1. CONJUNTO DE DATOS

   Viajes en taxi dentro de una ciudad, donde se proporciona información del lugar donde se tomó el taxi, el lugar donde terminó el viaje, entre otras características que se enlistan a continuación. El objetivo es utilizar el dataset para generar las tablas listas para entrenamiento y prueba, con el objetivo de predecir el monto de la tarifa (fare_amount).

key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	passenger_count	fare_class
2012-11-11 13:45:00.00000029	12.5	2012-11-11 13:45:00 UTC	-73.956322	40.813427	-73.959143	40.783220	1	low_fare
2010-01-26 00:13:00.00000050	11.3	2010-01-26 00:13:00 UTC	-73.982577	40.746018	-73.980713	40.780807	1	low_fare
2014-10-07 19:24:00.000000235	19.5	2014-10-07 19:24:00 UTC	-73.972000	40.759470	-74.006190	40.708460	2	low_fare
2012-10-20 13:48:40.0000002	10.5	2012-10-20 13:48:40 UTC	-74.002701	40.728209	-74.013599	40.710990	1	low_fare
2014-04-29 20:27:00.000000186	7.0	2014-04-29 20:27:00 UTC	-73.982517	40.770782	-73.979932	40.754880	1	low_fare
...	...	...	...	...	...	...	...	...
2010-07-20 09:21:00.00000052	14.5	2010-07-20 09:21:00 UTC	-73.979120	40.746182	-74.004187	40.705937	1	low_fare
2013-02-20 15:16:27.0000004	3.5	2013-02-20 15:16:27 UTC	-73.955352	40.804620	-73.955352	40.804620	1	low_fare
2010-10-07 22:32:00.000000223	7.3	2010-10-07 22:32:00 UTC	0.000000	0.000000	0.000000	0.000000	2	low_fare
2010-09-13 13:28:00.000000138	4.1	2010-09-13 13:28:00 UTC	-73.973527	40.784913	-74.001877	40.740657	6	low_fare
2014-09-03 20:35:37.0000004	8.0	2014-09-03 20:35:37 UTC	-73.996540	40.747920	-73.979322	40.762122	1	low_fare

Variables:

- key : Variable que debe indicar el id del viaje, sin embargo no tiene información correcta , por lo cual no hace caso de dicha variable.
- fare_amount: Monto de la tarifa asociado a cada viaje, la tarifa incluye gastos de peaje.
- pickup_datetime : Fecha y hora en la que se comenzó el viaje
- pickup_longitude : Longitud donde comenzó el recorrido
- pickup_latitude : Latitud donde comenzó el recorrido
- dropoff_longitude : Longitud donde concluyó el recorrido
- dropoff_latitude : Latitud donde concluyó el recorrido
- passenger_count : Número de pasajeros durante el trayecto
- fare_class : Tipo de tafira que se cobró, una tarifa baja o una tarifa alta.

Objetivo : Predecir el monto de la tarifa (incluidos los peajes) para un viaje en taxi dadas las características. La variable "fare_class" se debe omitir en este problema ya que está completamente ligada al monto de la tarifa. Los puntos a cubrir en el desarrollo son los siguientes:

- Calidad de datos
- Análisis Exploratorio (cada hallazgo acompañado de una pequeña descripción)
- Outliers - Si considera que es viable, en caso contrario justificar su respuesta
- Missings - Si se considera que es viable, en caso contrario justificar su respuesta
- Cambio de la unidad muestral de la tabla - Si considera que es viable
- Ingeniería de variables
- Reducción de dimensiones/Selección de características - Si se considera que algún método es viable, en caso contrario justificar su respuesta

La salida final debe ser el conjunto de entrenamiento y prueba con las variables viables a ocupar para un modelo, adicional a la variable objetivo (X_train,X_test)

ENTREGABLE:

- Código en python (Notebook) , limpio, ordenado , comentado y bien estructurado, sin errores en el código. En el notebook deben mostrarse todas las gráficas que contiene el PDF además de los entrenamientos.

Consideraciones

- Todos los recursos necesarios para probar el procespo deben añadirse, si hace falta alguno se tomara como un examen incompleto.
- El examen debe resolverse completamente en Python
- La entrega debe encontrarse en perfecto orden y de forma entendible
- El código no debe contener ningún error