



정보물리 3조-최종발표

python을 이용한 축구 데이터 분석

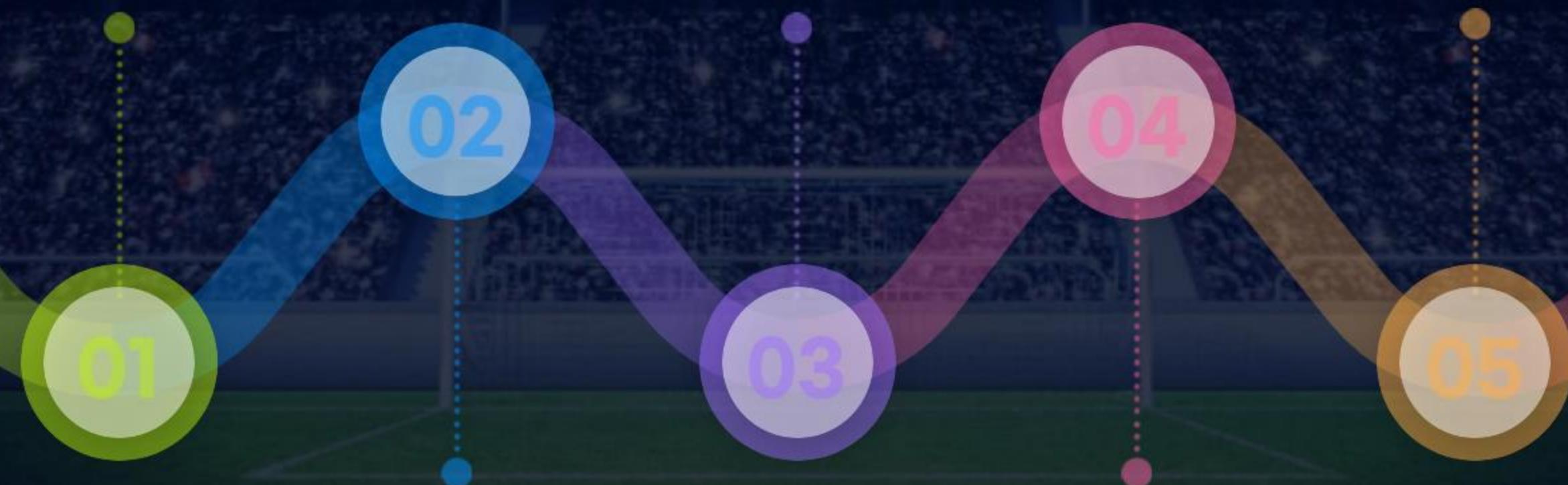
물리학과 2021102349 이규현
물리학과 2022102376 여대운
물리학과 2023102038 엄태빈

Tap to Start

주제선정 및 데이터 전처리

다경기 분석

최종 발표



한경기 분석

최종보완



분석 목적
데이터 설명



주요 코드
분석함수



분석 결과



결론

분석목적

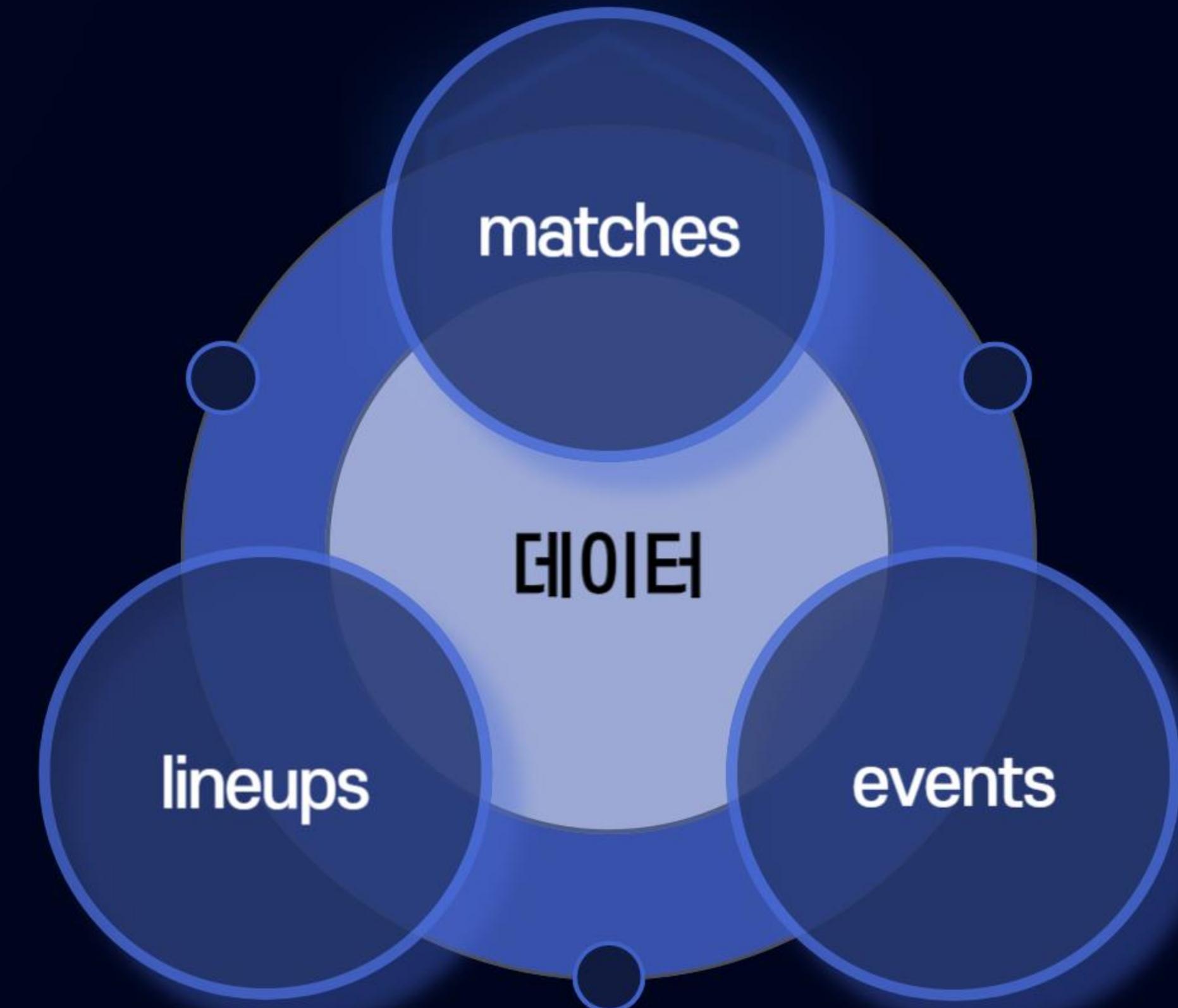
복잡한 데이터 분석

- 통계적으로 유의미한 결과 도출

매우 다양한 정보가 담겨 있는 축구 경기 데이터

- 필요한 정보만 추출
- 경기의 내용을 시각화, 통합
- 여러가지 지표와 경기의 승패 사이의 연관성 통계적으로 검증

데이터 분석



데이터 분석

File	matches	lineups	events
Data	리그, 시즌, match_id(match_id.json), home_score, away_score	lineup, starting_Xi, player[player_name]	한 경기에 대한 상세정보 (time, location, pass, shot)

데이터 형태

리스트 안에 여러 딕셔너리가 중첩된 상태
딕셔너리의 value로 새 dictionary나 string이 오기도 함

```
print(type(events_data))
print(type(events_data[0]))
print(type(events_data[0]["type"]))
print(type(events_data[0]["type"]["name"]))
```

```
<class 'list'>
<class 'dict'>
<class 'dict'>
<class 'str'>
```

핵심 코드

- 핵심 코드

json.load(f) : json 파일을 list로 불러옴

```
with open(events_file_path, 'r', encoding='utf-8') as f  
    events_data = json.load(f)
```

for 문 : events_data 의 각 행을 읽어들이는 루프문 (event는 dictionary)

```
for event in events_data:
```

if 문 : for 루프로 각 행을 읽으며 조건에 맞는 데이터 추출

```
if event["type"]["name"] == "Pass" and event["t"]  
    team_passes.append(event)
```

분석함수

- 분석 함수

- `draw_soccer_field(ax, side, field_dimen=(120, 80))`
- `draw_pass_network(events_file_path, lineup_file_path, side)`
- `draw_shot_map(events_file_path, lineup_file_path, side="home")`
- `extract_match_data(events_file_path)`
- `create_match_table(match_data, team_names, possession_percentages)`
- `extract_record_data(events_file_path)`
- `find_shot_events(events, team_name)`
- `get_locations(events_file_path, shot_events_id, team_name)`
- `draw_event_chain(events_file_path, lineup_file_path, side)`
- `draw_heatmap(data_imported, team_name, type)`

분석결과(한 경기 분석)

Most Dribbler

Player	Dribble
Lionel Andrés Messi Cuccittini	10

Most Shooter

Player	Shot
Lionel Andrés Messi Cuccittini	8

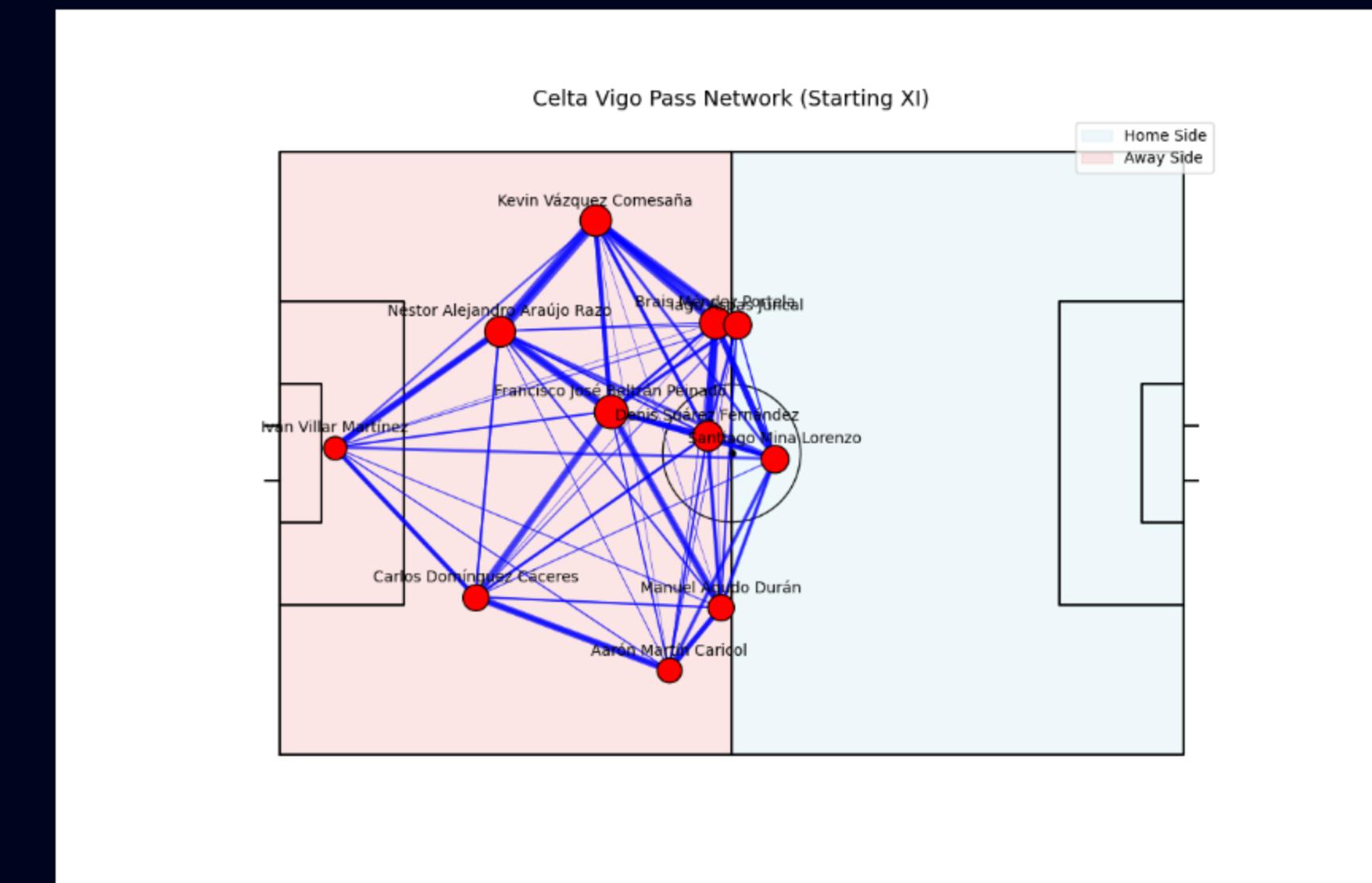
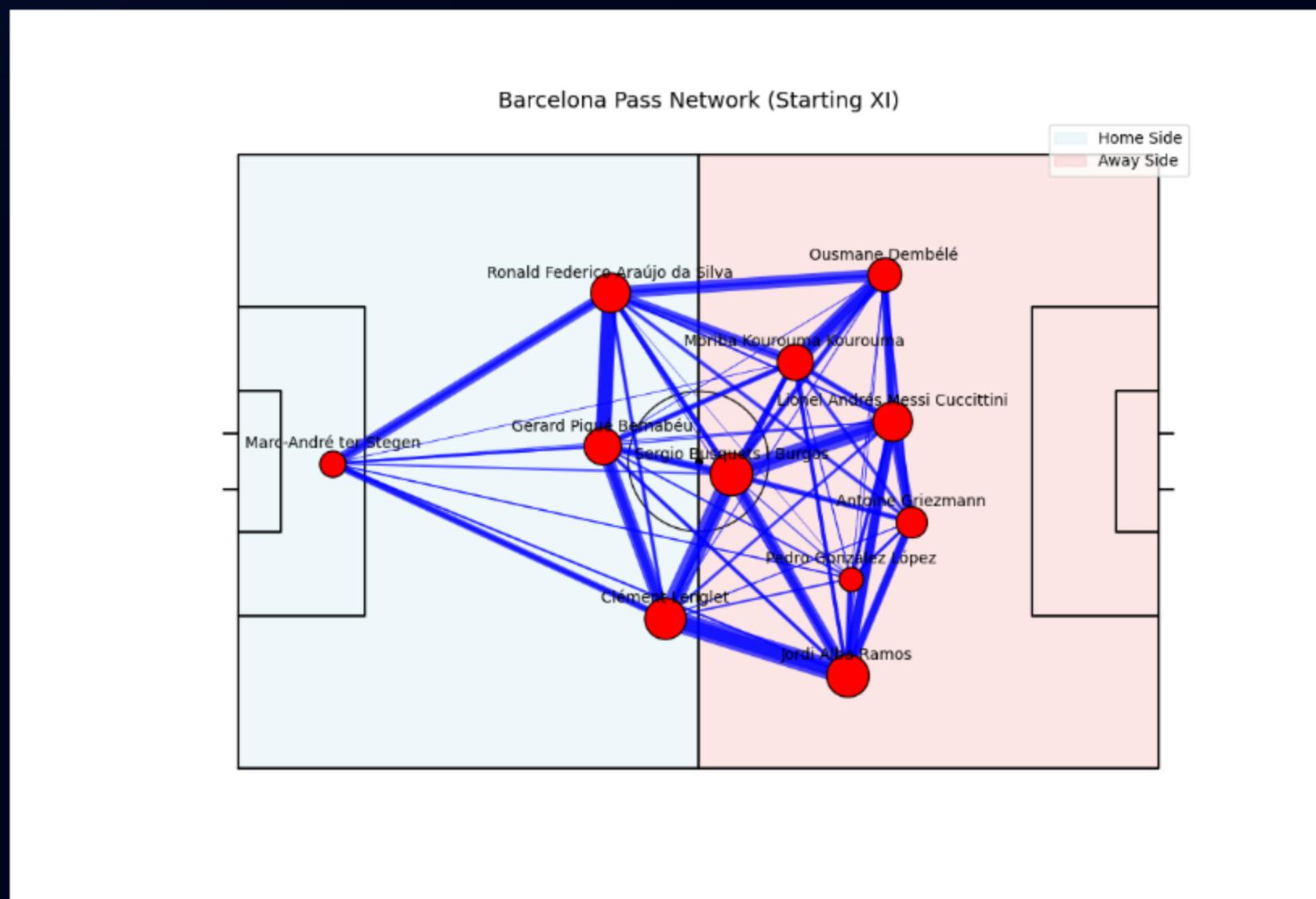
Most Passer

Player	Pass
Sergio Busquets i Burgos	86

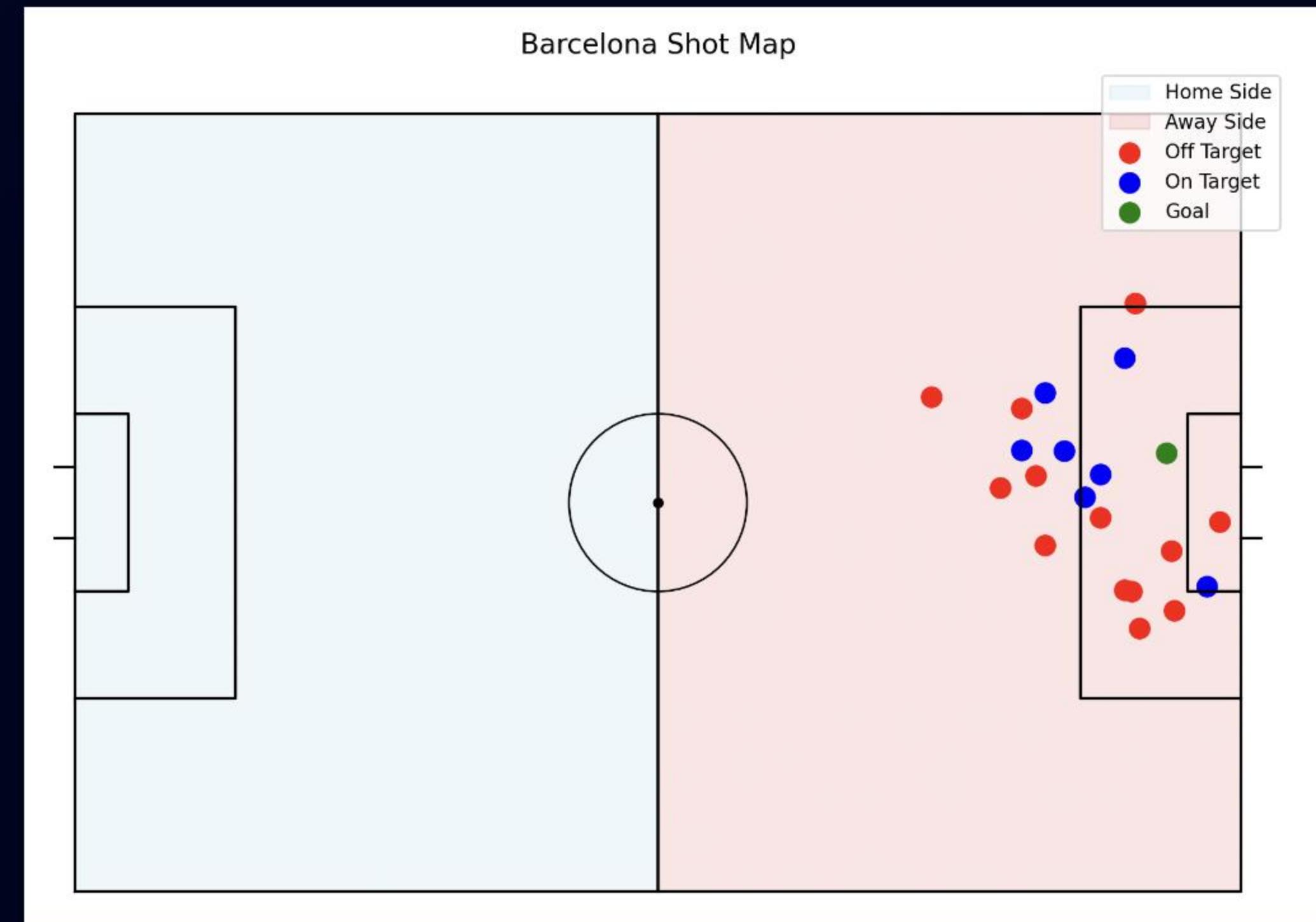
Match Statistics

Category	Celta Vigo	Barcelona
Shots	4	21
On Target	4	8
Possession	33.94%	66.06%
Passes	440	761
Pass Accuracy	80.91%	90.41%
Foul	25	9
Yellow Card	2	2
Red Card	0	1
Offside	1	0
Corners	1	1

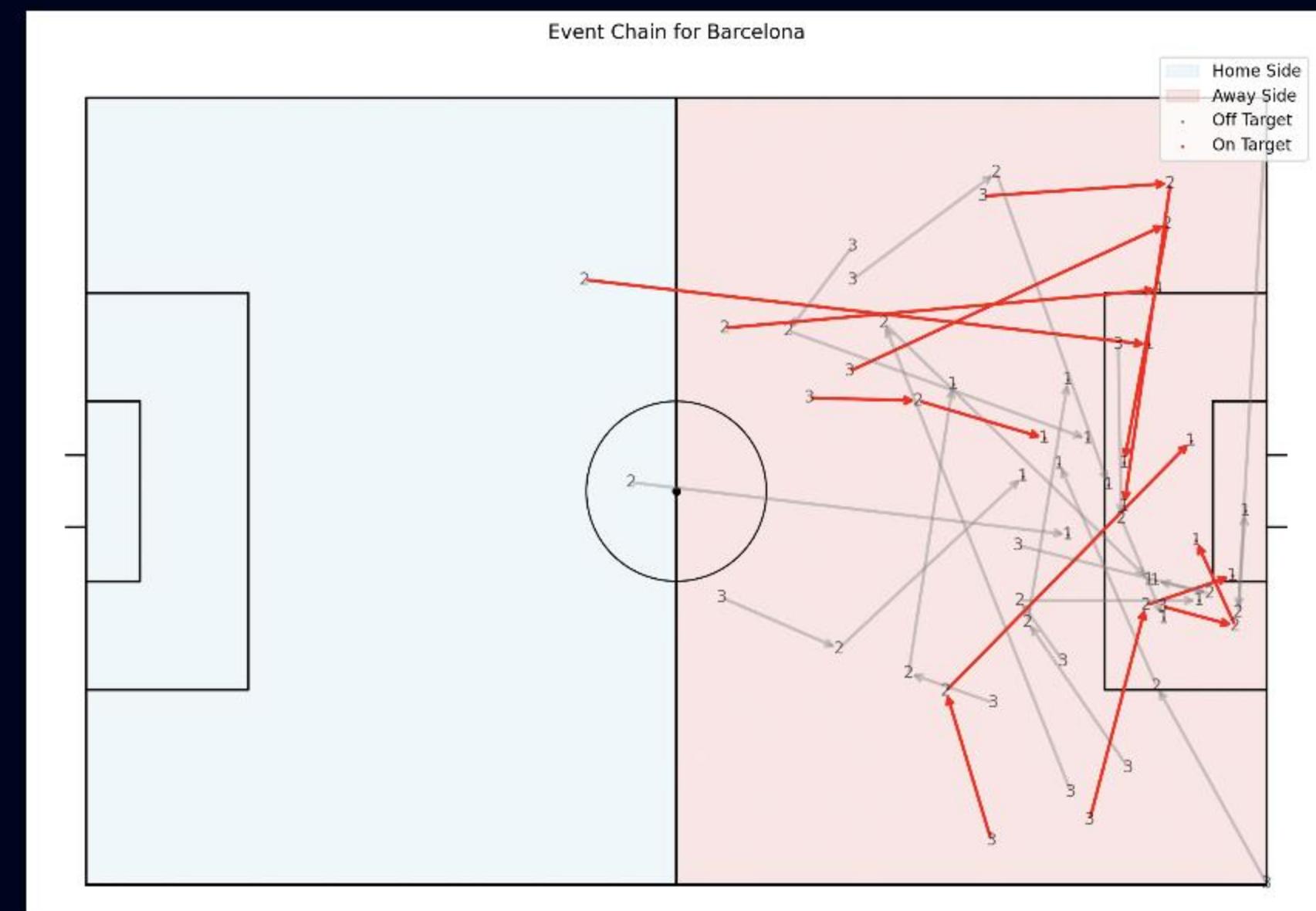
분석결과(단)



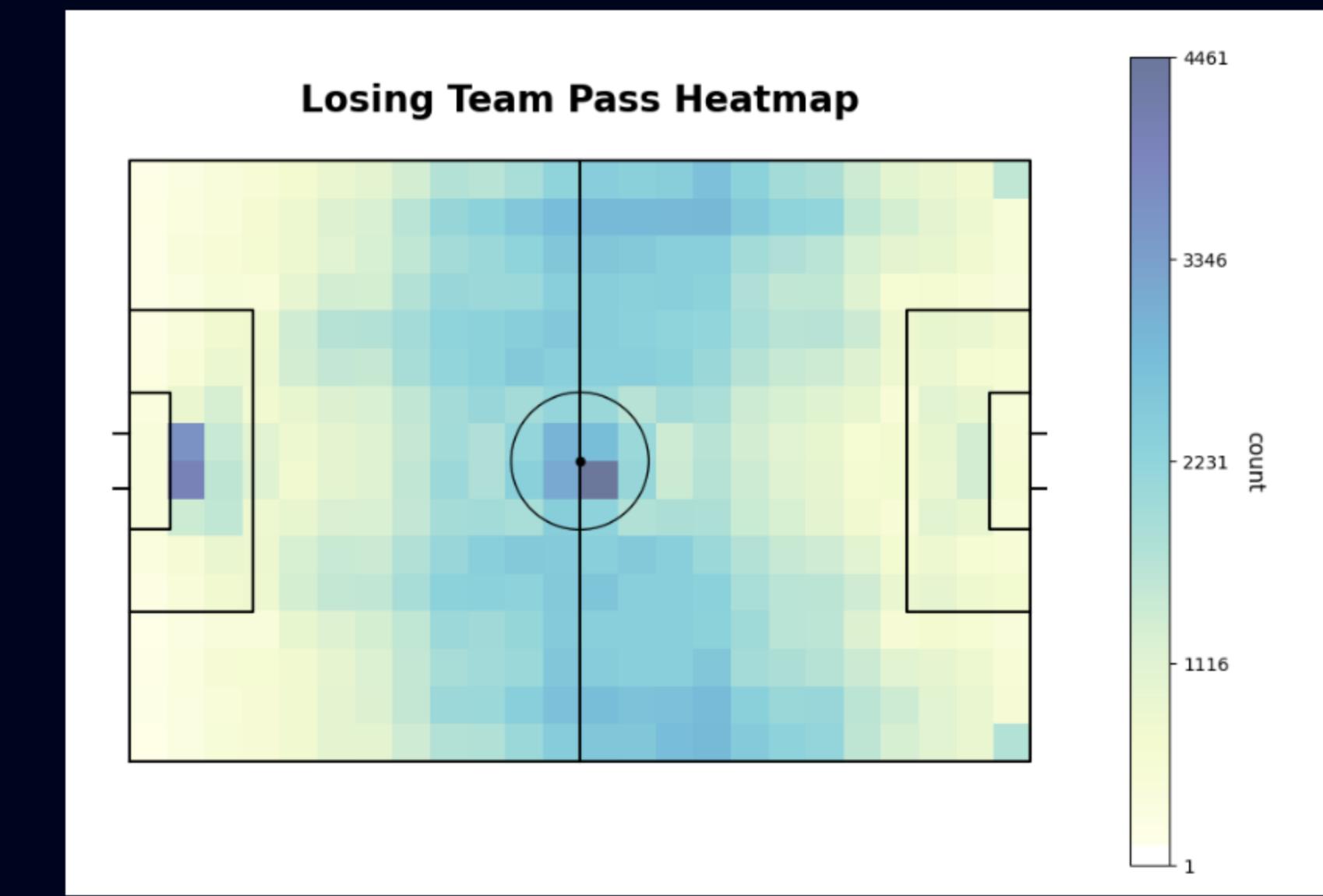
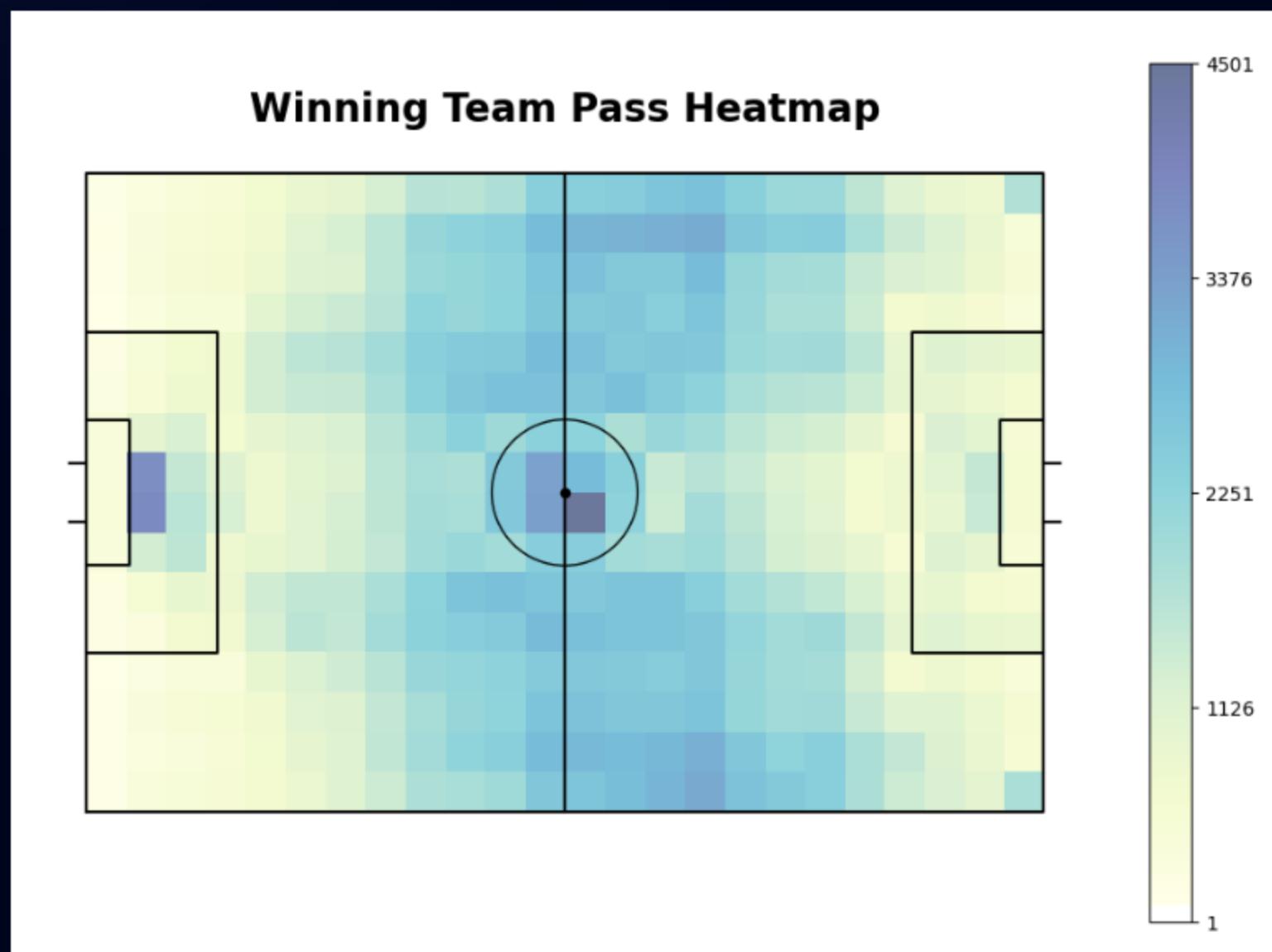
분석결과(단)



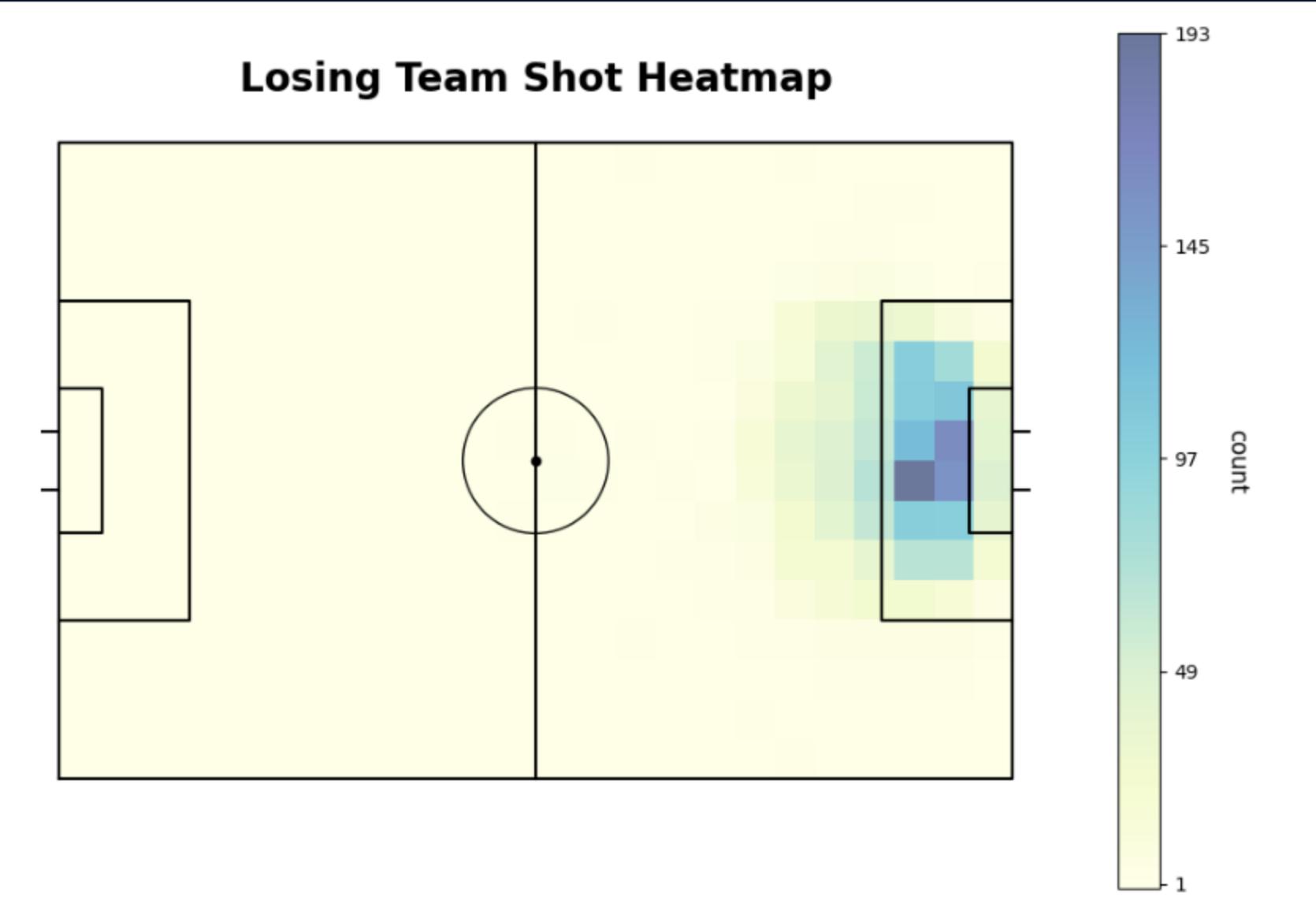
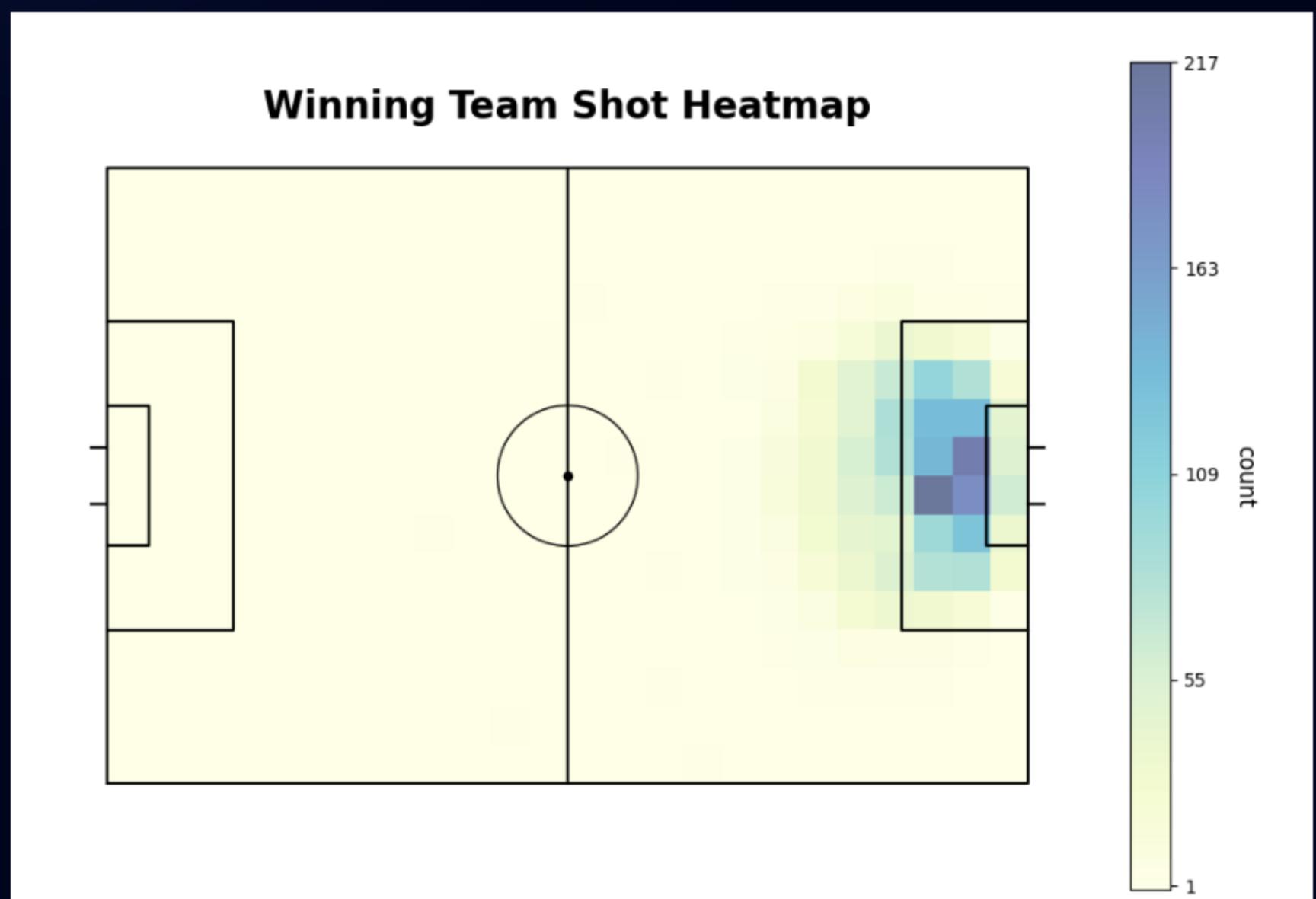
분석결과(단)



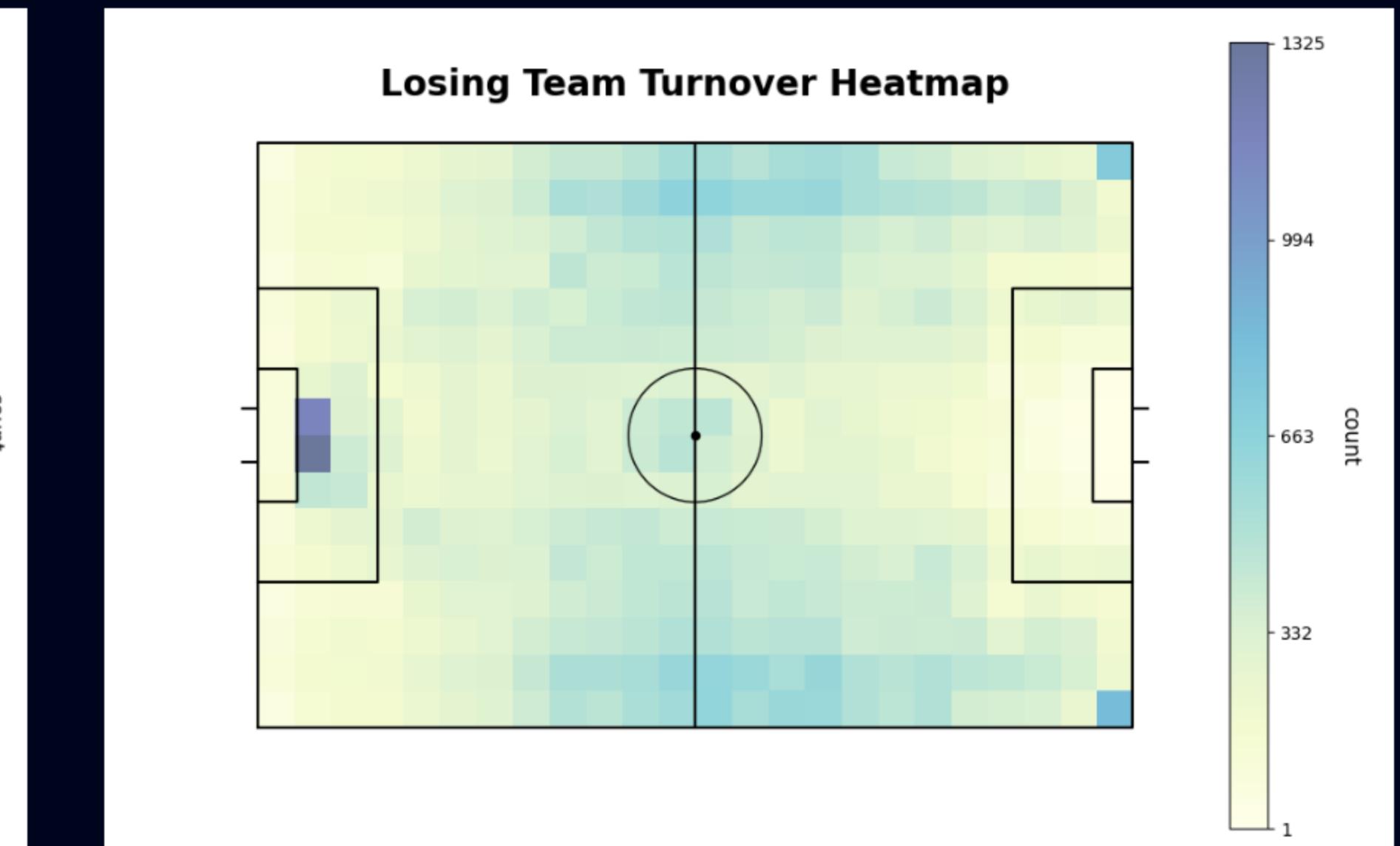
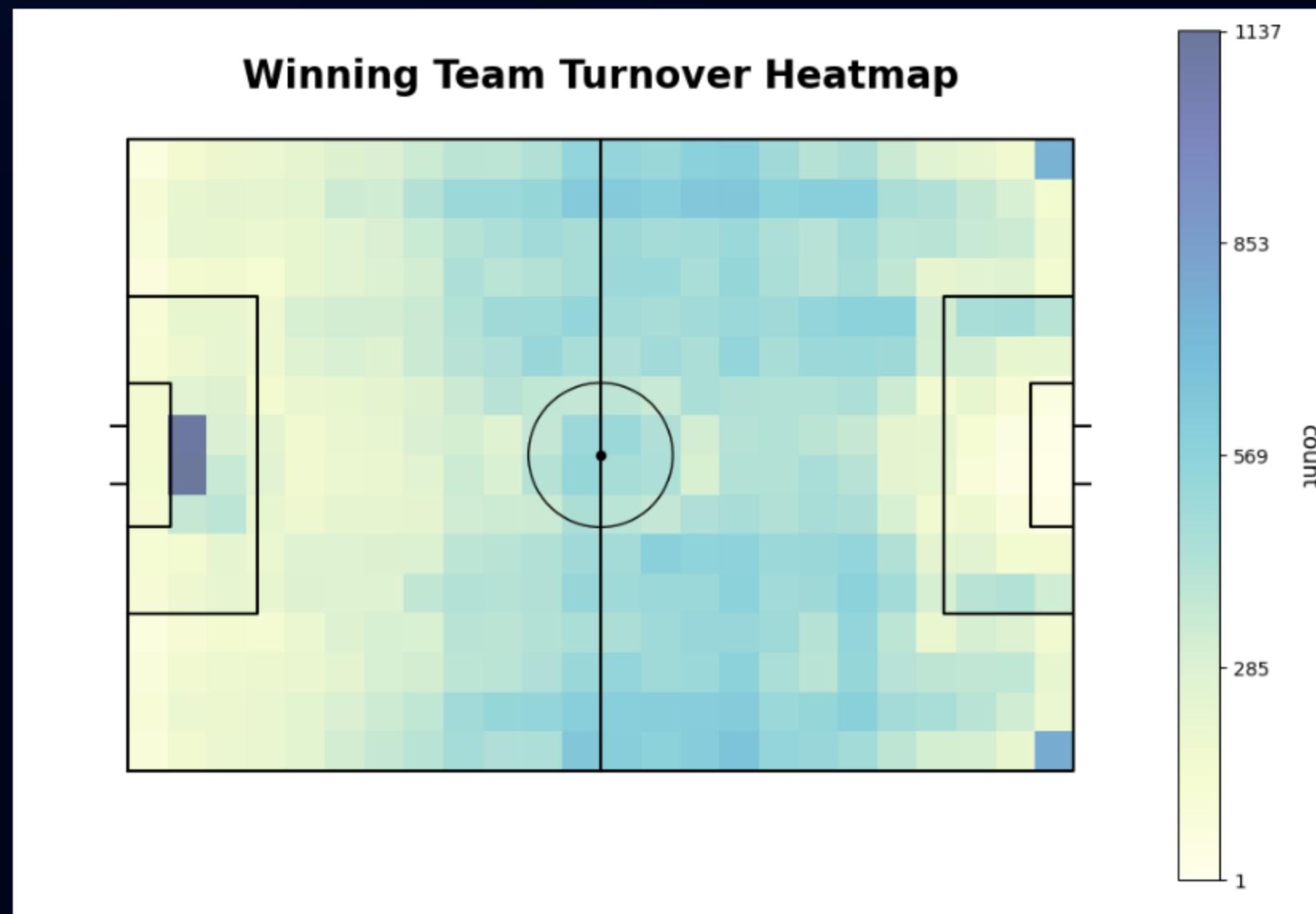
분석결과(다)



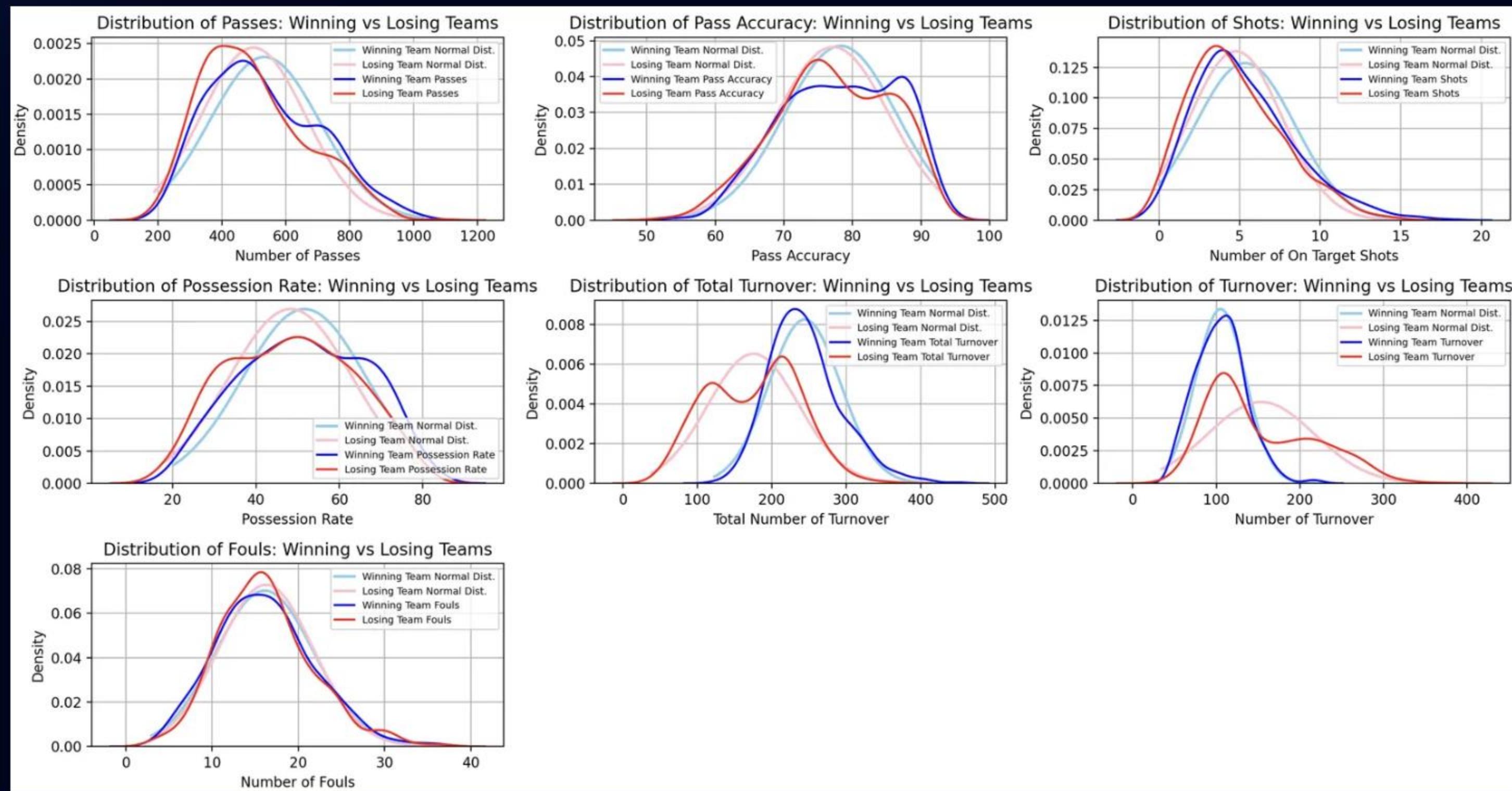
분석결과(다)



분석결과(다)



분석결과(다)



분석결과(다)

```
Pass
533.4640980735552 498.3397548161121
pass t-statistic: 3.5294291189295044
pass p-value: 0.00043305741545518955
패스 개수와 경기 승패 간에 유의미한 차이가 존재합니다.

-----
Pass Accuracy
78.28740805604197 77.16516637478107
Pass Accuracy t-statistic: 2.2981615385933742
Pass Accuracy p-value: 0.021732938743250278
패스 성공율과 경기 승패 간에 유의미한 차이가 존재합니다.

-----
Shots
5.33800350262697 4.768826619964973
Shots t-statistic: 3.2055502636135196
Shots p-value: 0.0013854699729391844
유효슛과 경기 승패 간에 유의미한 차이가 존재합니다.

-----
Possession Rate
51.67315174826997 48.326848251730055
Possession Rate t-statistic: 3.8059587310502256
Possession Rate p-value: 0.0001487661181868542
점유율과 경기 승패 간에 유의미한 차이가 존재합니다.

-----
Total fouls
105.25218914185639 175.24868651488617
Total Turnover t-statistic: 21.169763254974413
Total Turnover p-value: 3.7443568545865476e-84
전체 턴오버 개수와 경기 승패 간에 유의미한 차이가 존재합니다.

-----
Turnover
105.25218914185639 153.50262697022768
Turnover t-statistic: -16.33251986463857
Turnover p-value: 4.8388719083920595e-54
턴오버와 경기 승패 간에 유의미한 차이가 존재합니다.

-----
Fouls
16.089316987740805 16.231173380035028
fouls t-statistic: -0.4288752943706047
fouls p-value: 0.6680950087903101
파울 개수와 경기 승패 간에 유의미한 차이가 존재하지 않습니다.
```

결론

분석 인사이트

한 경기 결과에 대한 시각화
공격 진영에서의 점유율, 유효슛과 패스 수를 높이는 것이 승리에 유리

한계점

선수의 오프더볼 움직임에 대한 데이터가 없어 더 구체적인 분석 불가
더 다양한 지표(패스 정확도, 수비와 관련된 데이터)를 통계적으로 검증하지 못함

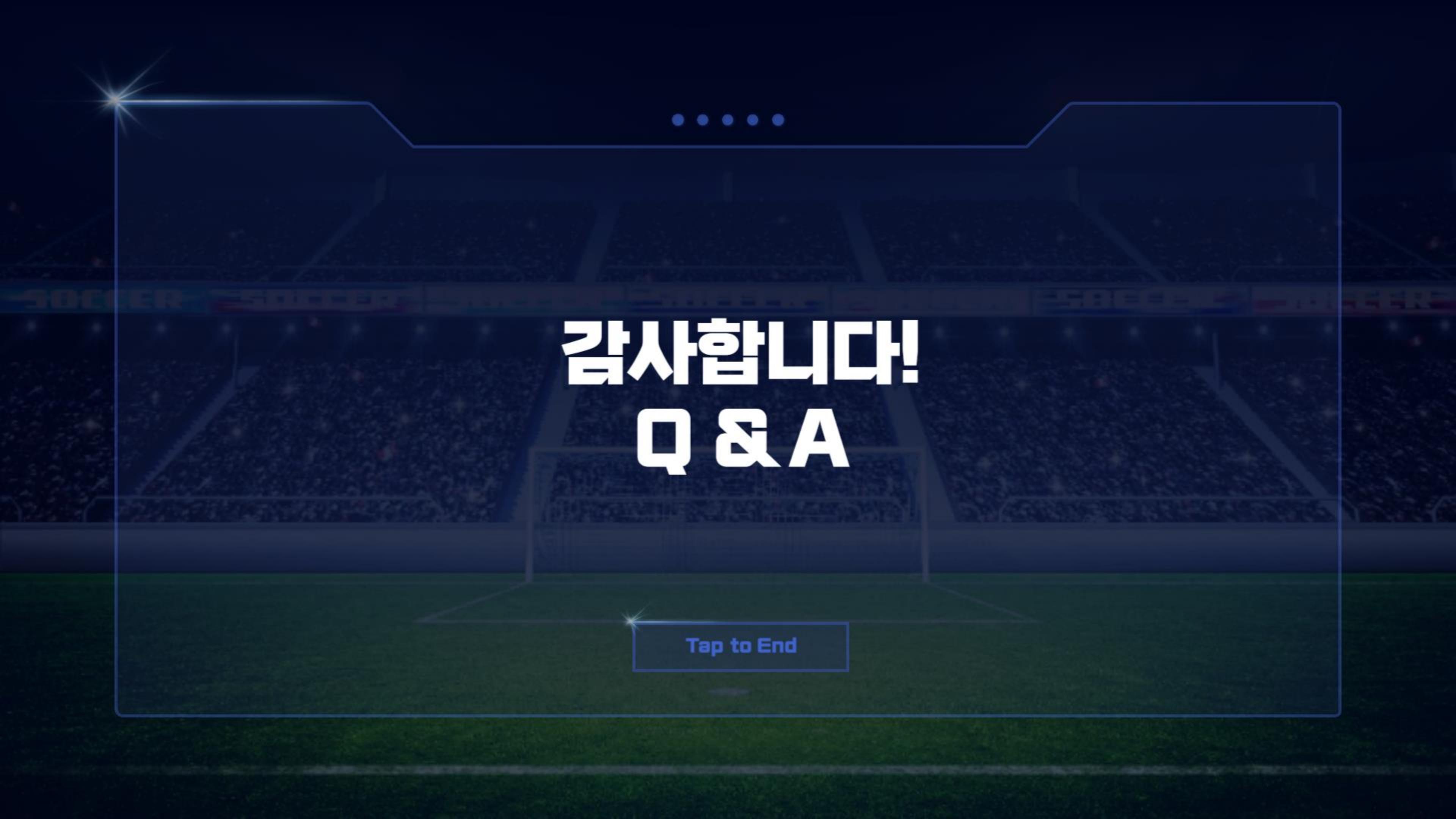
결론

한계점

선수의 오프더볼 움직임에 대한 데이터가 없어 구체적인 분석 불가

논문 : 축구 경기의 경기력 분석 (한국데이터정보과학회, 2015)

수비 요인이 승패에 중요한 역할을 하며, 특히 수비 간격과 좌우 폭이 좁을수록 실점 확률이 낮아짐.



감사합니다!
Q & A

Tap to End