Se agregaron las siguientes columnas:

Promedio

Popularidad

Predictor Flag

Fecha de inicio y Fecha de fin

Se retiraron las siguientes columnas:

Usuarios recurrentes: porque tenía muchos ceros al comienzo

La última fila que tenía decimales

Se detectó que la columna “Tráfico de móviles” y “Trafico móvil y de Tablet” Son Iguales, a retirar una de las 2.

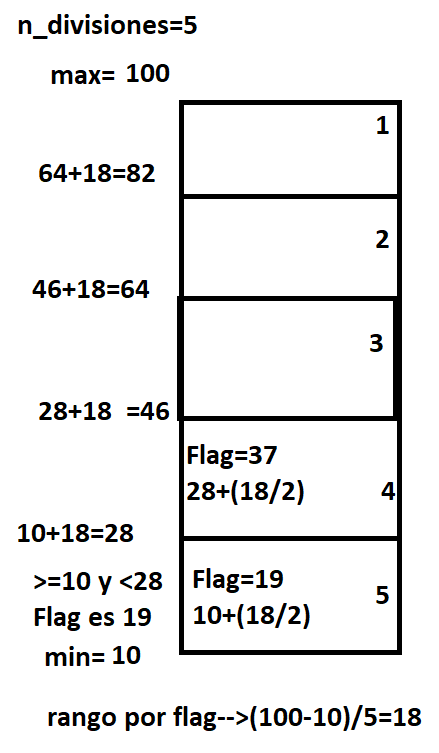
# Explicación de cálculo cada columna:

#promedio por fila

promedio\_Fila=dataset.mean(axis=1)

# agrego como una columna extra

dataset['promedio']=promedio\_Fila

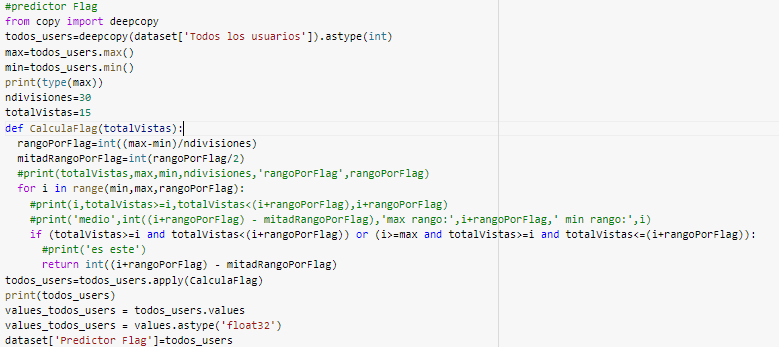
Predictor Flag:

Elijo el total de divisiones 30

Entre el rango del mínimo y el máximo de la columna “Todos los usuarios”

De ahí el valor medio para cada Flag a predecir.

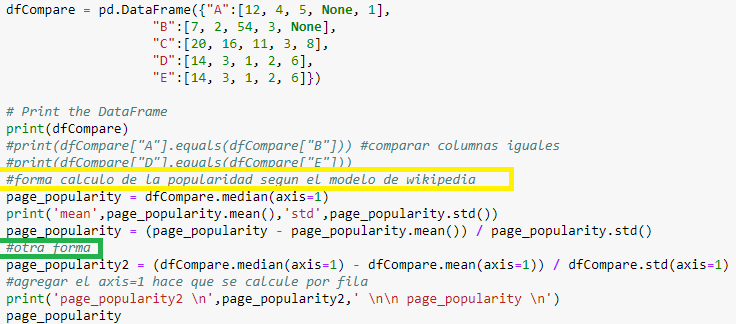
Elegí 30 divisiones Nose si agregar más o menos.

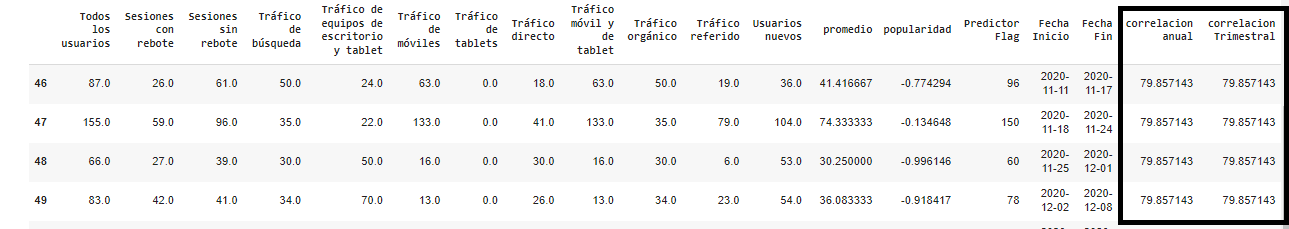
Luego creo una columna donde por cada valor de “Todos los usuarios” veo a que Flag pertenece.

Columna Popularidad:

El Cálculo es: (mediana-media)/desvió estándar

La duda es la forma de cómo hacer el cálculo



Columnas de Correlación Anual y Trimestral (quaterly)

Utilizando las columnas de Fechas de inicio/fin

Calculo un valor promedio cada 3 meses para el trimestral para los que están dentro del mismo trimestre

Calculo un valor promedio Anual para los que están dentro del mismo año.

Retiro las columnas de fechas al final.