# Versiones

# General

* Selección variables
* Usar algoritmos y tunearlos
* Comparalos
* Ensemble

## Que voy a necesitar

* Selección de variables
* Metricas

Metric -> no quedarme con la acuracy solo, como de bien predice cada categoría

ROC -> como predice por debajo de la curva

* Puede que se haga mejor con tenicas de desmbalaceo

## V01

# Fecha\_ 16/05/2019

#Objetivos: importacion de datos y exploracion inicial

* Pocos missings
* No hay variables lineales, salvo tiempo
* Categoria deny
  + # no yes
  + # 2096 285
  + #Desbalance 7.33

## [Binary Classification under Sample Selection Bias](https://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262170055.001.0001/upso-9780262170055-chapter-3)

* Pocos valores en categorías: Mirar la frecuencia de las variables amphiso, subset
* ~~Según en análisis no tiene variables discretas~~
* Que es un QQplot

# Selección de variables

Las varaibles que elija estarán en todos los modelos ¿

Como elijo las variables ? modelos tentativos, compararlos ,elegir el mejor, tunearlo

Todo se basa en tener un conjunto de modelo, desde modelo robusto pero menos sesgo hasta modelo con menos sesgo y más varianza

Aproximacion lineal

Aproximacion no lineal