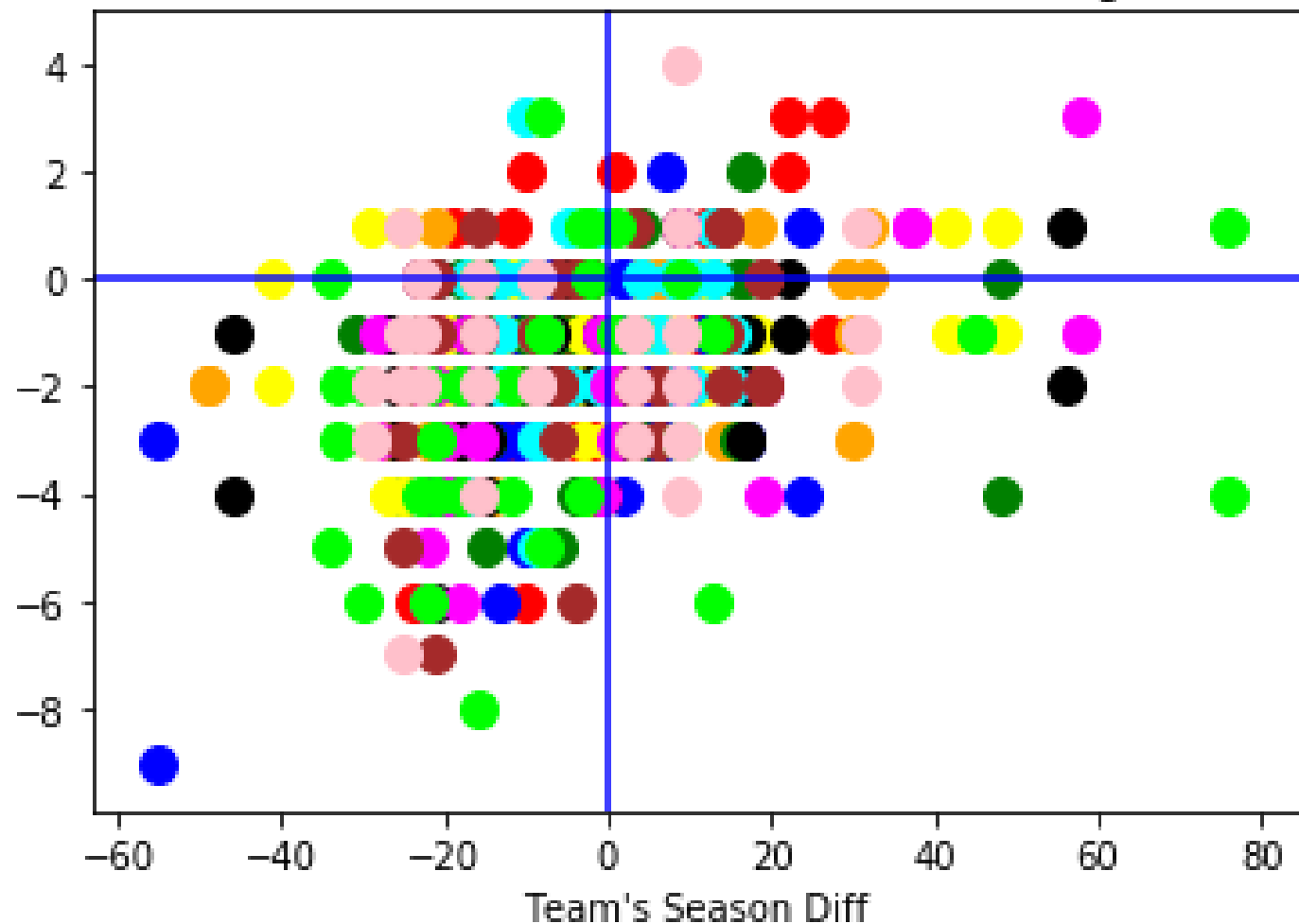
- `df_complete = pd.merge(all_games, next_game.shift(-1).fillna(0), left_index=True, right_index=True)`
- Select 332 rows with game against Dominant Team and Next Game

Comparing Dominant Teams across Leagues by Goal Differentials



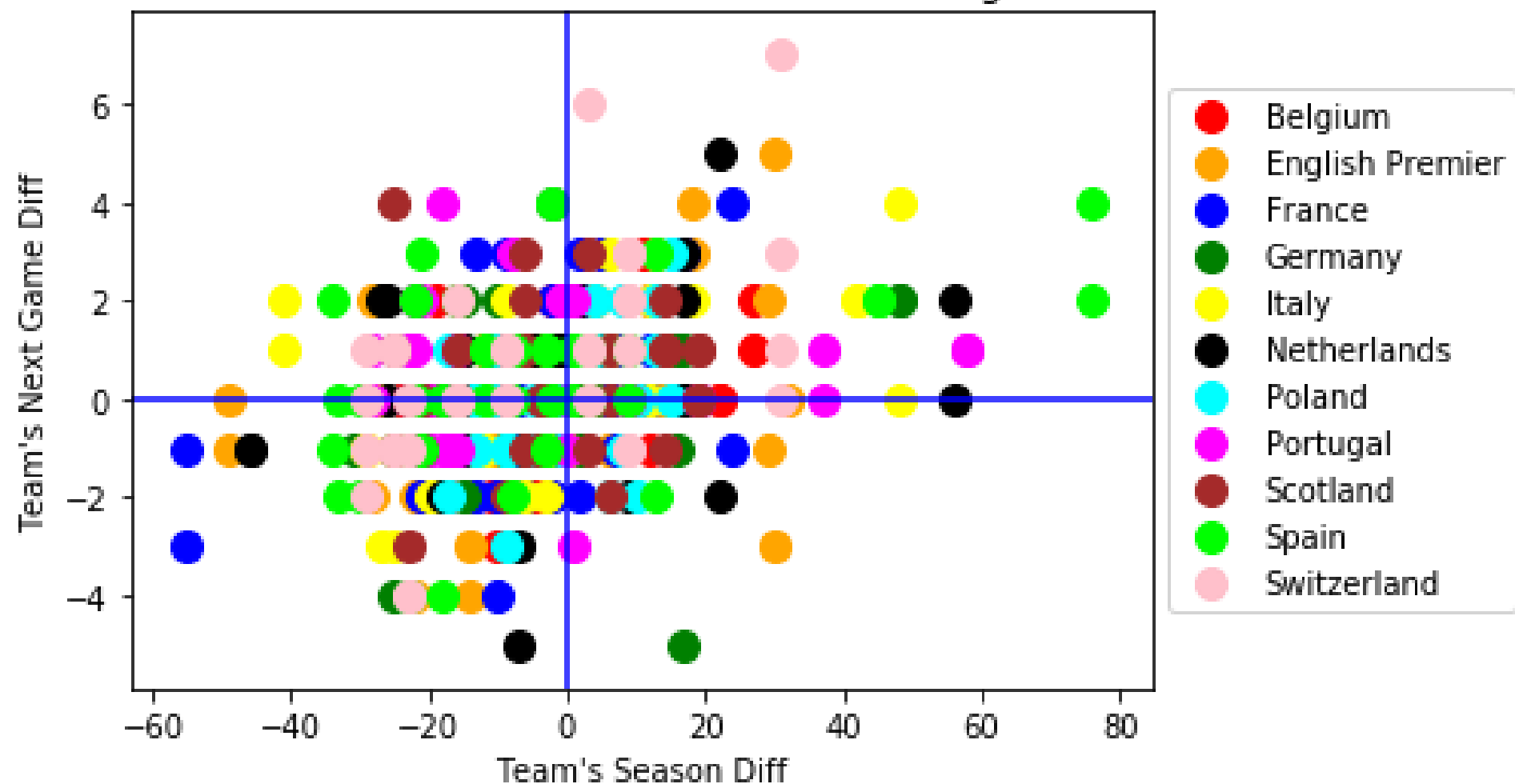
# Dom Game Performance on Team Strength

Team's Diff against Dominant Team



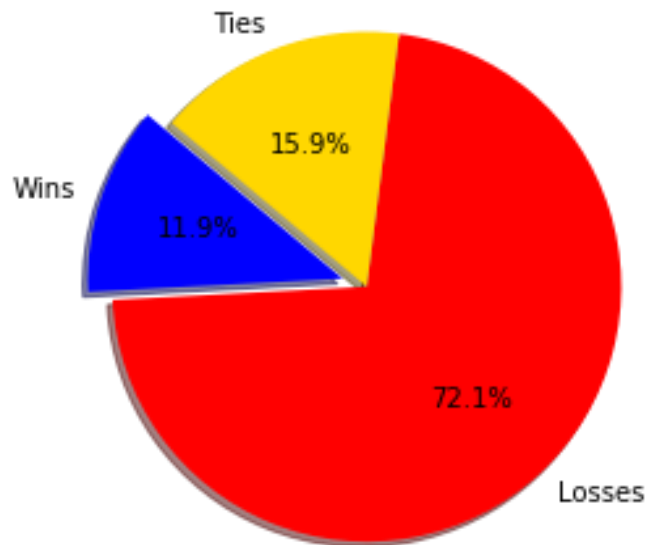
- Belgium
- English Premier
- France
- Germany
- Italy
- Netherlands
- Poland
- Portugal
- Scotland
- Spain
- Switzerland

Next Game Performance on Team Strength

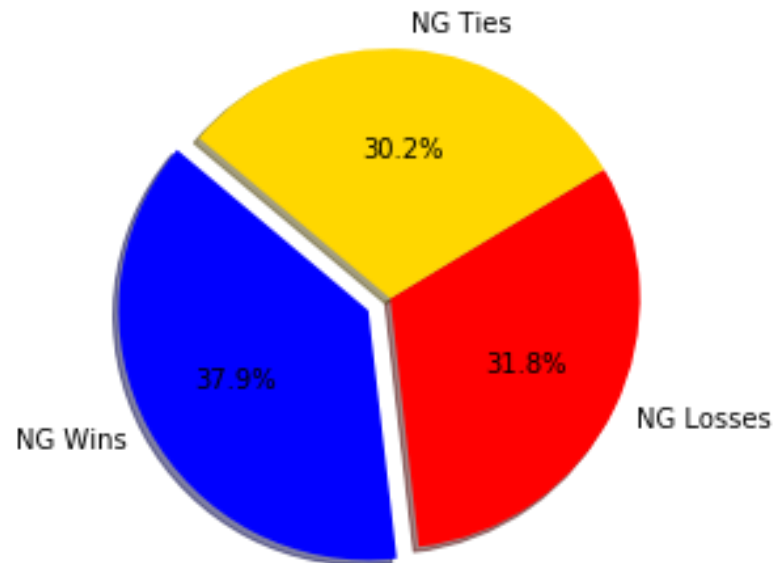


# W/L/T against Dominant Team and in the Following Game

W/L/T Percentage for Games Against Dominant Teams in Every League

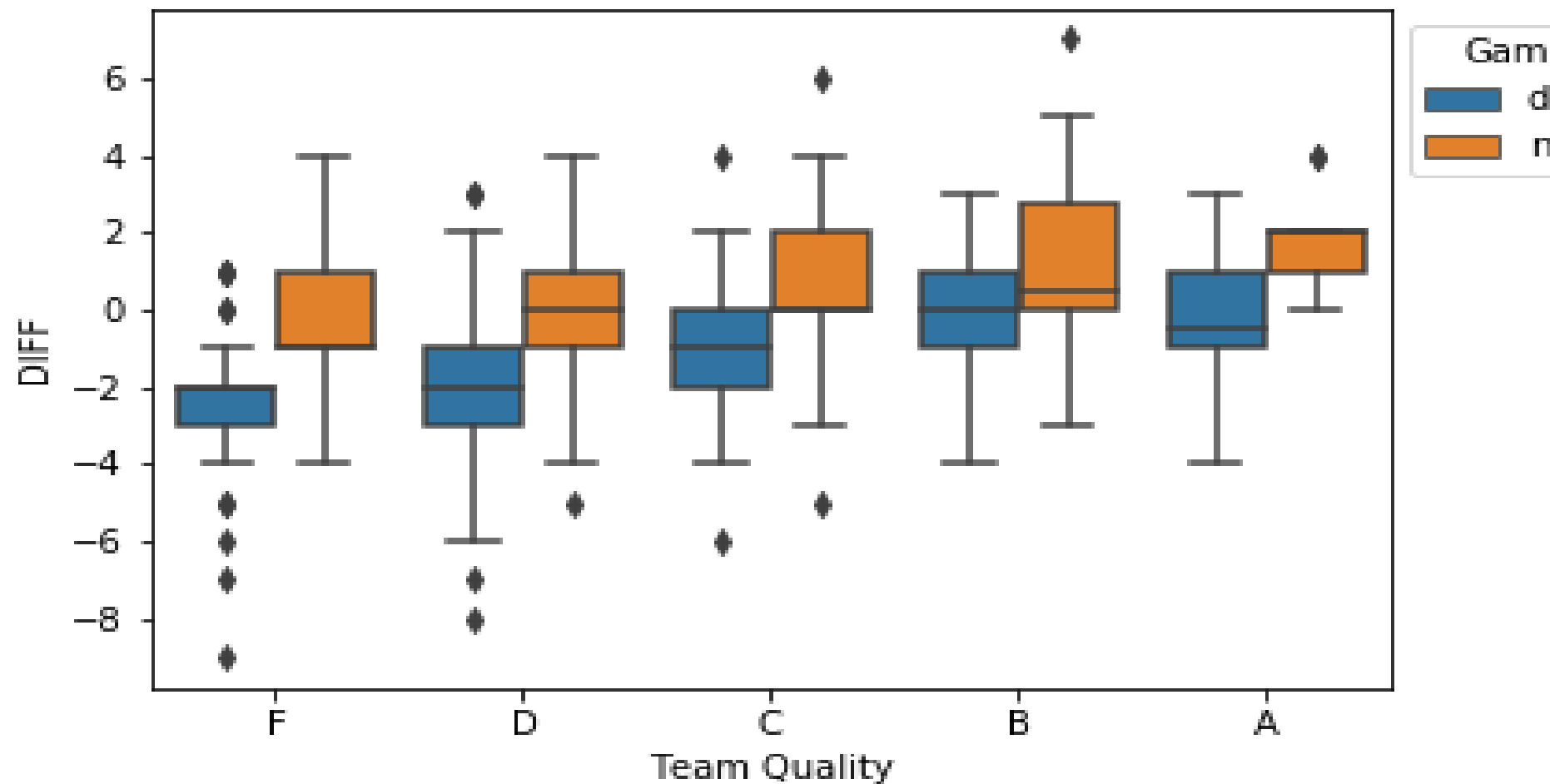


Next Game W/L/T After Playing a Dominant Team

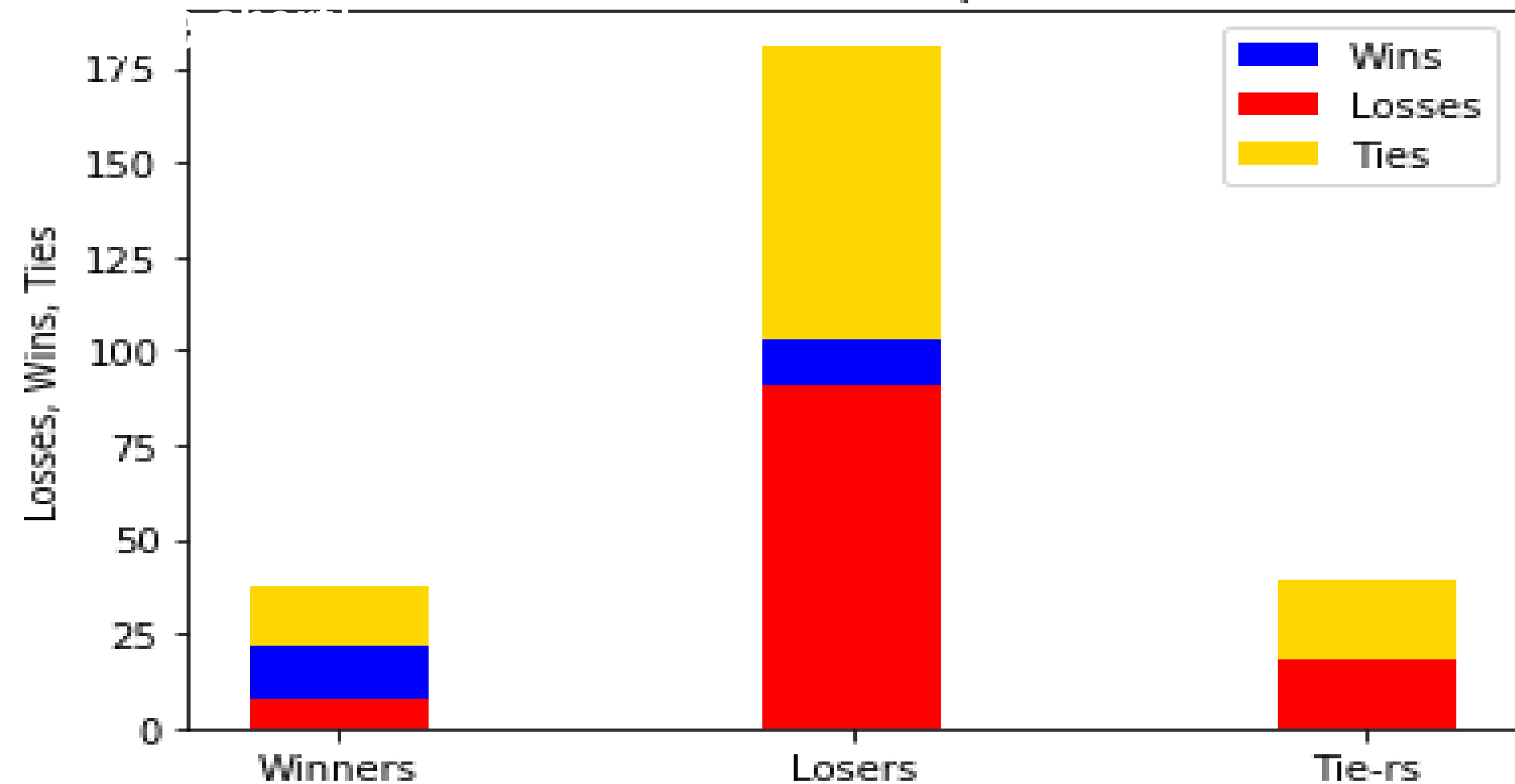




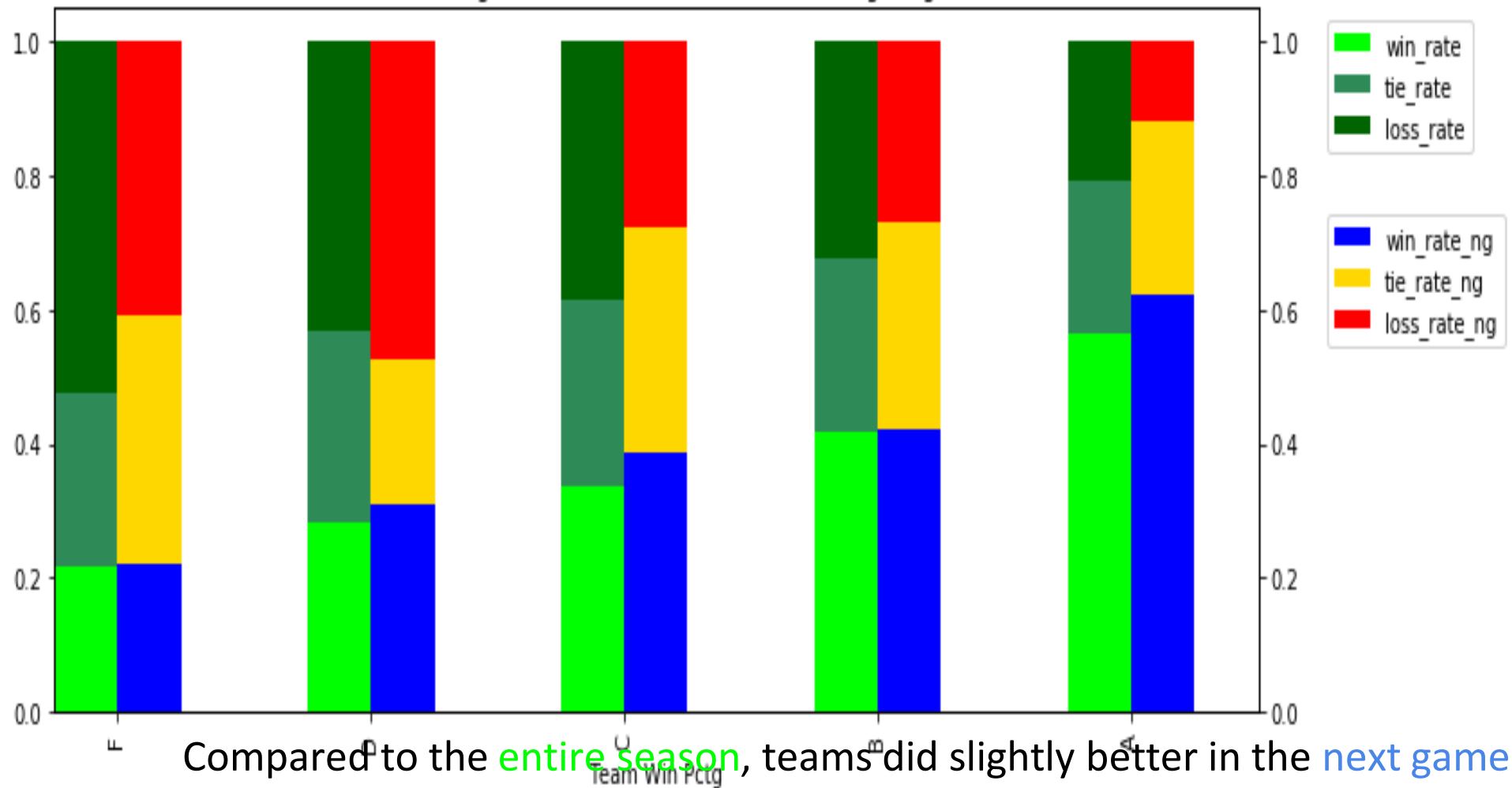
How teams fared against Best Team and in Next Game



How Winners, Losers, & Tie-rs perform in NextGame



Season winning % (tri-color) Vs Next Game winning % (green)



## What we should/could have done

- Compare outcome in next game vs odds in next game based on two teams' diff's
- Histogram for team winning percentages (skewed vs. normal distribution for Major League Baseball)
- Could have used multiple seasons
- Used additional dimensions (time of possession) for team quality including for the best team