A close up of a sign

Description automatically generated15 de septiembre del 2020

Investigación de Operaciones II

Ricardo González

Ceci Cojulún

**Premier League**

**Simulación**

Daniela Domínguez

20180365e

Índice

[Data sets 3](#_Toc51106569)

[Flujograma 4](#_Toc51106570)

[Archivos utilizados 6](#_Toc51106571)

[Resultados 6](#_Toc51106572)

# Data sets

La data utilizada fue obtenida de “data.world”, un sitio online donde las personas pueden compartir y encontrar datasets fácilmente.

Se descargaron los datos de la Premier League desde el 2013 hasta el año pasado. Esto fueron proporcionado por Chas Peacock y por Andrés Hincapié. Estos datasets contenían información muy vasta, incluyendo detalles como las tarjetas rojas, amarillas, tiros de esquina, fabols, el nombre del arbitro de cada partido, entre muchos otros, pero no todos fueron utilizados para este análisis.

Para este trabajo las variables que se utilizaron fueron: hometeam (nombre del equipo que jugaba en casa), awayteam (nombre del equipo que jugaba de visitante), fthg (cantidad total de goles del equipo en casa a lo largo de todo el partido), ftag (cantidad total de goles del equipo de visita a lo largo de todo el partido), hst (home shots on target), ast (away shots on target).

Llama la atención que no se utilizó hs (home shots) ni as (away shots), sino que hst y ast. Esta decisión se temó con base al artículo “Analysis: Total Shots vs Shots on Target” publicado en SBNation el 11 de enero del 2013. En este se analizaba la correlación entre los tiros de los equipos o los tiros a la porterilla con la cantidad de goles que metía cada equipo. El articulo demuestra con datos históricos que existe una mayor correlacion entre hst y la cantidad de goles, que hs y los goles. Es por eso que se decidió usar las variables hst y ast.

# Flujograma

La estructura de trabajo se organizo originalmente en una servilleta.

A continuación, se muestra una foto, seguido de otro flujograma mas organizado.

A close up of text on a white background

Description automatically generated

Screen of a cell phone

Description automatically generated

La forma de hacer la simulación es siempre comparando dos equipos en un partido. Para el equipo 1, se busca si el juego se está dando en su propio estado, es decir si está en *Home* (H), o si está *Away* (A). Dependiendo de esto tenemos las diferentes probabilidades que el equipo tiene de ganar (G), perder (P), o quedar empate (E). Para esto se usan probabilidades discretas. Estas probabilidades de G, P o E, van cambiando para cada equipo dependiendo de contra qué equipo se esté jugando, y también, como ya se mencionó antes, depende de si está H o A.

Cuando ya se determinó si el equipo 1 G, P o E, el equipo tiene una probabilidad de meter cierta cantidad de tiros hacia la portería. Estos tiros no son lo mismo que goles. Estos tiros siguen una distribución normal, por lo que para sacarlos se tiene una media y una desviación estándar. Nuevamente, estos dependen de H, A, G, P, E y el equipo contra quienes estén jugando.

Cuando ya se tienen los tiros que hará el equipo, se calcula la cantidad de tiros acertados, es decir cantidad de goles. Para calcular estos se sacó la media de la razón de goles/tiros, y la desviación estándar. El número resultante es un número con decimales, por lo que se redondea al entero más cercano. La cantidad de goles depende de todo lo anterior mencionado: H, A, G, P, E, tiros, y el equipo contra el que estén jugando.

Cuando ya se tienen todos estos datos para el equipo 1, se hace el mismo proceso para obtener los mismos datos para el equipo 2.

Cuando ya se tiene la cantidad de goles de ambos equipos, se compara y se dan los puntos respectivos. El equipo que gana recibe 3 puntos, el equipo que pierde recibe 0 punto, y en el caso que queden empate, ambos reciben 1 punto.

Estos puntos se van

# Archivos utilizados

Para la simulación se utilizó una mezcla entre Excel y R.

Por primera vez puse en práctica lo que nos enseñaban en varias clases de la universidad, que en la vida real el hacer análisis de algo, el 80% del tiempo se va únicamente en limpiar datos, ordenarlos y ponerlos de una forma que sean útiles antes de poder programar o hacer cualquier otra cosa. Sinceramente pensaba que esta era una gran exageración, pero no me daba cuenta de que siempre me daban datos ya *digeridos* y relativamente fáciles de usar. Con este trabajo específicamente me atrevería a decir que mas del 90% del tiempo me la pasé arreglando datos y el resto del tiempo programando.

Entre los problemas que me encontré al realizar este trabajo fue que en mis datos históricos, había veces que un equipo siempre ganaba frente a otro en su casa, por lo que no tenia probabilidades y medias o desviaciones para los demás casos (que perdiera o empatara).

Otra cuestión que surgió fue que de tres equipos no tenía datos históricos, por lo que se tomaron promedios para los datos que hacían falta.

También a medio camino de trabajar, me di cuenta que los tiros y goles eran decimales, por lo que se decidió redondear únicamente los goles hacia el entero mas cercano. También con este tema había veces que la cantidad de goles salía negativo, por lo que en esos casos la cantidad se reemplazó con cero.

# Resultados

A screenshot of a cell phone

Description automatically generated