Preparación de Datos.

Tras la investigación de los datos, hemos encontrado que hay ciertos DataSets (**df_distance_km**, **df_distance_min** y **df_historic_order_demand**) que requerían de ciertas imputaciones de datos.

Para **'df_distance_km'**, vemos que hay distancias de 0 entre diferentes clientes cuando la distancia es mayor. Para solventar este problema, hemos decidido hacer uso de la **fórmula de la distancia euclidiana**, que obtiene una distancia entre 2 coordenadas (lat, long) aproximada en kilómetros.

Para **'df_distancia_min'**, tenemos tiempos de 0, dandose el mismo "fallo" que en el 'df_distance_km', por lo que hemos imputado los nuevos datos gracias a las nuevas distancias corregidas. Para calcular los tiempos necesitamos la velocidad, como no tenemos una velocidad exacta y puede variar dependiendo de los tramos, usaremos la **fórmula de la velocidad implícita** para sacar una "velocidad media". Con las distancias en km divididas entre la velocidad implícita, podremos obtener una distancia en minutos aproximada, ya que la velocidad es media y puede variar y los datos de las distancias de km también son aproximadas.

Para **'df_historic_order_demand'**, hay ciertos valores nulos en diferentes datos de la columna **'order_demand'**, por lo que hemos imputado los valores medios de ese cliente en el mes indicado.