POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Systems and Methods for Big and Unstructured Data Project

Author(s): **Cristiano Battistini**
Carlo Cardignan 11001182

Group Number: **GroupNumber**

Academic Year: 2023-2024

Contents

1 | Introduction

1.1. Introduction

The advent of data-driven strategies in soccer has revolutionized the sport, providing insights that drive decisions, from player selection to game tactics. There are indeed many examples of clubs that have scouted top talent by leveraging data analysis. What's more, the millions of fans of the sport love to follow the statistics, to see the analytics of their idols in smartphone apps.

Our project builds on this data-centric approach, leveraging a detailed data set that encapsulates the multifaceted nature of soccer, including player statistics, club data, and competition history. We chose MongoDB because its non-relational structure is particularly well suited to handle the diversity and volume of data we are dealing with. Unlike traditional relational databases, MongoDB's flexible data model allows us to store and process different types of data without the need for a predefined schema.

MongoDB, being a document-oriented database, allows each player, club or competition to be represented as a document with a rich and dynamic set of attributes. In addition, MongoDB's horizontal scalability through automatic sharding is critical for handling large volumes of data, such as those generated in modern soccer. The performance of this technology is in fact optimal even as the data size increases.

All of this allows us to dynamically adapt to the evolving nature of the dataset, reflecting the real-world fluidity of soccer team compositions and league structures, keeping in mind that nowadays there is always a competition game to watch.

Our goal is to extract meaningful patterns and insights that can influence various factors, such as talent scouting but also match analysis. The dataset includes detailed details on players, with their appearances and ratings, club situations, and the competitive landscape of the major leagues.

2 | Data Wrangling/Data Generation

2.1. Chapter Introduction

A cleanup and standardization of football-related data was necessary. This process included replacing long, detailed descriptive strings with one- or two-character abbreviations to make the data more manageable and optimize the size of the database. The operation was done manually using the "find and replace" function of the Visual Studio Code software. Specifically, the "description" attribute within the "game_events" dataset was simplified by removing unnecessary details such as the total number of season goals or the specification of tournament goals, keeping only the essentials such as the type of shot that led to the goal. This process made the data more streamlined and focused on relevant aspects for later analysis. The following tables describe what was changed in the initial datasets.

2.1.1. Players

Original Dataset	Attribute	Old Value	New Value
players	sub_position	Attacking Midfield	AM
players	sub_position	Defensive Midfield	DM
players	sub_position	Goalkeeper	G
players	sub_position	Centre-Forward	CF
players	sub_position	Centre-Back	CB
players	sub_position	Central Midfield	CM
players	sub_position	Left Winger	LW
players	sub_position	Right Winger	RW
players	sub_position	Right-Back	RB
players	sub_position	Left-Back	LB
players	sub_position	Left Midfield	LM
players	sub_position	Right Midfield	RM
players	sub_position	Second Striker	ST
players	position	Defender	D
players	position	Midfield	C
players	position	Attack	A
players	position	Missing	M
players	position	GoalKeeper	G
players	foot	right	R
players	foot	left	L

Table 2.1: Your caption here

2.1.2. Game events

Original Dataset	Attribute	Old Value	New Value
game_events	description	Right-footed shot	R
game_events	description	Penalty	P
game_events	description	Direct free kick	F
game_events	description	Left-footed shot	L
game_events	description	Header	H
game_events	description	Tap-in	T
game_events	description	Deflected shot on goal	D
game_events	description	Long distance kick	K
game_events	description	Own goal	O
game_events	description	Solo run	S
game_events	description	Counter attack goal	C
game_events	description	Chest	Q
game_events	description	Penalty rebound	B
game_events	description	Direct corner	N

Table 2.2: Your caption here

2.1.3. club games

Original Dataset	Attribute	Old Value	New Value
club_games	hosting	Home	H
club_games	hosting	Away	A

Table 2.3: Your caption here

2.2. Original Data

Initially, the data were distributed in several datasets: 'clubs', 'club_games', 'competitions', 'games', 'game_events', 'player', 'player_valuations' and 'appearances'. To optimize organization and accessibility, a restructuring into three main collections in the 'football' database was adopted. These are the initial datasets and their attributes:

Dataset name	Attributes
players	player_id, first_name, last_name, name, last_season, current_club_id, player_code, country_of_birth, city_of_birth, country_of_citizenship, date_of_birth, sub_position, position, foot, height_in_cm, market_value_in_eur, highest_market_value_in_eur, contract_expiration_date, agent_name, image_url, url, current_club_domestic_competition_id, current_club_name
player_valuations	player_id, last_season, datetime, date, dateweek, market_value_in_eur, n, current_club_id, player_club_domestic_competition_id

Dataset name	Attributes
appearances	appearance_id, game_id, player_id, player_club_id, player_current_club_id, date, player_name, competition_id, yellow_cards, red_cards, goals, assists, minutes_played
competitions	competition_id, competition_code, name, sub_type, type, country_id, country_name, domestic_league_code, confederation, url, club_games, game_id, club_id, own_goals, own_position, own_manager_name, opponent_id, opponent_goals, opponent_position, opponent_manager_name, hosting, is_win

Dataset name	Attributes
games	game_id, competition_id, season, round, date, home_club_id, away_club_id, home_club_goals, away_club_goals, home_club_position, away_club_position, home_club_manager_name, away_club_manager_name, stadium, attendance, referee, url, home_club_name, away_club_name, aggregate, competition_type, clubs, club_id, club_code, name, domestic_competition_id, total_market_value, squad_size, average_age, foreigners_number, foreigners_percentage, national_team_players, stadium_name, stadium_seats, net_transfer_record, coach_name, last_season, url

Dataset name	Attributes
game_events	game_id, minute, type, club_id, player_id, description, player_in_id

The 'player' collection now integrates player information with related 'valuations' and 'appearances', thanks to the Python script 'players_complete_info.py'. The 'clubs' collection encapsulates within it the 'club_games', which in turn contain details about 'game_events', aggregated via the 'club_and_games.py' script. Finally, the 'competitions' collection was enriched with the associated 'games' and their respective 'game_events' through the use of the 'competitions_and_games.py' script. In merging multiple datasets, attributes that were repeated were removed both to save space and because they were unnecessary to the context outlined. In the following repository
(at this URL `https://github.com/cristianobattistini/smbud`) it is possible to observe the python code for the updates to the original datasets.

3 | Dataset

3.1. Dataset

Dataset The selected dataset is a large collection of football data, derived primarily from Transfermarkt. Updated regularly, it offers accurate data on more than 60,000 global competition matches, details on 400 clubs, and statistics on more than 30,000 players, including current and historical market values, physical characteristics, team membership, and individual performances. More than 1.2 million records detail competitive performances, such as appearances and cards. It is possible to observe the original one, saved in `www.kaggle.com` , at this link: `https://www.kaggle.com/datasets/thedevastator/football-data-competitions-clubs-players-statist` After the Data Wrangling/Data Generation changes, the MongoDB database, called football, has the following collections and statistics:

- 'clubs' collection: 411 documents, with an average size of 108.74 kB per document. Total size of indexes: 20.48 kB.
- 'Competitions' collection: 43 documents, with an average size of 974.35 kB per document. Total size of indexes: 20.48 kB.
- 'Players' collection: 28,459 documents, with an average size of 7.45 kB per document. Total size of indexes: 409.60 kB.

All the collections contain one or more arrays of sub-dcouments. Some sub-documents contain also other sub-documents.

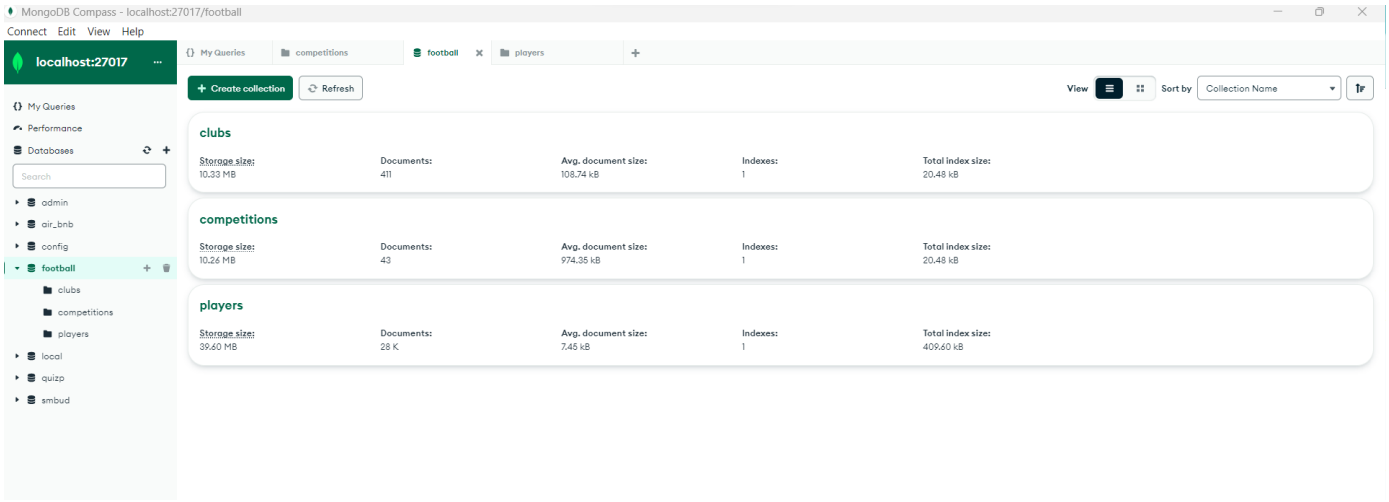

Figure 3.1: Football Dataset

3.2. Collections

3.2.1. Clubs

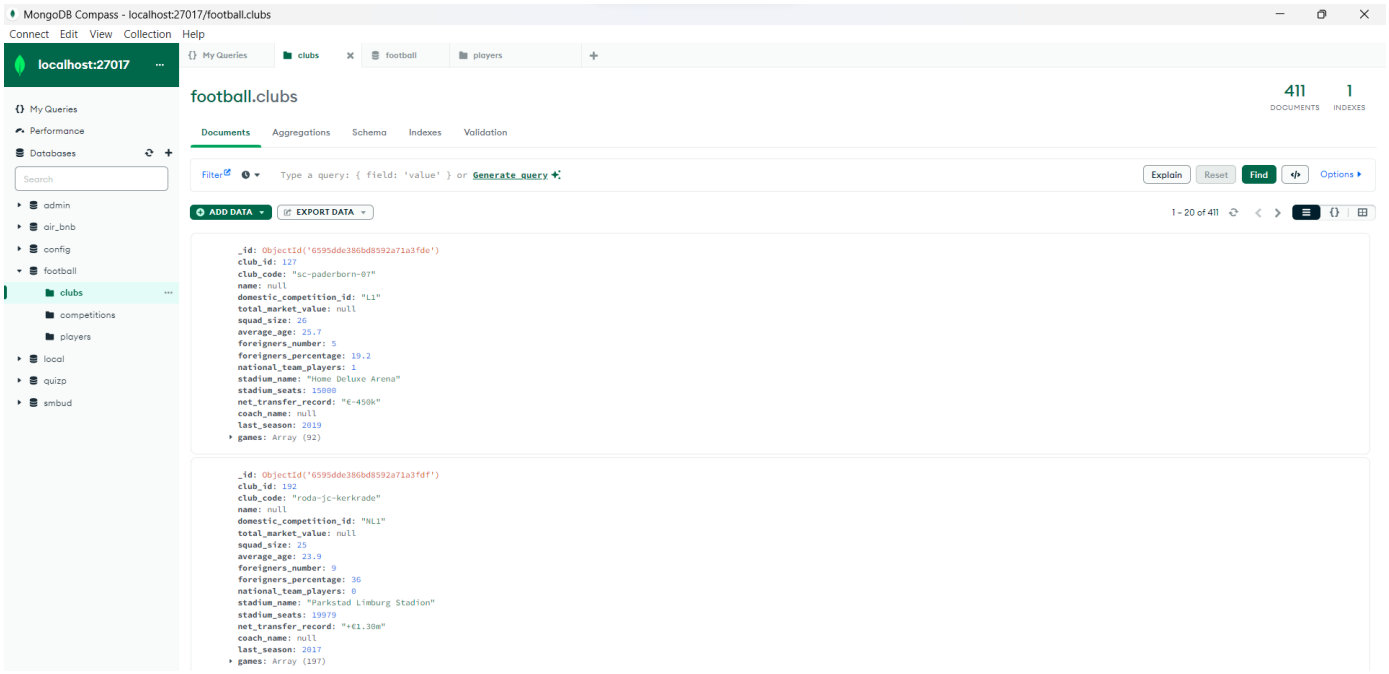

Figure 3.2: Clubs Collection

Attribute	Type	Description
_id	ObjectId	A unique identifier for the document.
club_id	Int32	An integer representing the club's unique ID.
club_code	String	A textual code that uniquely identifies the club.
name	String	The official name of the club.
domestic_competition_id	String	The ID of the domestic league in which the club competes.
total_market_value	Null	The total market value of the club, currently not available.
squad_size	Int32	The number of players in the club's squad.
average_age	Double	The average age of the players in the squad.
foreigners_number	Int32	The count of foreign players in the squad.
foreigners_percentage	Double	The percentage of foreign players relative to the total squad size.
national_team_players	Int32	The number of players who are also national team members.
stadium_name	String	The name of the club's home stadium.
stadium_seats	Int32	The seating capacity of the club's stadium.
net_transfer_record	Null	The net financial record of player transfers, currently not available.
coach_name	String	The name of the club's coach.
last_season	Int32	The most recent season the club competed in.
games	Array (of objects)	A textual code that uniquely identifies the club.

Game Object inside Club

Attribute	Type	Description
game_id	Int32	The unique identifier for the game.
own_goals	Int32	The number of goals scored by the club.
own_position	Int32	The league position of the club at the time of the game.
own_manager_name	String	The name of the club's manager.
opponent_id	Int32	The unique identifier for the opponent club.
opponent_goals	Int32	The number of goals scored by the opponent.
opponent_position	Int32	The league position of the opponent at the time of the game.
opponent_manager_name	String	The name of the opponent's manager.
hosting	String	A character indicating whether the club was hosting the game ('H' for home, 'A' for away).
is_win	Int32	Indicates the outcome of the game (e.g., 0 for loss or draw, 1 for win).
events	Array (of objects)	A list of significant events during the game, with each event as an object containing its own set of attributes.

Events inside Games inside Clubs

Attribute	Type	Description
minute	Int32	The match time in minutes when the event occurred.
type	String	The category of the event, e.g., "Substitutions" or "Goals"
player_id	Int32	The unique identifier of the player involved in the event.
description	Null/String	A detailed description of the event, if available.
player_in_id	Int32	The unique identifier of the player substituted into the game, relevant for substitution events.

3.2.2. Competition

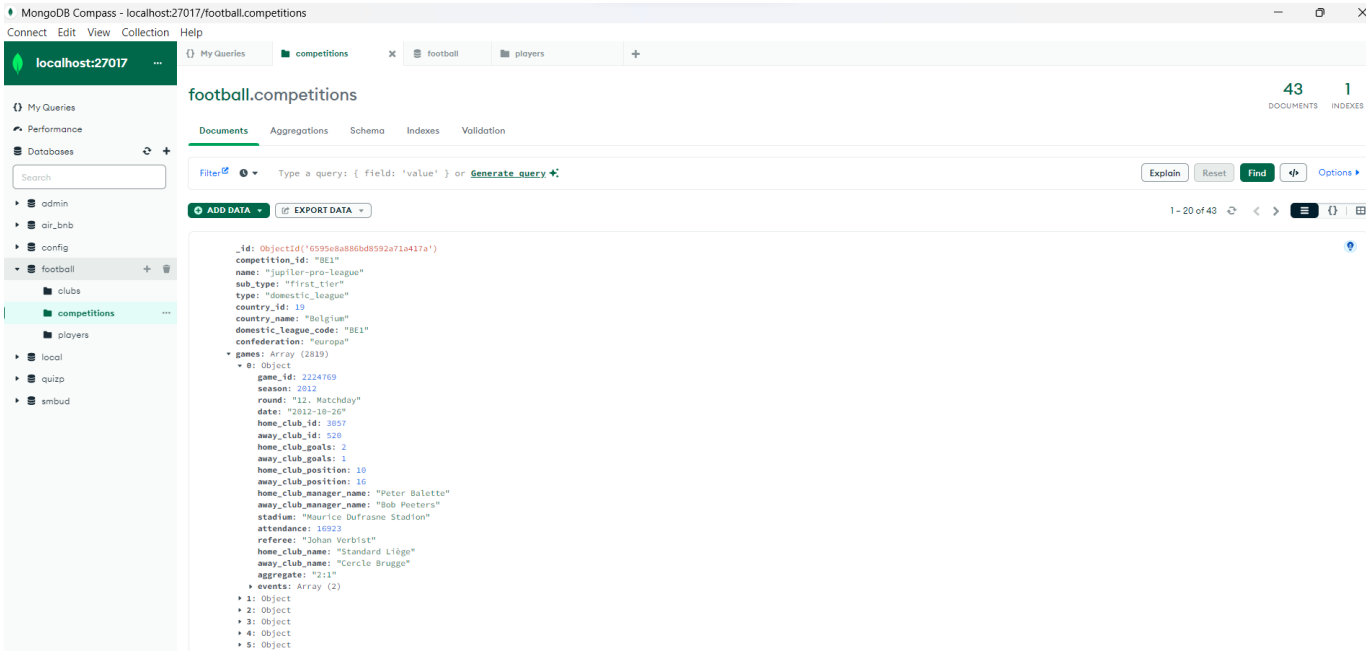

Figure 3.3: Competitions Collection

Attribute	Type	Description
_id	ObjectId	A unique identifier for the document.
competition_id	String	An identifier for the competition.
name	String	The official name of the competition.
sub_type	String	A category within the competition, such as "first_tier".
type	String	The nature of the competition, for example, "domestic_league".
country_id	Int32	A numeric identifier for the country associated with the competition.
country_name	String	The name of the country.
domestic_league_cod	String	A unique code representing the domestic league.
confederation	String	The football confederation to which the competition belongs.
games	Array	A collection of game records associated with the competition.

Game Object inside Competition

Attribute	Type	Description
game_id	Int32	The unique identifier for the game.
season	Int32	The year of the football season.
round	String	Distinct round: ['1. Matchday' '3. Matchday' '4. Matchday' '11. Matchday' ...
date	String	When the game was played.
home_club_id	Int32	Identifier for the home club.
away_club_id	Int32	Identifier for the away club.
home_club_goals	Int32	Goals scored by the home club.
away_club_goals	Int32	Goals scored by the away club.
home_club_position	Int32	League position of the home club at game time.
away_club_position	Int32	League position of the away club at game time.
home_club_manager_name	String	Name of the home club's manager.
away_club_manager_name	String	Name of the away club's manager.
stadium	String	Name of the stadium where the game was played.
attendance	Int32	Number of people who attended the game.
referee	String	Name of the referee of the game.
home_club_name	String	Name of the home club.
away_club_name	String	Name of the away club.
aggregate	String	Overall score
events	Array (of objects)	An array detailing significant events during the game.

Events inside Games inside Competions

Attribute	Type	Description
minute	Int32	The match time in minutes when the event occurred.
type	String	The category of the event, e.g., "Substitutions" or "Goals"
player_id	Int32	The unique identifier of the player involved in the event.
description	Null/String	A detailed description of the event, if available.
player_in_id	Int32	The unique identifier of the player substituted into the game, relevant for substitution events.

3.2.3. Players

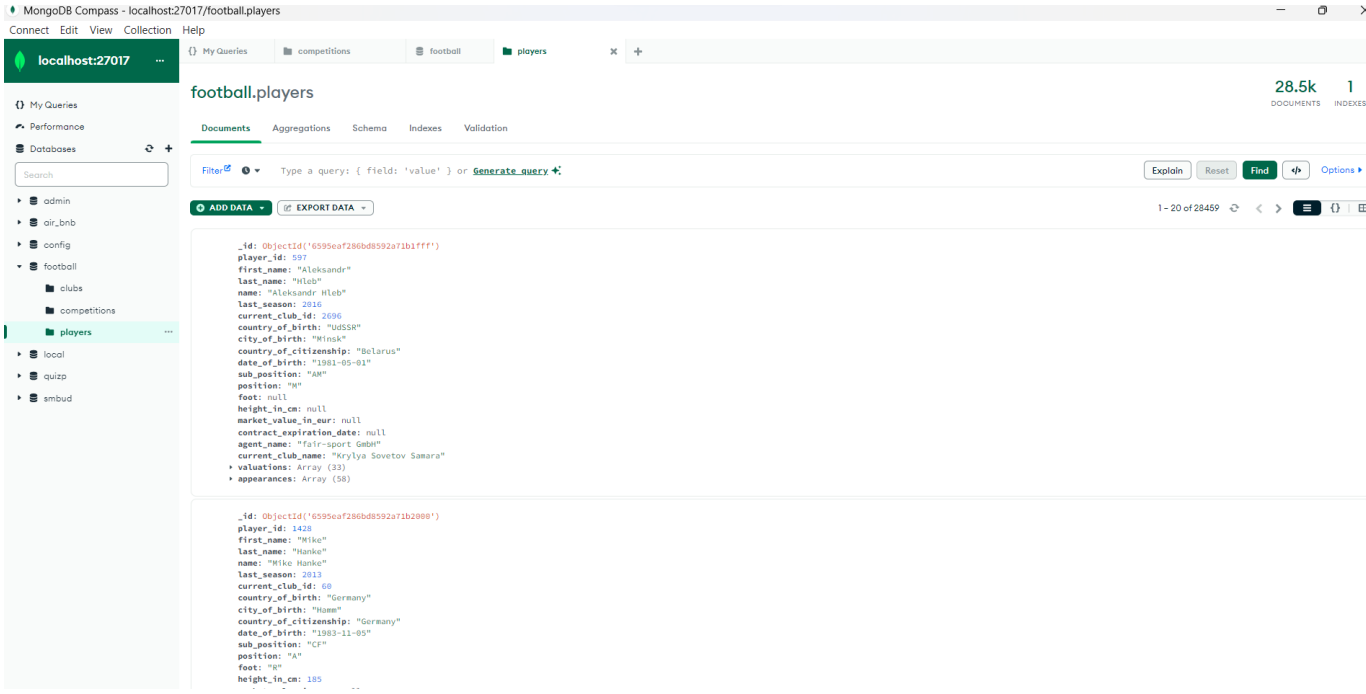

Figure 3.4: Players Collection

Attribute	Type	Description
_id	ObjectId	A unique identifier for the document.
player_id	Int32	A numeric identifier for the player.
first_name	String	The player's first name.
last_name	String	The player's surname.
name	String	The full name of the player.
last_season	Int32	The last active season of the player.
current_club_id	Int32	Identifier for the player's current club.
country_of_birth	String	The country where the player was born.
city_of_birth	String	The city where the player was born.
country_of_citizenship	String	The country of the player's citizenship.
date_of_birth	String	The player's birthdate.
sub_position	String	The player's specific position on the field.
position	String	The general position category the player occupies.
foot	Null/String	Preferred foot of the player.
height_in_cm	Null/Int32	The player's height in centimeters.
market_value_in_eur	Null/Int32	The player's market value in euros.
contract_expiration_date	String	When the player's contract is set to expire.
agent_name	String	The name of the player's agent.
current_club_name	String	The name of the player's current club.
valuations	Array (of objects)	A history of the player's market value evaluations.
appearances	Array (of objects)	A record of the player's appearances in games.

Player Valuations

Attribute	Type	Description
last_season	Int32	The season year of the valuation.
market_value_in_eur	Int32	The player's market value in euros at that time.
current_club_id	Int32	The identifier for the club the player was with during that season.

Player Appearances

Attribute	Type	Description
game_id	Int32	The identifier for the game.
date	String	The date when the game was played.
competition_id	String	The identifier for the competition in which the game took place.
yellow_cards	Int32	The number of yellow cards received by the player in the game.
red_cards	Int32	The number of red cards received by the player in the game.
goals	Int32	The number of goals scored by the player in the game.
assists	Int32	The number of assists made by the player in the game.
minutes_played	Int32	The number of minutes the player played in the game.

4 | Queries

4.1. Leagues analysed

The competitions analyzed will include the Champions League and the major national leagues of Europe: Ligue 1 (France), LaLiga (Spain), Premier League (England), Serie A (Italy), and Bundesliga (Germany). The Champions League is Europe's most prestigious international competition, featuring Europe's best teams. The other national leagues are among the most important in Europe, characterized by top players, significant economic power of the clubs, and a high degree of competitiveness and visibility at the international level.

LEAGUE TYPE	LEAGUE COUNTRY	LEAGUE CODE	LEAGUE NAME	
International (European) League		CL	Champions League	_id: ObjectId('6593e60f0f39fcbfdbd5cbc1') competition_id: "CL" name: "uefa-champions-league" sub_type: "uefa_champions_league" type: "international_cup" country_id: -1 country_name: null domestic_league_code: null confederation: "europa" ▸ games: Array (1360)
Domestic League	FRANCE	FR1	Ligue1	_id: ObjectId('6593e60f0f39fcbfdbd5cbc1') competition_id: "CL" name: "uefa-champions-league" sub_type: "uefa_champions_league" type: "international_cup" country_id: -1 country_name: null domestic_league_code: null confederation: "europa" ▸ games: Array (1360)
Domestic League	SPAIN	ES1	LaLiga	_id: ObjectId('6593e60f0f39fcbfdbd5cbc1') competition_id: "CL" name: "uefa-champions-league" sub_type: "uefa_champions_league" type: "international_cup" country_id: -1 country_name: null domestic_league_code: null confederation: "europa" ▸ games: Array (1360)

LEAGUE TYPE	LEAGUE COUNTRY	LEAGUE CODE	LEAGUE NAME	
Domestic League	ENGLAND	GB1	Premier League	_id: ObjectId('6593e60f0f39fcbfdbd5cbc1') competition_id: "CL" name: "uefa-champions-league" sub_type: "uefa_champions_league" type: "international_cup" country_id: -1 country_name: null domestic_league_code: null confederation: "europa" games: Array (1360)
Domestic League	GERMANY	L1	Bundesliga	_id: ObjectId('6593e60f0f39fcbfdbd5cbc1') competition_id: "CL" name: "uefa-champions-league" sub_type: "uefa_champions_league" type: "international_cup" country_id: -1 country_name: null domestic_league_code: null confederation: "europa" games: Array (1360)
Domestic League	ITALY	IT1	Serie A	_id: ObjectId('6593e60f0f39fcbfdbd5cbc1') competition_id: "CL" name: "uefa-champions-league" sub_type: "uefa_champions_league" type: "international_cup" country_id: -1 country_name: null domestic_league_code: null confederation: "europa" games: Array (1360)

4.2. Players Collection

4.2.1. Top Football Agents

This query lists agents according to the number of players they represent in the database, ordered from highest to lowest, clearly showing the most influential agents, or companies, in the world of football. In recent years, the figure of the agent or company, which looks after the interests of players, especially in terms of contracts, has had an enormous increase in power.

The agents or the most important companies manipulate the market trying to profit for the player but also for them: every transfer or contract in fact provides compensation for the agents.
In recent years, many agents have played the big game, cornering many clubs and earning huge amounts of money.
Simply, the query's behavior is to group by the attribute agent_name and then to compute the total players for each agent_name.

```
db.players.aggregate([
    { $match: { "agent_name": { $ne: null } } },
    { $group: { _id: "$agent_name", numberOfPlayers: { $sum: 1 } } },
    { $sort: { numberOfPlayers: -1 } }
  ])
```

Figure 4.1: Top Football Agents Result

4.2.2. Competitions statistics

Knowing the statistics of players in competitions is certainly important to identify those talents to be acquired during the football market phase. There are many characteristics that can be evaluated for a player: for example, his goals, assists, cards taken or minutes played in a specific competition. There are many competitions around the world, but only the most important ones will be listed: the five top European leagues (England, Spain, France, Italy, Germany) and the Champions League.

Top GoalScorers

The most coveted record or award is surely to become the top scorer in a competition. This query shows the best champions, usually strikers, to have scored the most goals. The query starts with $unwind to break down the appearances array of each player document, then filters the appearances for a given competition_id. Next, $group aggregates the data by player, adding up the goals scored and capturing the player's name. Finally, $sort and $limit sort the players from highest to lowest number of goals and limit the output to the top 10.

```
  db.players.aggregate([
{ $unwind: "$appearances" },
{ $match: { "appearances.competition_id": "<competition_id>" } },
{
  $group: {
    _id: "$_id",
    totalGoals: { $sum: "$appearances.goals" },
    playerName: { $first: "$name" } // Assumi che 'name' sia il campo che contien
    }
    },
    { $sort: { totalGoals: -1 } },
    { $limit: 10 }
  ])
```

- CL

```
> db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "CL" } },
    {
      $group: {
        _id: "$_id",
        totalGoals: { $sum: "$appearances.goals" },
        playerName: { $first: "$name" }
      }
    },
    { $sort: { totalGoals: -1 } },
    { $limit: 10 }
  ])
< {
    _id: ObjectId('6595eb0986bd8592a71b5032'),
    totalGoals: 74,
    playerName: 'Robert Lewandowski'
  }
  {
    _id: ObjectId('6595eb2686bd8592a71b85de'),
    totalGoals: 73,
    playerName: 'Cristiano Ronaldo'
  }
  {
    _id: ObjectId('6595eb2386bd8592a71b7f41'),
    totalGoals: 62,
    playerName: 'Lionel Messi'
  }
```

- IT1

```
> db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "IT1" } },
    {
      $group: {
        _id: "$_id",
        totalGoals: { $sum: "$appearances.goals" },
        playerName: { $first: "$name" }
      }
    },
    { $sort: { totalGoals: -1 } },
    { $limit: 10 }
  ])
< {
    _id: ObjectId('6595eb2386bd8592a71b8077'),
    totalGoals: 165,
    playerName: 'Ciro Immobile'
  }
  {
    _id: ObjectId('6595eb2786bd8592a71b865a'),
    totalGoals: 108,
    playerName: 'Gonzalo Higuaín'
  }
  {
    _id: ObjectId('6595eb1286bd8592a71b5fcc'),
    totalGoals: 106,
    playerName: 'Paulo Dybala'
  }
```

- ES1

```
> db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "ES1" } },
    {
      $group: {
        _id: "$_id",
        totalGoals: { $sum: "$appearances.goals" },
        playerName: { $first: "$name" }
      }
    },
    { $sort: { totalGoals: -1 } },
    { $limit: 10 }
  ])
< {
    _id: ObjectId('6595eb2386bd8592a71b7f41'),
    totalGoals: 231,
    playerName: 'Lionel Messi'
  }
  {
    _id: ObjectId('6595eafb86bd8592a71b3456'),
    totalGoals: 176,
    playerName: 'Luis Suárez'
  }
  {
    _id: ObjectId('6595eafd86bd8592a71b3ad0'),
    totalGoals: 161,
    playerName: 'Karim Benzema'
  }
```

- FR1

```
> db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "FR1" } },
    {
      $group: {
        _id: "$_id",
        totalGoals: { $sum: "$appearances.goals" },
        playerName: { $first: "$name" }
      }
    },
    { $sort: { totalGoals: -1 } },
    { $limit: 10 }
  ])
< {
    _id: ObjectId('6595eb1286bd8592a71b60f5'),
    totalGoals: 155,
    playerName: 'Kylian Mbappé'
  }
  {
    _id: ObjectId('6595eb1086bd8592a71b5cf0'),
    totalGoals: 122,
    playerName: 'Edinson Cavani'
  }
  {
    _id: ObjectId('6595eb1486bd8592a71b6505'),
    totalGoals: 111,
    playerName: 'Wissam Ben Yedder'
  }
```

- GB1

```
> db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "GB1" } },
    {
      $group: {
        _id: "$_id",
        totalGoals: { $sum: "$appearances.goals" },
        playerName: { $first: "$name" }
      }
    },
    { $sort: { totalGoals: -1 } },
    { $limit: 10 }
  ])
< {
    _id: ObjectId('6595eafb86bd8592a71b3748'),
    totalGoals: 203,
    playerName: 'Harry Kane'
  }
  {
    _id: ObjectId('6595eb0986bd8592a71b5200'),
    totalGoals: 134,
    playerName: 'Jamie Vardy'
  }
  {
    _id: ObjectId('6595eb0b86bd8592a71b5629'),
    totalGoals: 133,
    playerName: 'Mohamed Salah'
  }
```

- L1

```
> db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "L1" } },
    {
      $group: {
        _id: "$_id",
        totalGoals: { $sum: "$appearances.goals" },
        playerName: { $first: "$name" }
      }
    },
    { $sort: { totalGoals: -1 } },
    { $limit: 10 }
  ])
< {
    _id: ObjectId('6595eb0986bd8592a71b5032'),
    totalGoals: 238,
    playerName: 'Robert Lewandowski'
  }
  {
    _id: ObjectId('6595eafb86bd8592a71b3464'),
    totalGoals: 97,
    playerName: 'Andrej Kramaric'
  }
  {
    _id: ObjectId('6595eb2786bd8592a71b87e6'),
    totalGoals: 96,
    playerName: 'Timo Werner'
  }
```

4.2.3. Best Assist-men

Making an assist means putting a teammate in a position to put the ball in the net. This is also a very important statistic, often peculiar to side or full-back defenders, midfielders or wingers. The query starts with *unwindtobreakdowntheappearancesarrayofeachplayerdocument*, *the* aggregates the data by player, adding up the assists and capturing the player's name. Finally, there is a descending sort and a limit for the best 10 assist-men.

```
    db.players.aggregate([
  { $unwind: "$appearances" },
  { $match: { "appearances.competition_id": "<competition_id>" } },
  {
    $group: {
      _id: "$_id",
      totalAssists: { $sum: "$appearances.assists" },
      playerName: { $first: "$name" } // Aggiunge il nome del giocatore
    }
  },
  { $sort: { totalAssists: -1 } },
  { $limit: 10 }
])
```

Figure 4.2: Best Assist-men Executed with CL competition

4.2.4. Players with multiple yellow cards

There are also many players who make impetuousness and confrontation their strong point. The statistics on yellow cards say a lot about those players who exploit their physical strength, but because of this attitude are prone to committing fouls, punishable by yellow cards. The query starts with $unwind to break down the appearances array of each player document, then fi
aggregates the data by player, adding up the yellow cards taken and showing the player's name.

```
    db.players.aggregate([
  { $unwind: "$appearances" },
  { $match: { "appearances.competition_id": "<competition_id>" } },
  { $group: { _id: "$_id", totalYellowCards: { $sum: "$appearances.yellow_cards" },  pla
  { $sort: { totalYellowCards: -1 } },
  { $limit: 10 }
])
```

```
> db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "IT1" } },
    { $group: { _id: "$_id", totalYellowCards: { $sum: "$appearances.yellow_cards" },  playerName: { $first: "$name" }  } },
    { $sort: { totalYellowCards: -1 } },
    { $limit: 10 }
  ])
< {
    _id: ObjectId('6595eb0986bd8592a71b50ce'),
    totalYellowCards: 79,
    playerName: 'Tomás Rincón'
  }
  {
    _id: ObjectId('6595eb2186bd8592a71b7a2c'),
    totalYellowCards: 74,
    playerName: 'Marcelo Brozovic'
  }
  {
    _id: ObjectId('6595eb1d86bd8592a71b742f'),
    totalYellowCards: 69,
    playerName: 'Nicolò Barella'
  }
```

Figure 4.3: Players with multiple yellow cards Executed with IT1 competition

4.2.5. Players with multiple red cards

In contrast to yellow cards, which can happen in the course of a match, getting a red card means, most of the time, having committed something serious, such as a bad foul, violent conduct or repeated protests to the referee. This statistic shows those players who struggle most to maintain control on the pitch and whose attitudes risk leaving the team one down. The query starts with *unwindtobreakdowntheappearancesarrayofeachplayerdocument*, *thenfilters* aggregates the data by player, adding up the red cards for each player.

```
    db.players.aggregate([
  { $unwind: "$appearances" },
  { $match: { "appearances.competition_id": "<competition_id>" } },
  { $group: { _id: "$_id", totalRedCards: { $sum: "$appearances.red_cards" }, playe
  { $sort: { totalRedCards: -1 } },
  { $limit: 10 }
])
```

```
> db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "GB1" } },
    { $group: { _id: "$_id", totalRedCards: { $sum: "$appearances.red_cards" }, playerName: { $first: "$name" } } },
    { $sort: { totalRedCards: -1 } },
    { $limit: 10 }
  ])
< {
    _id: ObjectId('6595eb0786bd8592a71b4e22'),
    totalRedCards: 4,
    playerName: 'David Luiz'
  }
  {
    _id: ObjectId('6595eb1d86bd8592a71b757e'),
    totalRedCards: 4,
    playerName: 'Granit Xhaka'
  }
  {
    _id: ObjectId('6595eb0e86bd8592a71b56cc'),
    totalRedCards: 3,
    playerName: 'Fernandinho'
  }
```

Figure 4.4: Players with multiple red cards Executed with IT1 competition

4.2.6. Players with the most minutes played

This analysis shows those players who are certainties for their clubs: the so-called immovable starters. These players, workaholics par excellence, are usually the players who are almost always at the top, avoiding injuries. The query breaks down appearances, then it filters the appearances in a certain league, and then sums up the minutes played by each player, showing also the name of him.

```
db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "<competition_id>" } },
    { $group: { _id: "$_id", totalMinutesPlayed: { $sum: "$appearances.minutes_playe
    { $sort: { totalMinutesPlayed: -1 } },
    { $limit: 10 }
  ])
```

```
> db.players.aggregate([
    { $unwind: "$appearances" },
    { $match: { "appearances.competition_id": "ES1" } },
    { $group: { _id: "$_id", totalMinutesPlayed: { $sum: "$appearances.minutes_played" }, playerName: { $first: "$name" } } },
    { $sort: { totalMinutesPlayed: -1 } },
    { $limit: 10 }
  ])
< {
    _id: ObjectId('6595eb2086bd8592a71b79e1'),
    totalMinutesPlayed: 26361,
    playerName: 'Jan Oblak'
  }
  {
    _id: ObjectId('6595eb0986bd8592a71b5098'),
    totalMinutesPlayed: 26359,
    playerName: 'Dani Parejo'
  }
  {
    _id: ObjectId('6595eb1486bd8592a71b6455'),
    totalMinutesPlayed: 24491,
    playerName: 'Koke'
  }
```

Figure 4.5: Players with the most minutes played Executed with IT1 competition

4.3. PAGINA 28

A | Appendix A

If you need to include an appendix to support the research in your thesis, you can place it at the end of the manuscript. An appendix contains supplementary material (figures, tables, data, codes, mathematical proofs, surveys, ...) which supplement the main results contained in the previous chapters.

List of Figures

List of Tables