

Visualització de dades: Activitat 7

Selecció del conjunt de dades

Gisela Claret Tortajada - 3 de maig del 2021

1. El conjunt de dades “Lahman’s Baseball Database” i la rellevància del conjunt de dades en el seu context

a. Introducció al conjunt de dades

El conjunt de dades que s’ha seleccionat és l’anomenat “Lahman’s Baseball Database”. Es tracta d’un conjunt de 27 arxius CSV que contenen la informació més important de les lligues de beisbol professional dels EUA entre els anys 1871 i el 2020. Concretament s’inclouen les dades de les lligues de beisbol ja no existents anomenades “National Association”, “American Association”, “Players League”, “Union Association” i “Federal League”, així com les dades de les dues lligues majors de beisbol actuals “American League” (AL) i “National League” (NL) agrupades dins de l’organització MLB (Major League Baseball).

b. Font de les dades

Les dades han estat extretes de la pàgina web del mateix autor de la base de dades, Sean Lahman. L’enllaç on es pot trobar el conjunt de dades és el següent:

<http://www.seanlahman.com/baseball-archive/statistics>

La base de dades va estar creada per Sean Lahman el 1996 i l’ha estat actualitzant des de llavors. Cal mencionar que el conjunt de dades es tracta d’un seguit de col·laboracions de Sean Lahman amb diferents reporters i persones importants dins el món del beisbol, entre ells Pete Palmer, un estatista d’esports americà, o Sean Forman, el creador de la famosa pàgina web sobre beisbol baseball-reference.com.

El treball de Lahman està sotmès a la llicència **CC BY-SA 3.0** que permet la còpia i redistribució del material en qualsevol mitjà i format i també permet barrejar, transformar i construir sobre el material per qualsevol finalitat, fins i tot comercial [1].

El conjunt d’arxius original, el conjunt de dades creat a partir dels 27 arxius i el codi python creat pel tractament de les dades han estat presentats en el repositori **Github** següent:

https://github.com/giselacaret/PRACTICA_visualitzacio_de_dades

c. L’autor de les dades i la rellevància en el seu context

Sean Lahman és un periodista dels EUA que ha treballat per diaris com USA Today Network, New York Sun i Rochester Democrat and Chronicle, com a membre dels seus equips d’investigació. Ha guanyat diversos premis pels seus articles de denúncia contra violència armada o abusos sexuals a menors per part de la Església catòlica, entre d’altres. Actualment, a part de ser col·laborador del Rochester Democrat and Chronicle, treballa per la “Society for American

Baseball Research” coordinant projectes de recollida de dades. La seva base de dades “Lahman’s Baseball Database” i bases de dades seves d’altres esports han estat utilitzades per la reedició d’enciclopèdies de baseball, bàsquet, futbol i tennis. La seva base de dades sobre el beisbol ha resultat tan important que s’ha inclòs com a “R Package” [2][3], per tal de facilitar la descàrrega de les dades a l’hora d’utilitzar el llenguatge de programació R.

d. Motius per l’elecció de les dades

Els motius pels quals he decidit escollir aquesta base de dades són, en primer lloc, el meu interès personal per aquest esport. En segon lloc, per la importància que té aquesta base de dades en el món del beisbol, ja que es tracta de la base de dades més completa i que, a més a més, conté informació des d’abans de la creació de la “Major League Baseball”. En tercer lloc, a l’hora d’escollir aquesta base de dades he tingut en compte la seva complexitat, ja que resulta un conjunt de dades molt extens que requereix una mica de treball previ per seleccionar i filtrar les dades rellevants per la realització de les visualitzacions. Per últim, aquesta base de dades consta de tota la informació que requereixo per respondre totes les preguntes que vull respondre en les visualitzacions que presentaré en la segona part de la pràctica, i que explicaré en els següents apartats.

e. Un breu resum del beisbol

El beisbol és un esport en el qual s’enfronten 2 equips amb 9 jugadors cada un. Els equips prenen torns per jugar de forma ofensiva, colpejant la bola amb el bat i corrent per les bases, i de forma defensiva, jugant al camp i llençant la pilota perquè l’altre equip colpegi la bola. El jugador encarregat de llençar la pilota s’anomena *pitcher*. L’objectiu del joc és que durant el torn ofensiu els jugadors colpegin la pilota de tal manera que els defensors no la puguin atrapar fàcilment i córrer al voltant del camp. Cada vegada que un jugador fa una volta sencera al camp s’atorga un punt a l’equip. Al final del partit, l’equip que té més punts és el guanyador. La duració del partit no es determina per temps sinó per torns, cada vegada que els dos equips han jugat de forma ofensiva i defensiva constitueix un inning. Un partit consisteix en 9 innings, i en cada un dels innings, els torns acaben quan els defensors eliminen 3 batejadors.

En l’Annex s’ha inclòs una definició més extensa d’alguns dels termes més importants per entendre amb profunditat les variables de la base de dades.

2. Contingut de la base de dades: els 27 arxius.

La base de dades proporcionada per Sean Lahman consta de 27 arxius CSV, que es poden trobar al repositori GitHub proporcionat en el primer apartat. Aquests 27 arxius de mides diferents presenten estadístiques de tota mena referents al beisbol professional d’EUA. Els arxius són:

- a. **AllStarFull.csv** – Dades sobre els jugadors que han participat en partits All-Star. Es tracta d’un conjunt de 5.375 entrades i 8 columnes. Pertanyent als anys 1933 – 2019.
- b. **Appearances.csv** – Dades sobre les posicions dels jugadors per equip i any. Es tracta d’un conjunt de 108.717 entrades i 21 columnes, pertany als anys 1871 – 2020.

- c. **AwardsManagers.csv** – Dades sobre els premis guanyats pels mànagers. Conté un total de 179 entrades amb 6 columnes i pertany als anys 1983 – 2016.
- d. **AwardsPlayers.csv** – Dades sobre els premis guanyats pels jugadors. Conté 6.236 entrades, 7 columnes i correspon als anys 1877 – 2017.
- e. **AwardsShareManagers.csv** – Dades sobre les votacions dels premis dels mànagers. Conté 425 entrades, 7 columnes i correspon als anys 1983 – 2016.
- f. **AwardsSharePlayers.csv** – Dades sobre les votacions dels premis dels jugadors. Conté 6.879 entrades, 7 columnes i correspon als anys 1956 – 2016.
- g. **Batting.csv** – Dades estadístiques sobre el batejadors. Conté 108.789 entrades, 22 columnes i correspon als anys 1871 – 2020.
- h. **BattingPost.csv** – Dades dels batejadors dels partits de posttemporada. Conté un total de 15.460 entrades, 22 columnes i correspon als anys 1884 – 2020.
- i. **CollegePlaying.csv** – Llistat de jugadors amb les universitats on van anar. Conté 17.350 entrades, 3 columnes i pertany als anys 1864 – 2014.
- j. **Fielding.csv** – Dades sobre les estadístiques dels jugadors de camp. Conté un total de 144.768 entrades, 18 columnes i pertany als anys 1871 – 2020.
- k. **FieldingOF.csv** – Dades sobre els jugadors de camp de tipus outfielder. Conté 12.028 entrades, 6 columnes i pertany als anys 1871 – 1955.
- l. **FieldingOFsplit.csv** – Dades sobre els jugadors de camp de tipus infielders. Conté 33.801 entrades, 13 columnes i pertany als anys 1954 – 2020.
- m. **FieldingPost.csv** – Dades sobre els jugadors de camp en partits de posttemporada. Conté 14.657 entrades, 17 columnes i pertany als anys 1903 – 2020.
- n. **HallOfFame.csv** – Dades sobre les votacions del Hall of Fame. Conté 4.191 entrades, 9 columnes i pertany als anys 1936 – 2018.
- o. **HomeGames.csv** – Dades sobre els partits jugats en l'estadi propi dels diferents equips. Conté 3.108 entrades, 9 columnes i pertany als anys 1871 – 2019.
- p. **Managers.csv** – Dades estadístiques dels mànagers. Conté 3.567 entrades, 10 columnes i pertany als anys 1871 – 2020.
- q. **ManagersHalf.csv** – Dades dels mànagers a mitjans de temporada. Conté 93 entrades, 10 columnes i pertany als anys 1892 – 1981.
- r. **Parks.csv** – Dades sobre els estadis dels equips. Conté 255 entrades i 6 columnes.
- s. **People.csv** – Informació sobre els jugadors, com el nom i les dades biogràfiques. Conté 20.153 entrades i 24 columnes.
- t. **Pitching.csv** – Dades sobre les estadístiques dels pitchers. Conté 48.399 entrades, 30 columnes i pertany als anys 1871 – 2020.
- u. **PitchingPost.csv** – Dades sobre les estadístiques dels pitchers en la posttemporada. Conté 6.120 entrades, 30 columnes i pertany als anys 1884 – 2020.
- v. **Salaries.csv** – Dades sobre el sou anual dels jugadors. Conté 26.430 entrades, 5 columnes i pertany als anys 1985 – 2016.
- w. **Schools.csv** – Dades llista de les escoles a les quals van anar els jugadors. Conté un total de 1.208 entrades i 5 columnes.
- x. **SeriesPost.csv** – Dades sobre la sèrie mundial de beisbol. Conté 358 entrades, 9 columnes i pertany als anys 1884 – 2020.
- y. **Teams.csv** – Dades sobre els equips, incloent-hi el nombre de partits guanyats i perduts per any. Conté un total de 2.954 entrades, 48 columnes i pertany als anys 1871 – 2020.
- z. **TeamsFranchises.csv** – Dades de les franquícies dels equips. 120 entrades, 4 columnes.
- aa. **TeamsHalf.csv** – Dades dels equips a mitjans de temporada. Conté 54 entrades, 10 columnes i solament pertany a l'any 1981.

Tal com es pot observar és una base de dades molt extensa, que permet una gran quantitat d'anàlisi diferent. És per això que cal tenir ben clar quina és la informació que es voldrà mostrar en les visualitzacions i cal realitzar un tractament previ de les dades. Utilitzant solament els arxius necessaris, agrupant-los i eliminant les columnes que no siguin necessàries.

3. Preguntes a respondre amb la Visualització

En les visualitzacions que es realitzaran en la segona part de la pràctica es volen respondre les preguntes següents:

- Quina ha estat l'evolució temporal del sou dels pitchers?
- Què condiona que un pitcher tingui un salari major o menor?

Així doncs, les consideracions que s'han tingut en compte a l'hora de crear l'arxiu **baseball_dataset.csv** presentat en la pràctica han estat les següents:

- Cal que contingui solament les dades dels jugadors que tenen la posició de pitcher.
- Cal que contingui el sou anual
- Cal que contingui les estadístiques dels jugadors, per exemple el nombre de partits guanyats o el nombre de jugadors eliminats.
- Els anys que s'han tingut en compte són entre el 1985 i el 2016 per dos motius:
 - o Les dades que disposem del sou solament corresponen als anys 1985 – 2016
 - o Cal que dins d'aquests anys no s'hagin realitzat canvis de reglaments importants que afectin els pitchers [4]

En el cas d'aquest conjunt de dades no s'ha pogut tenir en compte la perspectiva de gènere, però, com veurem, s'avaluarà el lloc de naixement dels jugadors, per avaluar si hi ha algun condicionant d'aquest tipus.

4. Complexitat del conjunt de dades – agrupament i selecció.

Per tal de respondre a les preguntes plantejades en l'apartat 3, s'han seleccionat solament els arxius que contenen les dades dels jugadors que ocupen la posició pitcher. Per tant, s'han utilitzat els següents arxius:

- **Pitching.csv**
- **Salaries.csv**
- **People.csv**
- **Teams.csv**

Així amb l'ajuda del llenguatge de programació python s'ha procedit a agrupar les dades i eliminar les files o columnes que no cal tenir en compte segons les consideracions presentades en l'apartat anterior.

El notebook creat ha estat adjuntat en el repositori GitHub creat per la pràctica, presentat en l'apartat 1.b.

El procediment seguit pel tractament de les dades en el notebook ha estat el següent:

En primer lloc s'han carregat les dades utilitzant la llibreria pandas.

Seguidament s'ha passat a tractar les dades corresponents a l'arxiu dels **salaris**. Aquest arxiu constava de les següents columnes:

- **yearID** – Any al qual corresponen les dades
- **teamID** – Equip en el qual jugava el jugador
- **lgID** – Lliga en la qual juga l'equip
- **playerID** – Id del jugador
- **salary** – Sou anual en dòlars

S'ha eliminat la columna lgID, ja que aquesta informació ja serà inclosa mitjançant l'arxiu Teams.

La mida final del dataframe **df_salaries** és de (26.428, 4)

A continuació s'ha tractat el dataframe corresponent als **equips**. Aquest arxiu constava inicialment de 48 columnes en les quals s'especificaven: *l'any, la lliga, l'identificador de l'equip, l'identificador de la franquícia, l'identificador de la divisió (W, C, E), el rank del equip al final de temporada, el nombre de partits jugats, els partits jugats a casa, els partits guanyats, els partits perduts, si han estat els guanyadors de la divisió, si han estat els guanyadors de la Wild Card, si han estat els guanyadors de la lliga, si han estat els guanyadors de la World Series, les runs realitzades, nombre de jugadors al bat, hits, doubles, triples, homeruns, walks, strikeouts, bases robades, atrapats robant base, batejadors colpejats pel pitcher, sacrifice flies, runs dels oponents, earned runs allowed, earned runs average, jocs acabats, shoutouts, saves, outs pitched, hits dels oponents, homeruns dels oponents, walks dels oponents, strikeouts dels pitchers, errors, jugades dobles, Fielding percentatge, nom complet de equip, nom de l'estadi, assistència del públic, batting park factor, pitcher park factor, identificador de l'equip a baseball-reference.com, id del equip a la versió 4.5 de la base de dades de Lahman i id del equip a Retrosheet.*

En aquest cas s'han eliminat totes les dades corresponents als jugadors de camp i també les corresponents als pitchers, ja que aquí s'inclouen les dades totals de tots els pitchers. Les columnes que s'han mantingut han estat les següents:

- **yearID** – Any al qual corresponen les dades
- **lgID** – Lliga en la qual juga l'equip
- **teamID** – Identificador de l'equip
- **divID** – Divisió en la qual juga l'equip (W, C, E)
- **Rank** – Classificació final de l'equip en la lliga i divisió (numèric, de 1 a 5).
- **W** – Nombre de partits guanyats
- **L** – Nombre de partits perduts
- **Name** – Nom complet de l'equip.

També s'han eliminat les dades que no corresponen als anys entre 1985 i 2016 i s'han tornat a anomenar algunes columnes incloent la paraula team, perquè no es confongui amb les estadístiques dels jugadors. W -> team_W, L -> team_L i name -> team_name.

La mida final d'aquest dataframe ha estat (918, 8)

Pel que fa a l'arxiu referent als jugadors, consta de 24 columnes, entre les quals hi ha: *identificador de naixement, l'any de naixement, el mes de naixement, el dia de naixement, el país de naixement, l'estat de naixement, la ciutat de naixement, l'any de defunció, el mes de defunció, el dia de defunció, el país de defunció, l'estat de defunció, la ciutat de defunció, el nom, el cognom, el segon nom, el pes, l'altura, la mà amb la qual colpeja, la mà amb la qual llença la pilota, la data de debut, la data de l'últim partit, identificador per retrosheet i l'identificador per Baseball reference website.*

Pel cas de **df_people** solament he mantingut les columnes:

- **playerID** – l'identificador del jugador
- **birthCountry** – El país de naixement
- **nameFirst** – El nom del jugador
- **nameLast** – El cognom del jugador
- **weight** – El pes del jugador (en lliures)
- **height** – L'altura del jugador (en polsades)
- **bats** – la mà amb la qual colpeja
- **throws** – la mà amb la qual llença la pilota

A més a més, s'han creat dues columnes més en les que s'indica l'índex de massa corporal (**IMC_jugador**) i també el continent de naixement (**birth_continent**).

Finalment pel cas del dataframe corresponent a les estadístiques dels pitchers **df_pitching** que consta de 30 columnes, entre les quals hi ha *l'identificador del jugador, l'any, l'ordre d'aparicions de la temporada, ID del equip, els partits guanyats, els partits perduts, el nombre total de partits, el nombre de partits que ha començat com a pitcher, el nombre total de partits complets, el nombre de shoutouts, els saves, outs pitched, hits, earned runs, homeruns de l'altre equip, walks, strikeouts, la mitjana dels batejadors oponents, mitjana de earned run, intentional walks, wild pitches, batejadors colpejats pel pitcher amb la bola, balks, nombre de batejadors als que s'ha enfrontat, partits acabats, runs permeses, sacrificis by opposing batters, sacrifice flies by opposing batters, grounded into double plays by opposing batters.*

En aquest cas he decidit mantenir totes les columnes, ja que les estadístiques dels jugadors són les que més interessava mantenir. Tot i això hi ha algunes d'aquestes columnes que crec que tindran més rellevància que altres, aquestes són:

- **ERA** – Earned run average, és la xifra estadística més important que es té en compte amb els pitchers, com menor sigui aquesta xifra, millor és el pitcher.
- **El nombre de partits guanyats i perduts**
- **El nombre de partits començats** (és a dir si és o no el pitcher designat per començar el partit, anomenat starting pitcher)
- **El nombre de Walks**
- **El nombre de Strikeouts**
- **El nombre de Wild pitches**
- **El nombre de jugadors colpejats amb la pilota pel pitcher**
- **El nombre de runs permeses.**

Seguidament s'han agrupat aquests quatre dataframes utilitzant la funció `df.merge()` i finalment s'ha desat el dataset final amb el nom de **baseball_dataset**.

Aquest conjunt de dades final conté 11.953 files i 45 variables, entre les quals hi ha dades qualitatives i quantitatives. Els noms de les variables són les següents:

Nom columna	Significat
<i>playerID</i>	<i>Identificador del jugador</i>
<i>yearID</i>	<i>Any</i>
<i>stint</i>	<i>Ordre d'aparicions de la temporada</i>
<i>teamID</i>	<i>ID del equip</i>
<i>lgID</i>	<i>Lliga (AL o NL)</i>
<i>W</i>	<i>Nombre de partits guanyats</i>
<i>L</i>	<i>Nombre de partits perduts</i>
<i>G</i>	<i>Nombre de partits jugats</i>
<i>GS</i>	<i>Nombre de partits essent starting pitcher</i>
<i>CG</i>	<i>Nombre de partits acabats</i>
<i>SHO</i>	<i>Nombre de shoutouts</i>
<i>SV</i>	<i>Nombre de saves</i>
<i>IPouts</i>	<i>Nombre de outs pitched</i>
<i>H</i>	<i>Nombre de hits dels oponents</i>
<i>ER</i>	<i>Nombre de earned runs permesos pel pitcher</i>
<i>HR</i>	<i>Nombre de homeruns fets al pitcher</i>
<i>BB</i>	<i>Nombre de walks</i>
<i>SO</i>	<i>Nombre de strikeouts</i>
<i>BAOpp</i>	<i>Mitjana dels batejadors oponents</i>
<i>ERA</i>	<i>Mitjana de Earned Runs</i>
<i>IBB</i>	<i>Nombre de Walks intencionals</i>
<i>WP</i>	<i>Nombre de Wild Pitches</i>
<i>HBP</i>	<i>Nombre de batejadors colpejats amb la pilota a causa del pitcher</i>
<i>BK</i>	<i>Nombre de balks</i>
<i>BFP</i>	<i>Nombre de batejadors als que s'ha enfrontat</i>
<i>GF</i>	<i>Nombre de partits acabats pel pitcher</i>
<i>R</i>	<i>Nombre de runs permeses</i>
<i>SH</i>	<i>Nombre de sacrificis by opposing batters</i>
<i>SF</i>	<i>Nombre de sacrifice flies by opposing batters</i>
<i>GIDP</i>	<i>Nombre de vegades grounded into double plays by opposing batters.</i>
<i>birthCountry</i>	<i>País de naixement</i>
<i>nameFirst</i>	<i>Nom del jugador</i>
<i>nameLast</i>	<i>Cognom del jugador</i>
<i>weight</i>	<i>Pres (lliures)</i>
<i>height</i>	<i>Altura (polsades)</i>
<i>bats</i>	<i>Mà amb la qual colpeja la pilota</i>
<i>throws</i>	<i>Mà amb la qual llença la pilota</i>
<i>salary</i>	<i>Salari anual (dòlars)</i>
<i>divID</i>	<i>Divisió en la qual juga (W, C o E)</i>
<i>Rank</i>	<i>Classificació de l'equip al final de temporada (1-5)</i>
<i>team_W</i>	<i>Nombre de partits guanyats per l'equip</i>
<i>team_L</i>	<i>Nombre de partits perduts per l'equip</i>
<i>team_name</i>	<i>Nom de l'equip</i>
<i>birth_continent</i>	<i>Continent de naixement</i>
<i>IMC_jugador</i>	<i>IMC del jugador</i>

5. Originalitat del conjunt de dades

He escollit aquest conjunt de dades, ja que és un conjunt de dades molt extens que permet un gran nombre de visualitzacions diferents i es tracta d'una font fiable. Tot i que aquesta base de dades ha estat utilitzada anteriorment per realitzar molt bones visualitzacions com ara:

MLB Player performance:

<https://public.tableau.com/profile/brandon.fitzgerald#!/vizhome/MajorLeagueBaseballPlayerPerformances/Dashboard3>

MLB stats 1871-2018:

<https://public.tableau.com/profile/bart.eikelenboom#!/vizhome/MLBalltimebatterindex/Fielding>

I que s'han realitzat algunes visualitzacions sobre el sou dels jugadors:

30 years of MLB salaries:

<https://public.tableau.com/profile/gaunttchart#!/vizhome/30YearsofMLBSalaries/30YearsofMLBSalaries>

MLB salaries over the years:

<https://public.tableau.com/profile/gaurav.rohilla93#!/vizhome/MLB-SalaryAnalysis/MLB-Salaryovertheyears>

Salary vs. Performance:

<https://fathom.info/salaryper/>

No he trobat cap important visualització sobre el sou solament dels pitchers i com es veu afectat per les seves estadístiques.

6. Webgrafia

- [1] <https://creativecommons.org/licenses/by-sa/3.0/> visitat el 26/04/2021
- [2] <https://rdrr.io/cran/Lahman/man/Lahman-package.html> visitada el 27/04/2021
- [3] <https://cran.r-project.org/web/packages/Lahman/Lahman.pdf> visitada el 27/04/2021
- [4] <https://www.baseball-almanac.com/rulechng.shtml> visitada el 1/05/2021

Annex: Termes claus del beisbol

- **All-Star Game:** També anomenat “Midsummer Classic” és un partit de beisbol anual en el qual s’enfronten els millors equips de les dues lligues principals de beisbol l’American League (AL) i la National League (NL).
- **Balk:** Moviment del pitcher no permès segons les regles del beisbol.
- **Base Robada:** o stolen base. Una base és robada quan un jugador de l’equip ofensor avança a la següent base mentre la pilota està en moviment, però no ha estat colpejada pel batejador.
- **Batting Park Factor:** Indica la diferència entre les voltes dels jugadors al camp (o runs) fetes en l’estadi propi i les voltes fetes els estadis dels oponents.
- **Division:** Les dues lligues majors d’EUA, American League (AL) i National League (NL) divideixen els equips en tres divisions West, Central i East. En cada una d’aquestes divisions hi ha 5 equips.
- **Double:** Es realitza el que s’anomena double, o 2B, quan un jugador colpeja la pilota i arriba a segona base sense que l’eliminin.
- **Earned run:** Qualsevol *run* feta a l’equip defensiu sense que s’hagin produït *errors* de l’equip defensor. Altrament, una *unearned run* es tracta d’una *run* feta amb *errors* de l’equip defensor.
- **Eliminar un jugador:** Hi ha diferents maneres d’eliminar un jugador. Realitzar tres strikes quan el jugador ha de colpejar la pilota amb el bat, que els defensors agafin la pilota colpejada a l’aire, és a dir, sense que la pilota toqui el terra, que un defensor amb la pilota a la mà toqui l’atacant mentre és fora d’una de les tres bases o que els defensors es passin la pilota i arribi a la base abans que el jugador.
- **Error:** S’anomena error quan un jugador defensor realitza una mala maniobra que permet als jugadors atacants seguir avançant a la següent base. Per exemple quan se li escapa la pilota de les mans en agafar-la.
- **Grounded:** o **Ground out** és l’acció d’eliminar un batejador abans que arribi a primera base, sigui passant-se la pilota o agafant-la a l’aire.
- **Hit:** Quan un jugador colpeja la pilota i aconsegueix arribar a una base sense que l’eliminin.
- **Infielder:** Els jugadors defensius que juguen més propers al pitcher, a prop de les bases.

- **Inning:** És el terme utilitzat per definir la duració d'un partit. Un partit consta de 9 innings i cada inning està constituït pels dos equips tornant-se per colpejar la pilota i defensar.
- **Intentional Walks:** (IBB) Quan un pitcher fa que el jugador que bategi avanci a primera base sense que el jugador tingui l'oportunitat de colpejar la pilota. Això es pot realitzar tocant el jugador que ha de batejar amb la pilota o llençant quatre vegades la pilota fora de la zona de strike.
- **Lligues:** A la MLB (Major Leagues Baseball) hi ha dues lligues de beisbol diferents. La National League (NL) i l'American League (AL)
- **Outfielder:** Els jugadors defensius que juguen més allunyats del pitcher, a la part més externa del camp.
- **Out:** Quan s'elimina un jugador atacant.
- **Pitcher:** El jugador encarregat de llençar la pilota a la persona de l'altre equip que ha de colpejar la pilota amb el bat.
- **Pitcher Park Factor:** Un valor estadístic que indica la diferència entre el nombre de runs que ha realitzat l'equip contrari l'estadi propi i els que han realitzat en altres estadis.
- **Playoffs:** Seguit de partits disputats entre la lliga American League i la National League decidir el millor equip de la temporada.
- **Run:** Quan un jugador realitza una volta sencera al camp i se li atorga un punt a l'equip.
- **Sacrifice flies:** Es produeix un sacrifice fly quan encara no hi ha dos jugadors de l'equip atacant eliminat i un d'ells colpeja la pilota permetent que un altre jugador del seu equip acabi una volta (run) però el jugador que ha colpejat resulta eliminat.
- **Saves:** Els saves s'atorguen al pitcher que finalitza un partit quan aquest es guanya sota unes condicions concretes, per exemple, que al començament del novè inning, hi hagi una diferència de tres o menys punts amb l'altre equip.
- **Shoutout:** (SHO) A un pitcher se li atorga un shoutout quan aquest juga el partit sencer i l'equip contrari no marca cap punt, és a dir, no fa cap run.
- **Starting Pitcher:** El pitcher que comença el partit.
- **Strike:** Succeeix un strike quan un jugador intenta batejar la pilota i falla.
- **Strike zone:** Espai en què la pilota ha de passar per tal que, si el jugador no la colpeja, es denomini strike. Aquesta zona es determina com l'espai per sobre de la base i l'espai entre els genolls i el punt mitjà del tors del batejador.

- **Strikeout:** Quan el jugador que ha de batejar queda eliminat al realitzar tres strikes.
- **Triple:** o 3B, quan un jugador colpeja la pilota i arriba a la tercera base sense que l'eliminin.
- **Walk:** Quan un jugador avança a primera base sense haver colpejat la pilota.
- **Wild Card:** Es tracta del partit que inicia els playoffs.
- **Wild Pitch:** El fet que un pitcher llenci la pilota fora de la zona de strike.
- **World Series:** Partit en què s'enfronten els dos guanyadors de les dues lligues de beisbol pertanyents a MLB