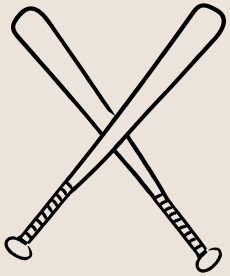
MAJOR LEAGUE BASEBALL



CONTEXT AND BUSINESS BACKGROUND

01

- **MLB is a professional baseball league in North America** with 30 teams split into the National League (NL) and the American League (AL).

02

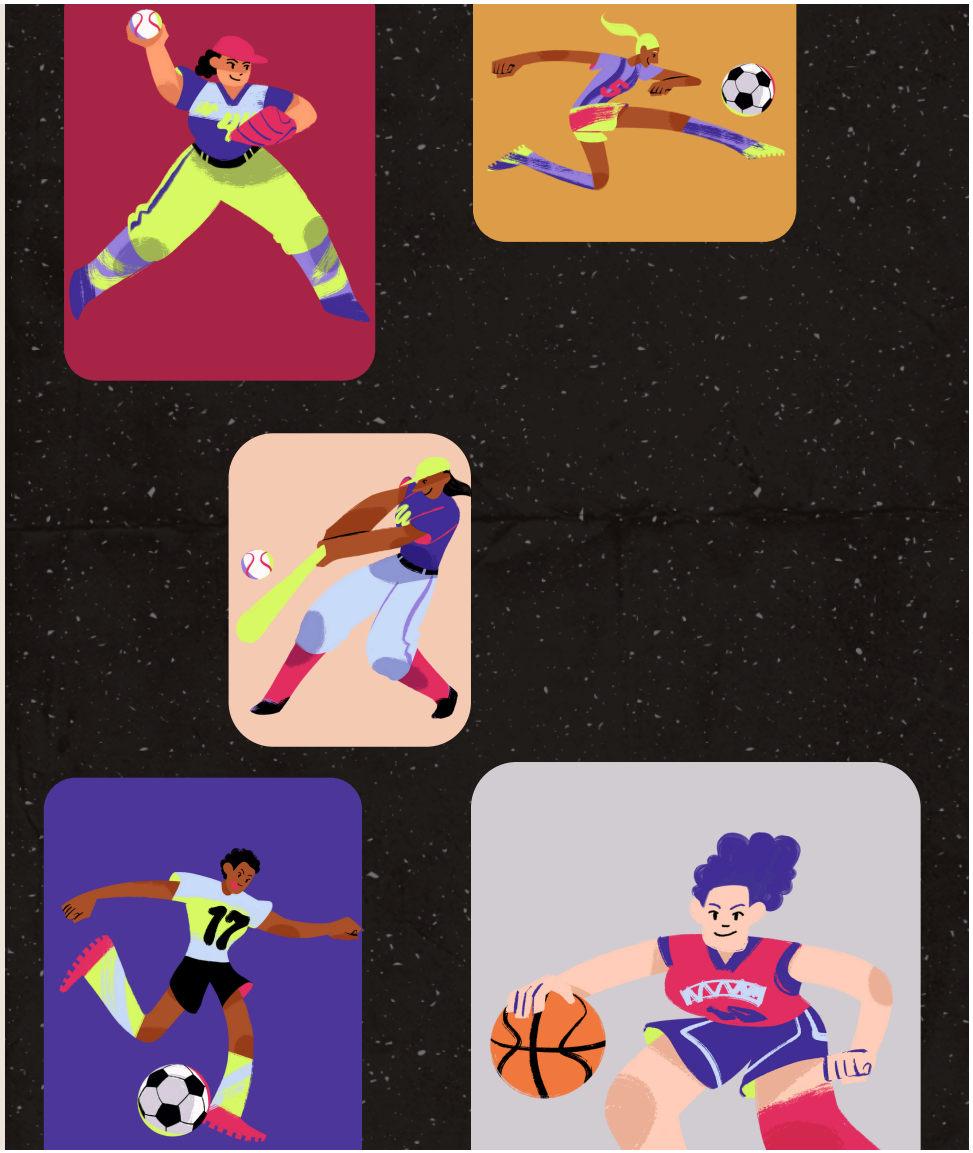
- Its business **operates through revenue streams** like broadcasting rights, ticket sales, **merchandise**, sponsorships, and licensing agreements.
- League earns significant revenue from media contracts with TV networks and digital streaming platforms for broadcasting games.

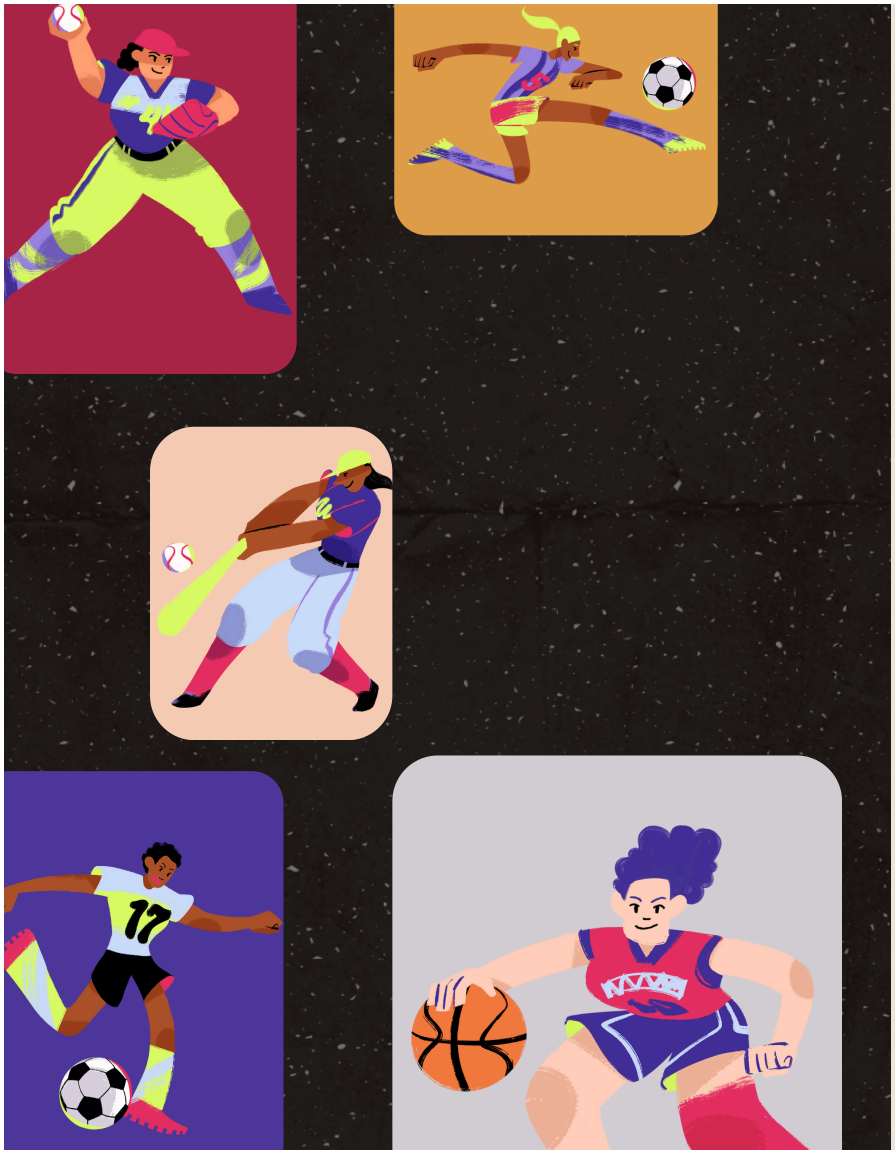
03

- MLB plans to **create a database of 10 years of historical data for partners** like TOPPS and 47 Brand to use for marketing ideas.
- The league aims **to transition from contracting** products **to establishing its own MLB wholesale brands**.

METADATA

Table Name	Attribute	Type/Constraints	Description
Ballpark	ParkID	Alphanumeric, unique	Unique identifier for each ballpark
	Name	Text string	Name of the ballpark
	City	Text string	City where the ballpark is located
	State	Text string	State where the ballpark is located, valid abbreviation
	Start	Date (YYYY-MM-DD)	Date the ballpark started being used
	End	Date (YYYY-MM-DD)	Date the ballpark stopped being used



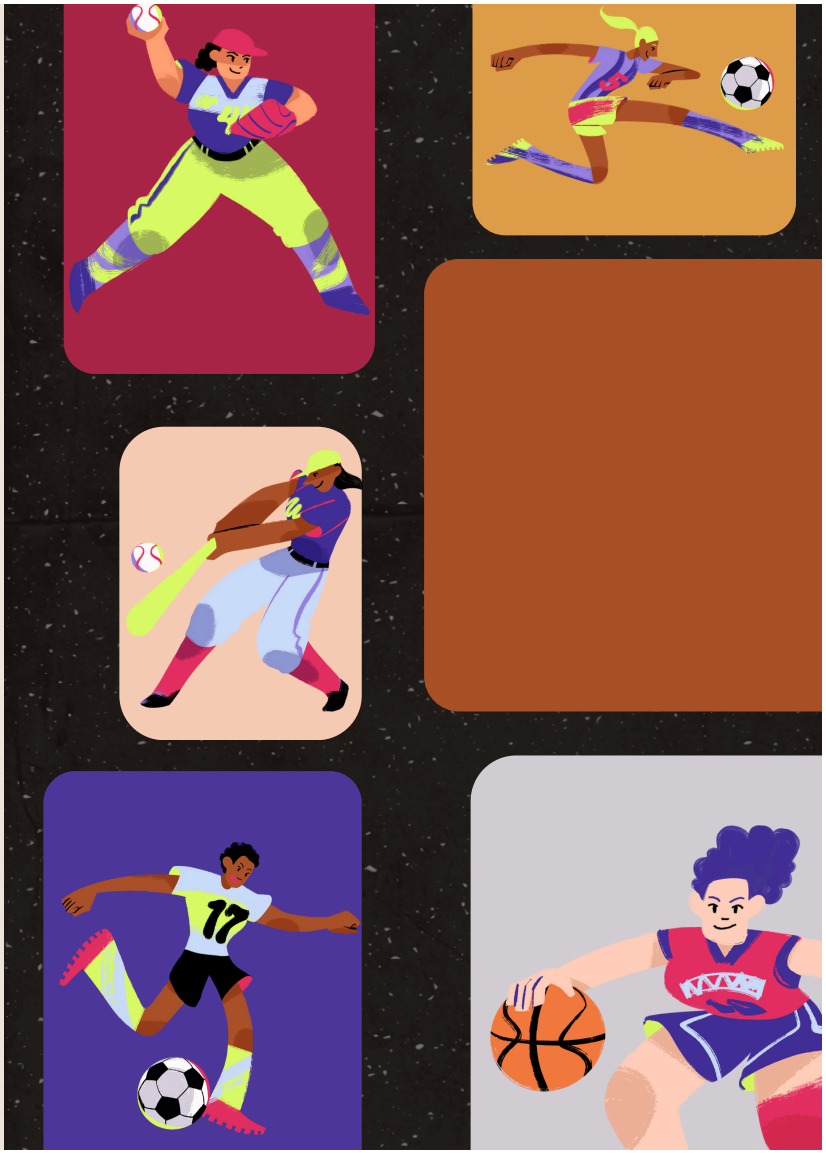


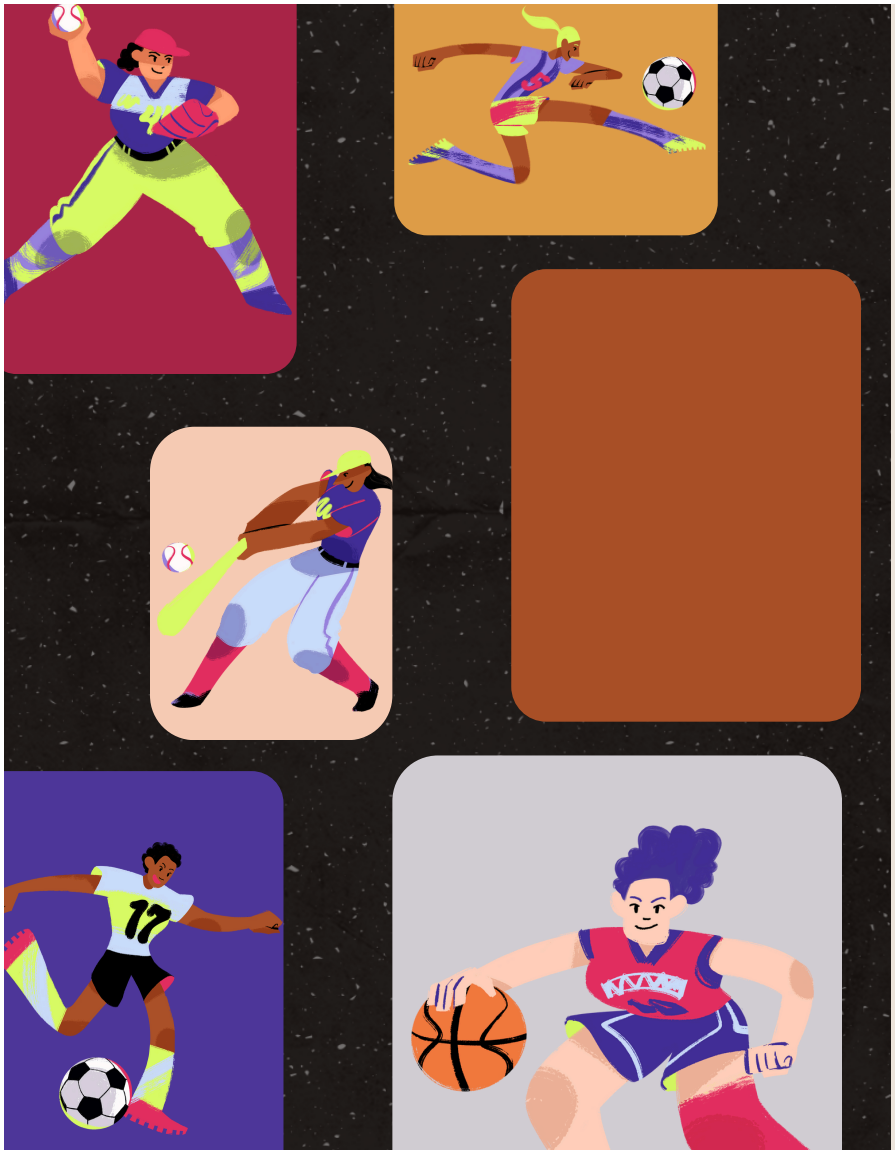
METADATA

Table Name	Attribute	Type / Constraints	Description
RosterOriginal	RosterID	Alphanumeric, unique	Unique identifier for each roster entry
	PlayerID	Alphanumeric, links to BioData	Identifier of a player
	NameLast	Text string	Last name of the player
	NameFirst	Text string	First name of the player
	Position	Text string, predefined	Position the player plays
	Theyear	Integer	Year the roster entry is relevant

METADATA

Table Name	Attribute	Type/Constraints	Description
WorldSeries	ParkID	Alphanumeric, links to Ballpark	Identifier of the ballpark for the World Series game
	BioID	Alphanumeric, links to BioData	Identifier of a player's biographical data
	RosterID	Alphanumeric, links to RosterOriginal	Identifier of the roster entry
	GameID	Alphanumeric, links to Games	Identifier of the game
	TeamID	Alphanumeric, links to Teams	Identifier of the team



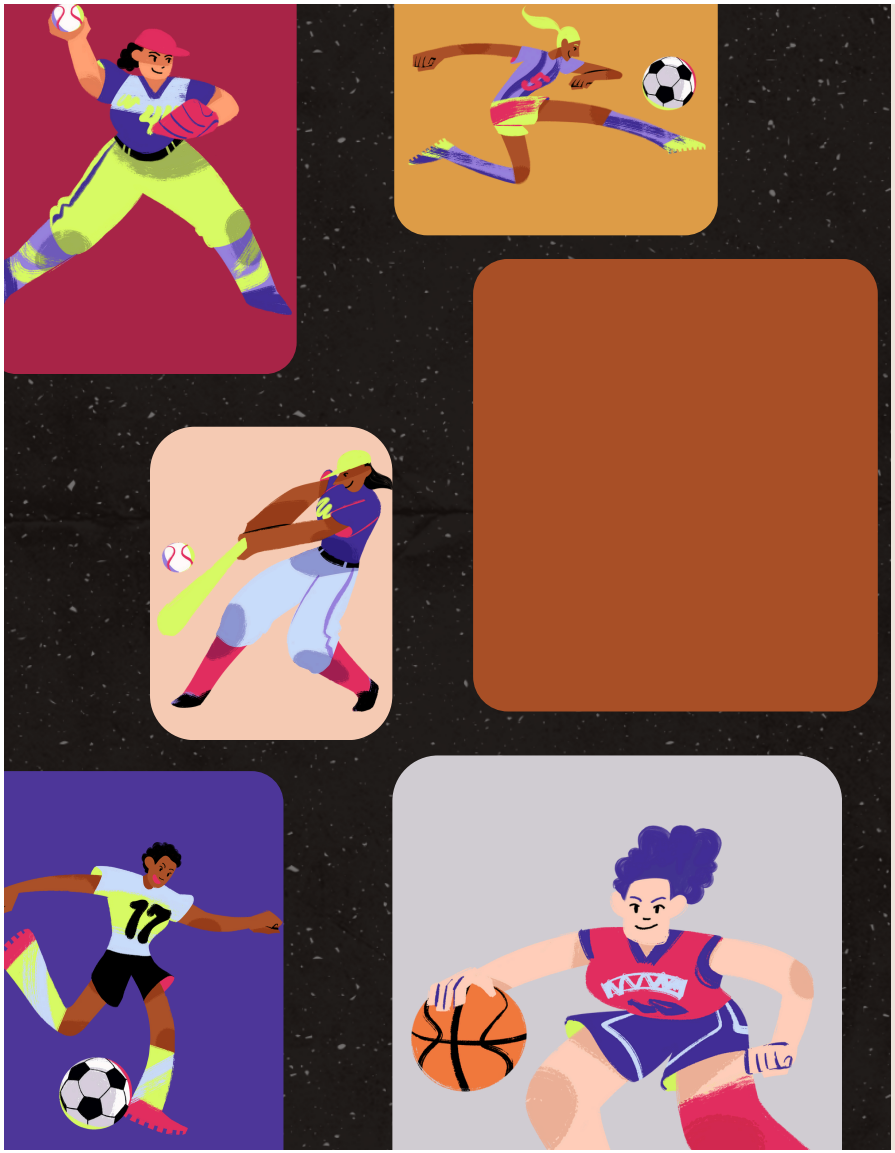


METADATA

Table Name	Attribute	Type/Constraints	Description
BioData	BioID	Alphanumeric, unique	Unique identifier for a player's biographical data
	NameFirstMiddle	Text string	The player's first and middle names
	NameLast	Text string	The player's last name
	Birthdate	Date (YYYY-MM-DD)	The player's date of birth
	Player_debut	Date (YYYY-MM-DD)	The date the player debuted in the league
	Height_inches	Non-negative integer	The player's height in inches

METADATA

Table Name	Attribute	Type/Constraints	Description
Games	GameID	Alphanumeric, unique	Unique identifier for each game
	Date	Date (YYYY-MM-DD)	The date the game was played
	Dayofweek	Text string	The day of the week the game was played
	Home_game no	Positive integer	The home game number in the season
	ParkID	Alphanumeric, links to Ballpark	Identifier of the ballpark where the game was played



METADATA

Table Name	Attribute	Type/Constraints	Description
GameStats	GameStatsID	Alphanumeric, unique	Unique identifier for each game statistics entry
	GameID	Alphanumeric, links to Games	Identifier of the game the statistics are for
	Visitor_score	Non-negative integer	The score of the visiting team
	Home_score	Non-negative integer	The score of the home team
	Winning_pitcher_id	Alphanumeric, links to RosterOriginal	Identifier of the winning pitcher in the roster

METADATA

Table Name	Primary Key	Attributes	Foreign Keys
Ballpark	ParkID	parkID, Name, City, State, Start, End	None
RosterOriginal	RosterID	rosterID, playerID, nameLast, nameFirst, position, theyear	playerID (links to BioData)
WorldSeries	None	parkID, [year], BioID, rosterID, gameID, teamID	parkID (links to Ballpark), BioID (links to BioData), rosterID (links to RosterOriginal), gameID (links to Games), teamID (links to Teams)
BioData	BioID	BioID, nameFirstMiddle, nameLast, birthdate, player_debut, height_inches	None
Games	GameID	gameID, date, dayofweek, home_gameno, parkID	parkID (links to Ballpark)
GameStats	GameStatsID	GameStatsID, gameID, visitor_score, home_score, winning_pitcher_id	gameID (links to Games), winning_pitcher_id (links to RosterOriginal)

RELATIONSHIPS AND BUSINESS RULES

1. Ballpark Rules:

- Each park must have a unique Park ID
- Deletion of a ballpark record requires handling related game records.
- Ballpark records become immutable once the usage period has ended.
- No updates allowed after End Date.

---

2. Roster Original Rules:

- Each roster entry must have a unique Roster ID.
- Player IDs in the roster must correspond to valid players in the Bio Data entity.
- The position attribute should adhere to predefined positions (e.g., pitcher, catcher, infielder, outfielder).

---

3. World Series Rules:

- A World Series entry must have a unique combination of year, BioID, Roster ID, Game ID, and Team ID.
- The BioID and Roster ID must correspond to valid entries in the Bio Data and Roster Original entities, respectively.
- Game ID and Team ID must correspond to valid entries in the Games and Teams entities, respectively.

4. Bio Data Rules:

- Each player's BioID must be unique.
- Birthdates should be valid dates, and players debut dates must be before the current date.
- Height should be recorded in inches and must be a non-negative integer.

---

5. Games Rules:

- Each game must have a unique gameID.
- Dates must be valid and formatted correctly.
- The park ID must correspond to a valid entry in the Ballpark entity.

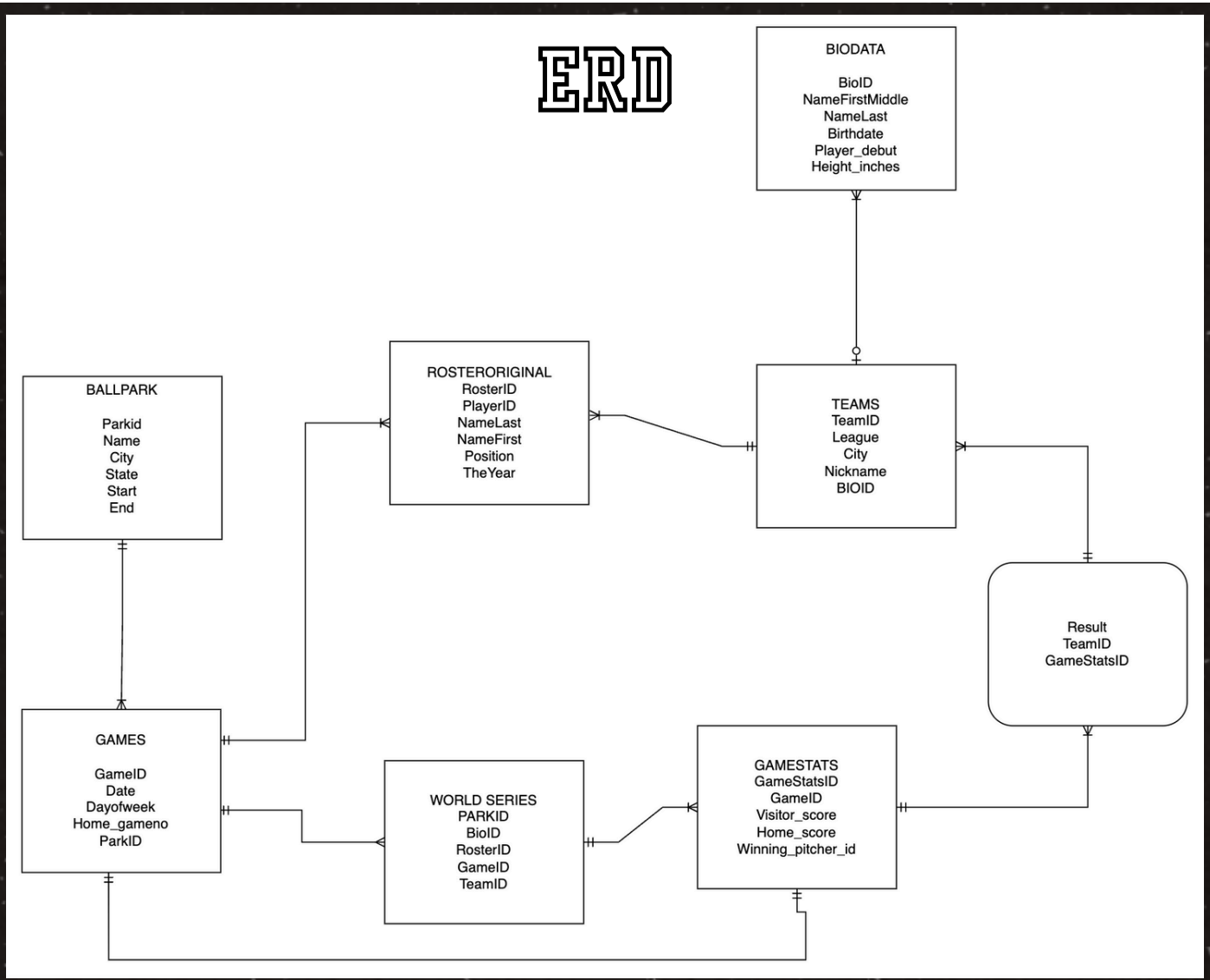
---

6. Game Stats Rules:

- Each Game Stats entry must have a unique GameStatsID.
- The gameID in GameStats must correspond to a valid gameID in the Games entity.
- The winning_pitcher_id must correspond to a valid player ID in the Roster Original entity.

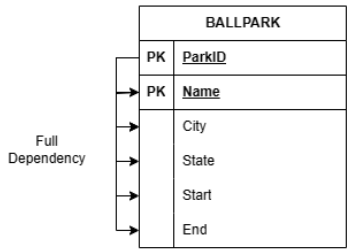
ENTITY-RELATIONSHIP DIAGRAM

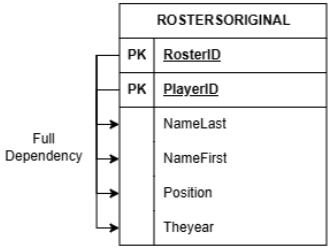
- The Entity-Relationship Diagram (ERD) visually represents the relationships between entities in the database.
- Key entities include Ballpark, Biodata, RostersOriginal, Teams, GameStats, Games, and World Series.
- Relationships are depicted with lines connecting entities, indicating how they are related (e.g., one-to-many, many-to-many).
- Attributes of each entity are listed, such as parkID, name, city, state for Ballpark, and teamID, league, city, and nickname for Teams.
- The ERD helps in understanding the database structure and how data is interconnected.

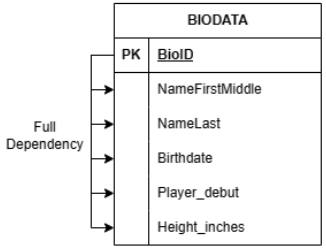

ERD
BIODATA
BioID
NameFirstMiddle
NameLast
Birthdate
Player_debut
Height_inches
ROSTERORIGINAL
RosterID
PlayerID
NameLast
NameFirst
Position
TheYear
TEAMS
TeamID
League
City
Nickname
BIOID
BALLPARK
Parkid
Name
City
State
Start
End
Result
TeamID
GameStatsID
GAMES
GameID
Date
Dayofweek
Home_gameno
ParkID
WORLD SERIES
PARKID
BioID
RosterID
GameID
TeamID
GAMESTATS
GameStatsID
GameID
Visitor_score
Home_score
Winning_pitcher_id

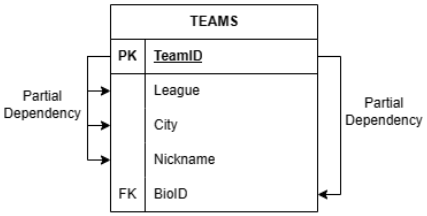
RELATIONAL DATA MODEL

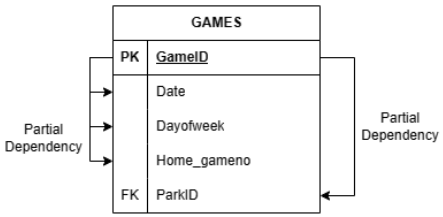
- The relational data model describes the database in terms of tables (relations), rows, and columns.
- Each entity in the ERD corresponds to a table in the relational model.
- For example, the Ballpark entity translates to a Ballpark table with columns for parkID, name, city, state, etc.
- Relationships between entities are implemented using foreign keys that reference primary keys in related tables.
- The relational model provides a more detailed view of how data is stored and how tables are linked through keys.

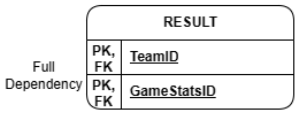

BALLPARK
PK ParkID
PK Name
City
State
Start
End
Full Dependency

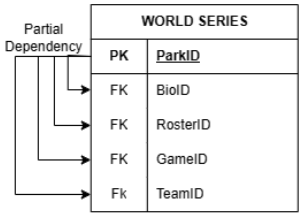

ROSTERSORIGINAL
PK RosterID
PK PlayerID
NameLast
NameFirst
Position
Theyear
Full Dependency

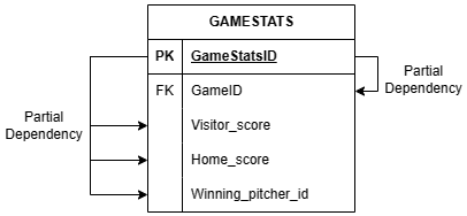

BIODATA
PK BioID
NameFirstMiddle
NameLast
Birthdate
Player_debut
Height_inches
Full Dependency

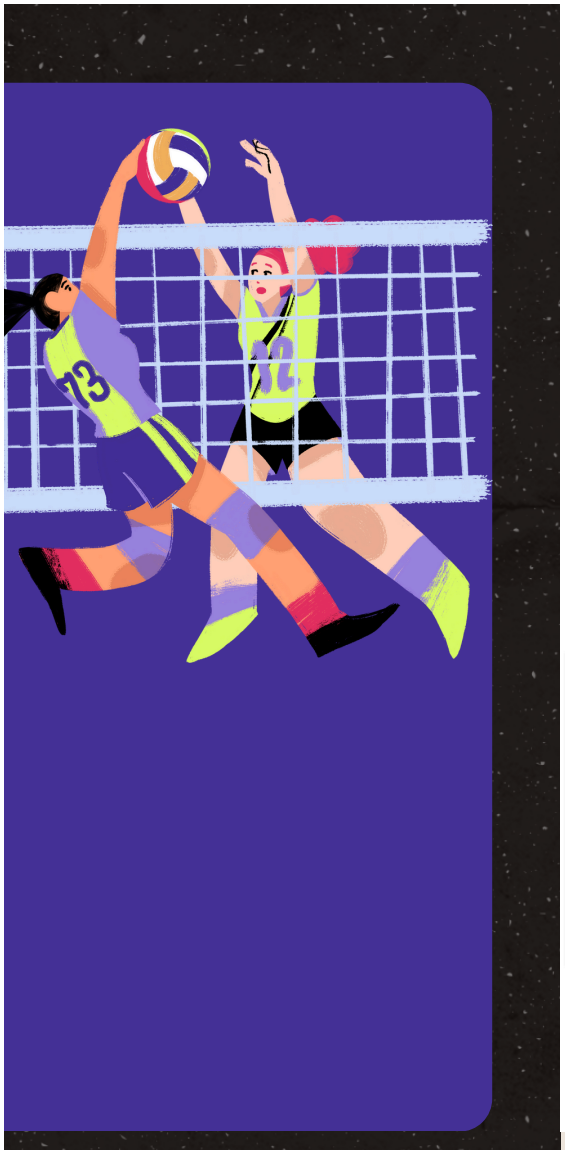

TEAMS
PK TeamID
League
City
Nickname
FK BioID
Partial Dependency
Partial Dependency

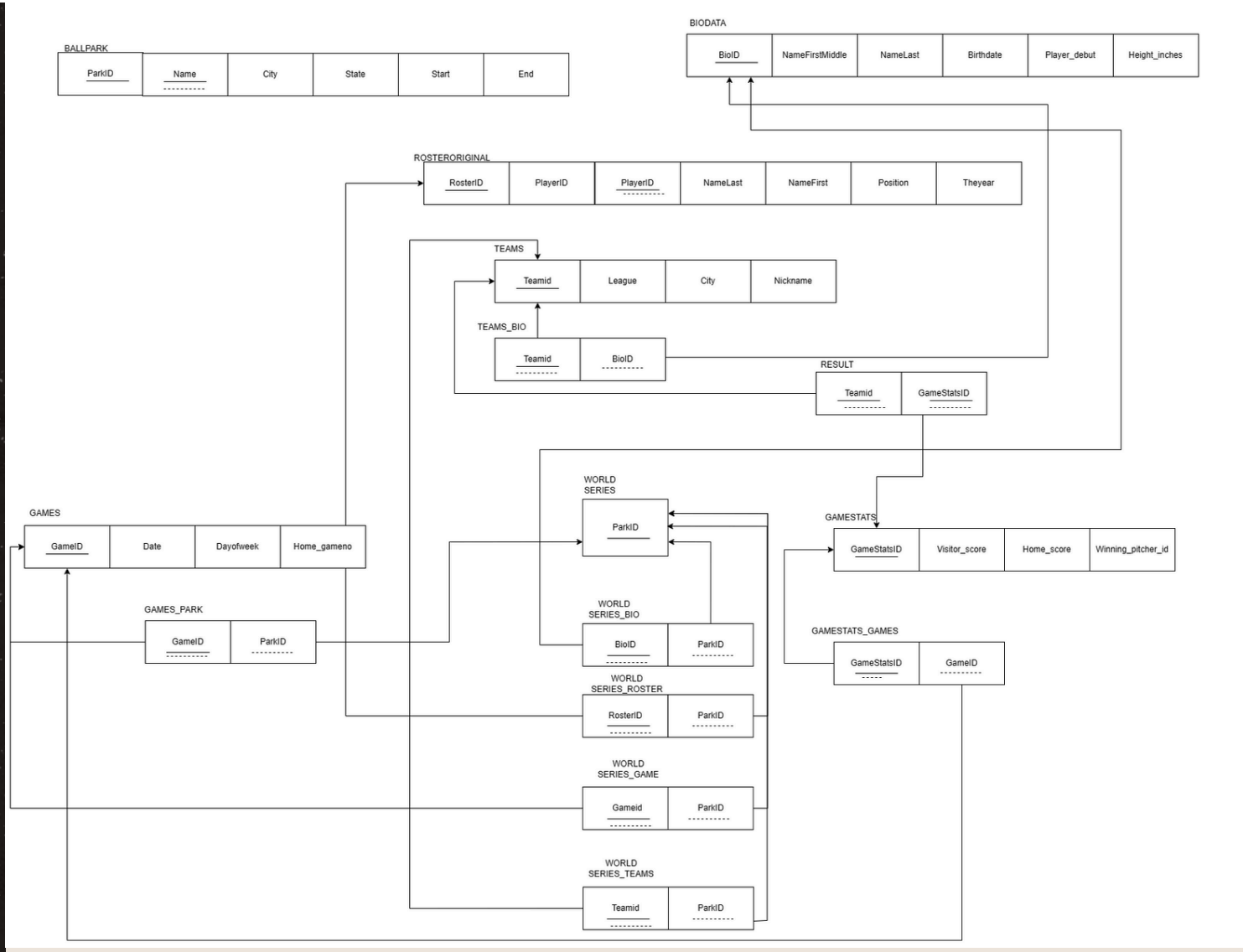

GAMES
PK GameID
Date
Dayofweek
Home_gameno
FK ParkID
Partial Dependency
Partial Dependency


RESULT
PK, FK TeamID
PK, FK GameStatsID
Full Dependency


WORLD SERIES
PK ParkID
FK BioID
FK RosterID
FK GameID
Fk TeamID
Partial Dependency


GAMESTATS
PK GameStatsID
FK GameID
Visitor_score
Home_score
Winning_pitcher_id
Partial Dependency
Partial Dependency




BALLPARK
ParkID
Name
City
State
Start
End
BIODATA
BioID
NameFirstMiddle
NameLast
Birthdate
Player_debut
Height_inches
ROSTERORIGINAL
RosterID
PlayerID
PlayerID
NameLast
NameFirst
Position
Theyear
TEAMS
Teamid
League
City
Nickname
TEAMS_BIO
Teamid
BioID
RESULT
Teamid
GameStatsID
GAMES
GameID
Date
Dayofweek
Home_gameno
WORLD SERIES
ParkID
GAMESTATS
GameStatsID
Visitor_score
Home_score
Winning_pitcher_id
GAMES_PARK
GameID
ParkID
WORLD SERIES_BIO
BioID
ParkID
GAMESTATS_GAMES
GameStatsID
GameID
WORLD SERIES_ROSTER
RosterID
ParkID
WORLD SERIES_GAME
GameId
ParkID
WORLD SERIES_TEAMS
Teamid
ParkID

DATA TABLES IN 3RD NORMAL FORM

BALLPARK

ParkID	Name	City	State	Start	End
ALB01	Riverside Park	Albany	NY	1880-09-11	1882-05-30
ALT01	Columbia Park	Altoona	PA	1884-04-30	1884-05-31
ANA01	Angel Stadium of Anaheim	Anaheim	CA	1966-04-19	
ARL01	Arlington Stadium	Arlington	TX	1972-04-21	1993-10-03
ARL02	Rangers Ballpark in Arlington	Arlington	TX	1994-04-11	2019-09-29
ARL03	Globe Life Field in Arlington	Arlington	TX	2020-07-24	
ATL01	Atlanta-Fulton County Stadium	Atlanta	GA	1966-04-12	1996-09-23
ATL02	Turner Field	Atlanta	GA	1997-04-04	2016-10-02
ATL03	Suntrust Park	Atlanta	GA	2017-04-14	
BAL01	Madison Avenue Grounds	Baltimore	MD	1871-07-08	1871-07-08
BAL02	Newington Park	Baltimore	MD	1872-04-22	1882-09-30

ROSTERORIGINAL

RosterID	PlayerID	NameLast	NameFirst	Position	Theyear
68518	avers001	Avery	Steve	P	1993
68519	bedrs001	Bedrosian	Steve	P	1993
68520	belli001	Belliard	Rafael	SS	1993
68521	berrd002	Berryhill	Damon	C	1993
68522	blauj001	Blauser	Jeff	SS	1993
68523	borbp001	Borbon	Pedro	P	1993
68524	breas001	Bream	Sid	1B	1993
68525	cabrf001	Cabrera	Francisco	1B	1993
68526	carar001	Caraballo	Ramon	2B	1993
68527	freem001	Freeman	Marvin	P	1993
68528	gantr001	Gant	Ron	OF	1993
68529	glavt001	Glavine	Tom	P	1993
68530	howe001	Howell	Jay	P	1993
68531	huntb001	Hunter	Brian	1B	1993
68532	jonec004	Jones	Chipper	SS	1993
68533	justd001	Justice	David	OF	1993
68534	klesr001	Klesko	Ryan	1B	1993
68535	lemkm001	Lemke	Mark	2B	1993
68536	lopej001	Lopez	Javy	C	1993

BIODATA

BioID	NameFirstMiddle	NameLast	Birthdate	Player_debut	Height_inch...
aardd001	David Allan	Aardsma	1981-12-27	2004-04-06	77
aaroh101	Henry Louis	Aaron	1934-02-05	1954-04-13	72
aarot101	Tommie Lee	Aaron	1939-08-05	1962-04-10	75
aasad001	Donald William	Aase	1954-09-08	1977-07-26	75
abada001	Fausto Andres	Abad	1972-08-25	2001-09-10	73
abadf001	Fernando Antonio	Abad	1985-12-17	2010-07-28	74
abadj101	John W.	Abadie	1850-11-04	1875-04-26	72
abbae101	Edward James	Abbaticchio	1877-04-15	1897-09-04	71
abbeb101	Bert Wood	Abbey	1869-11-11	1892-06-14	71
abbec101	Charles S.	Abbey	1866-10-14	1893-08-16	68.5
abboc001	Cory James	Abbott	1995-09-20	2021-06-05	73
abbod101	Leander Franklin	Abbott	1862-03-16	1890-04-19	71
abbof101	Harry Frederick	Abbott	1874-10-22	1903-04-25	70
abbog001	William Glenn	Abbott	1951-02-16	1973-07-29	78
abboj001	James Anthony	Abbott	1967-09-19	1989-04-08	75
abboj002	Jeffrey William	Abbott	1972-08-17	1997-06-10	74
abbok001	Lawrence Kyle	Abbott	1968-02-18	1991-09-10	76
abbok002	Kurt Thomas	Abbott	1969-06-02	1993-09-07	71
abboo101	Ody Cleon	Abbott	1886-09-05	1910-09-10	69

RESULT

TeamID	GameStatsID
BS1	9801115
CH1	2353165
CL1	2600284
FW1	6833022
NY2	1607896
PH1	3310752
RC1	6321704
TRO	6924929
WS3	7528238
BL1	2754744
BR1	9281355
BR2	2610621
MID	9112811
WS4	8561631
BL4	3597597
ELI	9126030
PH2	8065671
WS5	1786360
CH2	7462591

DATA TABLES IN 3RD NORMAL FORM

GAMES

GameID	Date	Day_of_Week	Home_Game_No
HOU202211050	11/5/2022 0:00	Sat	6
PHI202211030	11/3/2022 0:00	Thu	5
PH/202211020	11/2/2022 0:00	Wed	4
PH|202211010	11/1/2022 0:00	Tue	3
HOU202210290	10/29/2022 0:00	Sat	2
HOU202210280	10/28/2022 0:00	Fri	1
PHI202210230	10/23/2022 0:00	Sun	5
NYA202210230	10/23/2022 0:00	Sun	4
PHI202210220	10/22/2022 0:00	Sat	4
NYA202210220	10/22/2022 0:00	Sat	3

TEAMS

TeamID	League	City	Nickname
SFN	NL	San Francisco	Giants
PIT	NL	Pittsburgh	Pirates
PHI	NL	Philadelphia	Phillies
NYN	NL	New York	Mets
MIL	NL	Milwaukee	Brewers
MIA	NL	Miami	Marlins
SDN	NL	San Diego	Padres

GAMESTATS

GameStatsID	Visitor_score	Home_score	Winning_pitcher_id
9801115	0	2	mathb101
2353165	20	18	spala101
2600284	12	4	prata101
6833022	12	14	zettg101
1607896	9	5	spala101
3310752	18	10	zettg101

WORLD SERIES

ParkID
ALB01
ALT01
ANA01
ARL01

WORLD SERIES_BIO

ParkID	BioID
ALB01	adams102
ALT01	adamt001
ANA01	adamw001
ARL01	adamw002
ARL02	adamw101
ARL03	adcoj101
ATL01	adcon001
ATL02	addib101

GAMES_PARK

GameID	Park_ID
HOU202211050	HOU03
PHI202211030	PHI13
PH/202211020	PHI13
PH|202211010	PHI13
HOU202210290	HOU03
HOU202210280	HOU03
PHI202210230	PHI13
NYA202210230	NYC21
PHI202210220	PHI13
NYA202210220	NYC21

TEAMS_BIO

TeamID	Bioid
SFN	aardd001
PIT	aardd001
PHI	aardd001
NYN	aardd001
MIL	aardd001
MIA	aardd001
SDN	aardd001

GAMESTATS_GAMES

GameStatsID	GameID
9801115	FW1187105040
2353165	WS3187105050
2600284	RC1187105060
6833022	CH1187105080
1607896	TRO187105090
3310752	CL1187105110

WORLD SERIES_ROSTER

ParkID	RosterID
ALB01	32559
ALT01	24900
ANA01	23515
ARL01	22155
ARL02	27456
ARL03	39613

WORLD SERIES_TEAM

ParkID	TeamID
ALB01	BS1
ALT01	CH1
ANA01	CL1
ARL01	FW1
ARL02	NY2
ARL03	PH1
ATL01	RC1
ATL02	TRO

WORLD SERIES_GAME

ParkID	GameID
ALB01	FW1187105040
ALT01	WS3187105050
ANA01	RC1187105060
ARL01	CH1187105080
ARL02	TRO187105090
ARL03	CL1187105110
ATL01	CL1187105130

QUERY & FINDINGS

PRIMARY OBJECTIVES

↓

QUERIES

BUSINESS QUERY CASE STUDIES

1. MLB wants to **query Ballpark information to create a vintage ballpark graphic t-shirt line for ballparks that have participated in many World Series** (Yankees and Dodgers are some excellent teams that come to mind)

2. The MLB is making a documentary on the Astros cheating scandal of 2017 to clear up a lot of animosity and dust the league has held since the event. They wish to **query the 2017 roster that won the World Series through cheating for the documentary's producer**.

3. TOPPS wants to make some historic Game cards for patriotic holidays, i.e., Cinco de Mayo, 4th of July, etc.; **they want to query matchups that happened on those holidays.**

4. The MLB is conducting an internal investigation on the distribution of player ages from the 2022 season to see the age and locations of teams to market new adult beverage companies that they have acquired to the respective fan base, whether that may be veteran players or rookies. They want a base, whether that may be veteran players or rookies. They want to to **query the count of each birthday year from the 2022 season and their team/roster** (NOT distinct and outer join required).

5. Ice cream partner Silver Spoon is asking to **query for Summer attendance in Arizona between June 1st, 2022 and September 1st, 2022, but in Chase Field** (subquery).

QUERY #1 BALLPARK T-SHIRT CAMPAIGN

```sql
select p.name, COUNT(distinct ws.GameID) as
total_games
from world_series ws

join ballparks p on ws.PARKID = p.PARKID
group by p.PARKID, p.NAME
order by total_games desc
limit 5;
```

ballparks 1

```
select p.name as Name, COUNT(distinct ws.Gai  Enter a S
```

Grid | Text

	Name	Total_games
1	Minute Maid Park	12
2	Dodger Stadium	7
3	Globe Life Field in Arlington	6
4	Kauffman Stadium	6
5	Fenway Park	5

- Query **ballpark information** to create a vintage ballpark graphic t-shirt line for ballparks that have participated in many World Series.
- We will have to group ballparks and filter by ones that have played **at least one World Series game**. We will count the total games that the ballpark has had in all World Series in the last 29 years. Our finished query will display ballpark and count of World Series in that park in descending order.
- The value we gain: we can see the **top 5 organizations** and their ballpark-affiliated brands (Minute Maid, Dodgers, Globe Life). We can approach the brand to strike the clothing deal and begin designing and producing the clothing.

QUERY #2 2017 SCANDAL ROSTER

- MLB wants to film a documentary on the **2017 Astros** that were caught cheating. The producer needs a query of the roster from this 2017 World Series.

- The columns outputted will be the players first and last names, positions, and ages (2024 - year) from the **2017 championship team** that played the World Series last games.

- The value we gain: the producer and MLB expert can look at the positions of players and see who had the most significant roles in the **cheating scandal**, who they are to contact for interviews or statements, and also their age.

```sql
SELECT
    b.nameFirst AS "First Name",
    b.nameLast AS "Last Name",
    r.position AS "Position",
    (2024 - CAST(SUBSTRING(b.birthdate, 1, 4) AS UNSIGNED)) AS "Age"
FROM
    (SELECT * FROM Rosters WHERE theyear = 2017 AND theteam = 'HOU') AS r
JOIN
    (SELECT * FROM Biodata) AS b ON r.playerID = b.bioID;
```

	First Name	Last Name	Position	Age
1	Jose	Altuve	2B	34
2	Norichika	Aoki	OF	42
3	Carlos	Beltran	OF	47
4	Alexander	Bregman	SS	30
5	Juan	Centeno	C	35
6	Tyler	Clippard	P	39
7	Carlos	Correa	SS	30
8	Jonathan	Davis	3B	31
9	Christopher	Devenski	P	34
10	Dayan	Diaz	P	35
11	Michael	Feliz	P	31
12	Michael	Fiers	P	39

QUERY #3 HOLIDAY GAME CARDS

```sql
SELECT
    CONCAT(b.nameFirst, ' ', b.nameLast) AS player_name,
    t.` nickname` AS team_name,
    g.date AS game_date
FROM
    Games g
    JOIN Teams t ON g.visitor = t.teamID OR g.home = t.teamID
    JOIN Biodata b ON g.visitor_manager_id = b.bioID OR g.home_manager_id = b.bioID
WHERE
    MONTH(g.date) = 7 AND DAY(g.date) = 4
GROUP BY
    CONCAT(b.nameFirst, ' ', b.nameLast),
    t.` nickname`,
    g.date
ORDER BY
    player_name DESC,
    team_name DESC,
    game_date DESC;
```

SELECT CONCAT(b.nameFirst, ' ', b.nameLast) A | Enter a SQL expression

Grid | Text | Record

	Player_name	Team_name	Game_date
1	Walter Weiss	Rockies	2016-07-04
2	Walter Weiss	Rockies	2015-07-04
3	Walter Weiss	Rockies	2014-07-04
4	Walter Weiss	Rockies	2013-07-04
5	Walter Weiss	Giants	2016-07-04
6	Walter Weiss	Dodgers	2014-07-04
7	Walter Weiss	Dodgers	2013-07-04

- **TOPPS** wants to make some historic Game cards for patriotic holidays: Cinco De Mayo and 4th of July.

- We will group by team/roster and filter by date date: May 5, 2012-2022 and July 4, 2012-2022. Our output columns will be the team name, the player on the roster, and the date of the holiday of either **Cinco de Mayo** or 4th of July.

- The value we gain: we now have our complete list of players in descending alphabetical order with the players. We can hand this to TOPPS to make a decision based on a specific player they want so they can **create the baseball card** in their set.

QUERY #4 AGE MARKETING CAMPAIGN

- MLB wants to market adult beverages to a fan base but wants to find a team with **older players** to appeal to a fan base that is likely to be older rather than rookies with young fan followings.

- The columns outputted will be the team name, player name from the roster, city, and **count of the years** from birthday joined through team.

- The value we gain: we can see the count of birth years across teams. One example that we can see is the total number of 34-year-olds on the **Brewers**. Wisconsin has a lot of beer production in the area, so there is a big market that they can tap into there with the likely older fans following the older players (rookies are 18-22 for reference).

```
SELECT
    t.` nickname` AS team_name,
    CEIL(DATEDIFF(CURDATE(), STR_TO_DATE(b.birthdate, '%Y-%m-%d')) / 365) AS age,
    COUNT(*) AS player_count
FROM
    mlb_new.teams t
JOIN
    mlb_new.rosters r ON t.teamID = r.theteam
JOIN
    mlb_new.biodata b ON r.playerID = b.bioID
WHERE
    CEIL(DATEDIFF(CURDATE(), STR_TO_DATE(b.birthdate, '%Y-%m-%d')) / 365) > 21
    AND CEIL(DATEDIFF(CURDATE(), STR_TO_DATE(b.birthdate, '%Y-%m-%d')) / 365) <
GROUP BY
    t.` nickname`,
    CEIL(DATEDIFF(CURDATE(), STR_TO_DATE(b.birthdate, '%Y-%m-%d')) / 365)
ORDER BY
    player_count DESC,
    t.` nickname`,
    age;
```

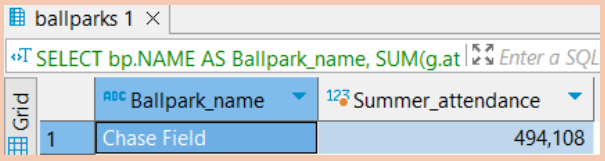
Grid | Text

	Team_name	Age	Player_count
1	Brewers	34	104
2	Brewers	32	82
3	Brewers	30	76
4	Marlins	34	71
5	Orioles	30	71
6	Pirates	32	70
7	Brewers	33	68
8	Angels	33	62
9	Tigers	34	61
10	Diamondbacks	34	60
11	Marlins	33	59
12	Blue Jays	34	58
13	Devil Rays	34	58
14	Mariners	34	57

QUERY #5 ICE CREAM DEMAND

- MLB partner wants to determine ice cream demand during hot days with the formula, but they need MLB's **attendance** data.
- The columns outputted will be summer attendance grouped by the three-month period of June 1 to August 1, 2022 for only the **Chase Field ballpark** in Arizona.
- The value we gain: we now have the total attendance of **494,108 people** from the Chase Field ballpark in a season of hot temperatures in Arizona, so Silver Spoon can now prepare production and distribution for the next season. It can add these seasonal insights to its model.

```sql
SELECT
    bp.NAME AS ballpark_name,
    SUM(g.attendance) AS summer_attendance
FROM
    Games g
JOIN
    mlb_new.ballparks bp ON g.parkID = bp.PARKID
WHERE
    bp.NAME LIKE '%chase%'
    AND g.date BETWEEN '2022-06-01' AND '2022-08-
GROUP BY
    bp.NAME;
```

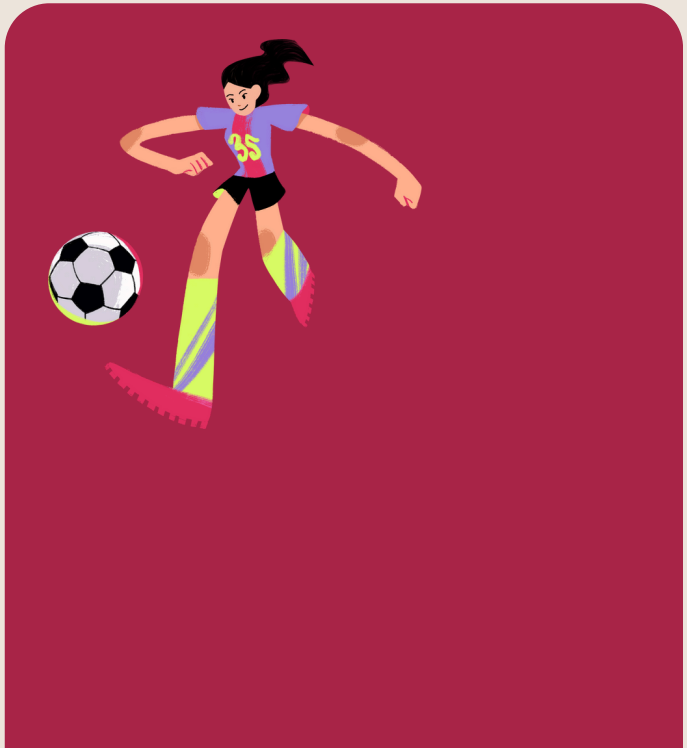


SCOPE FOR EXTENSION

- Implementing advanced SQL queries to analyze **key metrics** - Compute batting averages, ranking players, calculate team/league averages

- Player **Injury Tracking** - Create separate table to store injury, games missed, implement triggers to update

- Implementing mechanisms to provide **real-time updates** on game scores, player statistics, and league standings - Create message queues using Kafka, provide notification services based on alerts.



THANK YOU!!!