Messi at Barcelona

A Career in Numbers

HARSHIL RATHOD, KRUTI KOTADIA,

NISARG SHAH, YASHASWI PANDEY

SUBMITTED TO: PROFESSOR BENJAMIN NICHOLS

Table of Contents

Introduction

Sports as with other industries has started leveraging data to try and get an edge over the competition, but soccer compared to other sports is behind the curve of incorporating data into the day to day running of the team, its use mostly centred around player scouting and creating opposition reports. The teams that have an analytics department have started to outshine teams that have historically been powerhouses, the biggest example being Brentford from the English Premier League who have an analytical approach to decision making.

The aim of this project is to demonstrate the use of analytics in soccer, specifically how teams can scout players and prepare their players for upcoming matches by visualizing what a player is doing on the pitch. These uses will be demonstrated by analysing the career of Messi using event data from matches he played for Barcelona in the La Liga to show his evolution from a young prodigy to one of the best players of all time.

Data Preparation

Data Sources

The data for this project was provided by Statsbomb. The data was accessed using their python package, Statsbombpy and using mplsoccer's wrapper of their open data access. The data was gathered using four endpoints that provided us with the data for the competitions, matches, lineups, and events available.

The data sources can be accessed using the following links: -

- Statsbomb Open Data
- Statsbombpy
- MPL Soccer Wrapper

About the Data

The data is present in the form of JSON files but were parsed into data frame by the packages used to access the data.

The first file was the competitions data, which contained the competition name, season name, season is and competition id, country name, competition gender and some metadata columns.

Using the competition id and match id we accessed all the match data for the specific season of a competition, some of the columns are match id, match date, match week, season, home score, away score, home team name, home team managers name, away team name, away team managers name, stadium name, referee name.

Using the match id from the matches data frame we accessed the line up information that contained data on every player who were in both the teams for that match.

Using the season name we accessed the competition events data; this contained the data on all events that happened during that season, there are hundred and twenty-three columns in this data frame. The columns correspond to different actions that are recorded on the pitch along with meta data corresponding to the action.

Data Processing

We had to do a lot of preprocessing and processing before we finally had the data we needed. The first thing we needed was the competition and season ids for the La Liga matches, to get this we used the parsers competition() endpoint to get all the competitions whose data we have access to and saved the rows with competition_name = ‘La Liga’.

```python
#Load the competition file
competitions = sbo.competition()
```

Python

+ Code + Markdown

```python
competitions.head()
```

Python

	competition_id	season_id	country_name	competition_name	competition_gender
0	9	27	Germany	1. Bundesliga	male
1	16	4	Europe	Champions League	male
2	16	1	Europe	Champions League	male
3	16	2	Europe	Champions League	male

From the all the season data available for La Liga we selected the season ids for the seasons Messi played at Barcelona and used the ids to get all the match information using the match() endpoint. The data frame we got contained a lot of metadata columns that are of no use for our analysis, so we dropped these columns and selected only the rows in which Barcelona was either the home team or the away team.

```python
#Load the list of matches for this competition
col = temp.columns.tolist()
matches = pd.DataFrame(columns=col)
count = 0
for i in messiLaLiga:
    match = sbo.match(competition_id=comp_id,season_id=i)
    match['season'] = messiSeasons[count]
    count = count+1
    matches = pd.concat([matches, match])
```

Python

```python
matches = matches.reset_index()
```

Python

	index	match_id	match_date	kick_off	home_score	away_score	match_status
0	0	68314	2004-12-04	2004-12-04 20:00:00	4	0	available

The Barcelona data frame had values missing in the manager's name columns, to solve this issue, we save all the managers into two lists, one for Barcelona managers and one for opposition managers. For the scope of this project, we were interested in only filling in the values for the Barcelona managers name, so we created a logic that if for the match with a missing value, the name of the manager for the previous game and the game after were the same, then the missing value should be the same name.

We saved the match ids for all the Barcelona matches and used it to call the lineup() endpoint to access the line up information for the matches, for each match we checked if Messi was playing, saved the match ids for these matches and selected the rows corresponding to these ids from the Barcelona data frame.

We added a new column for the opposition team name and saved this data frame to a csv for use in some visualizations.

```python
messiMatchID =[]
for i in match_id:
    lineup = sbo.lineup(match_id=i)
    mid = lineup[lineup['player_nickname']=='Lionel Messi']['match_id'].to_l
    if(len(mid)!=0):
        messiMatchID.append(mid[0])
```

Python

```python
messiMatches = barca[barca['match_id'].isin(messiMatchID)]
```

Python

```python
messiMatches
```

Python

Using the season name column from our La Liga competition data frame and the competition_events() endpoint we made calls for all the events for all La Liga seasons except for the 1973/1974 season, for each call, we saved all the events related to Messi to a new data frame.

```python
col = tempE.columns.tolist()
eventsMessi = pd.DataFrame(columns=col)
i=0
for season in laLiga['season_name']:
    if(season!='1973/1974'):
        event = sb.competition_events(
            country="Spain",
            division= "La Liga",
            season=season
        )
        event['season'] = season
        eventsMessi = pd.concat([eventsMessi, event[event['player_id']==mess
    i=i+1
```

Python

In this new data frame, we had to split the columns pertaining to location information as they were in the form of lists, we created two columns, one for the x-axis and one for the y-axis for each of the location columns. After this we saved our event data frame to a csv.

```python
eventsMessi[['x_start','y_start']] = eventsMessi['location'].apply(pd.Series)
eventsMessi[['x_end','y_end']] = eventsMessi['pass_end_location'].apply(pd.Series)
eventsMessi[['x_carry_end','y_carry_end']] = eventsMessi['carry_end_location'].apply(pd.Series)
```

Python

```python
eventsMessi = eventsMessi.drop(['level_0'],axis=1)
```

Python

```python
eventsMessi.to_csv('eventsMessi.csv',index=False)
```

Python

Analysis

For the analysis we divided the data frame according to grouping of years to divide Messi's career into four phases.

The first phase is the Initial Years or Breakthrough, this phase covers the years where he broke into the first team and established himself as a regular starter for Barcelona.

The second phase is the Pep Years, this phase covers the time when Pep Guardiola was the manager of Barcelona and the team established itself as one of Europe's best, winning two champions leagues and getting to the semi-finals in every year of his tenure. This phase saw Messi transform from playing on the wings to playing centrally and getting to control more of the game.

The third phase starts with the arrival of Neymar and the formation of 'MSN' short for Messi, Neymar, and Suarez, this being one of the most dangerous strike partnerships in soccer history. The trio broke goalscoring and assist records both individually and as a team during the four years together.

The final phase is the Swan Song years, this period starts with the departure of Neymar from the team and Messi enters the final four years at Barcelona. This period saw him become captain, continue breaking individual records and establish himself as one of the greatest to play the game, before finally leaving his boyhood club after twenty-one years.

```
col = eventsMessi.columns.tolist()

initialYears = pd.DataFrame(columns=col)
pepYears = pd.DataFrame(columns=col)
msnYears = pd.DataFrame(columns=col)
swanSong = pd.DataFrame(columns=col)

dfList = []

for season in seasons:
    if(season=='2004/2005' or season=='2005/2006'
       or season=='2006/2007' or season=='2007/2008'):
        initialYears = pd.concat([initialYears,eventsMessi[eventsMessi['season']==season]])

    elif(season=='2008/2009' or season=='2009/2010'
       or season=='2010/2011' or season=='2011/2012' or season=='2012/2013'):
        pepYears = pd.concat([pepYears,eventsMessi[eventsMessi['season']==season]])

    elif(season=='2013/2014'or season=='2014/2015'
        or season=='2015/2016'or season=='2016/2017'):
        msnYears = pd.concat([msnYears,eventsMessi[eventsMessi['season']==season]])

    else:
        swanSong = pd.concat([swanSong,eventsMessi[eventsMessi['season']==season]])
```

Era Analysis

After dividing Messi's career into phases, we proceeded to analyze his playstyle in each of the phases or eras to see how he evolved as a player over the seventeen years he played first team football at Barcelona. Each era analysis contains visualizations of location of actions and analysis of different portions of his game to understand what he was doing on the pitch; these visualizations and analysis are: -

Visualizations

Action	Technique	Notes	Glossary

All actions	Heatmap, bin statistic	Locations of all Messi related events	Action – anytime players location is recorded in the event data.
Carries	Heatmap, bin statistic	Location of the starting and end location of carries	Carry – where a player controls the ball and moves.
Dribbles	Scatter plots, heatmap and bin statistics	Markers and based on outcome of dribble	Dribble – when a player attempts to beat an opponent.
Passes	Hexbin	Location of where Messi passes the ball and where the pass ends	Pass – when a player gives the ball to a teammate by kicking or throwing.
Danger Passes	Heatmap and bin statistics	Location of starting location of a dangerous pass	Danger Pass – a pass that ends up within fifteen yards of the center of the goal.
Assists	Scatter plot	Location of assists and assists from set pieces	Assist – a pass where the recipient ends up scoring a goal. Set-Piece – a penalty to the opposition where the team can pass or shoot the ball without direct or indirect interference from the opposition.

Goals	Scatter plot, heatmap and bin statistics	Location of goals, with size and color based on the xG, goals from set pieces	Goal – a shot that ends up inside the goalpost. xG – expected goals is the probability of the shot being a goal Shot - when a player kicks the ball with the intention of beating the goalkeeper and scoring a goal.

The visualizations were achieved by writing functions to reduce code writing.

Analysis

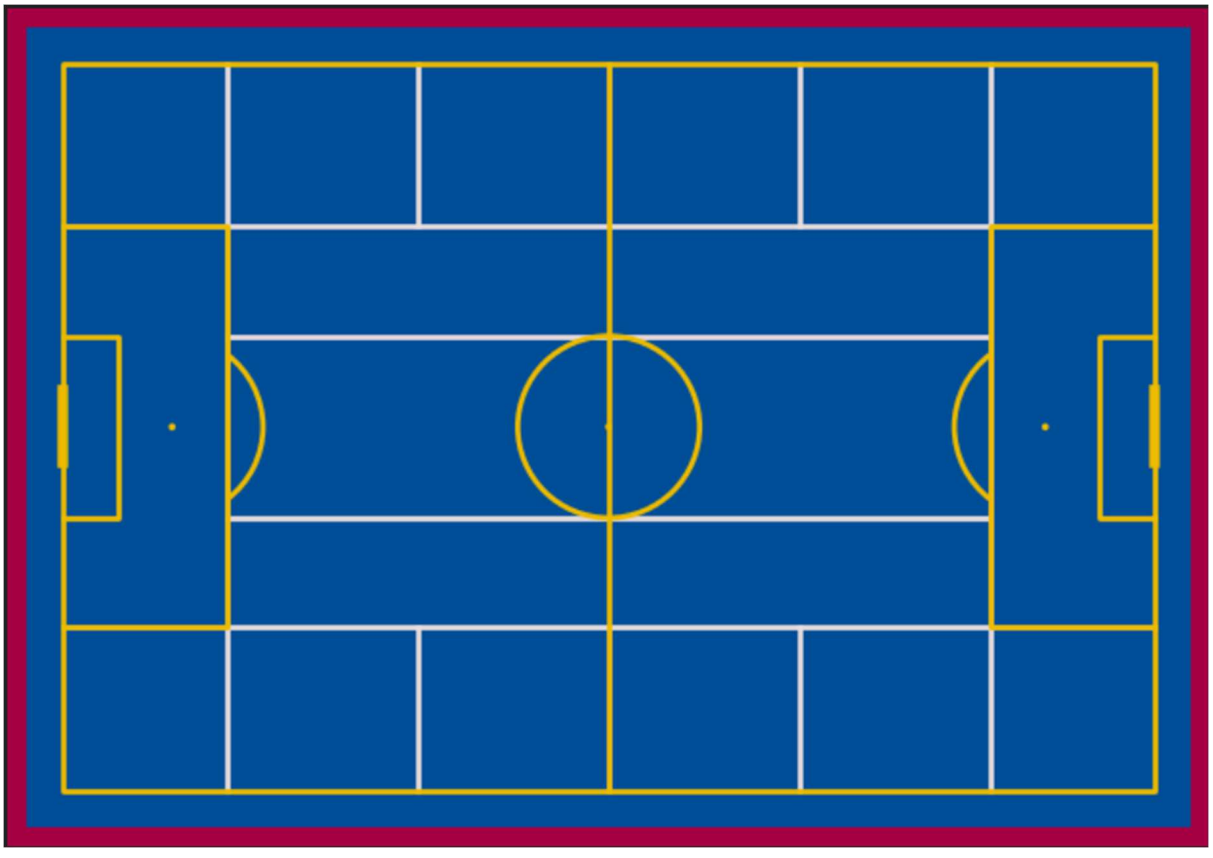
Action	Analysis
Pass	Top pass recipients
Assist	Top assisted players

The analysis and visualizations that were done can be the basis of scouting and preparing players for match ups, as we can visualize what the target player is doing on the pitch, and what his strengths and weaknesses are.

Evolution

The last phase of the analysis was comparing the analysis and visualizations to show the evolution of Messi's play style during the seventeen years playing professional football at Barcelona. The visualizations for this phase follow a similar form to those in era analysis but modified to make it easier to understand the differences.

To emphasize the evolution and how the location of actions changed over time, we will be dividing the pitch according to the philosophy of *Juego de Posición* ,also known as Positional Play. In this the pitch is divided vertically and horizontally so we can show the evolution by visualizing the change in zones where Messi operated.



. The evolution analysis contains these visualizations and analyses: -

Visualizations

Action	Technique	Notes
All actions	Scatter plot	Average location of all Messi related events for all the eras
Carries	Scatter plot	Average location of the starting and end location of carries for all eras
Dribbles	Scatter plot	Average location of successful dribble for all eras
Passes	Kdeplot	Starting and ending location of most common successful pass for all eras
Danger Passes	Scatter plot	Average location of starting location of a dangerous pass
Goals	Map	Location of cities with size of ring based on number of goals scored there.

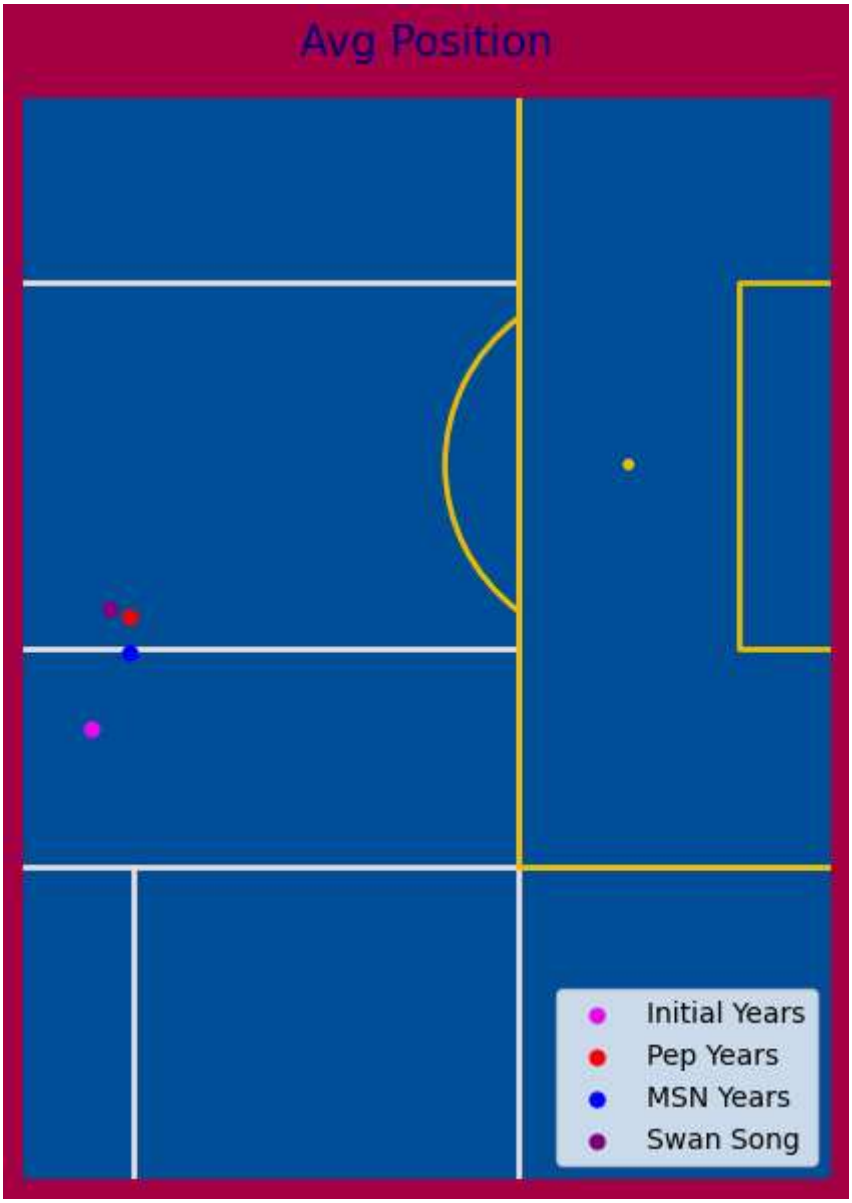
The visualizations were achieved by writing functions to reduce code writing.

Analysis

Action	Analysis
Carry Length	Average carry length for all the eras
Pass Length	Average pass length for all the eras
Passing Percentage	Percentage of passes successfully completed
Top Recipient	Players who Messi passed to most often
Assists	Number of assists, assists from outside the box, percentage of assists from outside the box, assists per match
Player Assisted	Players Messi assisted the most
Shots	Number of shots, number of shots from inside and outside the box, percentage of shots outside the box
Goals	xG, number of goals, number of shots, xG-overperformance, shooting percentage, goals per match
Goals according to Location	xG, number of goals from location, number of shots from location, xG-overperformance, percentage goals from location

The average position of Messi in each of the eras shows how he started out on the wings before moving centrally when Pep Guardiola took charge and then

shifting more to the wing to accommodate the arrival of Neymar and how he again moved to the center after Neymar left so he could impact the game more.



Movement

The movement subsection covers the actions of carries and dribbles, this emphasises the change in his position over the years we showed earlier.

The visualizations show the average starting and endling locations of carries, the average location of successful dribbles and the average length of carry length.

Passing

The passing subsection covers the most common pass start and end locations, the average length of passes in each era, the percentage of successful passes and the top recipients.

The ending location of passes shows us the change in style of play during the MSN era with Messi's pass finding Neymar showing up brightly telling us that this was one of the most common passes for Messi during this time. The second change that caught the eye was the drop in the average length of a successful pass during the Pep era, with the average length dropping to less that half of those in the other three eras.

	Era	Avg Successful Pass Length
0	Initial Years	5.232
1	Pep Years	2.625
2	MSN Years	5.859
3	Swan Song	5.358

Playmaking

The playmaking subsection displays Messi's evolution in creating chances and assisting his teammates. We analyse the average location of danger passes, the number of assists, assists from inside and outside the box, and the assists per match for each era.

Shooting

The shooting subsection covers the evolution of Messi's shooting, comparing the number of shots from inside and outside the box, the percentage of shots that end up in goals, the goals per match, the opposition against which most goals were scored.

The percentage of shots from outside the box increased from averaging around the mid-thirties in the first three eras to around fifty-two in the final era of his career and the percentage of goals from outside the box also increased from six

percent to fourteen percent to twenty percent and finally hitting around thirty percent goals during the four eras.

	Era	Number of Shots	Num Shots Inside Box	Num Shots Outside Box	% Shots Outside Box
0	Initial Years	176	113	63	35.80
1	Pep Years	774	515	259	33.46
2	MSN Years	683	427	256	37.48
3	Swan Song	723	349	374	51.73

	Era	xG	Num Goals Outside Box	Num Shots Outside Box	xG-OverPerformance	'%'Goals Outside Box
0	Initial Years	2.76	2	63	-0.76	6.45
1	Pep Years	12.63	27	259	14.37	14.67
2	MSN Years	12.95	27	256	14.05	20.15
3	Swan Song	19.67	36	374	16.33	28.80

Conclusion

After analyzing and comparing different parts of Messi's game during different eras while he was at Barcelona, we can conclude that when Messi started playing for the first team he was predominantly played on the right wing and would cut inside to get to more dangerous areas, the arrival of Pep Guardiola saw a stark change in his style of play as he was moved to the center of the pitch and given more responsibility with controlling the game, during this time his average pass length was the shortest and he had the highest goals per game ratio at 1.10. With the arrival of Neymar, Messi moved back to the wing to accommodate his new teammate but would continue to control the game by drifting to a more central position during the game, Messi continued his great form by averaging 0.99 goals per game during this era closely matching his ratio to the previous era, during this time he also scored 20% goals from outside of the box. The final era saw Messi continue performing at a high-level scoring 0.91 goals on average per game solidifying his status as one of the greatest to play the game.

Appendix A

Work done by group members: -

- Data Collection and Processing: - Yashaswi Pandey and Harshil Rathod
- Analysis: - Kruti Kotadia and Nisarg Shah
- Visualizations: - Everyone contributed