
EA SPORTS
FIFA 20
FIFA
OFFICIAL LICENSED PRODUCT
CHAMPIONS EDITION
Standard Chartered
LFC

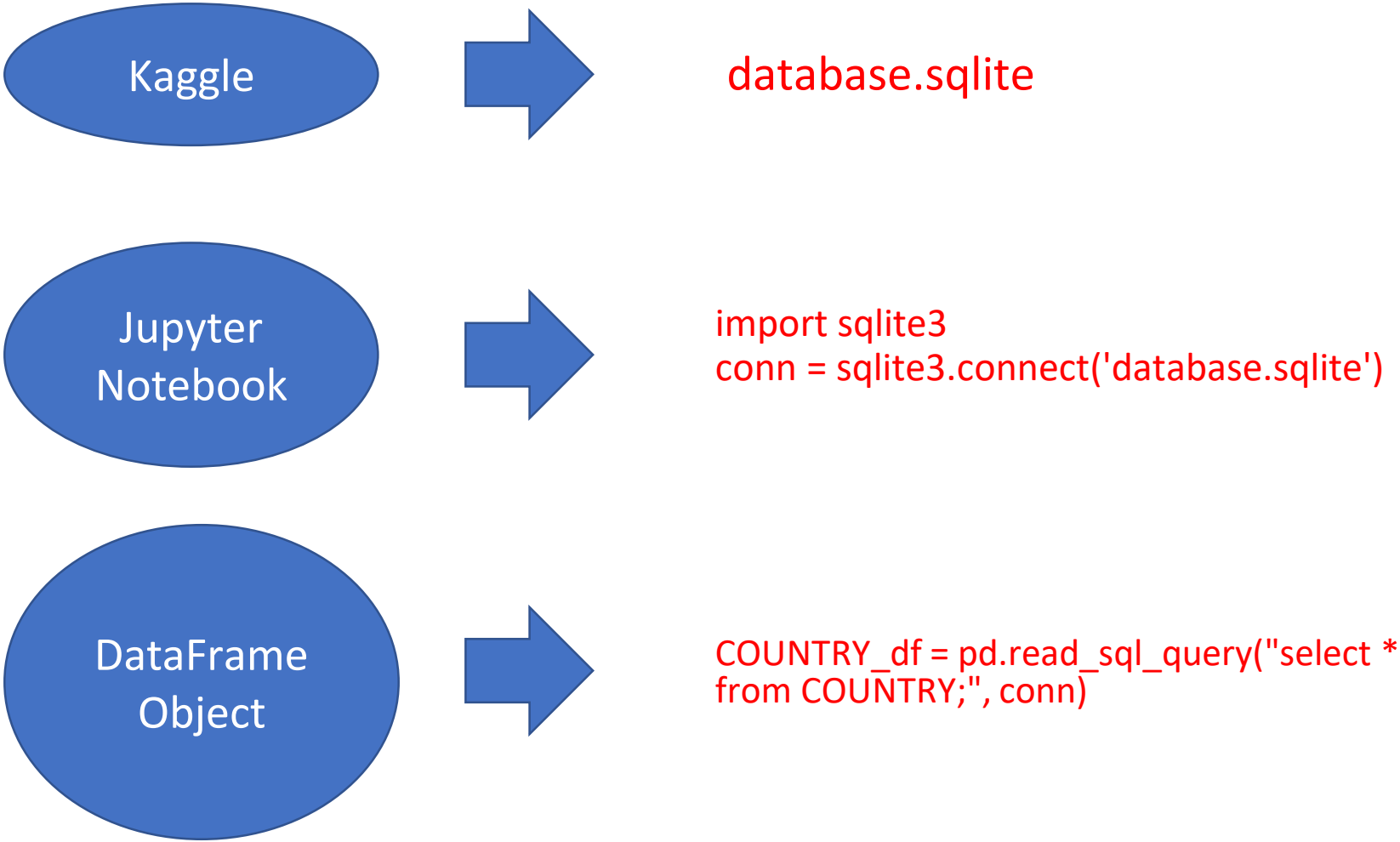
- Question: How do teams fare in the next game after playing the dominant team?
  - Are teams with better records more likely to beat the dominant team? Are they more likely to win the next game?
  - Has playing the dominant team changed their odds in the next game?

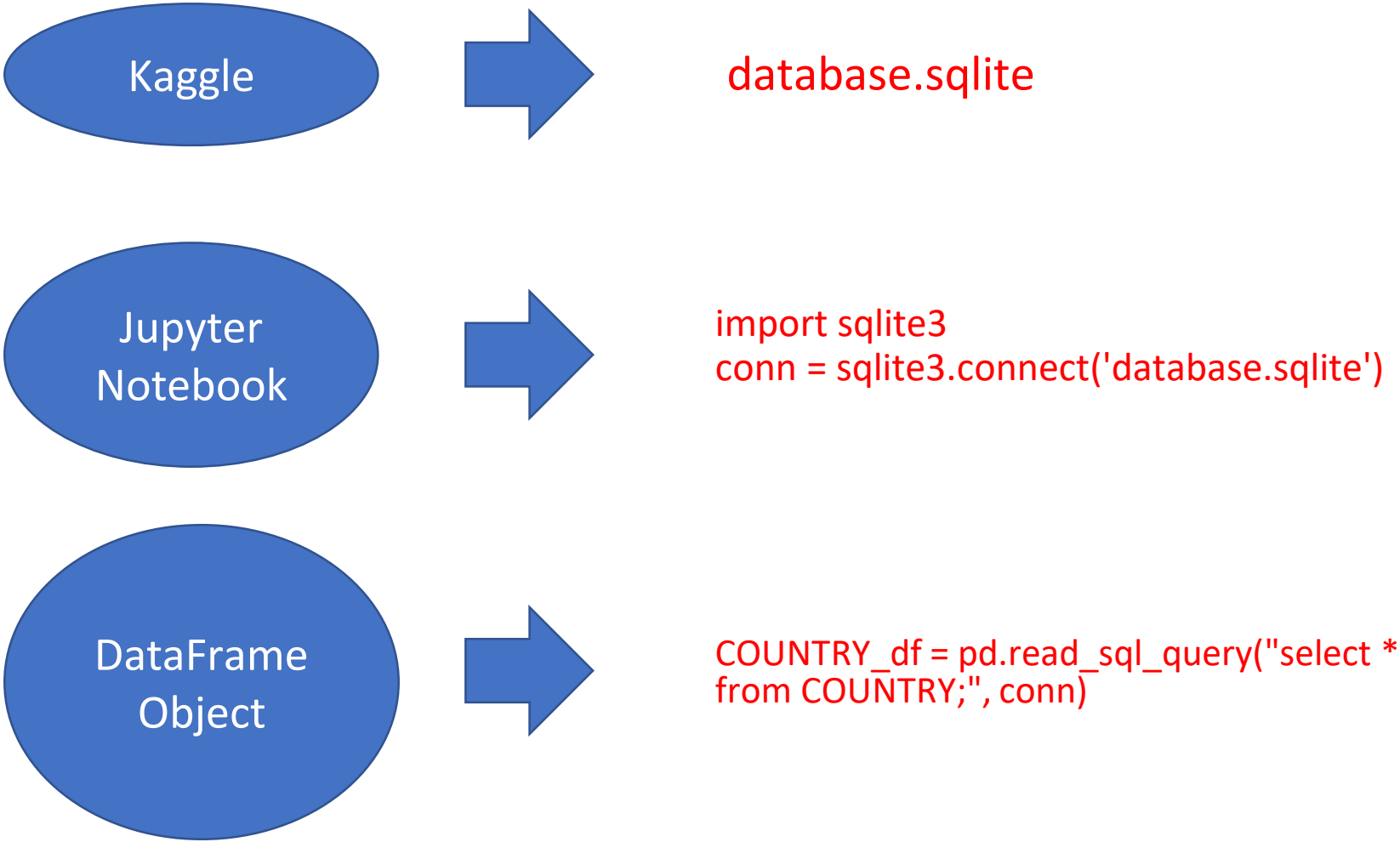
Data & Methodology

- **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

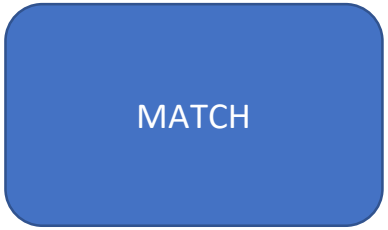
id	name
1	Belgium
1729	England
4769	France
7809	Germany
10257	Italy
13274	Netherlands
15722	Poland
17642	Portugal
19694	Scotland
21518	Spain
24558	Switzerland

TEAM

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

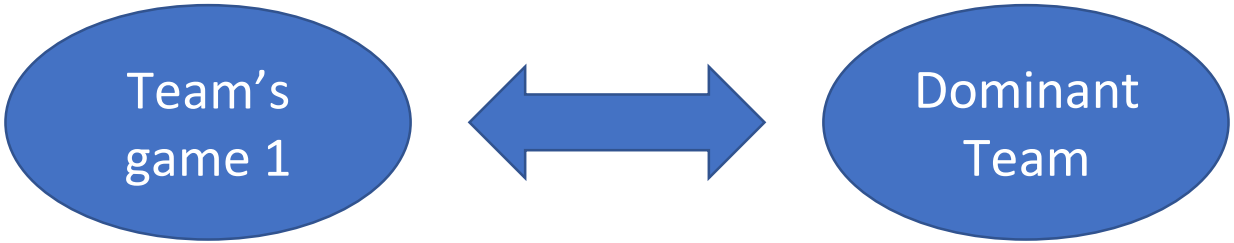
LEAGUE

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

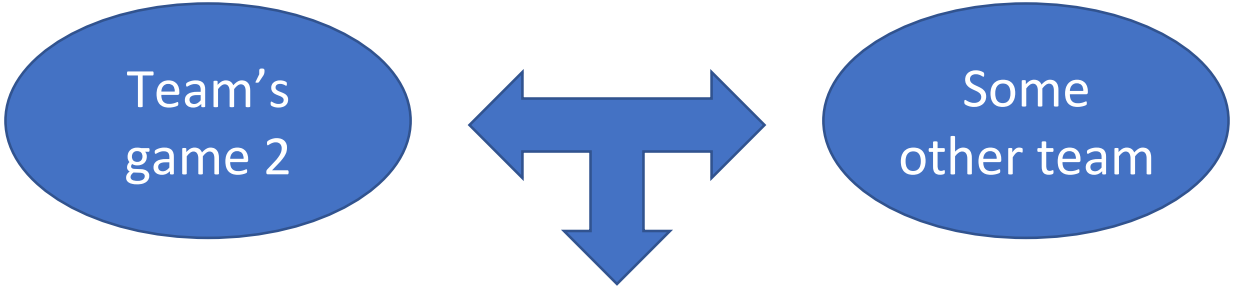


date	league_id	home_team_api_id	away_team_api_id	home_team_goal	away_team_goal
2015-07-24 00:00:00	1	9997	8342	2	1
2015-07-25 00:00:00	1	8571	9985	2	1
2015-07-25 00:00:00	1	9987	1773	3	1
2015-07-25 00:00:00	1	8573	8203	3	1
2015-07-25 00:00:00	1	10000	9994	3	1
2015-07-26 00:00:00	1	10001	9991	1	1
2015-07-26 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 …

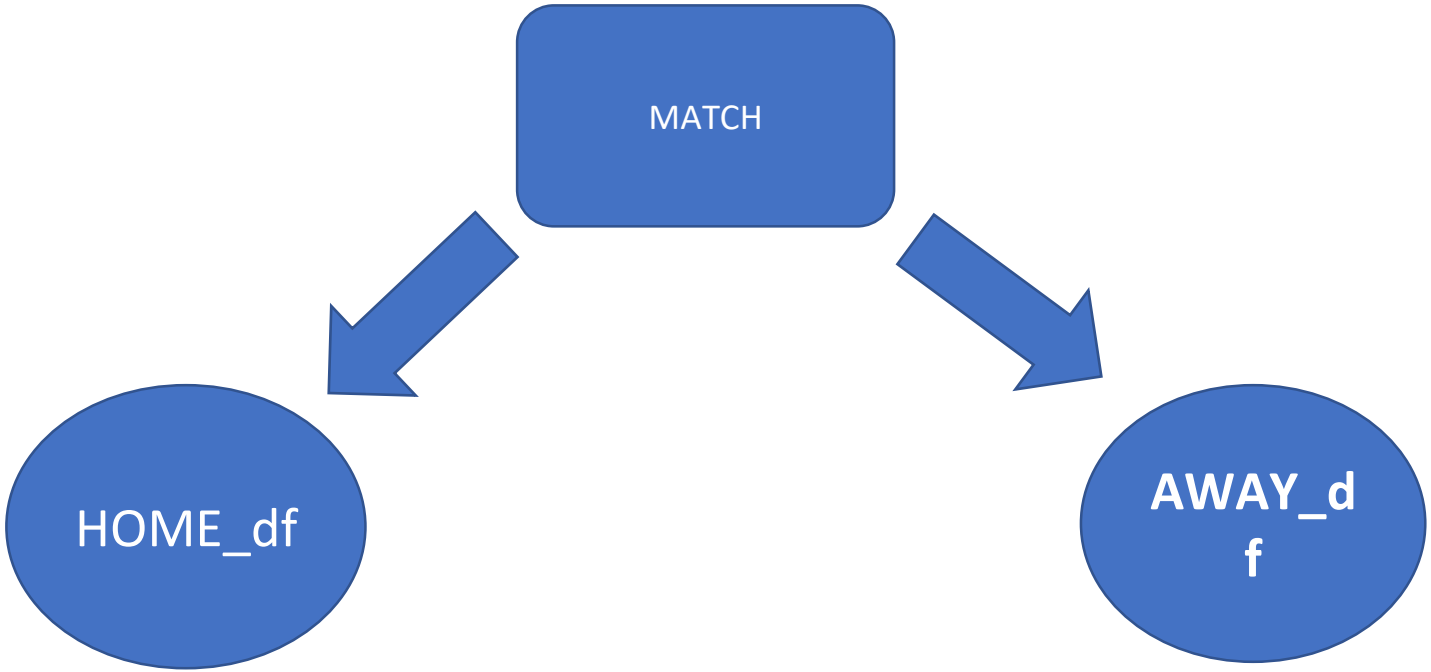


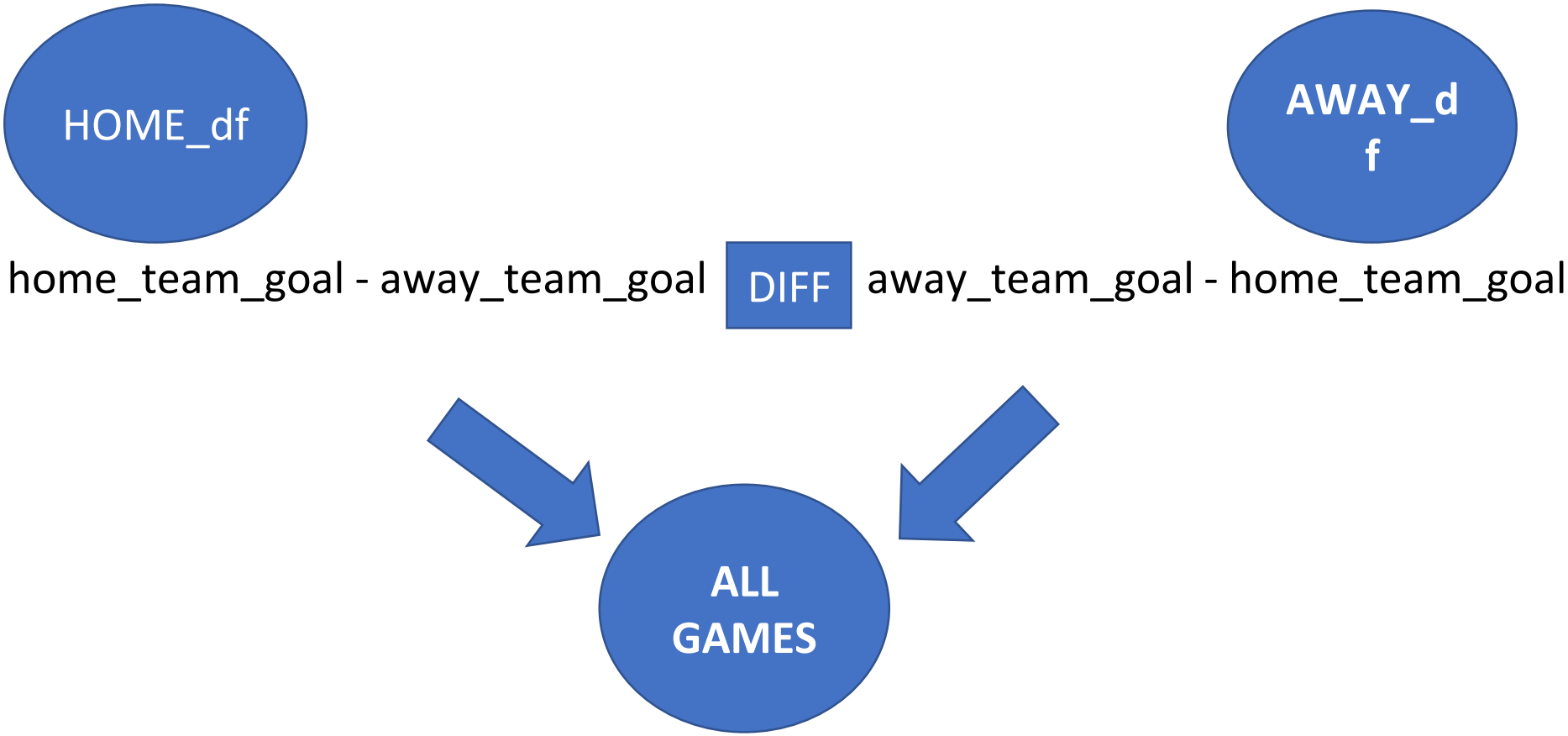
… what happened in the next game?



Win Loss or Tie? By how much?

Split out Home And Away Games




HOME_df
AWAY_df
home_team_goal - away_team_goal
DIFF
away_team_goal - home_team_goal
ALL GAMES
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 male a new dataframe
  Changes to next_game would apply to all_game

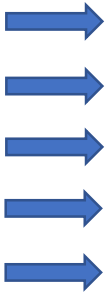
Use:

- 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

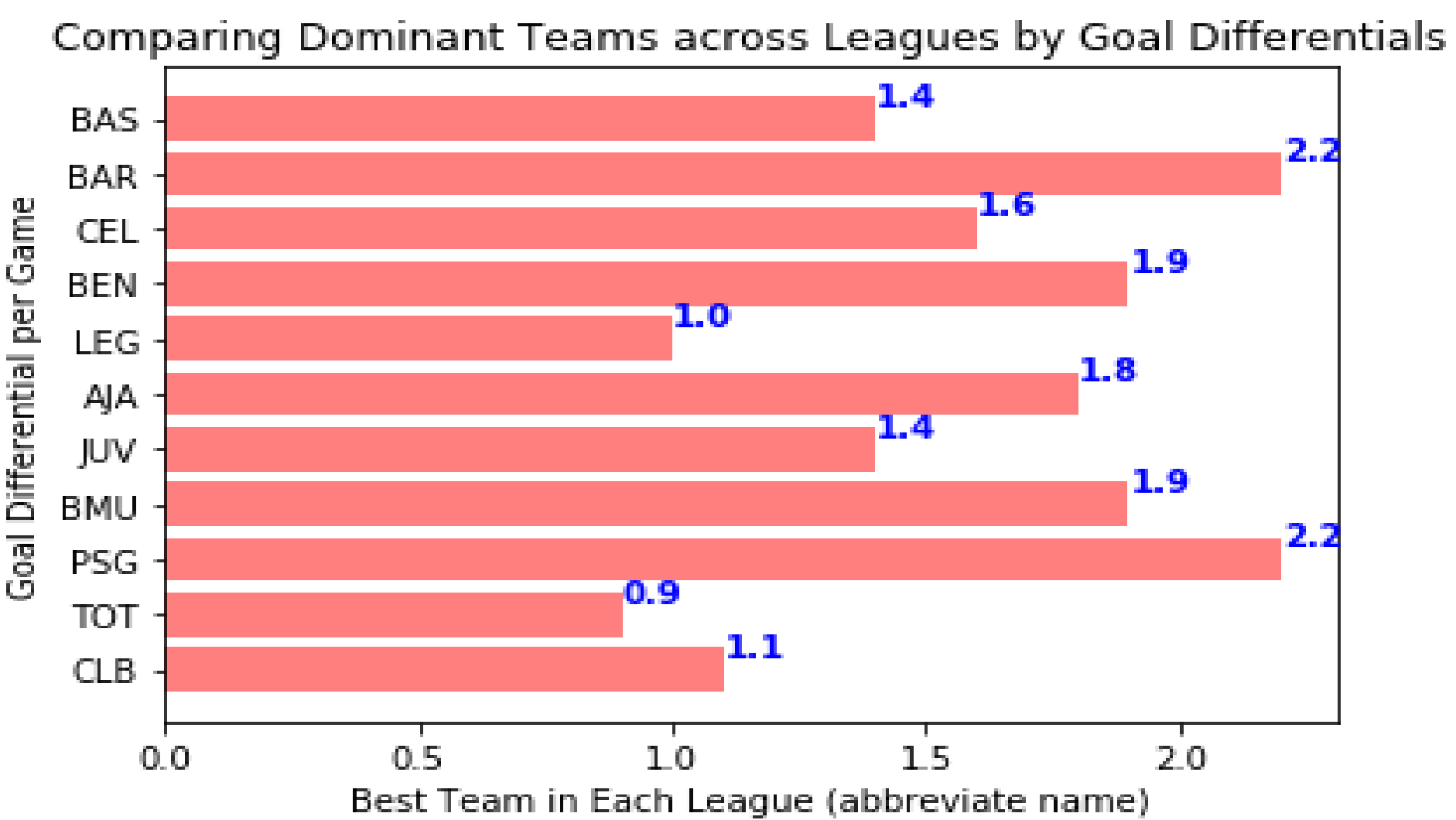
team_api_id	DIFF	league_id	opponent	W	L	T
1601	1	15722	8245	1	0	0
1601	-1	15722	1957	0	1	0
1601	-2	15722	8021	0	1	0
1601	-1	15722	2186	0	1	0
1601	0	15722	2182	0	0	1

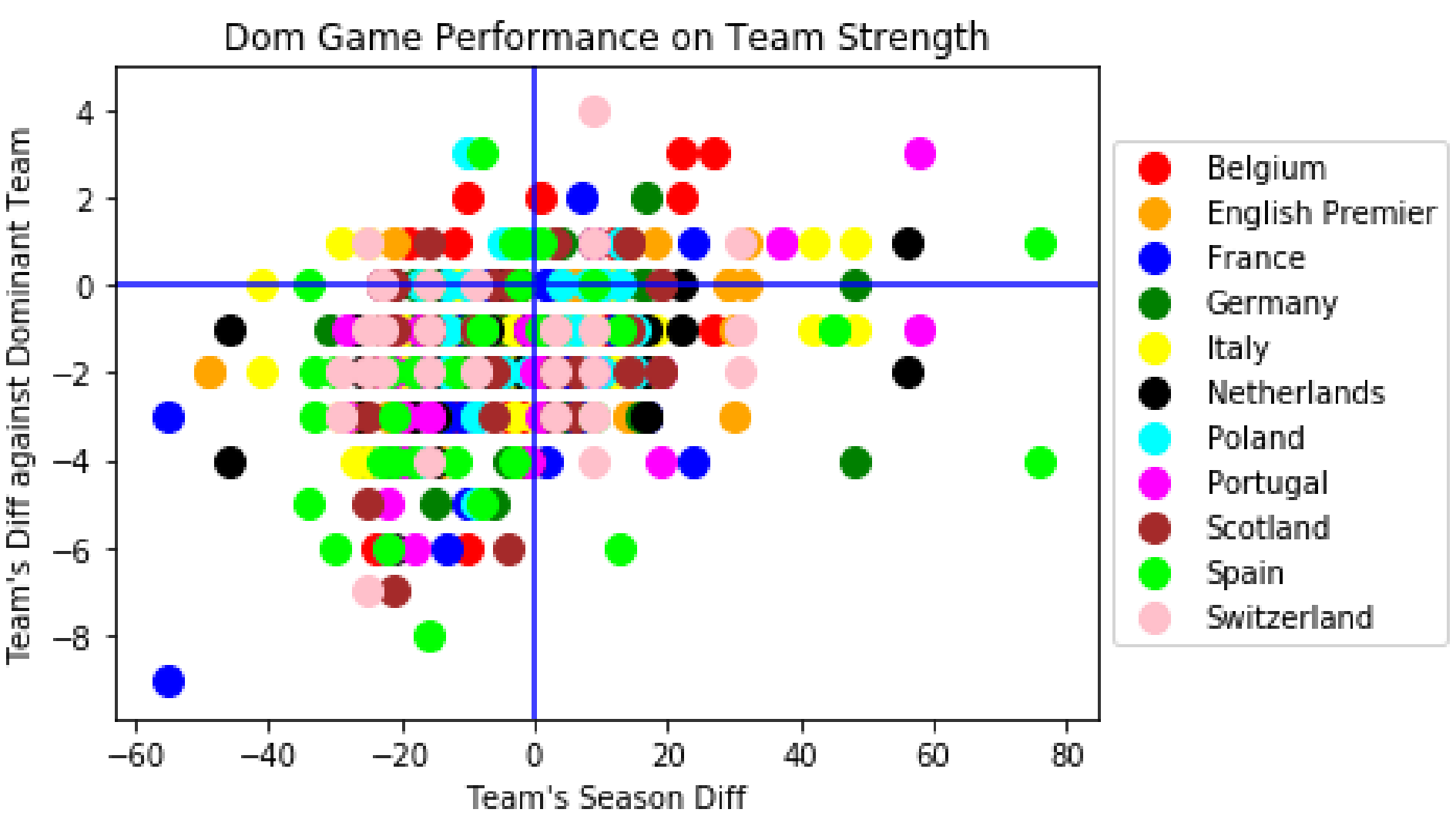


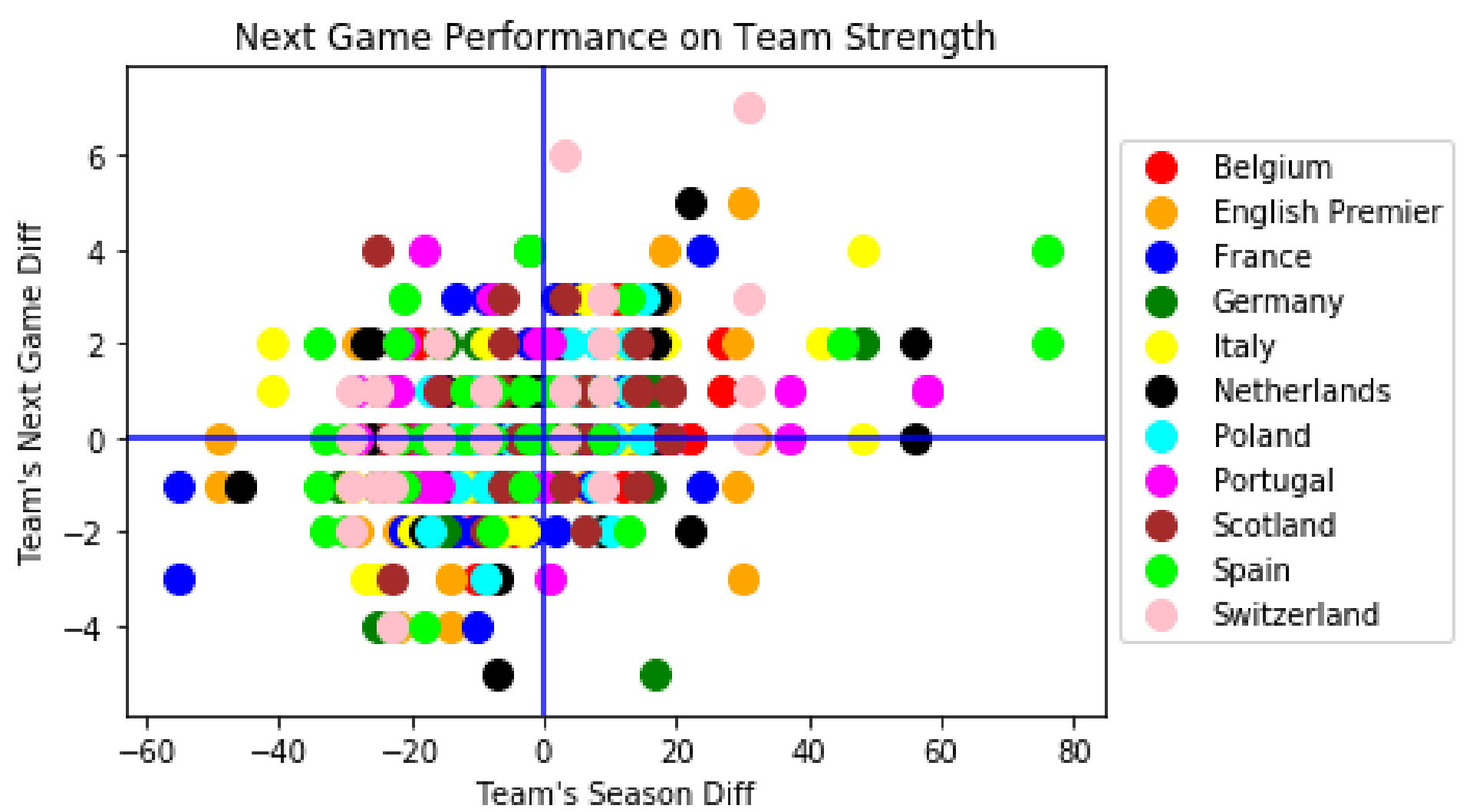
next_game

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

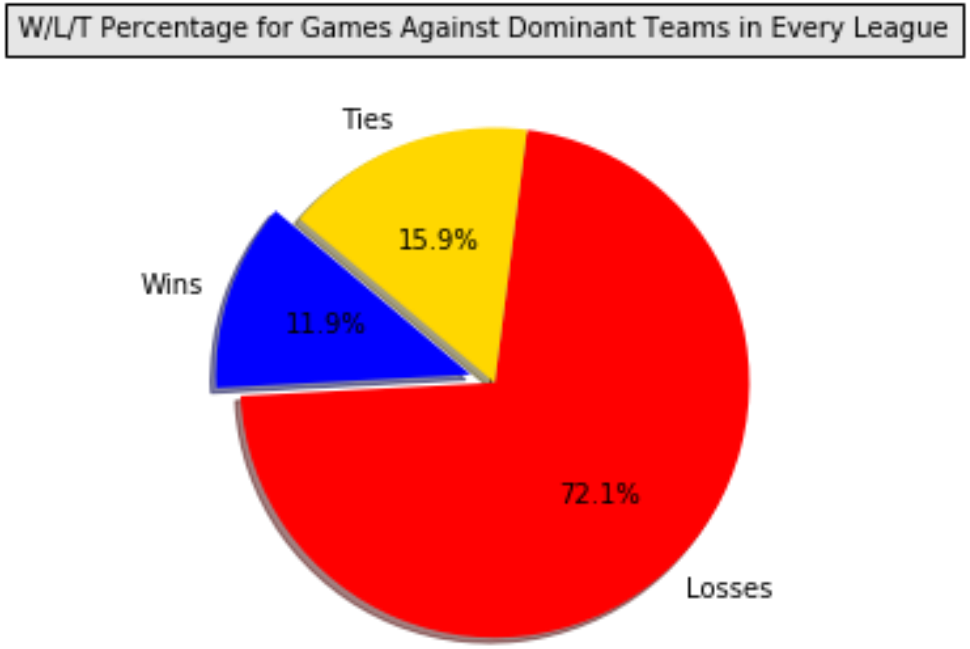
- df_complete = pd.merge(all_games, next_game.shift(-1).fillna(0), left_index=True, right_index=True)
- Select 332 rows with game against Dominant Team and Next Game

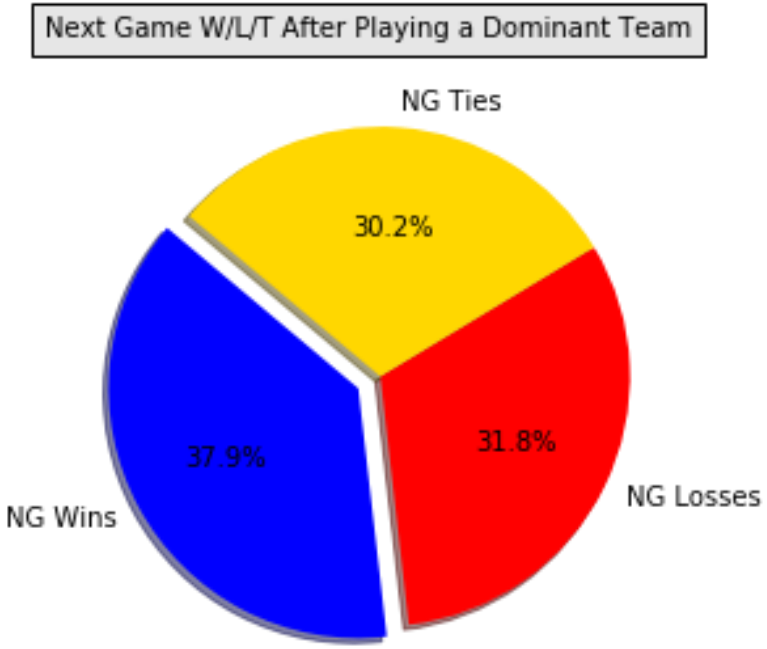

Comparing Dominant Teams across Leagues by Goal Differentials
Goal Differential per Game
BAS
BAR
CEL
BEN
LEG
AJA
JUV
BMU
PSG
TOT
CLB
1.4
2.2
1.6
1.9
1.0
1.8
1.4
1.9
2.2
0.9
1.1
0.0
0.5
1.0
1.5
2.0
Best Team in Each League (abbreviate name)

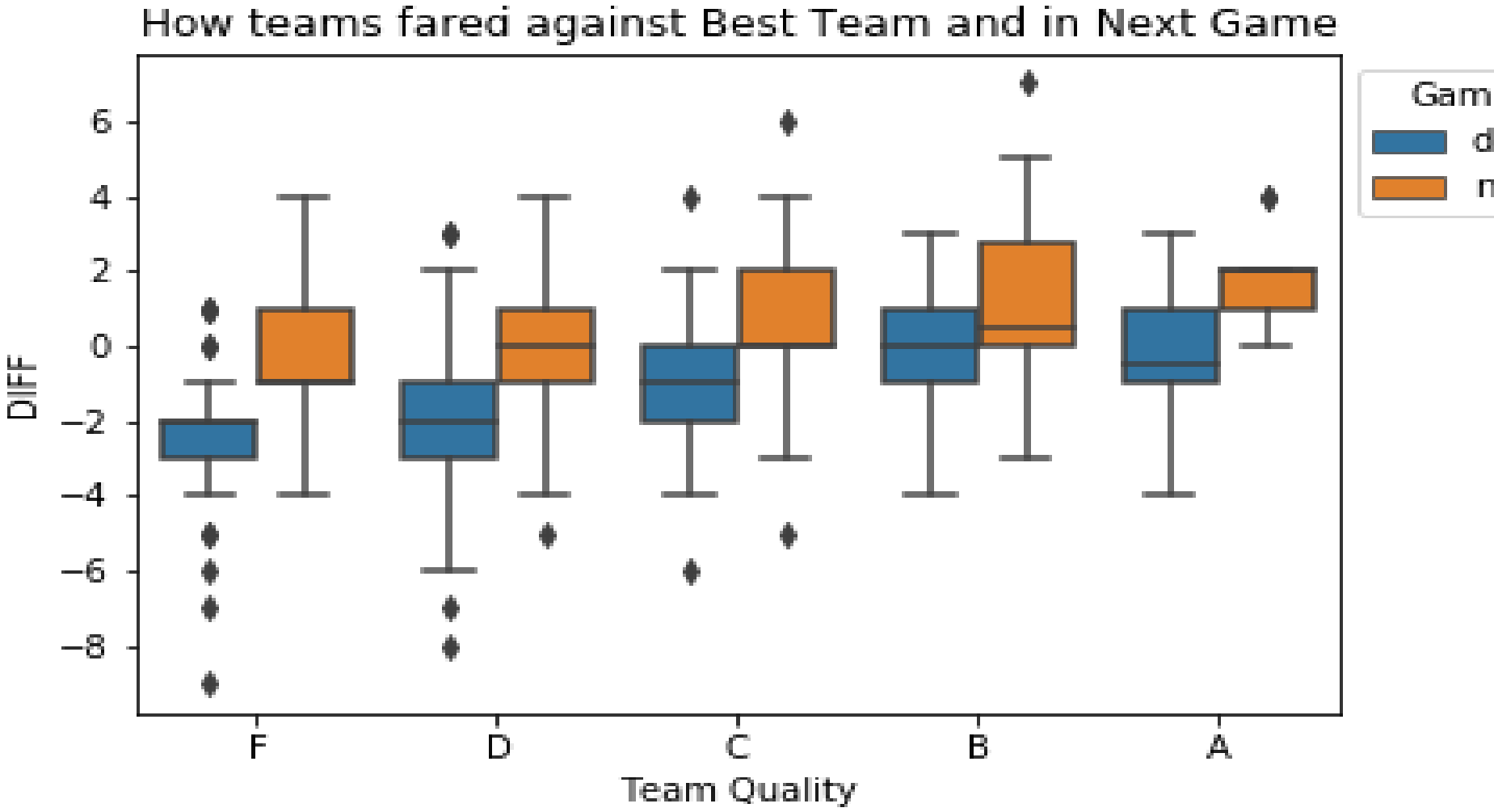

Dom Game Performance on Team Strength
Team's Diff against Dominant Team
Team's Season Diff
4
2
0
−2
−4
−6
−8
−60
−40
−20
0
20
40
60
80
Belgium
English Premier
France
Germany
Italy
Netherlands
Poland
Portugal
Scotland
Spain
Switzerland

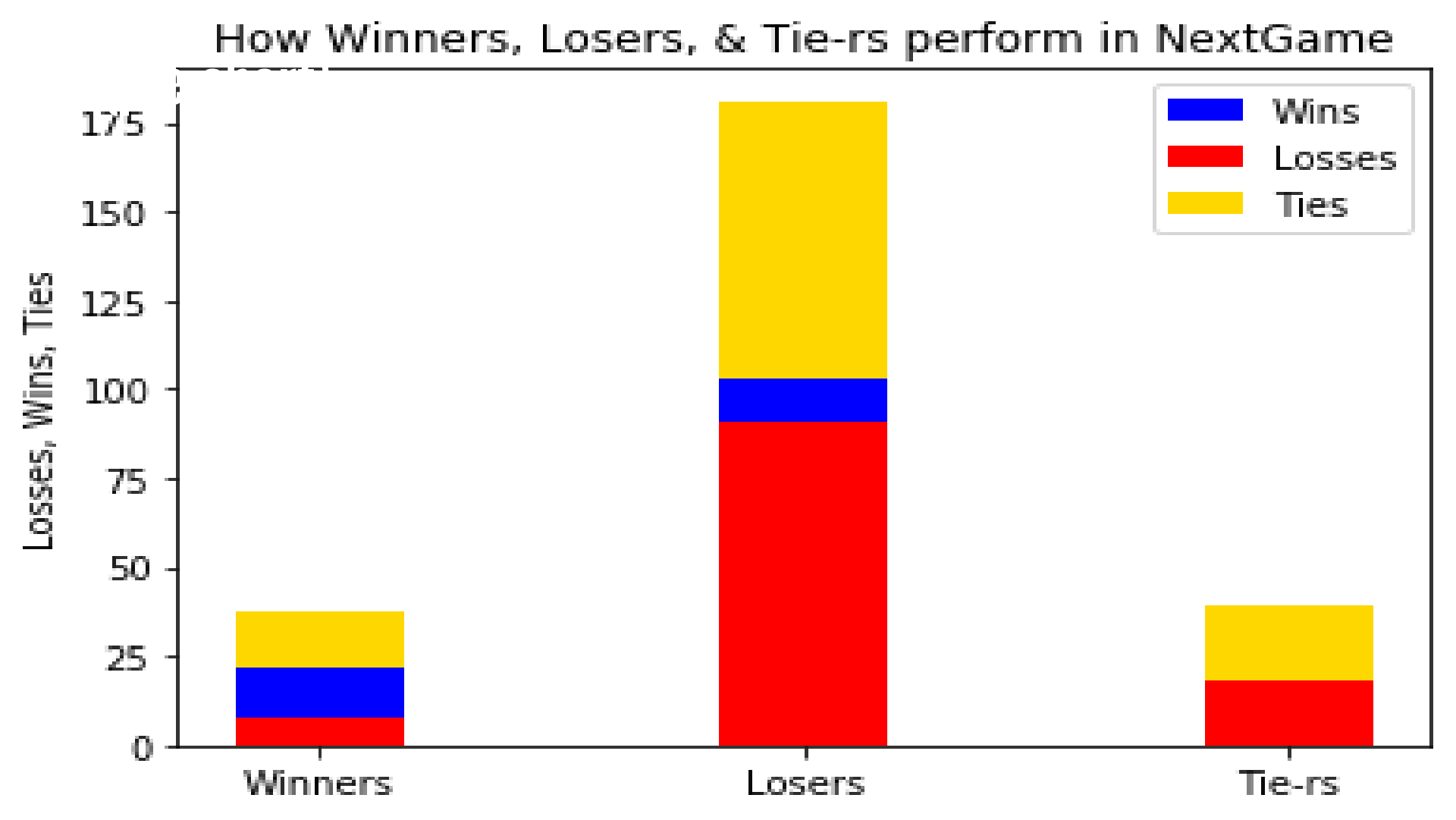

Next Game Performance on Team Strength
Team's Next Game Diff
Team's Season Diff
6
4
2
0
−2
−4
−60
−40
−20
0
20
40
60
80
Belgium
English Premier
France
Germany
Italy
Netherlands
Poland
Portugal
Scotland
Spain
Switzerland

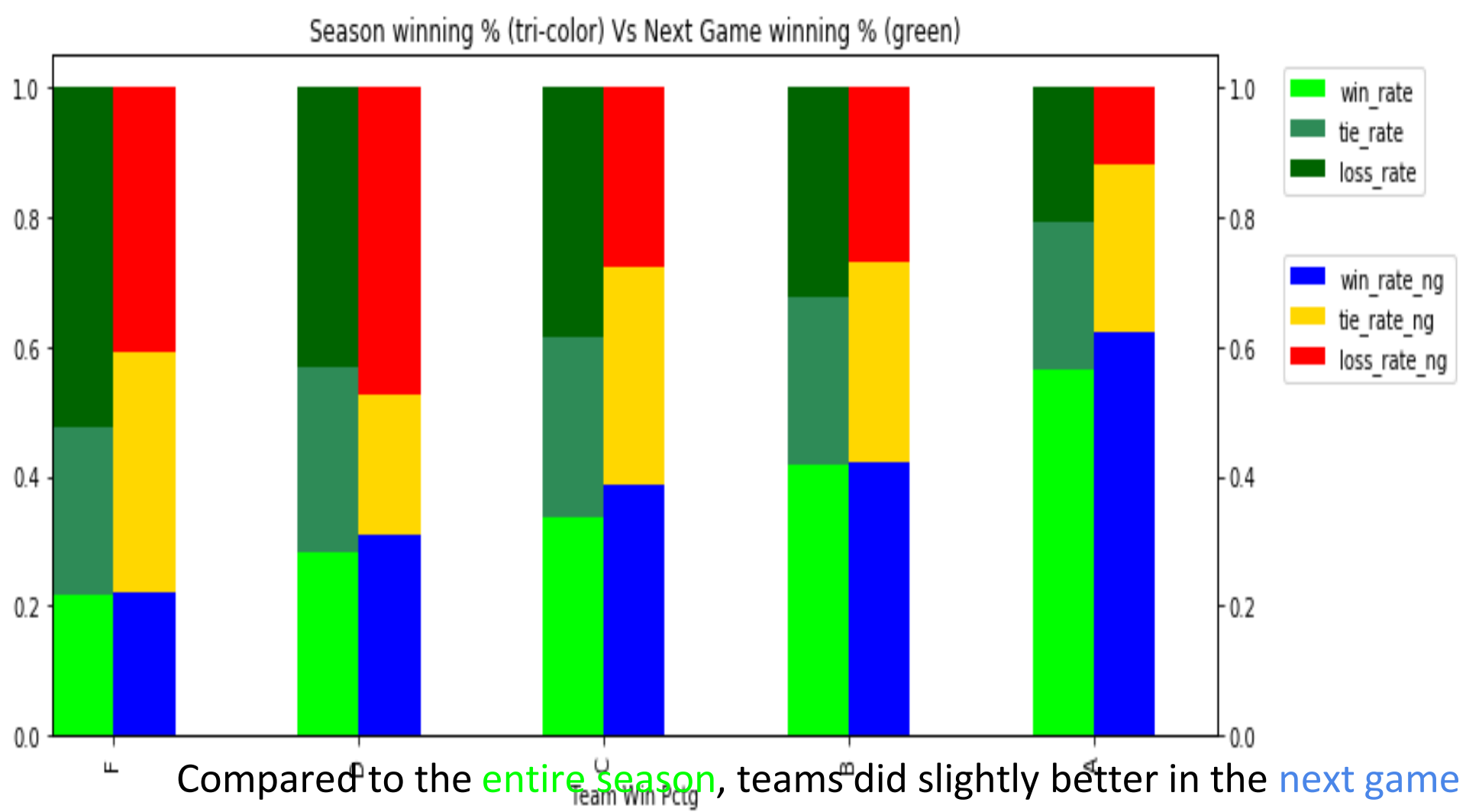
W/L/T against Dominant Team and in the Following Game






How teams fared against Best Team and in Next Game
DIFF
6
4
2
0
−2
−4
−6
−8
F
D
C
B
A
Team Quality
Gam
d
n


How Winners, Losers, & Tie-rs perform in NextGame
Losses, Wins, Ties
0
25
50
75
100
125
150
175
Wins
Losses
Ties
Winners
Losers
Tie-rs



Compared to the entire season, teams did slightly better in the next game

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