

## **Dataset: Datos de carrera de 5k de McMenamins Edgefield, Troutdale, OR el 2 de febrero de 2019**

Diego Armando Cale Pillco

11 de noviembre del 2019

### **Descripción**

El conjunto de datos generado como parte de esta actividad práctica reúne diferentes características de las carreras de 5k de McMenamins Edgefield, Troutdale, obtenidos el 2 de febrero de 2019 de la página <https://www.hubertiming.com/> sitio web que permite visualizar el cortometraje de las carreras desde 2009 y actualmente cronometra más de 60 carreras al año. Presenta un amplio conocimiento técnico, una comprensión profunda de la gestión de eventos y las operaciones de carrera, y un alto nivel de atención a los detalles. Algunas de las variables que se recogen en el conjunto de datos son el sitio, babero, nombre, genero, años, ciudad, edad, estado, tiempo de chip, ritmo de la viruta, lugar de género, grupo de edad, lugar del grupo de edad, tiempo de empezar, tiempo arma.

### **Imagen identificativa**



Figura 1: figura del evento 2019 El peor día del año 5K McMenamins Edgefield, Troutdale, OR 2 de febrero de 2019 y de los creadores del sitio web.

### **Contexto**

Como se ha comentado, la materia del conjunto de datos se corresponde a los datos de carrera 5k El conjunto de datos que tuvo lugar en McMenamins Edgefield, Troutdale, OR el 2 de febrero de 2019. Del siguiente sitio web: <https://www.hubertiming.com/results/2019WorstDay>

Específicamente, analizará el rendimiento de los corredores de 5 km y responderá preguntas como:

- ¿Cuál fue el tiempo promedio de finalización para los corredores?

- ¿Los tiempos de finalización de los corredores siguieron una distribución normal?
- ¿Hubo alguna diferencia de rendimiento entre hombres y mujeres de varios grupos de edad?

## Contenido

Para cada corredor, el cual se corresponde con un registro en el conjunto de datos, se recogen las siguientes características:

- Place.- Lugar de llegada de corredores
- Bib.- Identificación del corredor
- Name.- Nombre de los corredores
- Gender.- Género de los corredores
- Age.- Años de los corredores
- City.- Ciudad
- State.- Estado
- Chip Time.- Tiempo de chip
- Chip Pace.- Ritmo de la viruta
- Gender Place.- Lugar del género
- Age Group.- grupo por edad
- Age Group Place.- Lugar del grupo por edad
- Time to Start.- tiempo de inicio del recorrido
- Gun Time.- tiempo de disparo

Los autores de la web hubertiming llevan recopilando información sobre permite visualizar el cortometraje de las carreras desde 2009 y actualmente cronometra más de 60 carreras al año. Presenta un amplio conocimiento técnico, una comprensión profunda de la gestión de eventos y las operaciones de carrera, y un alto nivel de atención a los detalles. (Sitio web <https://www.hubertiming.com/>)

## Agradecimientos

Los datos han sido recolectados desde la base de datos online hubertiming. Para ello, se ha hecho uso del lenguaje de programación Python y de técnicas de Web Scraping para extraer la información alojada en las páginas HTML.

## Inspiración

El presente conjunto de datos podría utilizarse en ámbitos muy diversos. Uno de ellos podría ser conocer el ranking deportivo de los corredores, predicción del ritmo de carrera en una maratón, puede ayudarnos a monitorizar el progreso en el entrenamiento, aumentar

la confianza en las competiciones, proporcionarnos una estrategia de carrera muy valiosa e, incluso, prevenir lesiones.

También podría ser de gran utilidad en el campo de la minería de datos, a la hora de elaborar modelos predictivos (como por ejemplo árboles de decisión o redes neuronales). Así, se podría querer elaborar un modelo que permita predecir el ganador de las carreras 5k.


## Licencia

- La licencia escogida para la publicación de este conjunto de datos ha sido CC BY-SA 4.0 License. Los motivos que han llevado a la elección de esta licencia tienen que ver con la idoneidad de las cláusulas que esta presenta en relación con el trabajo realizado:
- Se debe proveer el nombre del creador del conjunto de datos generado, indicando los cambios que se han realizado. De esta manera, se reconoce el trabajo ajeno y en qué medida se han realizado aportaciones en relación con el trabajo original.
- Se permite un uso comercial. Esto haría que incrementen las probabilidades de que una empresa utilice los datos generados y realicen trabajos de calidad que reporten cierto reconocimiento al autor original.
- Las contribuciones realizadas a posteriori sobre el trabajo publicado bajo esta licencia deberán distribuirse bajo la misma. Esto hace que el trabajo del autor original continúe distribuyéndose bajo los términos que él mismo planteó.

## Código fuente y dataset

Tanto el código fuente escrito para la extracción de datos como el dataset generado pueden ser accedidos a través de este enlace.

### Contribuciones

Contribuciones	Firma
Diego Armando Cale Pillco	

## Recursos

1. Lawson, R. (2015). Raspado web con Python . Packt Publishing Ltd.
2. Mitchell, R. (2018). Web Scraping con Python: Recopilando más datos de la web moderna . "O'Reilly Media, Inc."
3. Thomas, D. M., & Mathur, S. (2019, June). Data Analysis by Web Scraping using Python. In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 450-454). IEEE.
4. Vanden Broucke, S. y Baesens, B. (2018). Raspado web práctico para la ciencia de datos: mejores prácticas y ejemplos con Python . Apress.
5. Bisong, E. (2019). Google Colaboratorio. En Building Machine Learning y Deep Learning Models en Google Cloud Platform