# Churn Norris – ML para mantener a tus clientes

PASS SQLSATURDAY MADRID | 28 SEP 2019

#SQLSatMadrid


Pablo Corral

Pau Sempere

pcorral@solidq.com

psempere@solidq.com

# BIG Thanks to SQLSatMadrid sponsors

Platinum

plain concepts

KABEL

aura

channeladvisor

Gold

SolidQ

encamina
PIENSA EN COLORES

東京
TOKIOTA

alesonITC
data thinking

zaltor
Software Distribution

Suricatta

Silver

devscope

SentryOne

Venue

Universidad Politécnica de Madrid
ETSI SISTEMAS INFORMÁTICOS

Global

Microsoft

PASS
SQLSATURDAY
ORGANIZER

# About us

## Pablo Corral

- Data Scientist @ SolidQ

- Email: pcorral@solidq.com
- Twitter: @pcorral23
- LinkedIn: https://www.linkedin.com/in/pablo-corral-llorca

## Pau Sempere

- Working with SQL Server since 2009
- Microsoft Machine Learning MCSA

- Mentor @ SolidQ

- Email: psempere@solidq.com
- Twitter: @pausempere
- LinkedIn: https://www.linkedin.com/in/pausempere/

# Agenda

- What is Churn?
- In balance there is virtue
- KISS
- How did we get here?

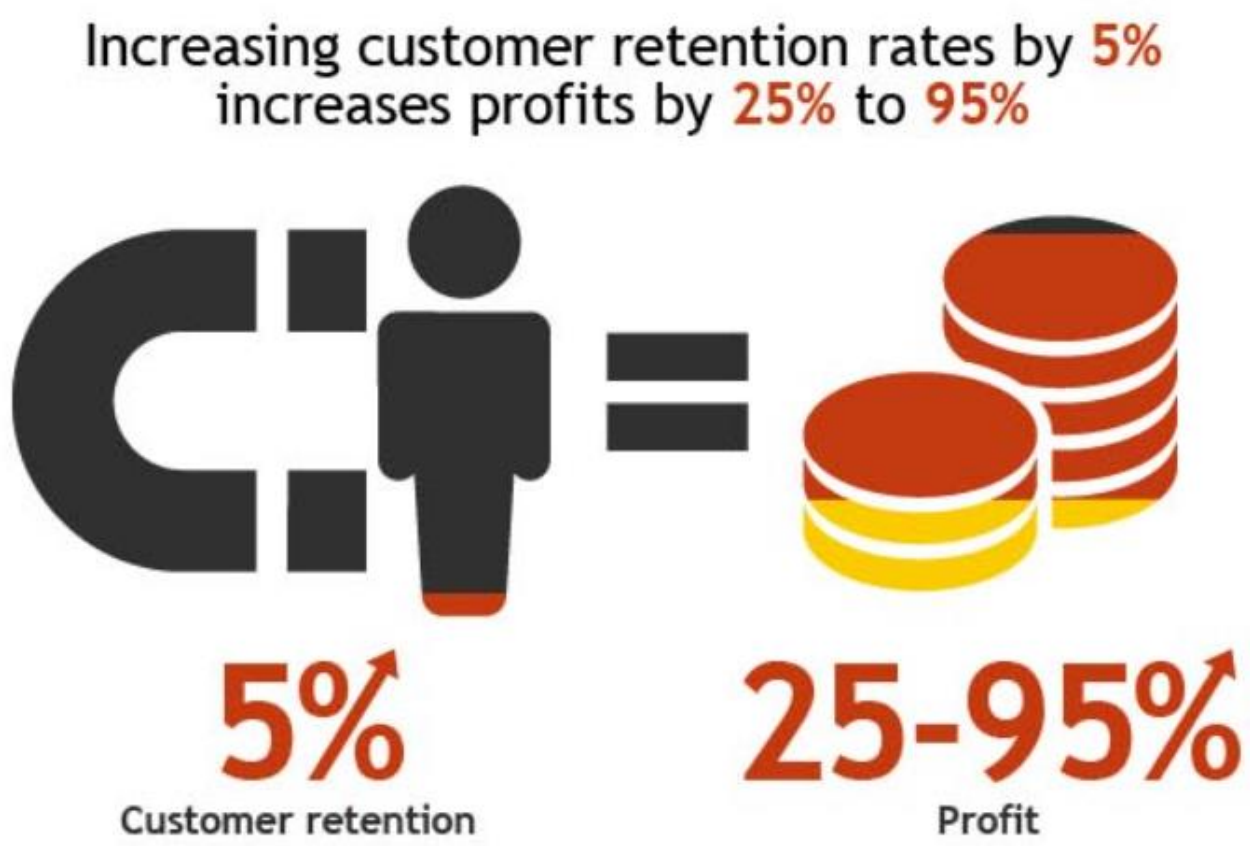
# What is Churn?

# What is Churn?

# What is Churn?

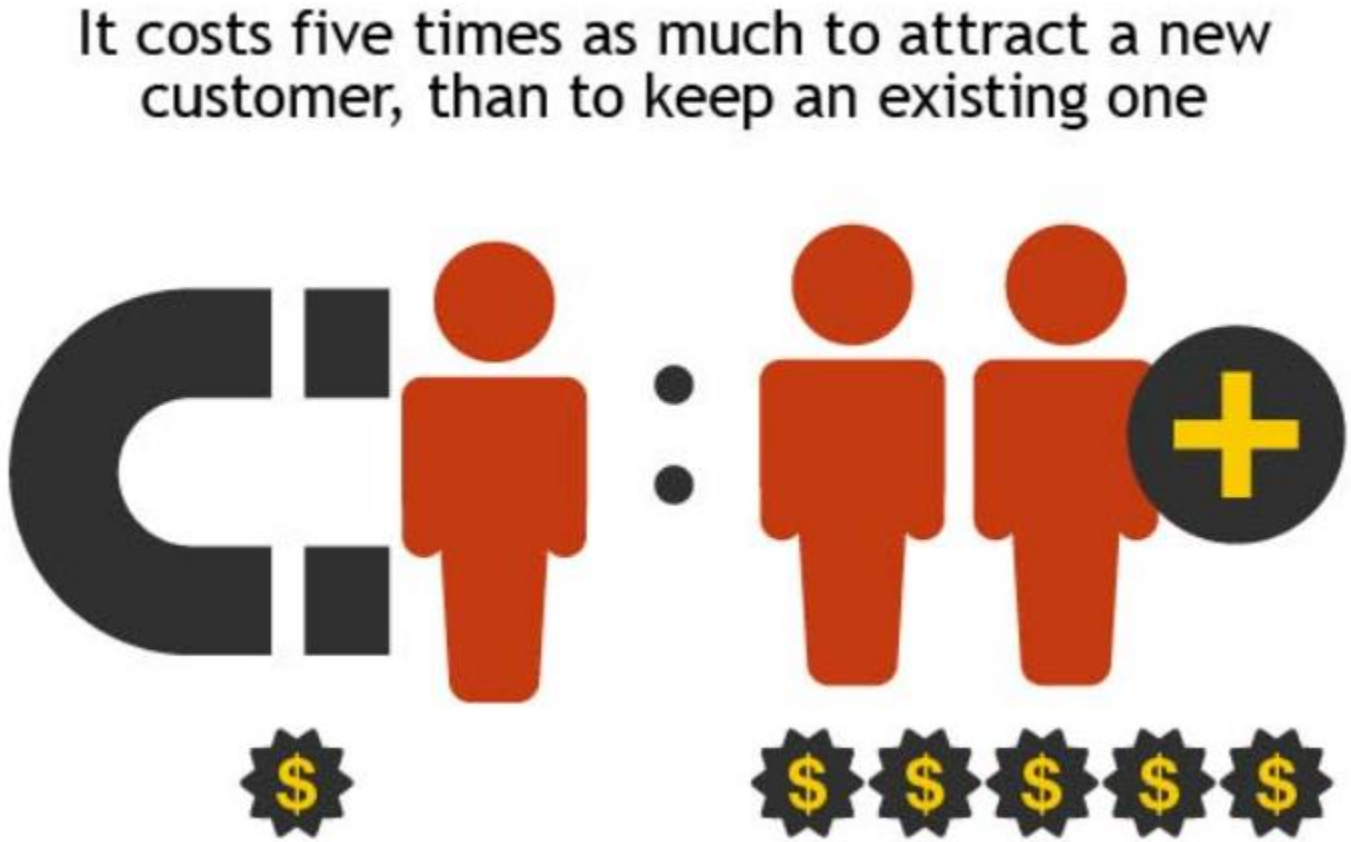
# What is Churn?

# What is Churn?




https://www.invespcro.com/blog/customer-acquisition-retention/

# What is Churn?




https://www.invespcro.com/blog/customer-acquisition-retention/

# What is Churn?

# What is Churn?

# Key concepts about Churn

CLEAR DEFINITION

HISTORICAL DATA

CONTEXT

# DEMO

## Basic Churn

# Looking for a better model

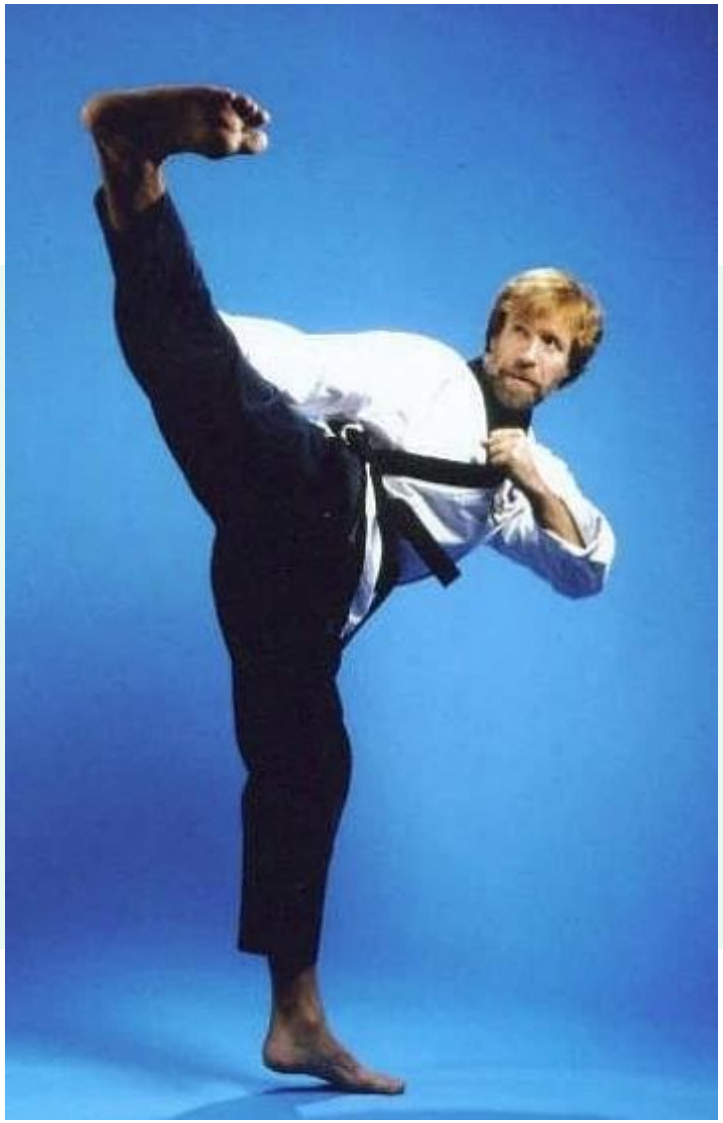
Random forest has:

- N_estimators
- Max_Depth
- Max_leaf_nodes
- …

We need to find a better model exploring such combinations

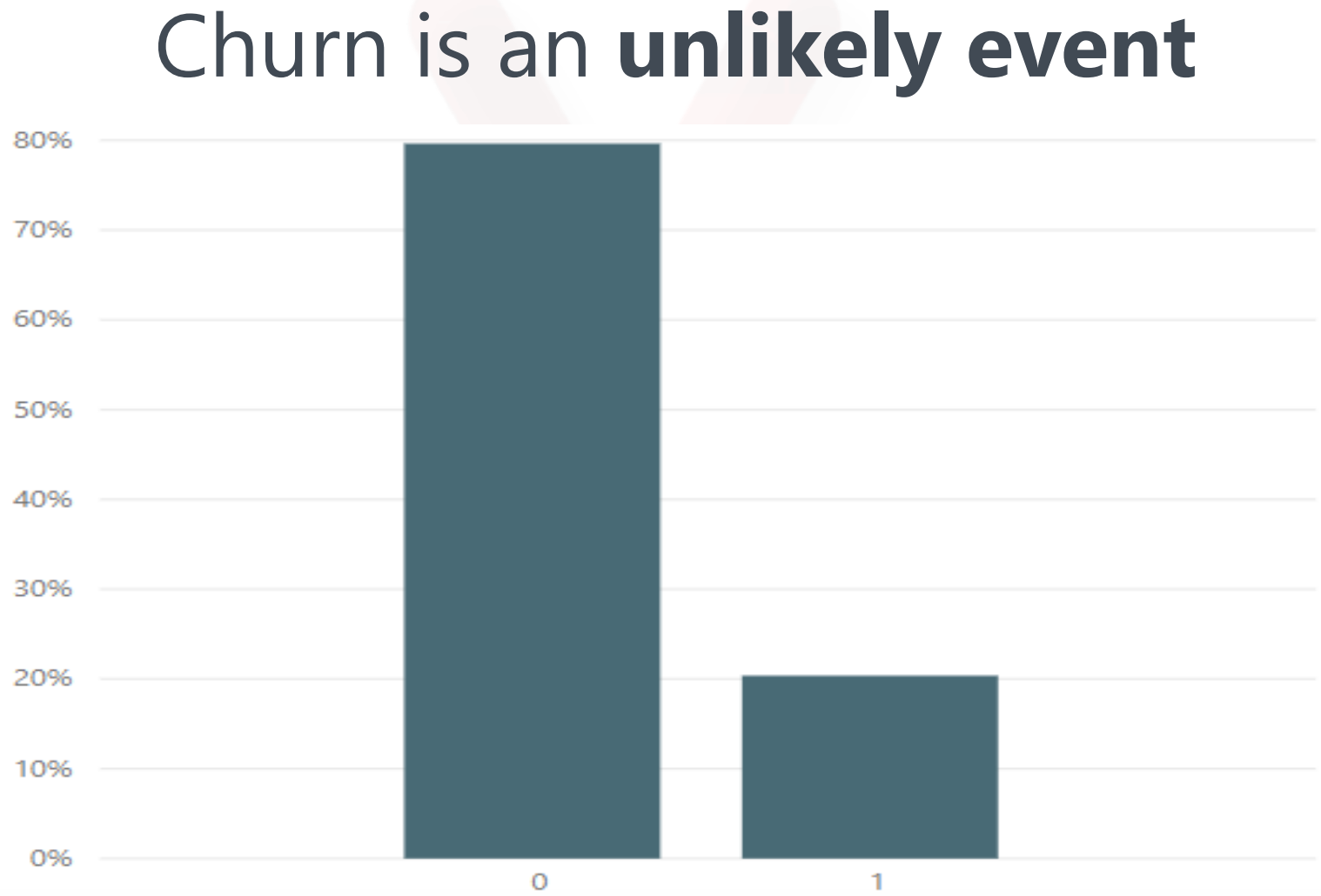
# DEMO

## Tuned Churn

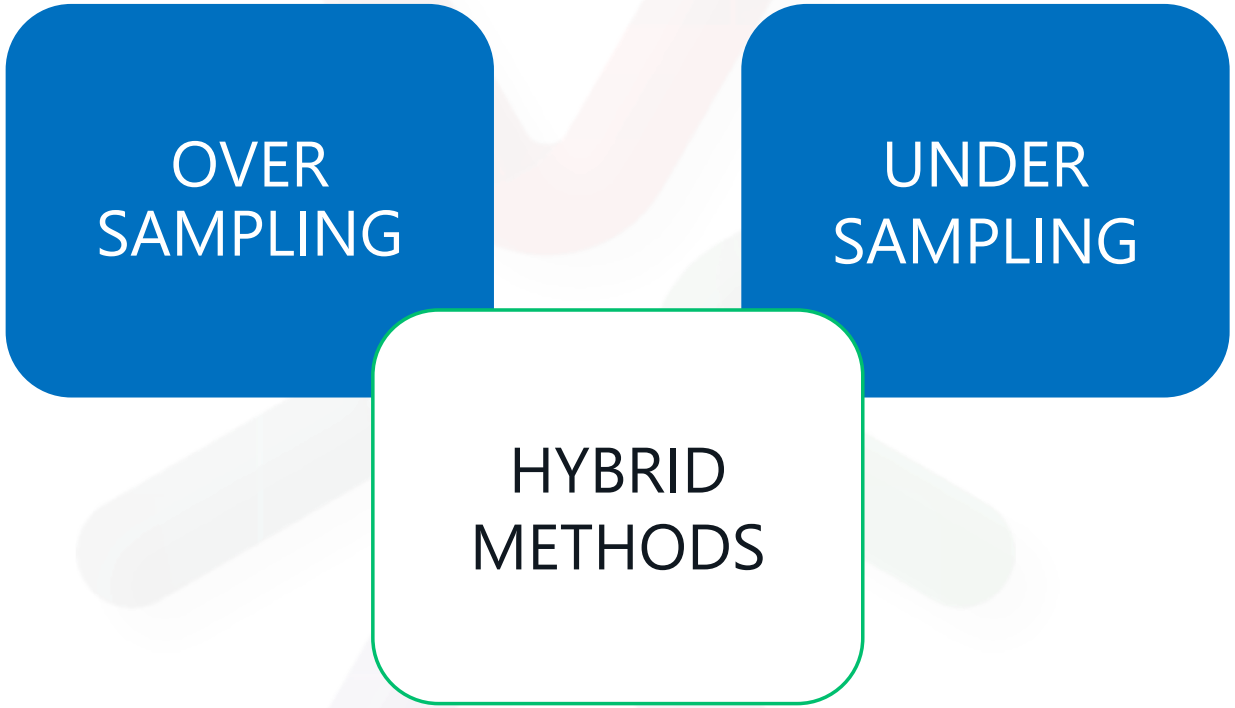


# In balance there is virtue

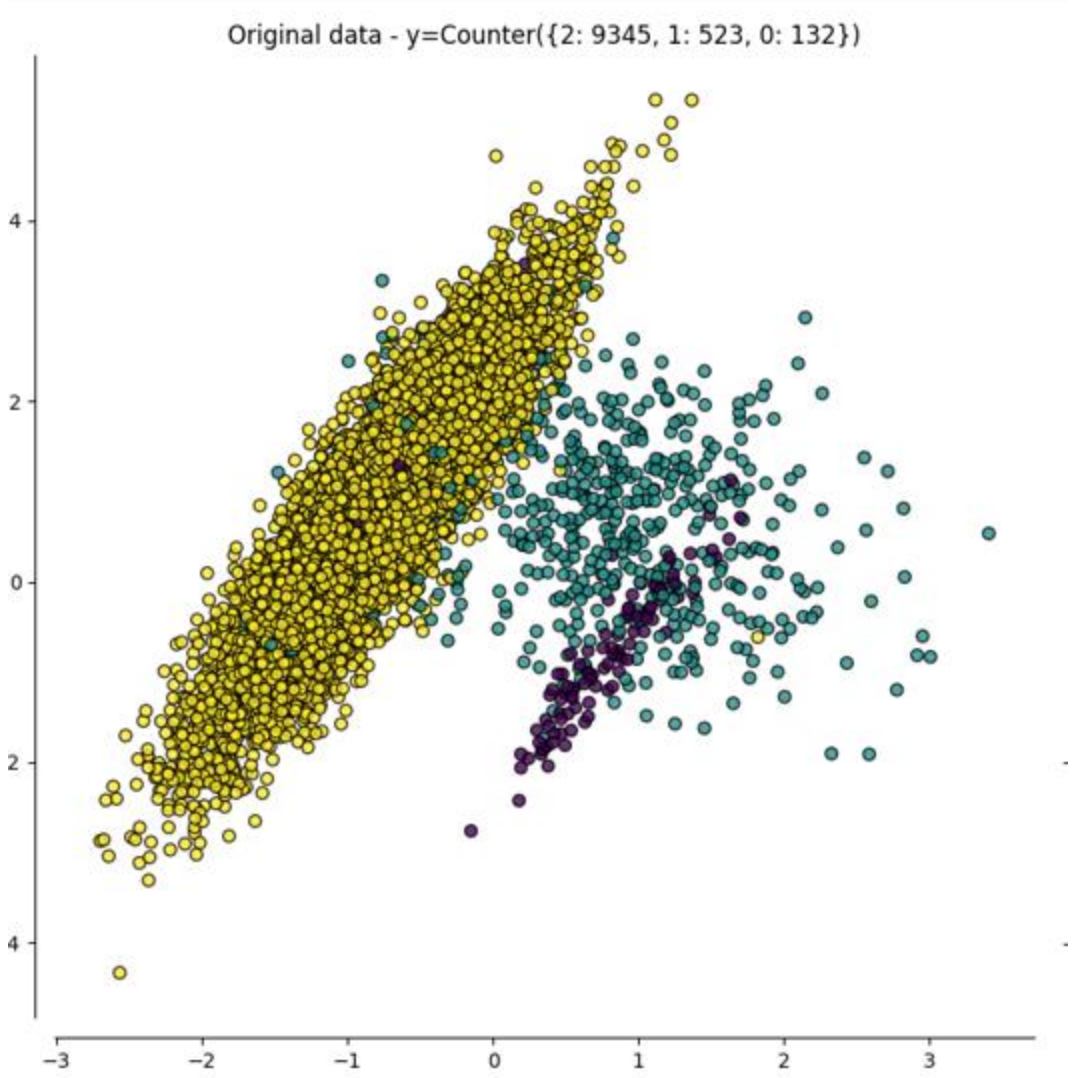
# In balance there is virtue



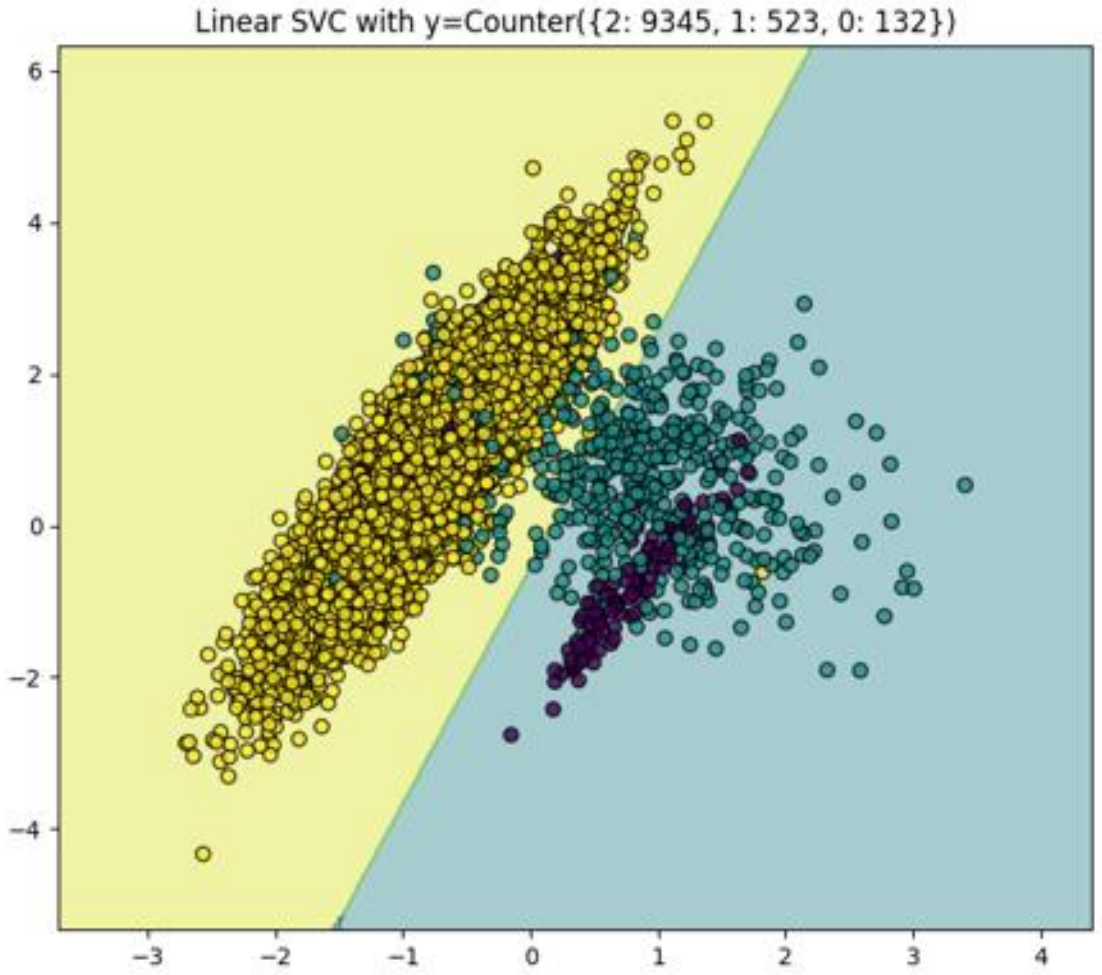
# In balance there is virtue



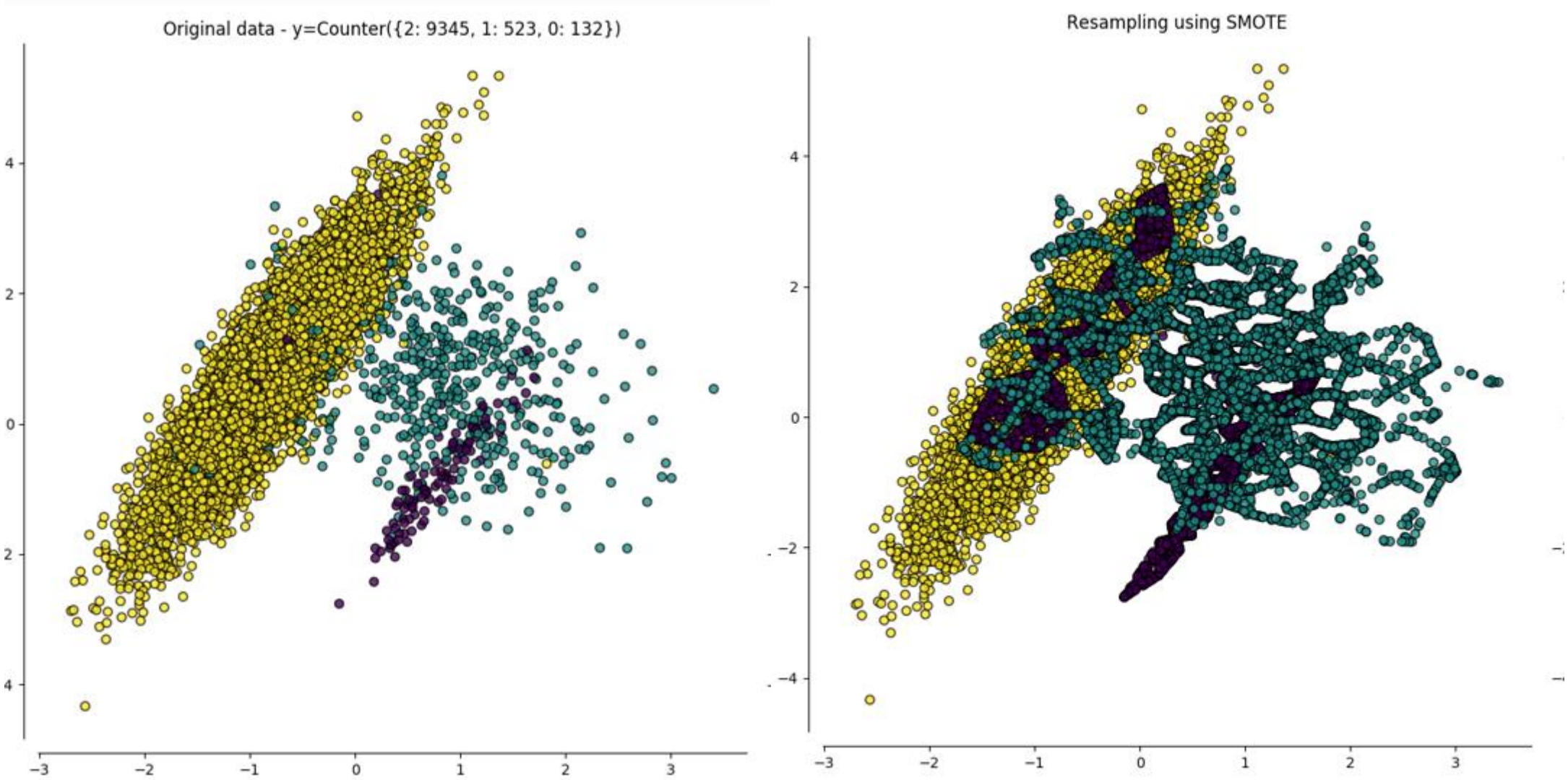
# In balance there is virtue



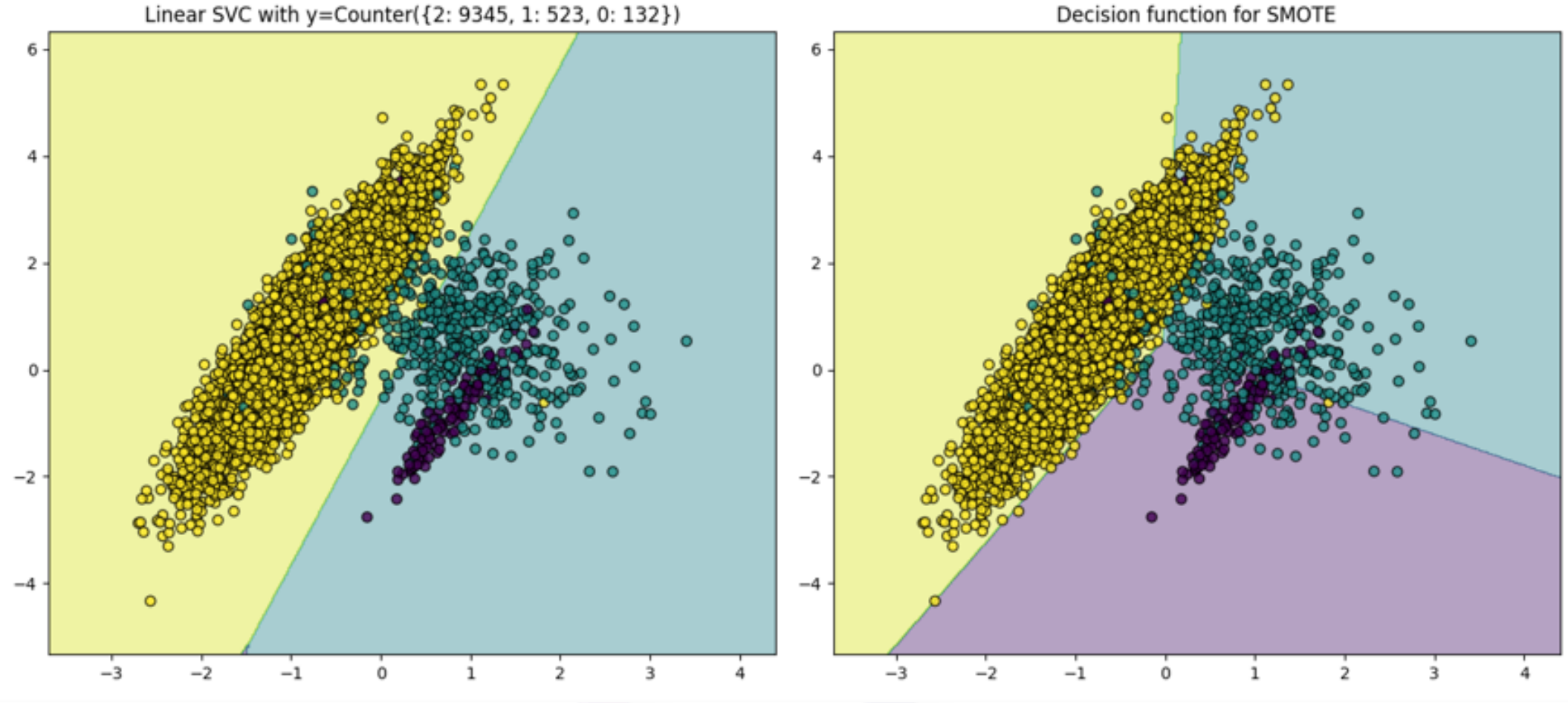
# In balance there is virtue



# In balance there is virtue



# In balance there is virtue



# In balance there is virtue

## Extra techniques (algorithm based)

WEIGHT VECTORS

BALANCED EXPLORATION

# DEMO

## Balanced Churn

# KISS

# KISS


**Simple models with significant features are usually better models**

# Tree-based methods

Decision Trees

Random Forests

Boosted Trees (lightgbm, xgboost, rxFastTrees…)

GAIN

COVER

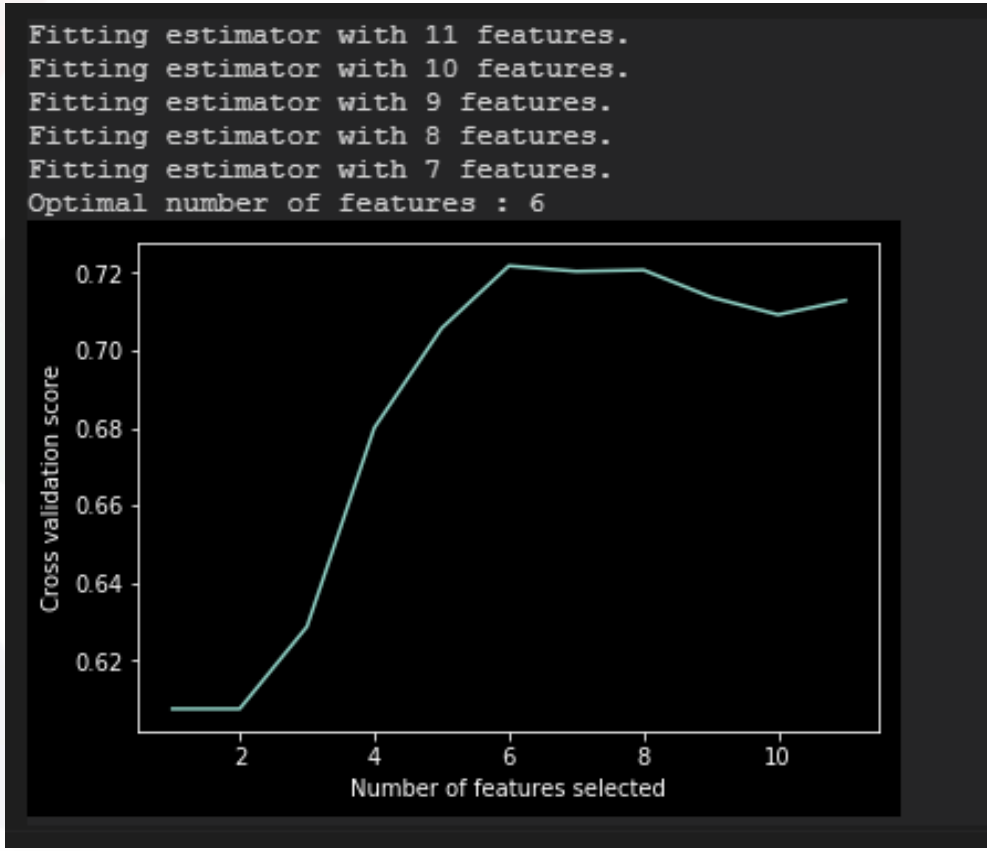
FREQUENCY

**DISCARD THE LESS USED FEATURES**

# Recursive Feature Elimination

Different combinations of features are tried **iteratively** and performance data is gathered

Requires an estimator (model)



# DEMO

## A simpler Churn

# How did we get here?

# How did we get here?

Churn is a **business problem**

**Business is key**

If our model is ONLY precise, **is not enough**

# Why do we need to interpret?

To understand the results we are receiving from our models

Debug and improve models

Avoid bias

Prescriptive analytics

# What can we see?

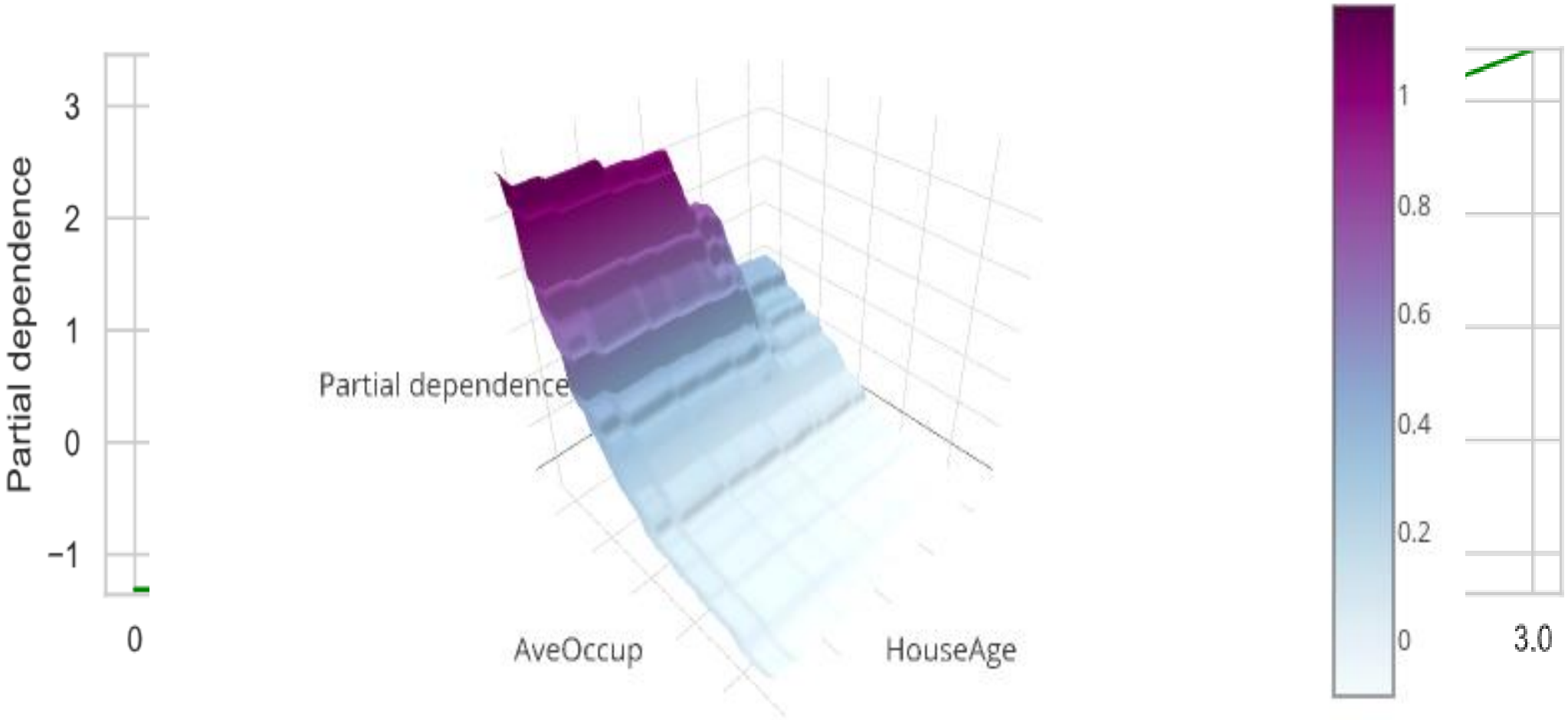
Returned values

- Regression
- Classification

Measures Can we rely ONLY in measures?

# Clever Hans

# Partial dependency plots



# DEMO

## Show me the model!

# QUESTIONS?

# THANK YOU!

Sigue #SQLSatMadrid