

Proyecto I
Semillero Análisis de Datos (AD)
Semestre 2024-1

Trabajo con un archivo del tipo json (Java Script Object notation).

Parte 1.

Cargar los modulos Numpy y Pandas. Verificar que el archivo del tipo json no puede ser cargado con read_json. Leemos la primera linea del archivo. Leemos todas sus categorías.

Parte 2

Separamos las zonas horarias que están en la característica 'tz' en una Series de Pandas. Miramos la categorización de esta característica. Los valores sin categorizar los categorizamos como unknown. Construir un barplot con Seaborn para ver las primeras cinco componentes de 'tz'.

Parte 3.

Consideremos la característica 'a'. Esta característica relacionada con el browser o dispositivo con el cual se hizo la ubicación. Extraigamos la informacion que hay en 'a' para dos valores: [1], [50].

Parte 4.

Creación de una variable 'os' (sistema operativo) para separar los usuarios que emplean y no emplean Windows.

Parte 5.

Averiguar las zonas horarias con mayor numero de miembros.

Parte 6.

Realizar la normalización de las variables por zona horaria.