Objectiu

Un mecanisme automàtic per obtenir un boxscore a partir d'un play-by-play. Això és bastant fàcil, llavors hauré de fer més coses com:

1. Generalitzar-ho perquè pugui agafar un play-by-play de qualsevol font (això ho hauré de fer quasi 100%)
2. Pensar noves variables per a posar al boxscore (ja en tinc pensada una)
3. Generar alguna visualització. Per exemple, si el play-by-play indica la distància del llançament, fer alguna gràfica on es pugui veure info sobre els llançaments dins/fora i la distància d'aquests
4. Fer una visualització on es pugui veure l'evolució d'un partit d'una forma dinàmica (un play-by-play més visual)
5. Poder fer el boxscore a partir d'una retransmissió com <https://www.marca.com/baloncesto/euroliga/td-systems-baskonia-real-madrid/directo/2020/10/02/5f77574fca4741d0638b4629.html> , però ho veig francament complicat

* Preparar-ho per a llegir minuts

PbP estàndard -> Boxscore

|  |  |
| --- | --- |
| **Aprendre a llegir** | **Per a** |
| Llançament de 1, 2 o 3 | Tir, punts, assistència |
| Rebot | Rebots |
| Pèrdua | Pèrdua |
| Robatori | Pèrdues, robatori |
| Tap | Tirs |
| Falta | Falta, pèrdua |
| Jugada d’equip: “Turnover by Team (shot clock)” | Qualsevol cosa (fila per a l’equip) |
| Canvi | +/-, minuts jugats |
| Altres | Guardar-ho en una llista per a poder veure si tot bé |

PbP -> PbP estàndard

|  |  |
| --- | --- |
| **Aprendre a llegir** | **Per a** |
| Llançament de 1, 2 o 3 | Tir, tap |
| Rebot | Rebot |
| Pèrdues | Pèrdua, robatori |
| Falta | Falta |
| Jugada d’equip: “Turnover by Team (shot clock)” | Qualsevol cosa (per a cap jugador) |
| Canvi | Canvi |
| Altres | Guardar-ho en una llista per a poder veure si tot bé |

A basketball reference, les implicacions ja estan indicades:

* Tir fallat de X (tap de Y)
* Pèrdua de X (robatori de Y)
* Pèrdua de X (tap de Y). Rebot ofensiu/defensiu de X/Y/Z/equip
* Falta en atac de X (rebuda per part de Y). Pèrdua de X (falta en atac)

Idea: Quan entra un jugador, es guarda el temps de joc en el qual entra i quan surt, es fa la resta. Pel que fa als inicials, si no hi ha temps inicial, es considera que han començat al minut 0.

Crec que la millor forma és usar un diccionari dels jugadors que hi ha al camp. Inicialment el diccionari està buit. Si se’n va un jugador que no és al diccionari, es dedueix que jugava des d’inici i s’afegeix al diccionari el jugador que ha entrat. La resta entre temps final i inicial i la suma entre aquest temps i el temps de joc provisional es fa:

Cal anar verificant si un jugador és a pista o no. Només tenir-ho en compte quan surt de pista seria possible, però si hi ha un jugador que juga un quart sencer no funciona.

Per a la construcció del PbP estàndard, millor guardar-me les jugades i fer una repassada printar-les:

* Basketball reference: 1a frase offensive foul. 2a frase turnover
* NBA:
  + 1a frase offensive foul. 2a frase turnover
  + 1a frase: tir fallat. 2a frase: tap
  + 1a frase: turnover. 2a frase: steal

Opcions rebots:

- Opció 1: indicar quin tipus de rebot és i confiar que està bé

- Opció 2: no indicar quin tipus de rebot és i esbrinar-ho segons el tir anterior

De fet, un possible error seria detectar si hi ha un rebot sense tir anterior, per tant s'acabaria mirant l'acció anterior i es podria fer 2 o 3

Error handling:

* Falta info
  + no hi ha el minut
  + no se sap l'equip d'una acció: possibilitat de saber-ho pel dorsal o per l'acció prèvia
  + no se sap l'autor d'alguna acció: afegir-ho al total i llavors total != suma
  + no se sap l'acció: eliminar-la
  + ni I ni O en un tir
* Info incorrecta:
  + El jugador no és a pista
  + El jugador no existeix
  + Tipus de rebot no correcte: en cas d'indicar si O o D (opció 1) mirar el tir anterior de qui ha estat
  + Rebot sense tir previ
* Jugada no interpretada pel programa

Implementació de boxscore parcial, del moment X al moment Y. És senzill, però hi ha el problema d'endevinar qui hi ha a pista

Solució 1: anar tirant enrere per veure qui ha anat entrant. Si no es poden obtenir 5 jugadors, algun és des d'inici.

Solució 2: Fer-ho al revés, es parteix del 5 inicial i es va canviant. Crec que és més fàcil

Variables:

* Crear diferenciació de tirs de 2 segons direcció cos respecte de la cistella:
  + Penetració
  + Tir vertical o fadeaway
* Crear diferenciació d'assistències segons el tir posterior
* Tir posterior a assistència o no
* +/-: <https://rompimientodefensivo.com/2016/02/15/guia-de-estadisticas-avanzadas-nba/>
* Punts/possessió, rebots /oportunitat de rebot (tirs errats)
* EFF (temporada)
* PER (temporada)

Game Score (simplificació per a un sol partit)

Classificació de jugades:

* Bag of words
* Transformer
* Clustering:
  + Classificació en k-means
  + DBSCAN (Density-Based Spatial Clustering of Applications with Noise) va millor que k-means perquè no usa distàncies, sinó densitat de punts. Sol anar millor que k-means en embeddings de text
  + HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise)
  + OPTICS (Ordering points to identify the clustering structure)

Reconeixement automàtic de noms de persones:

* <https://www.listendata.com/2018/05/named-entity-recognition-using-python.html>
* <https://stackoverflow.com/questions/20290870/improving-the-extraction-of-human-names-with-nltk>

Visualitzacions extres:

* Estadístiques a basketball reference
* Veure si els temps morts han aconseguit algun efecte