

Jennifer Paola Durango Monsalve

1128402240

Análisis de datos

Actividad 1

CODERHOUSE

Introducción

Actualmente la industria del fitness ha tomado fuerza de forma constante y vigorosa, dado que factores como el sedentarismo o el trabajo en casa, entre otros, han creado la cultura del autocuidado. Dentro de los elementos importantes de la industria del fitness podemos mencionar la ejecución de actividades físicas, el deporte de alto rendimiento y muy importante la calidad de la alimentación y/o nutrición. Dado el aumento de la demanda de los servicios y productos de la industria, se ha hecho necesario para los usuarios, ya sean entrenadores o asesorados/deportistas/atletas acceder a la información de manera ordenada y fácil para alcanzar sus objetivos.

Una de las herramientas más usadas para obtener y/o transferir la información son las APPs , las cuales permiten interacción coach/asesorado, y además permiten tener registros y conocer datos sobre el proceso.

El fondo de una APP implica tener un base de datos o un conjunto de datos, por lo cual se hace necesario integrar datos obtenidos de forma adecuada, para esto se propone en esta actividad tres DATASETS concernientes al fitness, dos de ellos son datos concernientes al entrenamiento físico y dos concernientes a la nutrición y suplementación deportiva.

De forma tentativa, uno de los objetivos del curso ANÁLISIS DE DATOS de CODERHOUSE, es realizar como proyecto final la base para la construcción de una APP que permita como coach/usuario tener un conjunto de datos asociados al entrenamiento físico y a la nutrición dentro del contexto del mismo.

Metodología

La fuente a través de la cual se obtuvieron los DATASETS es kaggle.com, con su motor de búsqueda las palabras claves fueron FITNESS DATASETS y NUTRITION DATASETS. Se obtuvieron los siguientes DATASETS con nombres literales: *megaGymDataset*, *cardioActivities*, *Sports Supplements*, *nutrition*. Todos se visualizaron a través de la función DATOS de excel, con los parámetros origen de archivo: 1252:Europeo occidental (windows), Delimitador: coma, Detección del tipo de datos: Basado en las primeras 200 filas.

megaGymDataset/ DATASET MEGAGYM

Para este DATASET se tenía inicialmente nueva atributos o columnas y 2916 tuplas o filas, dado que no contenía la fecha se le insertó de forma arbitraria un campo de fecha, quedando así un DATASET de diez columnas y 2916 filas. Se considera un DATASET muy completo y adecuado a la hora de consultar un ejercicio, dado que entre otros contiene el nombre del ejercicio, la forma de ejecutarlo, el músculo o grupo muscular trabajado, el tipo de habilidad trabajada, los equipos utilizados y el nivel del usuario. Esta información puede dar bases adecuadas para realizar una *base de datos relacional*, con algunos de dichos atributos como tablas.

cardioActivities

Para este DATASET se tienen 14 columnas o atributos y 508 tuplas o filas. Entre otros, la información que provee dicho DATASET son el tipo de actividad, los promedios entre distancias recorridas, los promedios de velocidad, las calorías gastadas, lo cual permite posteriormente que el usuario conozca cuánta actividad y qué tipo debe realizar para lograr sus objetivos. El DATASET no fue modificado, se presenta tal cual se descargó. Se presume que posteriormente se requiere de un uso y análisis muy inteligente de los datos para poder obtener una información adecuada.

Sports Supplements

Para este DATASET se tiene 46 columnas o atributos y 96 tuplas o filas. Si bien la cantidad de tuplas es muy limitada, se tiene un DATASET muy completo dado que contiene una gran cantidad de información sobre muchos suplementos deportivos y/o moléculas químicas que realizan funciones a nivel bioquímico. Se tiene una redundancia de atributos, dado que se referencian muchas fuentes de consultas y se comparan estudios sobre la eficacia de dichos suplementos, se presume que posteriormente se realizará una limpieza de datos para evitar la redundancia o el exceso de datos innecesarios.

nutrition

Para este DATASET se tiene 77 columnas o atributos y 8788 tuplas o filas. Es un DATASET bastante grande, con una gran cantidad de datos, que permite conocer el tipo de alimentos, su composición química y/o bioquímica, las cantidades de este, entre otros; infiriendo que es una buena información a la hora de construir tablas nutricionales y relaciones con las necesidades nutricionales del asesorado/deportista/atleta.

A continuación el link de la entrada Github con el presente pdf y los DATASETS referenciados en el trabajo <https://github.com/jpdurangom/FITNESS-DATASETS>

