



Instituto Tecnológico
de Buenos Aires

COMPETENCIA KAGGLE

Análisis Predictivo - 1Q 2022 - ITBA

Camila Collado

0,68463

SCORE OBTENIDO

4

Camila Collado

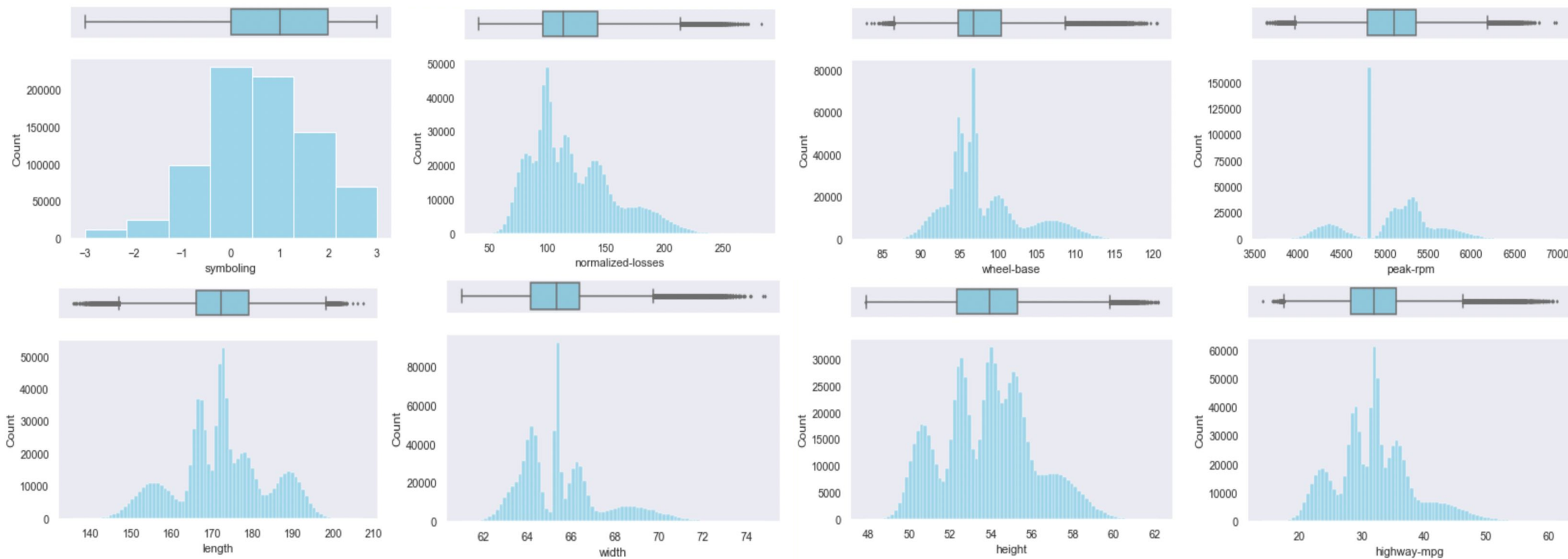


0.68463

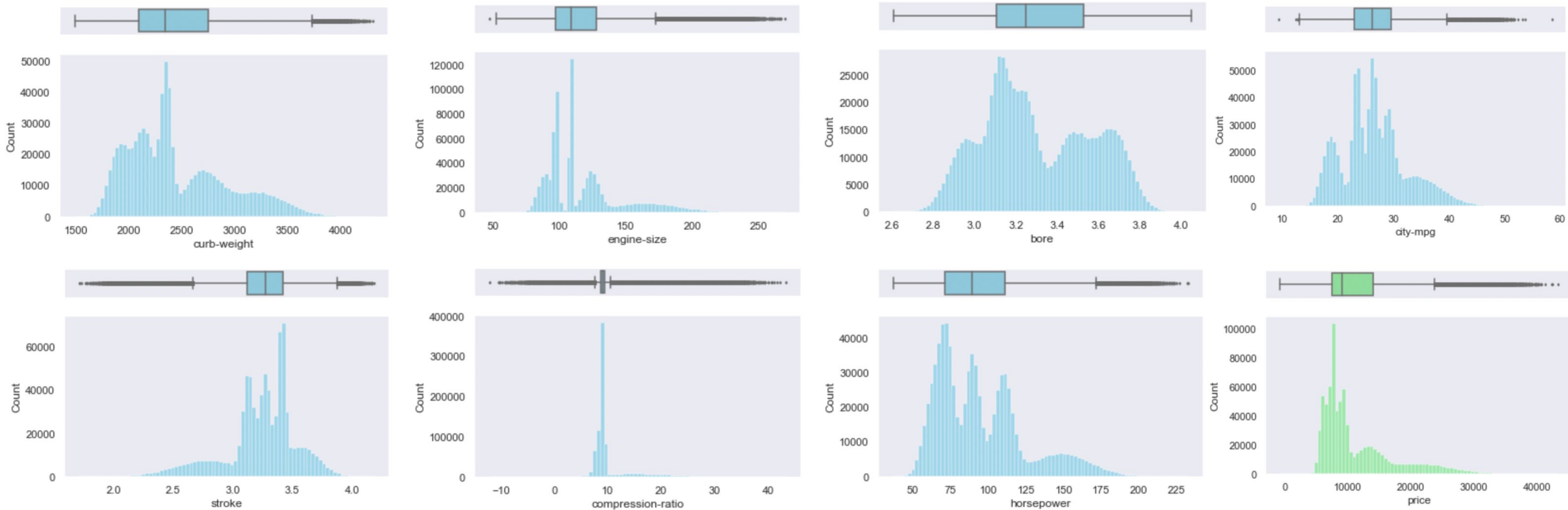
10

2h

ANÁLISIS EXPLORATORIO DE DATOS



ANÁLISIS EXPLORATORIO DE DATOS



Variable target

MODELO UTILIZADO

HISTOGRAM GRADIENT BOOSTING

- ✓ IMPLEMENTACIÓN DE CATBOOST PARA SKLEARN
- ✓ USO DE HISTOGRAMAS EN LUGAR DEL DATO PUNTUAL
- ✓ MEJOR SCORE
- ✓ EJECUCIÓN RÁPIDA
- ✓ AJUSTE DE HIPER PARÁMETROS

OTROS MODELOS

- 2. GRADIENT BOOSTING 67,6%
- 3. RANDOM FOREST 66,5%
- 4. ÁRBOL SIMPLE 59,1%
- 5. EXTRA TREES 53,9%
- 6. KNN 48,5%
- 7. MODELOS LINEALES 48,1%

SEPARACIÓN TRAIN Y TEST

Tipo de predicción
Regresión

Variable target
Price



Train 80% **Test 20%**

n = 640.000

n = 160.000

AJUSTE DE HIPER PARÁMETROS: GRID SEARCH

MAX_DEPTH

15



profundidad
máxima del árbol
default: none

MAX_ITER

192



máxima cantidad de
árboles
default: 100

MIN_SAMPLES_LEAF

200



mínimo de muestras por
hoja
default: 20

MAX_LEAF_NODES

201



máximo de hojas
por árbol
default: 31

Otros parámetros que quedaron como default

LEARNING_RATE = 0.1 LOSS = SQUARED ERROR

Score con parámetros
default: **0,65599**

GRACIAS!