PRA 1

Contexto

Los datos han sido extraídos de la página oficial de la FIFA (Federación Internacional de Fútbol Asociado). Tratándose de información sobre el mundial de futbol celebrado en Rusia en 2018 he considerado que la información suministrada por los propios organizadores del evento serian los más veraces.

Título

"World Cup Russia 2018 stats and reviews"

Descripción

El conjunto de datos contiene información sobre los partidos disputados en la copa del mundo de futbol celebrada en Rusia en 2018. Cada partido es descrito mediante un conjunto de variables cuantitativas como pueden ser por ejemplo, número de goles, faltas, tiros a puerta, etc. (stats) y por un conjunto de variables que podrían considerarse como cualitativas y que consisten en las reseñas hechas por los periodistas encargados de cubrir el evento (reviews).

Representación gráfica

Contenido

Cada registro se corresponde con un partido. Y para cada uno de ellos se han recogido los siguientes datos:

- **id**: Identificador del partido
- **name_home**: Nombre del equipo local
- **name_away**: Nombre del equipo visitante
- **code_home**: Código equipo local
- **code_away**: Código equipo visitante
- **group**: Ronda clasificatoria
- **stadium**: Estadio
- **venue**: Localización
- **datetime**: Fecha y hora
- **headline**: Titular resumen
- **summary**: Reseña neutral
- **summary_home**: Reseña del periodista local
- **summary_away**: Reseña del periodista visitante
- **goals_home**: Goles equipo local
- **goals_away**: Goles equipo visitante
- **attempts_home**: Lanzamientos equipo local
- **attempts_away**: Lanzamientos equipo visitante
- **on-target_home**: Lanzamientos entre los tres palos equipo local
- **on-target_away**: Lanzamientos entre los tres palos equipo visitante
- **off-target_home**: Lanzamientos fuera equipo local
- **off-target_away**: Lanzamientos fuera equipo visitante
- **blocked_home**: Paradas equipo local
- **blocked_away**: Paradas equipo visitante
- **woodwork_home**: Palos equipo local
- **woodwork_away**: Palos equipo visitante

Darío M. García Carretero

- **corners_home**: Saques de esquina equipo local
- **corners_away**: Saques de esquina equipo visitante
- **offsides_home**: Fueras de juego equipo local
- **offsides_away**: Fueras de juego equipo visitante
- **ball_possession_home**: Porcentaje de posesión equipo local
- **ball_possession_away**: Porcentaje de posesión equipo visitante
- **pass_accuracy_home**: Precisión pasa equipo local
- **pass_accuracy_away**: Precisión equipo visitante
- **passes_home**: Pases equipo local
- **passes_away**: Pases equipo visitante
- **passes_completed_home**: Pases completados equipo local
- **passes_completed_away**: Pases completado equipo visitante
- **distance_covered_home**: Distancia recorrida equipo local
- **distance_covered_away**: Distancia recorrida equipo visitante
- **balls_recovered_home**: Recuperaciones equipo local
- **balls_recovered_away**: Recuperaciones equipo visitante
- **tackles_home**: Entradas equipo local
- **tackles_away**: Entradas equipo visitante
- **blocks_home**: Cortes equipo local
- **blocks_away**: Cortes equipo visitante
- **clearances_home**: Despejes equipo Local
- **clearances_away**: Despejes equipo visitante
- **yellow_cards_home**: Tarjetas amarillas equipo local
- **yellow_cards_away**: Tarjetas amarillas equipo visitante
- **direct_red_cards_home**: Tarjetas rojas directas equipo local
- **direct_red_cards_away**: Tarjetas rojas directas equipo visitante
- **indirect_red_cards_home**: Tarjetas rojas indirectas equipo local
- **indirect_red_cards_away**: Tarjetas rojas indirectas equipo visitante
- **fouls_committed_home**: Faltas equipo local
- **fouls_committed_away**: Faltas equipo visitante
- **referee_name**: Nombre del árbitro
- **referee_conrtry**: País del árbitro
- **weather_description**: Descripción del tiempo
- **weather_temperature**: Temperatura (ºC)
- **weather_windspeed**: Velocidad del viento (km/h)
- **weather_humidity**: Humedad (%)

Agradecimientos

La propietaria de los datos es la FIFA. La Fédération Internationale de Football Association, más conocida por sus siglas **FIFA**, es la institución que gobierna las federaciones de fútbol en todo el planeta. Se fundó el 21 de mayo de 1904 y tiene su sede en Zúrich, Suiza. Forma parte del IFAB, organismo encargado de modificar las reglas del juego. Además, la FIFA organiza la Copa Mundial de Fútbol, los otros campeonatos del mundo en sus distintas categorías, ramas y variaciones de la disciplina, y los Torneos Olímpicos a la par del COI.

Me gustaría destacar la gran cantidad de información que la FIFA pone a disposición de los aficionados a través de su web y el orden y claridad con la que es presentada.

Inspiración

El conjunto de datos podría ser utilizado en diferentes ámbitos. Algunos de ellos podrían ser:

- **Periodismo:** Los datos podrían ser usados para, por ejemplo, realizar un reportaje sobre el mundial que sacara a relucir los datos más curiosos/destacados.
- **Almacén de datos:** Los datos podrían ser cruzados con datos pertenecientes a otras fuentes para enriquecer un almacén de datos estadísticos.
- **Minería de datos:** En cuanto a su uso para proyectos de minería de datos podría ser interesante estudiar:
    - **Sesgos en información:** Podrían estudiarse la diferencias o similitudes entre las reseñas de cada uno de los periodistas (local/visitante) y la reseña neutral para extraer conclusiones sobre sesgos en el periodismo.
    - **Generador de reseñas:** Utilizando en conjunto las variables cuantitativas y las reseñas se podría intentar construir un sistema que generase reseñas de manera automática. Aunque seguramente se necesitarían más datos, el conjunto podría ser un punto de partida para estudiar la viabilidad del

proyecto.

- **Estilos de juego:** Podrían estudiarse los diferentes estilos de juego de las selecciones (con técnicas de clustering) y estudiar cuales son los más efectivos.

Licencia

La licencia escogida ha sido **CC0: Public Domain License**. La razón por la que he escogido esta licencia, es que todo el trabajo ha sido realizado con el único objetivo de superar la asignatura "Tipología y ciclo de vida de los datos" por lo que si de alguna forma, alguien quiere usar los datos para cualquier finalidad, esta licencia le permitirá el uso de estos con el menor número de restricciones posibles.

Código

El código fuente se puede encontrar dentro de la carpeta "src".

Dataset

El dataset resultante se puede encontrar en la carpeta "csv".

Recursos

- Subirats, L., Calvo, M. (2019). Web Scraping. Editorial UOC.
- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.