- YEAR, MONTH, DAY, DAY_OF_WEEK: dates of the flight
- AIRLINE: An identification number assigned by US DOT to identify a unique airline
- ORIGIN_AIRPORT and DESTINATION_AIRPORT: code attributed by IATA to identify the airports
- SCHEDULED_DEPARTURE and SCHEDULED_ARRIVAL : scheduled times of take-off and landing
- DEPARTURE_TIME and ARRIVAL_TIME: real times at which take-off and landing took place
- DEPARTURE_DELAY and ARRIVAL_DELAY: difference (in minutes) between planned and real times
- DISTANCE: distance (in miles)

Preguntas sobre el modelo:

• ¿Cuál es el problema de regresión que están tratando de resolver?

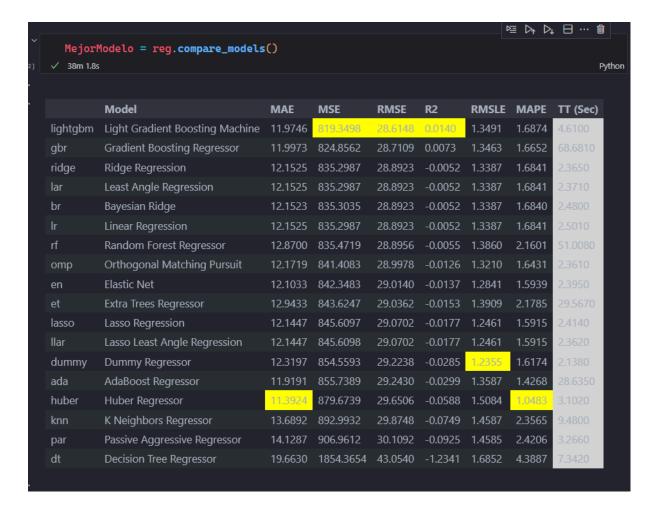
El objetivo es usar el mejor modelo de regresión para predecir el tiempo de retraso en la salida de los vuelos

• ¿Qué columna de los datos de retrasos de aerolíneas especificaron como la columna objetivo?

La columna de datos que es la columna objetivo es "DepDelay"

• ¿Qué modelo de regresión seleccionaron como el mejor y por qué?

Es el mejor debido a que da los mejores resultados en las pruebas de MSE RMSE y R2



• ¿Cómo evaluaron el rendimiento del modelo en el conjunto de pruebas? ¿Cuáles fueron las métricas de rendimiento que utilizaron para evaluar el modelo? ¿Cuáles fueron los resultados?

Primero modificamos el rendimiento del modelo en base a los datos obtenidos, tras esto introducimos las columnas que usaremos para trabajar los datos que son las siguientes:

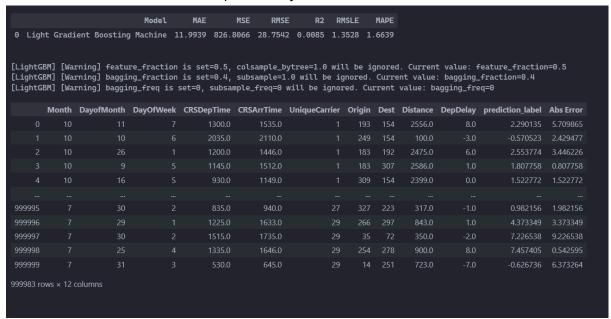
Una vez hecho esto vemos los resultados obtenidos usando un dataframe

```
Model
MAE
MSE
RMSE
R2
RMSLE
MAPE

0
Light Gradient Boosting Machine
11.9939
826.8066
28.7542
0.0085
1.3528
1.6639
```

• ¿Cómo visualizaron los resultados del modelo? ¿Qué información pueden extraer de la visualización?

Usamos nuestro modelo tuneado para trabajar el dataframe de los datos.



Lo visualizamos a través de una tabla de la que podemos ver los resultados obtenidos a través de las columnas del final

Distance	DepDelay	prediction_label	Abs Error
2556.0	8.0	2.290135	5.709865
100.0	-3.0	-0.570523	2.429477
2475.0	6.0	2.553774	3.446226
2586.0	1.0	1.807758	0.807758
2399.0	0.0	1.522772	1.522772
317.0	-1.0	0.982156	1.982156
843.0	1.0	4.373349	3.373349
350.0	-2.0	7.226538	9.226538
900.0	8.0	7.457405	0.542595
723.0	-7.0	-0.626736	6.373264

• ¿Cuáles fueron los hiper parámetros que utilizaron para mejorar el rendimiento del modelo?

Los hiper parámetros usados son los de la columna "DepDelay" menos la columna "prediction label"

• ¿Cuáles fueron las características más importantes del modelo?

Las características más importantes es que gracias a este modelo, tanto los clientes como las aerolíneas pueden ganar tiempo y dinero debido a que pueden saber si un vuelo tiene probabilidades de que se retrase y por tanto arreglar el retraso. Para ello, se debe conseguir predecir correctamente los datos que se le piden, para ello debemos de trabajar antes correctamente los datos y los tipos de datos.

Por ello, también se debe de tener en cuenta el error absoluto del modelo, para seguir trabajando con él para mejorarlo lo más posible.

• ¿Cómo desplegaron el modelo y cómo podría ser utilizado en una aplicación real?

Este modelo puede ser usado en una aplicación real para las propias aerolíneas puesto que esto puede ayudar tanto a los clientes como a las aerolíneas a la correcta organización de los vuelos y de los pasajeros.