Datos importantes que creo que afectan al valor total\_minutes (antes de realizar algún análisis de parámetros con alguna herramienta).

1. Distancia entre tienda y cliente (calcular usando Lat y Lng)
2. [quantity y buy\_unit] Cantidad de productos en KG (contar cuantas instancias iguales hay de order\_id en la tabla order\_product)
3. [quantity y buy\_unit] Cantidad de productos en unidades (contar cuantas instancias iguales hay de order\_id en la tabla order\_product)
4. Seniority del Shopper asignado
5. Picking\_speed del Shopper asignado
6. store\_id (esto puede afectar si hay tiendas que usualmente son lentas o congestionadas)
7. Estimacion de Gente en el Local (EGL). Una función E(dia, hora) = numero entre 0 y 1, indicando cuan lleno deben estar los supermercados. 0 indica nada de tiempo extra en colas. 1 indica infinito tiempo en colas.
8. Estimacion de Gente en el transporte (EGT). Una funciona ET(dia, hora, medio de transporte) = numero entre 0 y 1.
9. Opcional – numero de carros necesarios para realizar la compra. 1 no afecta. 2 afecta. Y mas de dos afecta muchísimo. (función de quantity y volumen de cada producto)

Cornershop trabaja actualmente en 6 ciudades completando miles de órdenes por día. Un elemento fundamental para determinar cuántas órdenes podemos aceptar, cuando debemos comenzar a trabajar en una orden y que horario le podemos prometer a los clientes es la **estimación de tiempo**.

En este caso te pedimos que entrenes el mejor estimador posible utilizando técnicas de Machine Learning y los datos disponibles en la carpeta *data* (cada tabla en un archivo formato csv). Puedes utilizar el lenguaje y librería que más te acomode. *Nosotros trabajamos con Python, Scikit Learn y Keras, pero si prefieres trabajar en R u otra librería, no tenemos problemas*.

A continuación el detalle de la información que te facilitaremos:

***Tabla Orders***:

* order\_id: Identificador de la orden.
* lat y lng: Latitud y Longitud de la dirección de entrega.
* dow: Día de la semana en que se prometió la orden. 0: Domingo 6: Sabado.
* promised\_time: Hora de entrega prometida al cliente.
* actual\_time: Hora en que se entregó el pedido al cliente.
* on\_demand: Si el pedido fue solicitado en “Menos de 90 minutos” o en una ventana horario en el futuro.
* picker\_id: Identificador del shopper que hizo de picker en el pedido.
* driver\_id: Identificador del shopper que hizo de driver en el pedido.
* store\_branch\_id: Identificador del local en que se realizó la compra.
* total\_minutes: Tiempo total que demoró completar la orden (desde la compra hasta la entrega).

***Tabla Order\_Product***:

* order\_id: Identificador de la orden.
* product\_id: Identificador del producto.
* quantity: Cantidad solicitada del producto por el cliente.
* quantity\_found: Cantidad encontrada del producto por el shopper.
* buy\_unit: Formato en que se vendió el producto (Unidades o KG).

***Tabla Shoppers***:

* shopper\_id: Identificador del Shopper.
* seniority: Clasificación de experiencia del Shopper.
* found\_rate: Porcentaje histórico de productos que el shopper encontró.
* picking\_speed: Velocidad histórica de recogida de productos.
* accepted\_rate: Porcentaje histórico de pedidos aceptado por el shopper.
* rating: Evaluación histórica del shopper.

***Tabla Storebranch***:

* store\_branch\_id: Identificador del local.
* store: Identificador de la tienda a la cual pertenece el local.
* lat y lng: Latitud y Longitud de la ubicación del local.

Todos los IDs y nombres han sido encriptados para mantener la anonimidad. Además, esto corresponde a un pequeño sampling de la operación de Cornershop.

El objetivo de este caso es que realices una estimación de cuánto tiempo tomará completar los pedidos que tienen el campo **total\_minutes** vacío. También te solicitaremos el script con el que entrenaste el modelo y un pequeño informe (**en 1 plana**) explicando cómo enfrentaste este problema en términos de:

* Generación de tu dataset.
* Transformaciones que hayas realizados a los datos.
* En qué entorno entrenaste los modelos.
* Que modelos utilizaste y con qué métricas evaluar el desempeño de tu modelo.

Puedes escribir un correo a [matias.sanchez@cornershopapp.com](mailto:matias.sanchez@cornershopapp.com) para cualquier consulta adicional.