



Objetivo:

Predecir el precio de un diamante según una serie de variables numéricas y categóricas.

He utilizado H2O para realizar la predicción.

Variables categóricas en H2O

En el momento de ajustar un modelo (*GLM, GBM, DRF, Deep Learning, K-Means, Aggregator, XGBoost*), **H2O** identifica automáticamente que variables son categóricas (y crea internamente las variables *dummy* correspondientes.)

Con esta herramienta no he conseguido ver qué pasa, es como una caja negra donde ocurren cosas dentro. En otras palabras, es como ir en modo “piloto automático”.



Resultados:

MEJOR MODELO: Gradient Boosting Machine (GBM)

```
ModelMetricsRegression: gbm  
** Reported on test data. **
```

MSE: 281443.28112484125

RMSE: 530.5122817851075

MAE: 275.73892584168584

RMSLE: 0.10101297032833796

Mean Residual Deviance: 281443.28112484125

Este resultado nos ha hecho quedar en la tabla de la siguiente manera:

