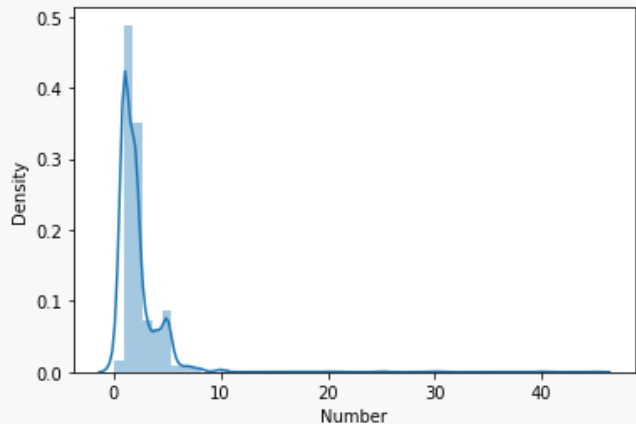


# Modelamiento predictivo de postulación de cupos en Move On

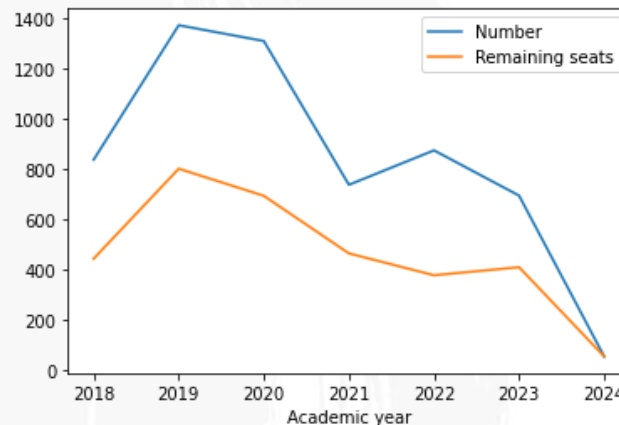
Ciencia de datos aplicada

- Análisis exploratorio
- Pautas principales
- Construcción de modelo base
- Modelos experimentales

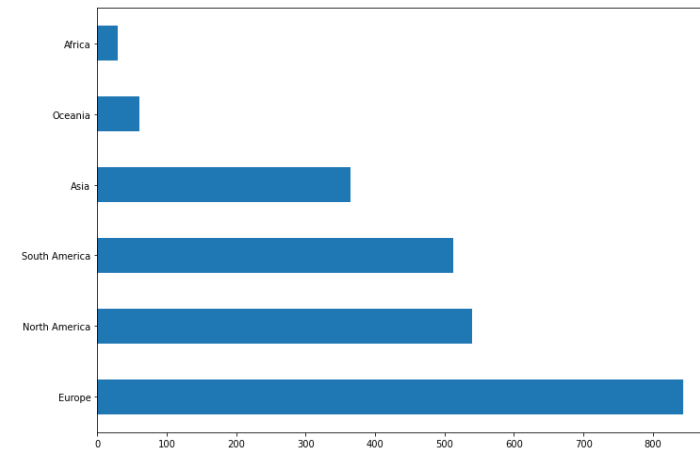
# Análisis exploratorio



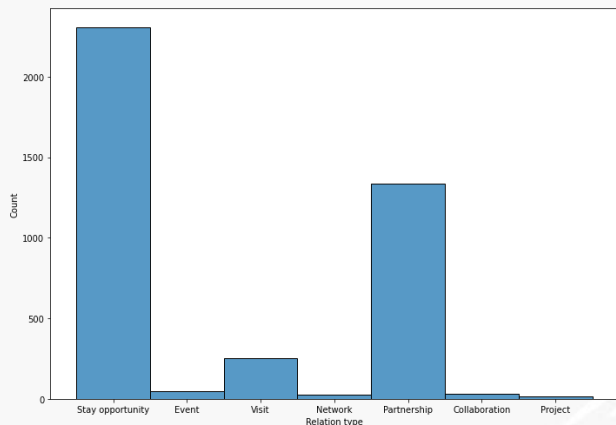
Distribución de cupos



Asignación de cupos



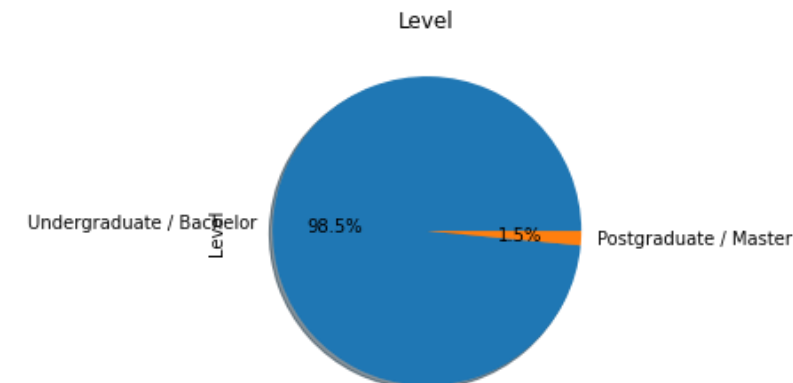
Distribución geográfica de instituciones



Convenios / Oportunidades de intercambio

	Academic year	Number	Remaining seats	Relation: Relation ID
Academic year	1.000000	-0.045528	-0.204352	-0.358787
Number	-0.045528	1.000000	0.792317	0.011644
Remaining seats	-0.204352	0.792317	1.000000	0.050139
Relation: Relation ID	-0.358787	0.011644	0.050139	1.000000

Correlación de variables para asignaciones (Seats)



Nivel académico de aspiraciones (Saty Wishes)

# Pautas principales

## Contexto

Se realiza unos modelos experimentales con un intento mayor de precisión

## Objetivo

Construir un modelo de RL que prediga el número óptimo de cupos de convenios en un semestre dado.

## Protección de datos

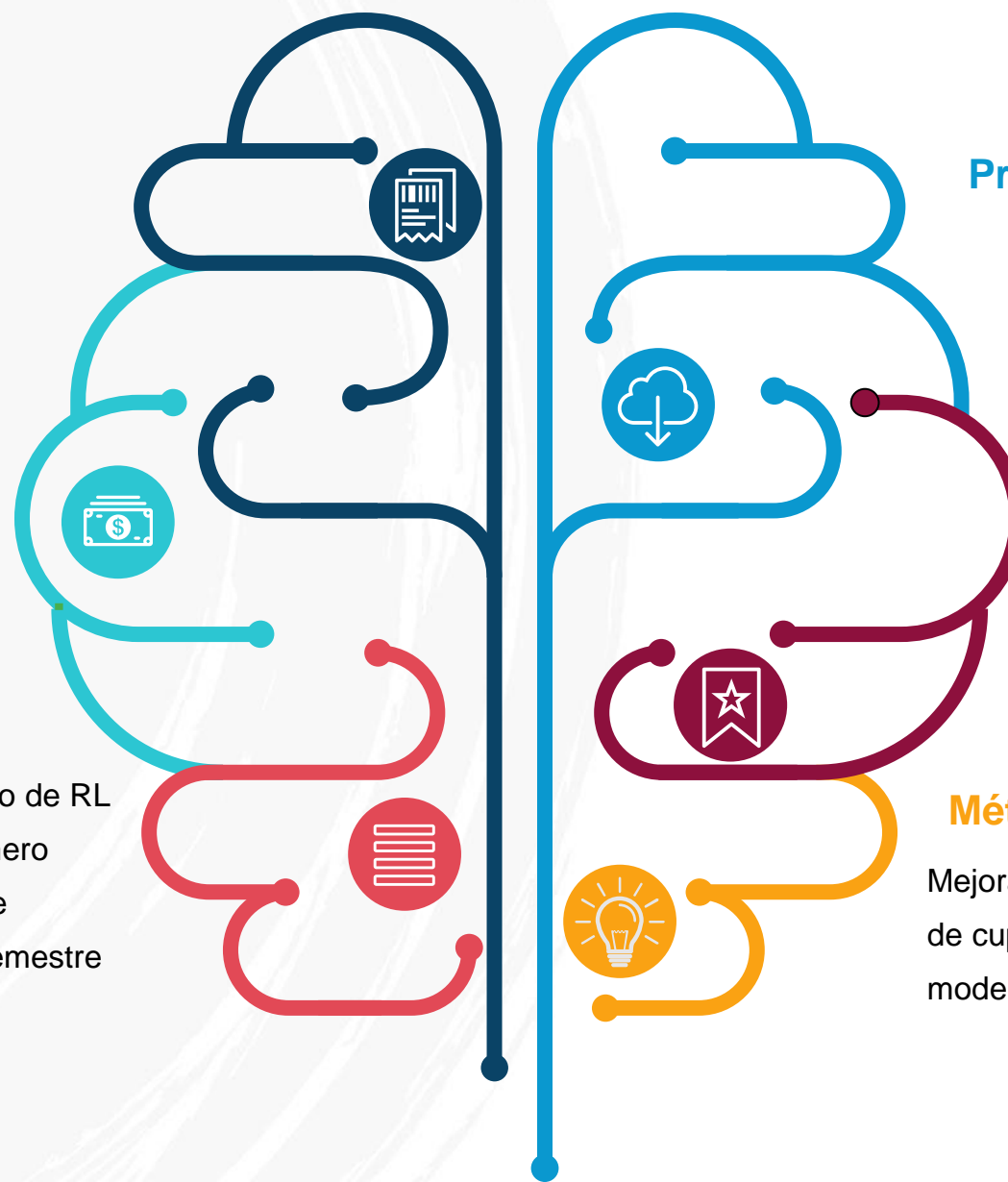
Anonimización de datos para proteger datos personales.

## Acotación de alcance

Entrega de una predicción lo más real posible

## Métrica

Mejorar el kpi de asignación de cupos mediante el modelo de regresión lineal.



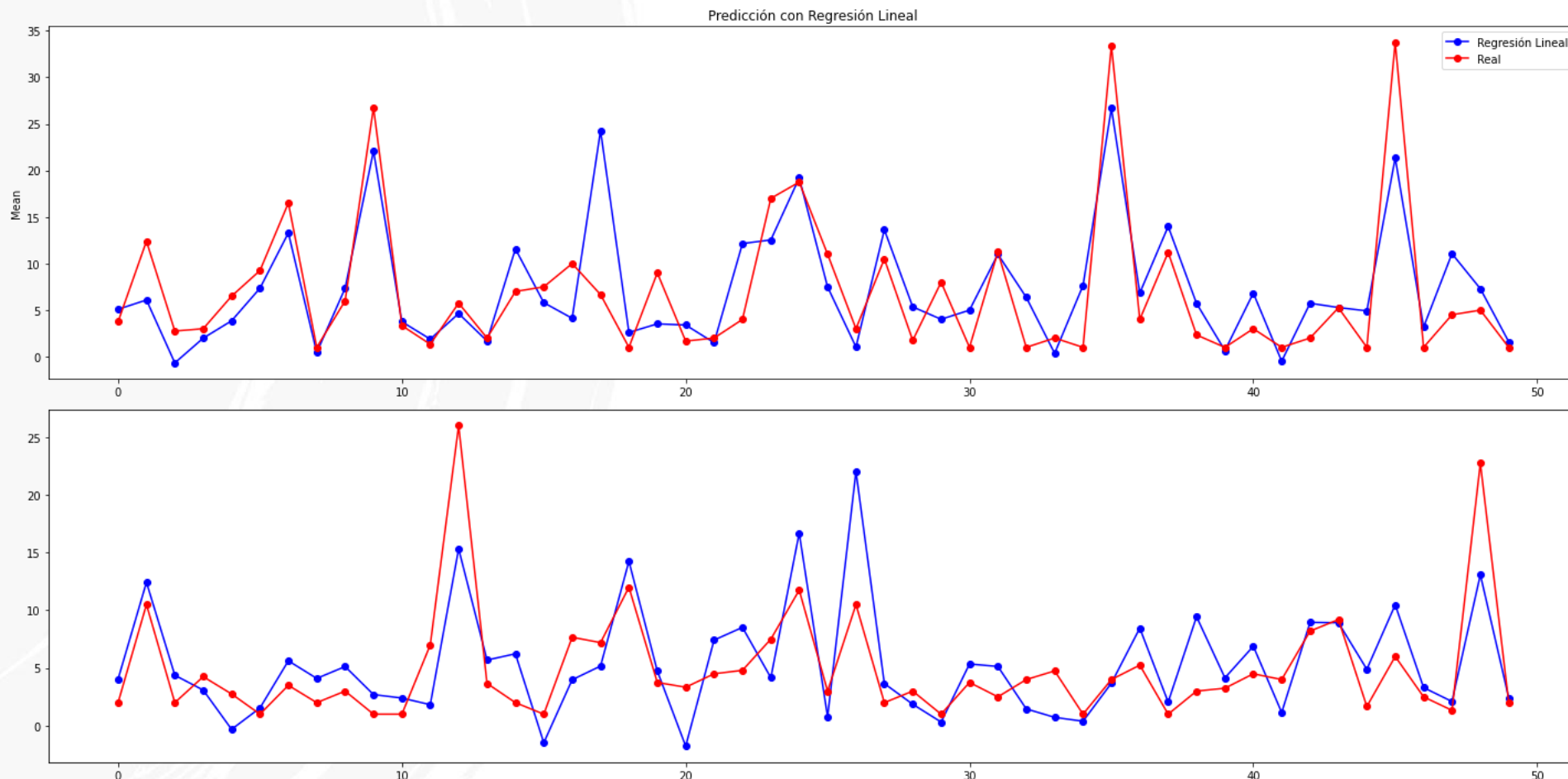
# Construcción del modelo base

## Variables

- Country.
- Main External Institutions.
- External institutions.
- Institution Name.
- Relation Type.
- Frameworks.
- Start Date.
- End Date.
- Status.
- Correlation Content.
- Last renewal.
- Degree programme.
- Internal institutions.
- Direction.
- Parent Relation.
- Level.
- Reach

¿Regularización?: No  
One Hot Encoder?: Si  
Variable a predecir: Mean

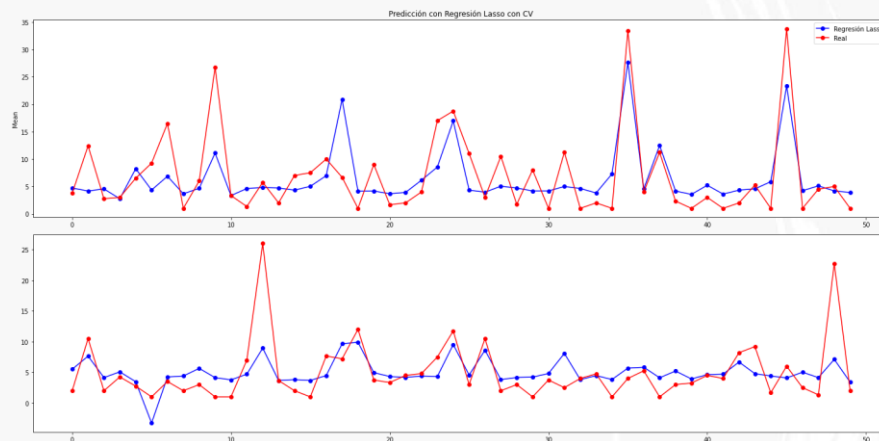
R2 entrenamiento: 0.69011  
R2 prueba: 0.43250



# Modelos experimentales

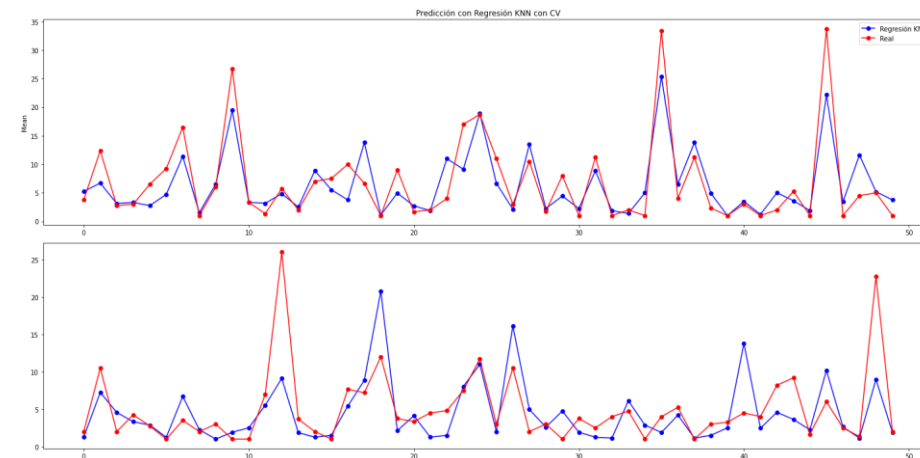
Usamos Grid search para encontrar los mejores parámetros.

*Polinomial  
Lasso*



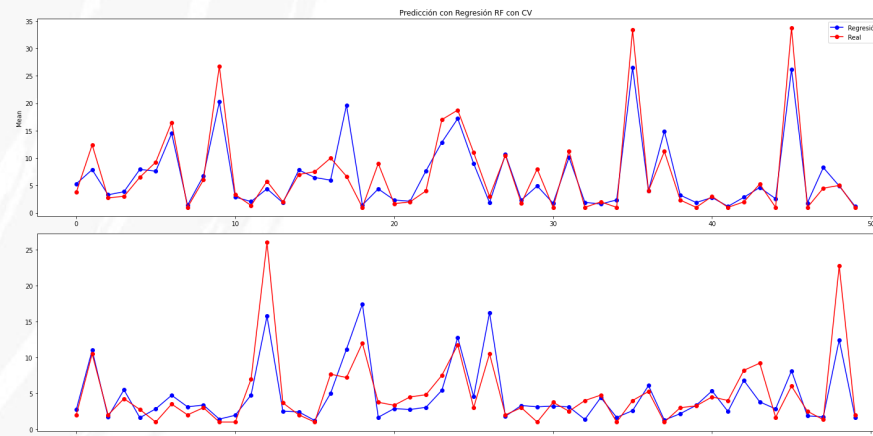
**R2 entrenamiento: 0.69011**  
**R2 prueba: 0.43250**

*Nearest  
Neighbor*



**R2 entrenamiento: 0.46750**  
**R2 prueba: 0.43785**

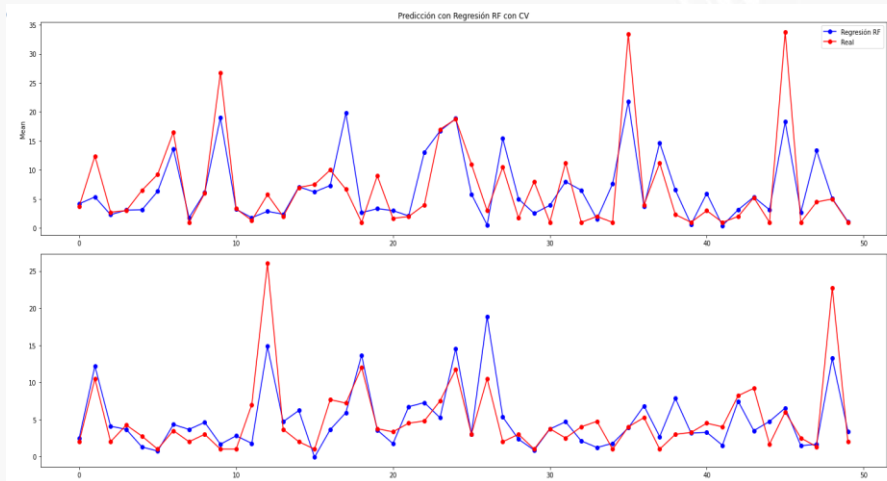
*Random  
Forest*



**R2 entrenamiento: 0.84211**  
**R2 prueba: 0.62212**  
**Mejor ajuste**

# Modelos experimentales

Support  
Vector  
Regresion



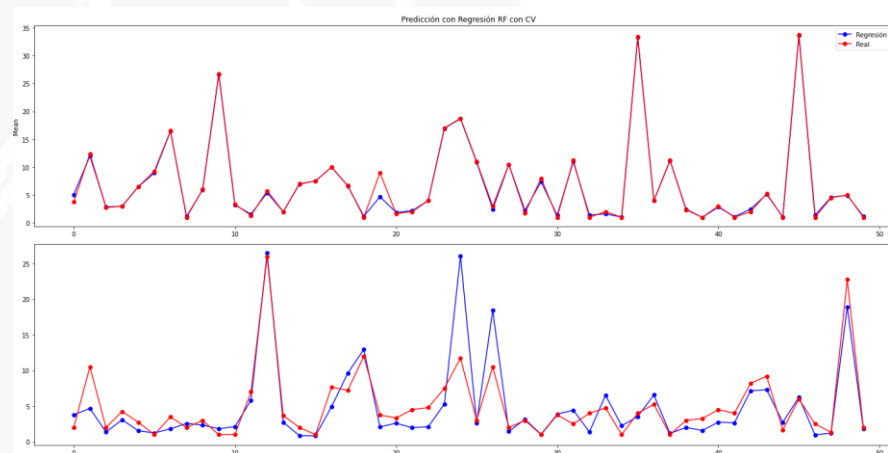
**R2 entrenamiento: 0.62367**  
**R2 prueba: 0.5326**

Decision  
Tree  
Regresor



**R2 entrenamiento: 0.87259**  
**R2 prueba: 0.07850**

XGBoost  
Regresion



**R2 entrenamiento: 0.99046**  
**R2 prueba: 0.5477**  
***SUPER-SOBREAJUSTE***

# Modelo final

Predicción con Regresión RF con CV

