

# UNIVERSITYHACK 2019<sup>®</sup> DATAATHON





## **Miguel Amezua Villegas**

- Graduado en Física
- Máster en Ciencia de Datos



## **Javier Aznar Palop**

- Graduado en Física
- Máster en Ciencia de Datos



## **Lucía Carrasco Ribelles**

- Graduada en Ingeniería Biomédica
- Máster en Ciencia de Datos



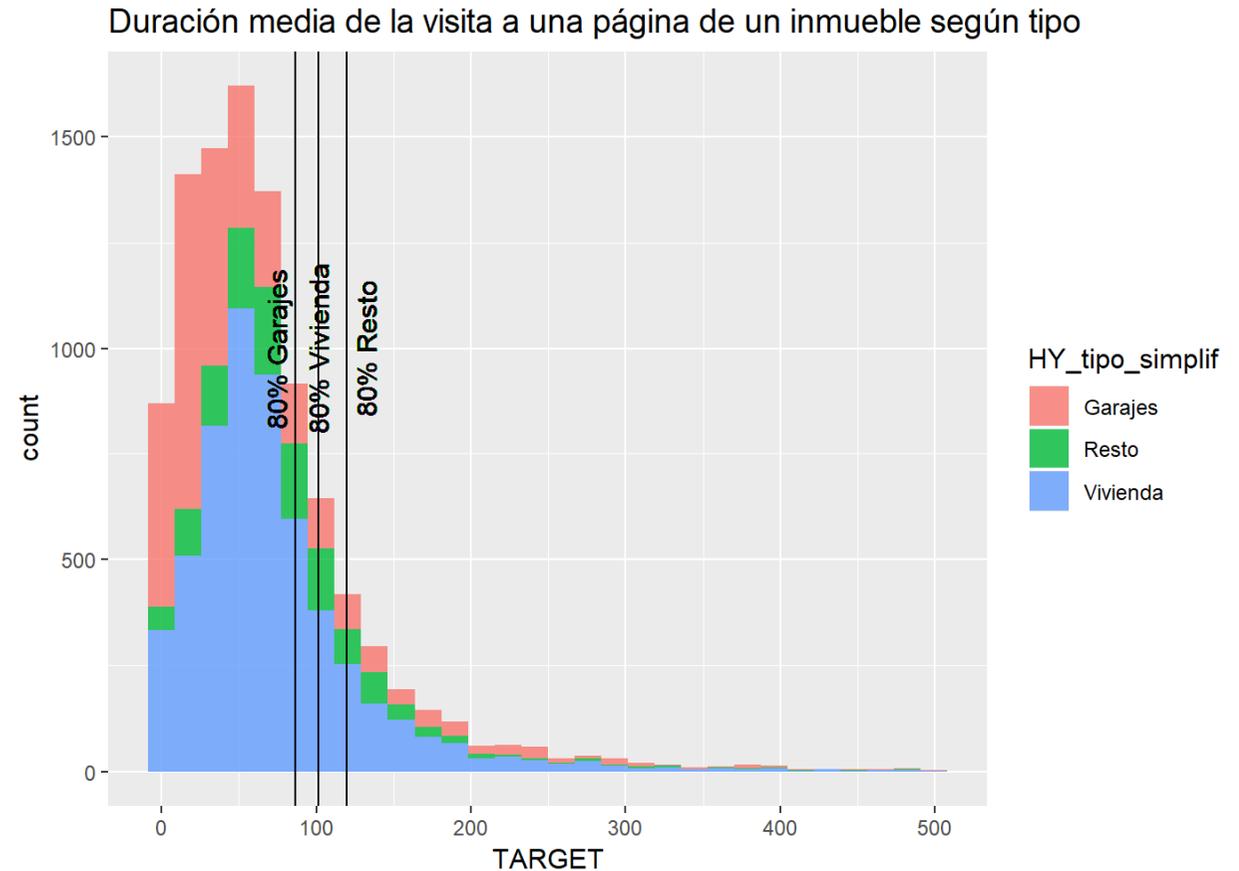
# Mongo (UV) / Herramientas



Cubist

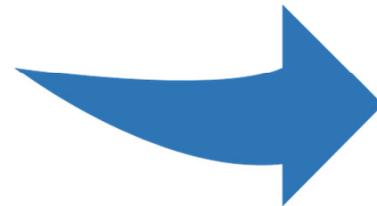


- Distribución de la variable objetivo
- Imputación de valores de IDEA
- Estudio de valores extremos
- Creación de variables nuevas



### Modelos probados con CV

- K-NearestNeighbors
- NeuronalNetwork
- RandomForest
- bartMachine
- XGBoost
- Cubist
- SVM

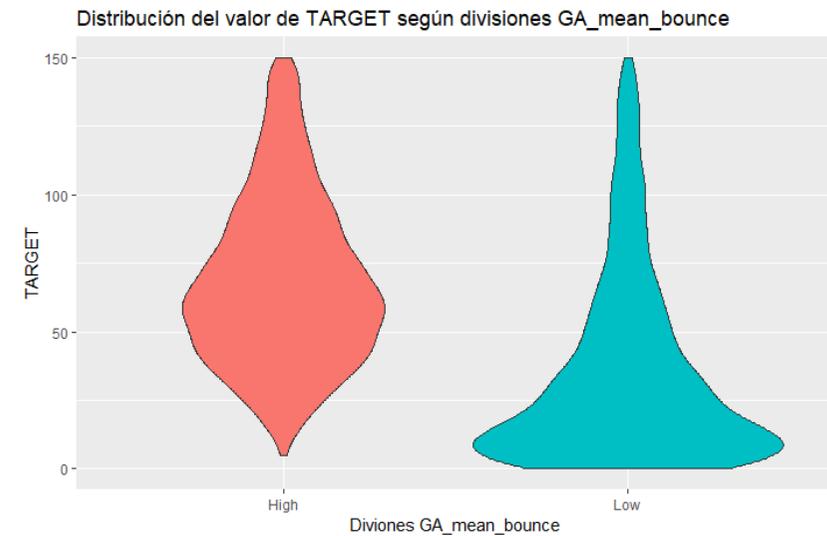
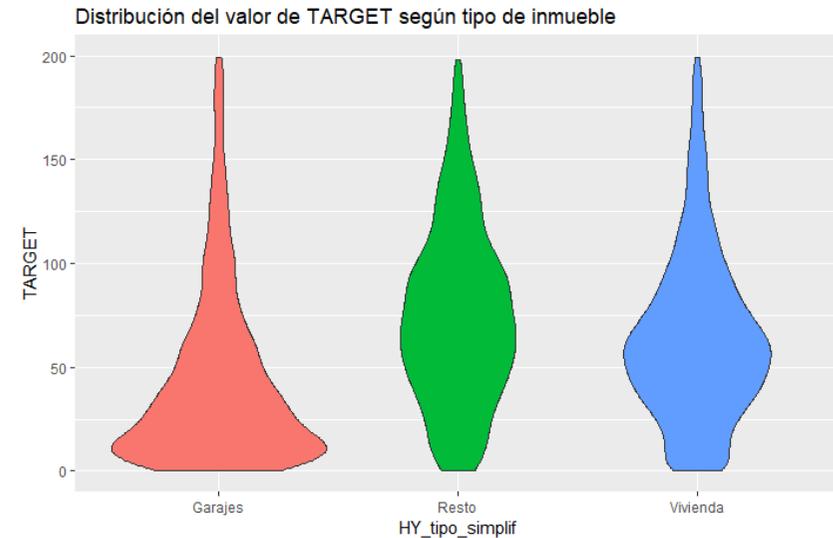


### Mejores modelos

SVM  
Cubist

- **Modelo Cubist**
  - *Árboles de decision + modelos lineales*
  - *Valores faltantes y variables categóricas*
  - *Extrapolation, committees, rules, unbiased*
- Cross Validation: 5 folds
- Filtrado del conjunto de train:
  - Observaciones con valores extremos de TARGET por tipo
  - Observaciones sin visitas o sin visualización real
- 28 variables
- **Métrica final: 15.076**

- Modelos según los valores de las **características**
  - Tipos de inmueble
  - GA\_mean\_bounce
- Modelos **ensemble**





La competición de analítica de datos  
más grande de España.

Muchas gracias.