


Project about “Барсиків”

Inna, Vasyl, Yurii, Oleh, Yulia, Dmytro

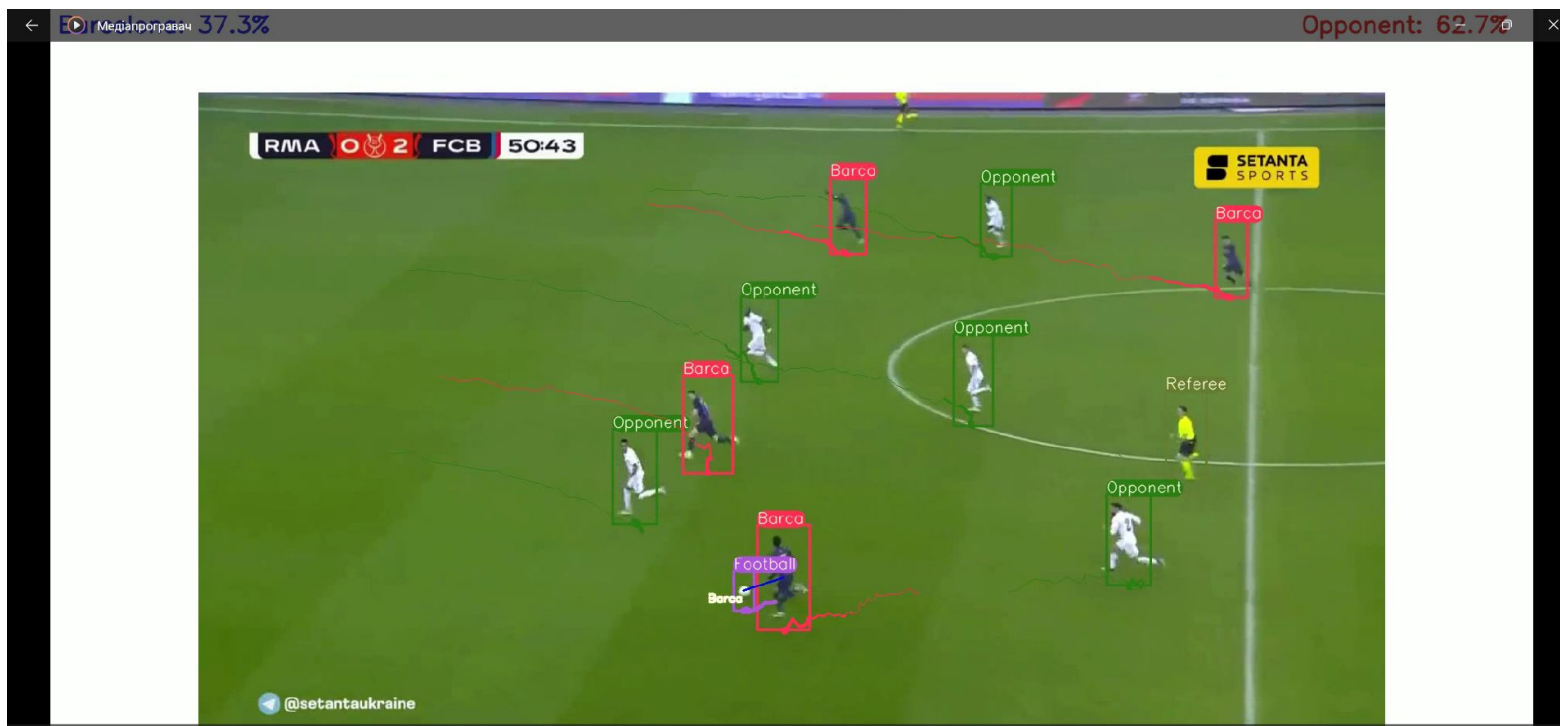


Agenda

- computer vision part;
- betting;
- regressions;
- some analytics;
- conclusions;

Computer vision part

Мета: визначення володіння м'ячем



Computer vision part

Video

1. Завантаження відео з youtube.
2. Анотація даних - встановлення міток для гравців команди Барселона, їх противників, рефері і м'яча
3. Експорт датасету розділеного на train, validation and test частини до проекту (використання Roboflow)
4. Навчання моделі YOLOv7 на створеному датасеті(визначено ваги і графіки ефективності, confusion matrix)
5. Визначення траєкторії польоту м'яча:
 - a) Спроба розпізнавання м'яча
 - b) Якщо м'яч не виявлено, використання інформації про напрямок і швидкість руху м'яча для передбачення ймовірної траєкторії, яку вважаємо правильним шляхом польоту м'яча. (1 секунда)
 - c) Визначення володіння м'ячем на основі попереднього фрейму, якщо м'яч не виявлено протягом 1 секунди.
6. Обрахунок володіння м'ячем(верхній лівий і правий кути)

Computer vision part

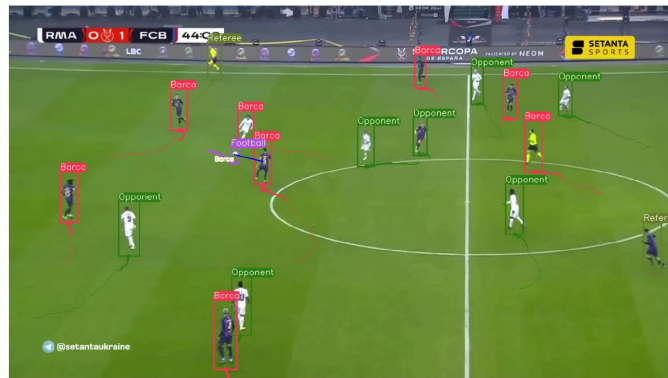
Проблеми при застосуванні YOLOv7:

1. При застосуванні моделі для визначення: обмежувальна коробка м'яча була прихована за обмежувальною коробкою гравців - намагалися вирішити перетреновуванням на окремих датасетах. Перший - рефері, гравці Барселони, противники, інший- м'яч.
2. М'яч пропадає в ногах у гравців - рішення жовтий квадрат є передбаченням куди полетить м'яч, спираючись на його напрямок і швидкість. Якщо м'яч зник то брати його знаходження з попереднього кадру

Примітка: для більшої універсальності використання бажано додати до датасету нарізки з різних матчів Барселони

Barcelona: 22.2%

Opponent: 77.8%



Impact of Unexpected Outcomes on Betting Strategy

Goals:

- detect how many Barcelona matches ended with “unexpected” results, i.e. result doesn’t match lowest betting odd.
- detect is there profitable long-run betting strategy(period - season), if user always bets against odds.
- detect correlation between number of “unexpected” results in season and profit/loss

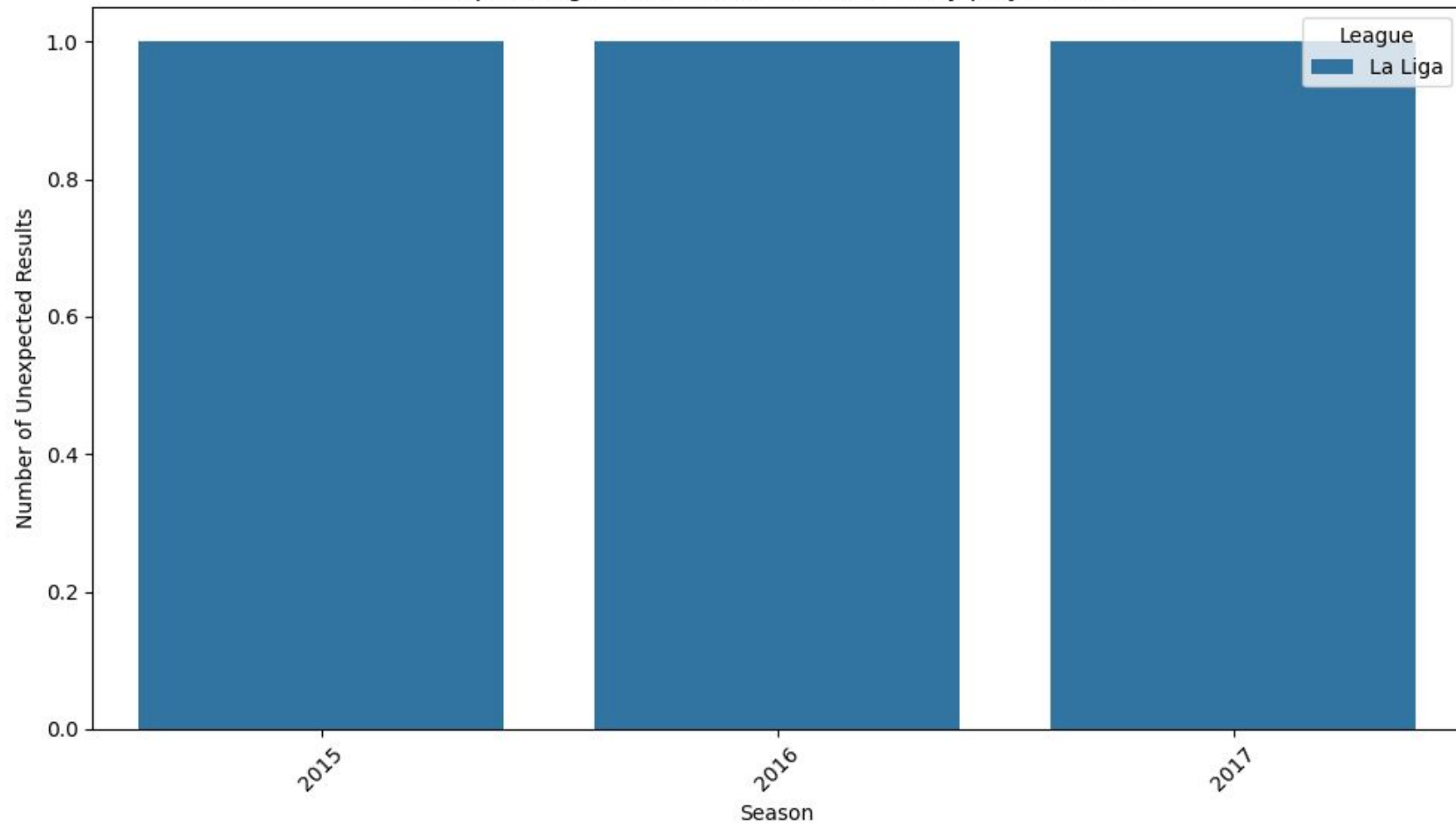
Input data: results of LaLiga in seasons 2014/2015 - 2020/2021



1%

Over the course of seven seasons, Barcelona exceeded expectations only three times, and interestingly, all three instances were in matches against Real Madrid. This leads to our first tip: Barcelona tends to overperform quite rarely.

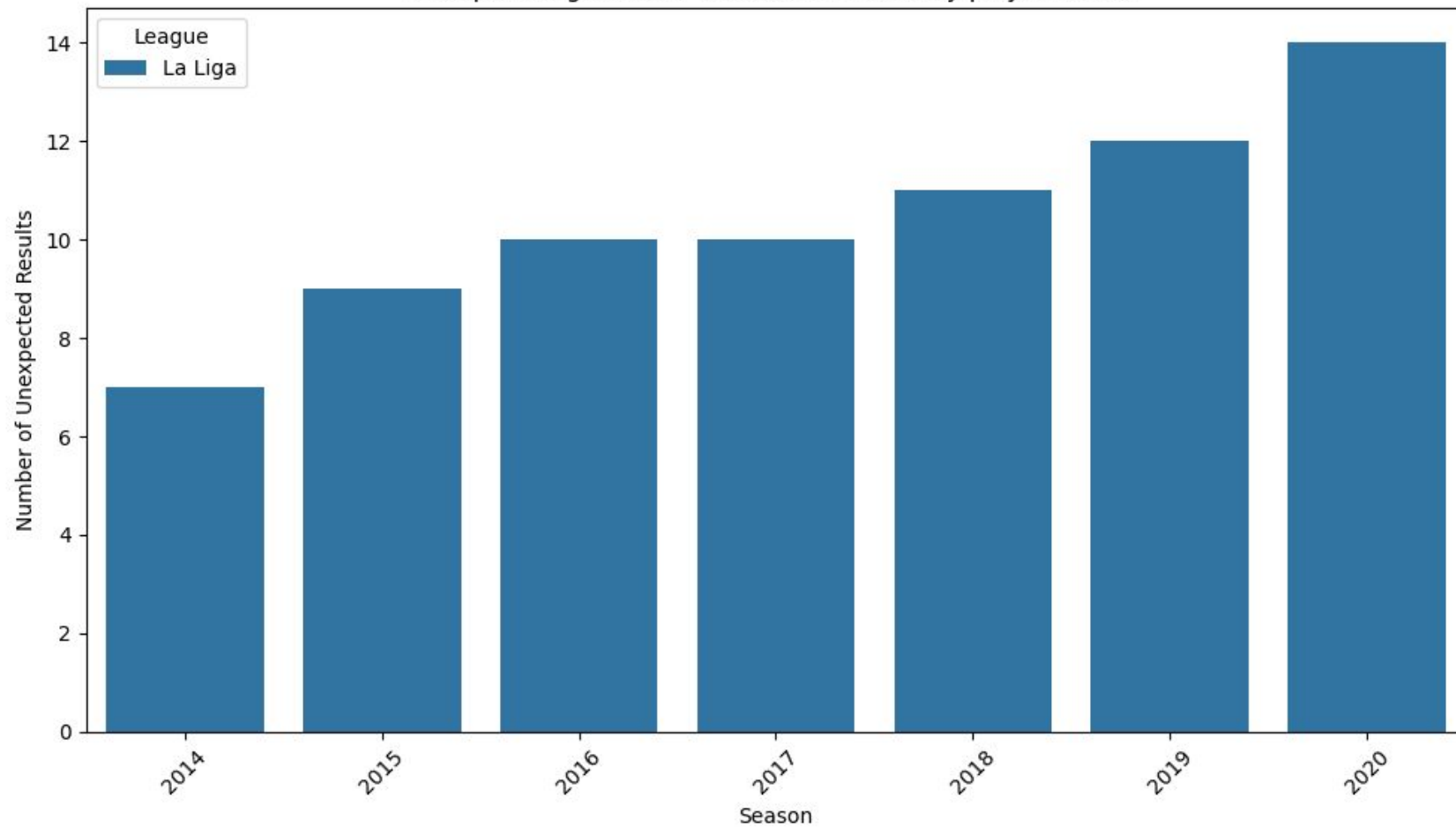
Unexpected games for Barcelona when they played better:



27%

Over a span of seven seasons, Barcelona underperformed compared to expectations on just 73 occasions. Notably, this number is on an upward trend.

Unexpected games for Barcelona when they played worse:



We analyzed two strategies for betting on unexpected results

- Linear - betting the same amount every game in season
- Catch-up strategy - betting amount is calculated by formula:

$$\mathbf{S = X + Y / K - 1}$$

S - is the stake or the amount of the required bet.

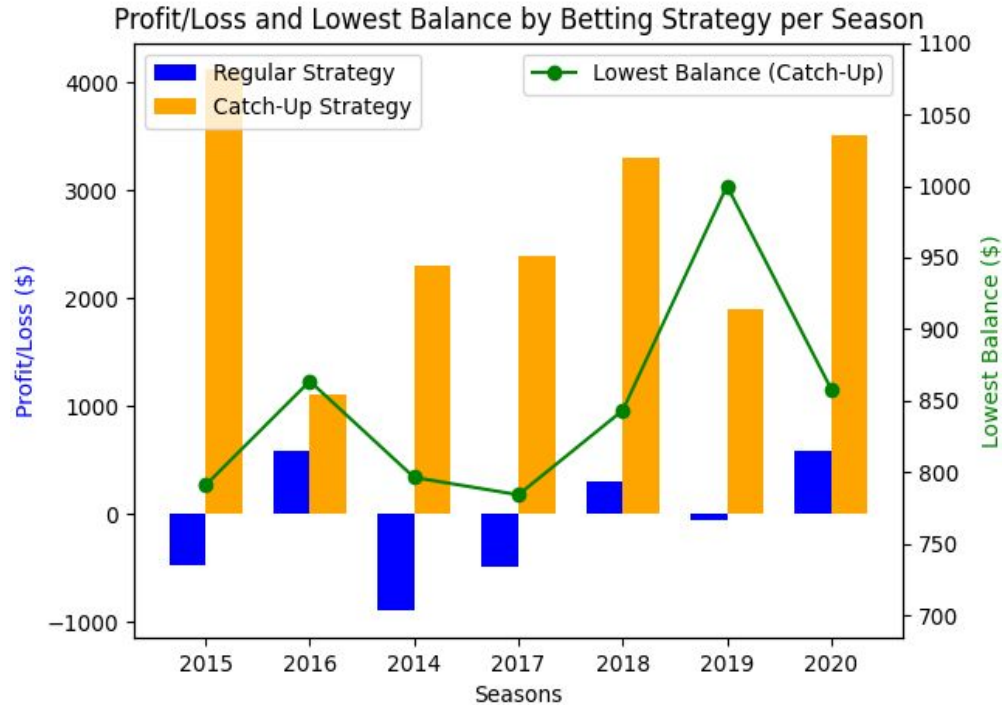
X - is the amount of potential profit from the first bet, minus the stake.

Y - is the sum of all previous losses.

K - is the odds of the upcoming event.

Conclusion

The linear betting strategy, as expected, resulted in losses in four out of seven seasons. In contrast, the catch-up strategy yielded profits in every season, with its least profitable season still outperforming the best season of the linear strategy by double. Interestingly, basing the betting strategy on 'unexpected' results did not correlate with overall profit. This is likely because profitability is more influenced by the sequence of 'unexpected' matches rather than their total count in a season.



Modelling match results

Approach Description

- 2 models were trained : Logistic regression and BradleyTerry2 (R package "BradleyTerry2")
- Approach for BradleyTerry taken from here :

<https://link.springer.com/article/10.1007/s10994-018-5741-1>

- Dataset : all matches from years 2019-2020 from the dataset at

<https://www.kaggle.com/datasets/technika148/football-database>

- Dataset size : 1.8K / 0.9K train/oout
- Target : match winning

- Final selected features (in both models):
- % of wins in the last year of the "away" team
- % of wins in the previous season of the "home" team
- Average corners per match of the "away" team
- Average PPDA ever by the "away" team
- Average number of goals per match in the current season by the "away" team
- Average shotsOnTarget by the "home" team
- Average number of corners in the current season by the "away" team
- Average PPDA in the current season by the "home" team

Results

Logistic Regression : good result

- AUC 0.75/0.76 train/oout

% correct predictions (sample : without draws)				
cnt	bads	avg(x_br)	win_rate	pc_correct_F
875	475	58%	54%	69%

Win rate , actual vs predicted (sample : without draws)				
	pred_bin	bads	avg(prediction)	br
1	349	267	0,79	0,765
2	350	176	0,53	0,503
3	176	32	0,243	0,182

Win rate , actual vs predicted (sample : with draws)				
	pred_bin	bads	avg(prediction)	br
1	456	266	0,773	0,583
2	498	171	0,516	0,343
3	231	38	0,256	0,165

BradleyTerry

Approach failed to show good result on the same set dataset and same set of features (maybe, due to incorrect settings)

Win rate , actual vs predicted (sample : without draws)					
bin_num	pred_bin	cnt	bads	avg(prediction)	br
1	[-1,807,-0,123]	292	143	-47%	49%
2	[-0,123, 0,117]	292	166	0%	57%
3	[0,117, 1,907]	291	166	51%	57%

xG, xGA Comparison (Home vs Away)

Задача:

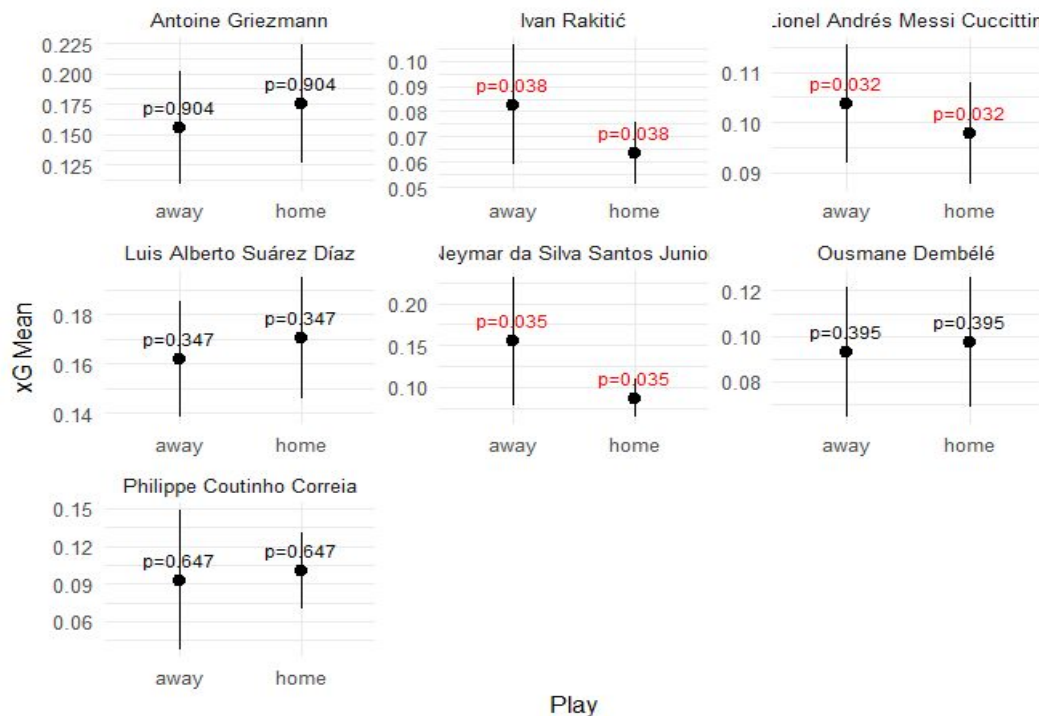
Порівняння показників гравців між домашніми та гостьовими матчами

Методологія:

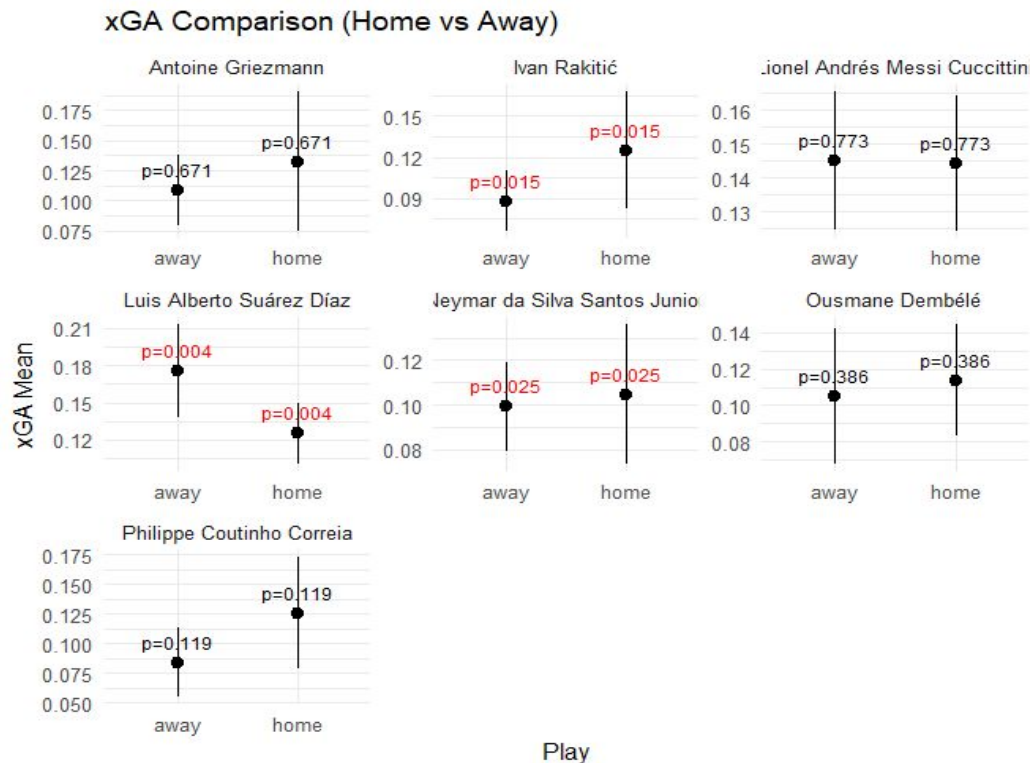
- Дані StatsBomb
- Матчі La Liga, 5 сезонів з 2016 до 2020
- Гравці з кількістю івентів для розрахунку xG, xGA більше 25 для домашніх та для гостьових ігор
- Без урахування пенальті
- Порівняння за t-test

xG Comparison (Home vs Away)

xG Comparison (Home vs Away)



xGA Comparison (Home vs Away)



“Hold the ball, read stats. Analytics helps you win”

- Yoda (maybe)



Datasets

- Transfermarkt dataset (for game events, players, etc.);
- StatsBomb (transfers);

Tasks

- find the best players group (connections graph);
- analyse young academy (Barcelona B players);
- some other приколы;

Player connections

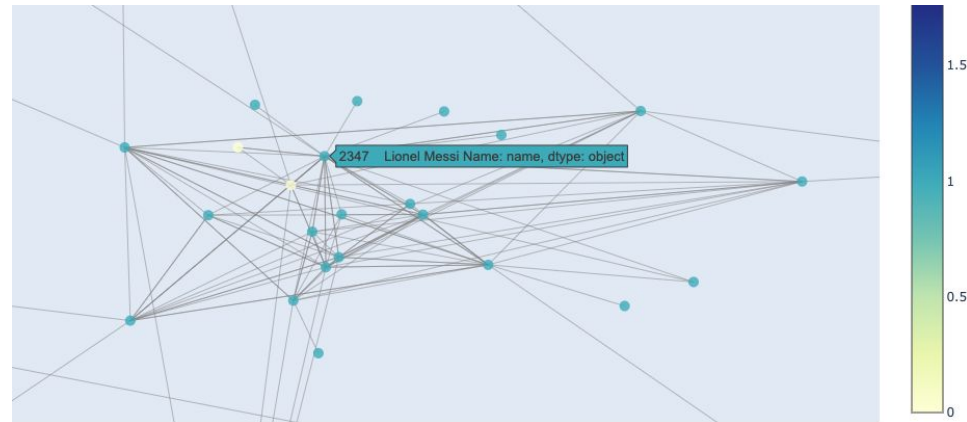
Tried to identify strong connections between some players

Used Leuvian method for building communities.

Results show that players that have more common seasons and games are more likely to be part of the same community.

Partitioning of 3 periods (Messi, Puyol, current one)

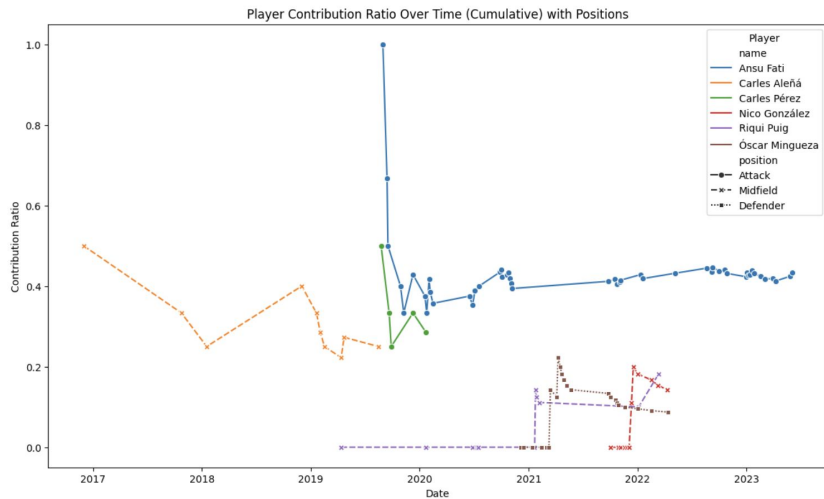
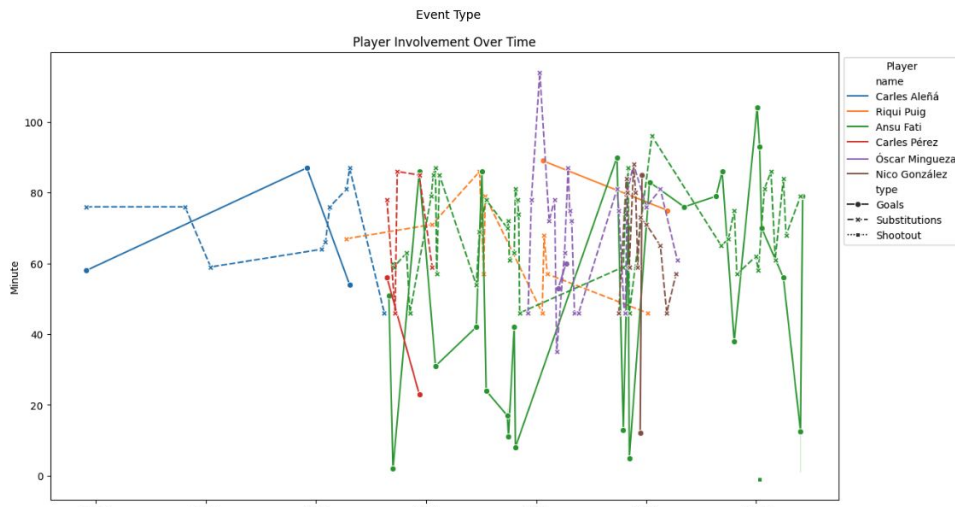
Potential use-case: analyse game lineups and potential performance



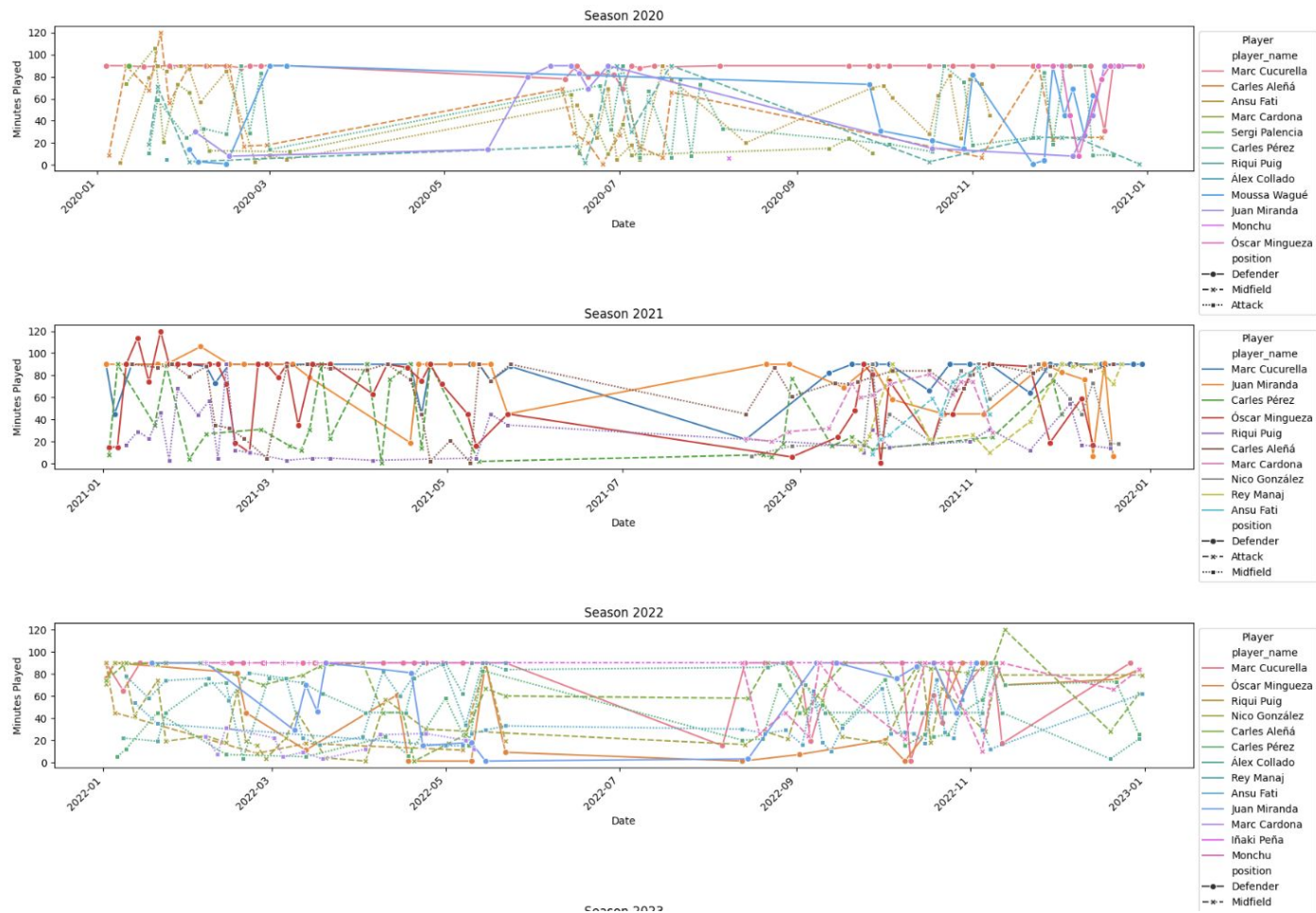
Barcelona B -> Barcelona A

Steps:

1. get all players in Barca A transferred from Barca B (18 players)
2. check out some things
3. players contribution rate

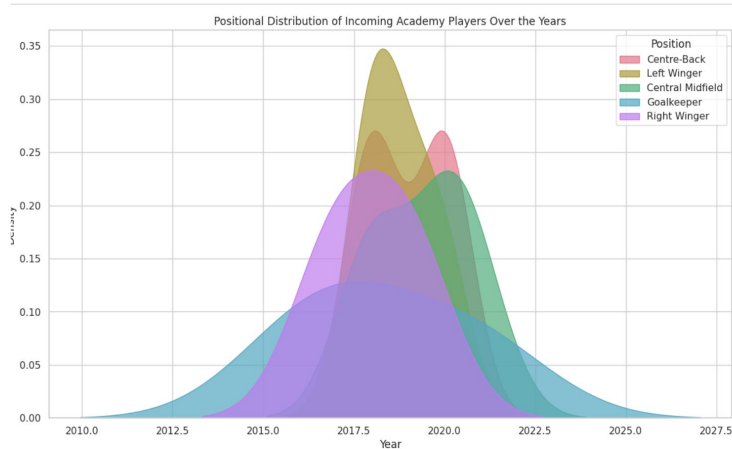


Minutes played and substitute strategies (Xavi)



Miscellaneous

- referee stuff
- pass maps
- positional distribution of incoming academy players (covid?)
- player performance index



Win Rate with Top Referees:

referee

Alejandro Hernández Hernández 0.687500

Jesús Gil Manzano 0.777778

José María Sánchez Martínez 0.823529

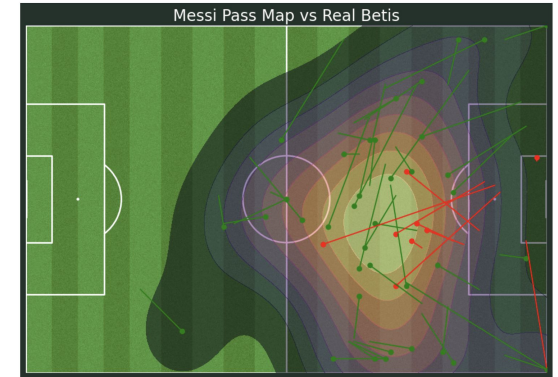
Juan Martínez Munuera 0.750000

Mateu Lahoz 0.647059

Name: result, dtype: float64

Overall Win Rate:

0.7704402515723271



Thanks :>