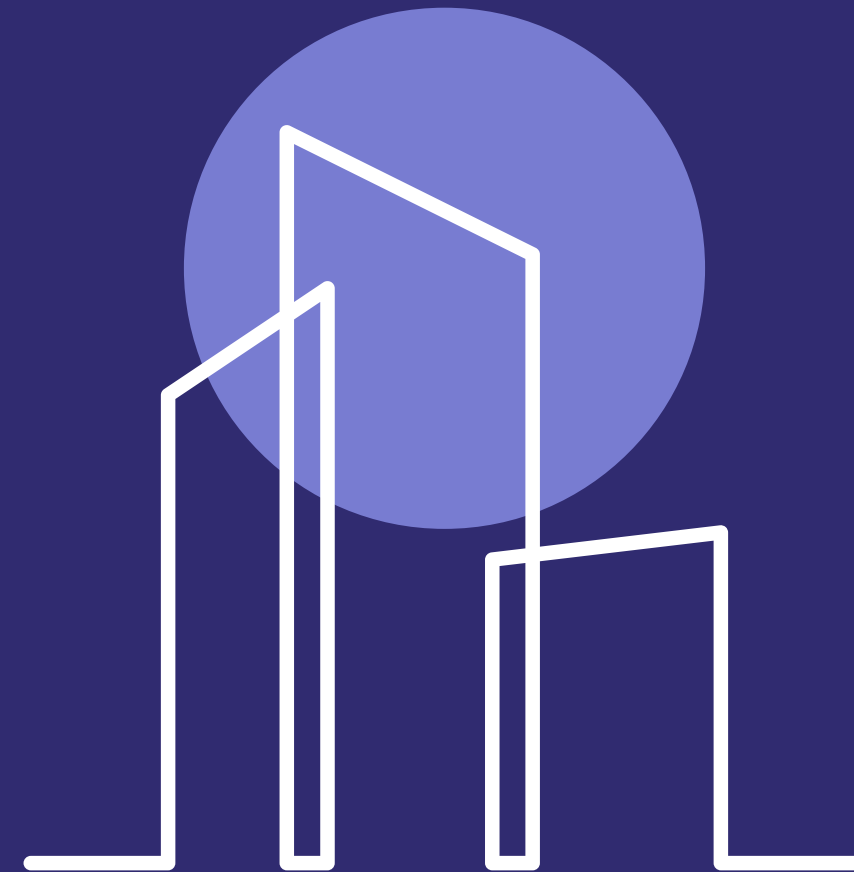


# NEGOCIO DE INMOBILIARIA

Predecir los precios de una propiedad



Laura Martinez y Diana Olano

# TABLA DE CONTENIDO



1

Contexto del negocio

2

Antecedentes de la  
solución

3

Información del dataset  
y pregunta de interés

4

Análisis exploratorio de  
los datos

5

Resultados de los  
modelos





# CONTEXTO



# Contexto

¿En que consiste la empresa?



Las agencias de bienes raíces tienen acceso a una gran cantidad de datos sobre el mercado inmobiliario.

Los precios de las propiedades son un buen indicador tanto de la condición general del mercado como de la salud económica de un país. Además proporcionan información sobre la oferta y la demanda del mercado inmobiliario.



# Objetivos



1

Líder del mercado



2

Sistema preciso de  
predicción de precios



3

Eficiencia en la venta



# Pregunta de interés



¿Cuáles son los factores con mayor influencia al establecer los precios de propiedades en Colombia durante el periodo 2023?



# ANTECEDENTES





Amazon

UBER

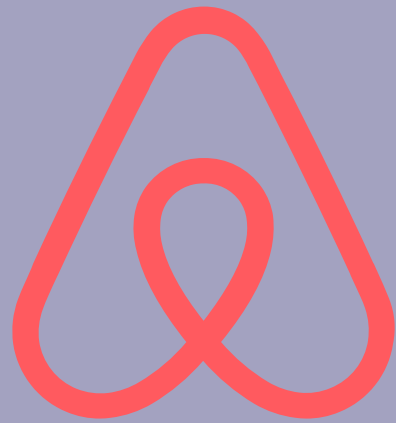
Uber



Aerolíneas



Bolsa de  
valores



Airbnb



Google Ads

Google Ads



Netflix



# INFORMACIÓN DEL DATASET



# Información del dataset

## Información general



El dataset contiene **anuncios inmobiliarios de Colombia**, con información sobre número de habitaciones, barrios, precios, etc. Incluyen casas, apartamentos, lotes comerciales, entre otros tipos de propiedades.

Los conjuntos de datos provienen del sitio web [Properati Data](#).



Contiene 1.000.000 de filas y 25 columnas de datos.



# Información del dataset

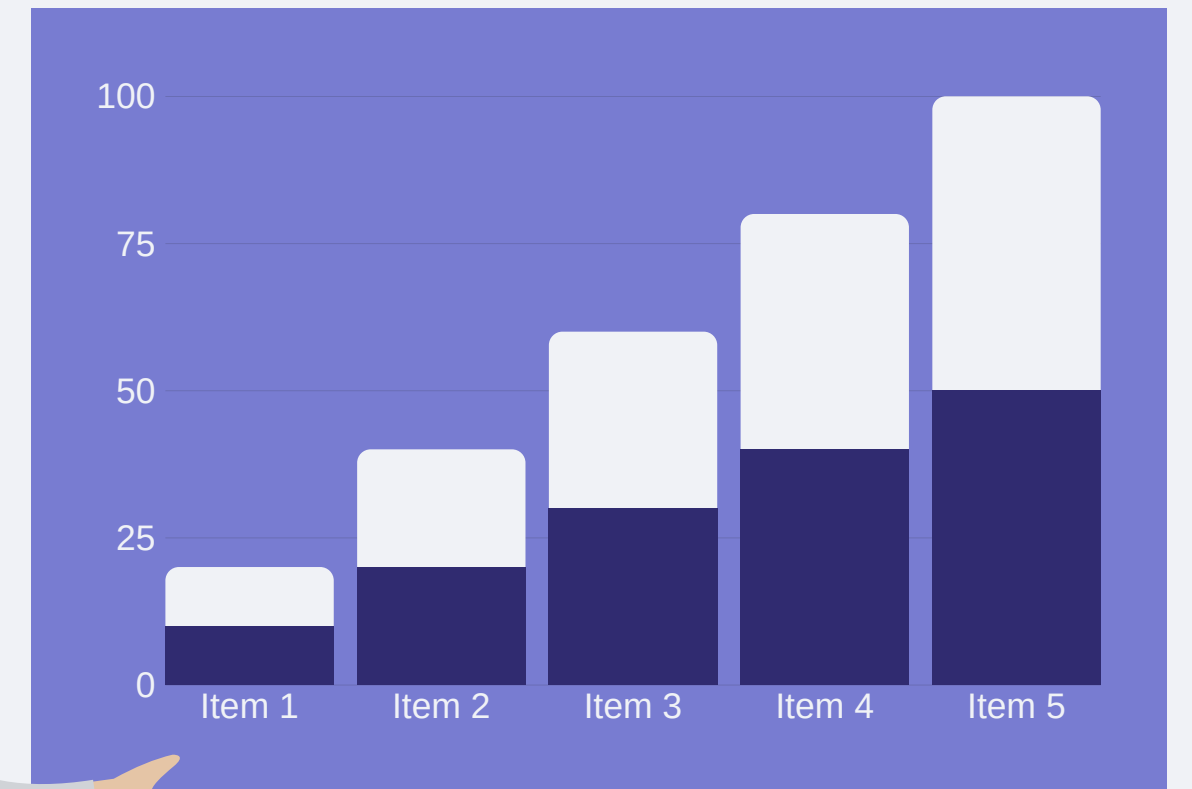
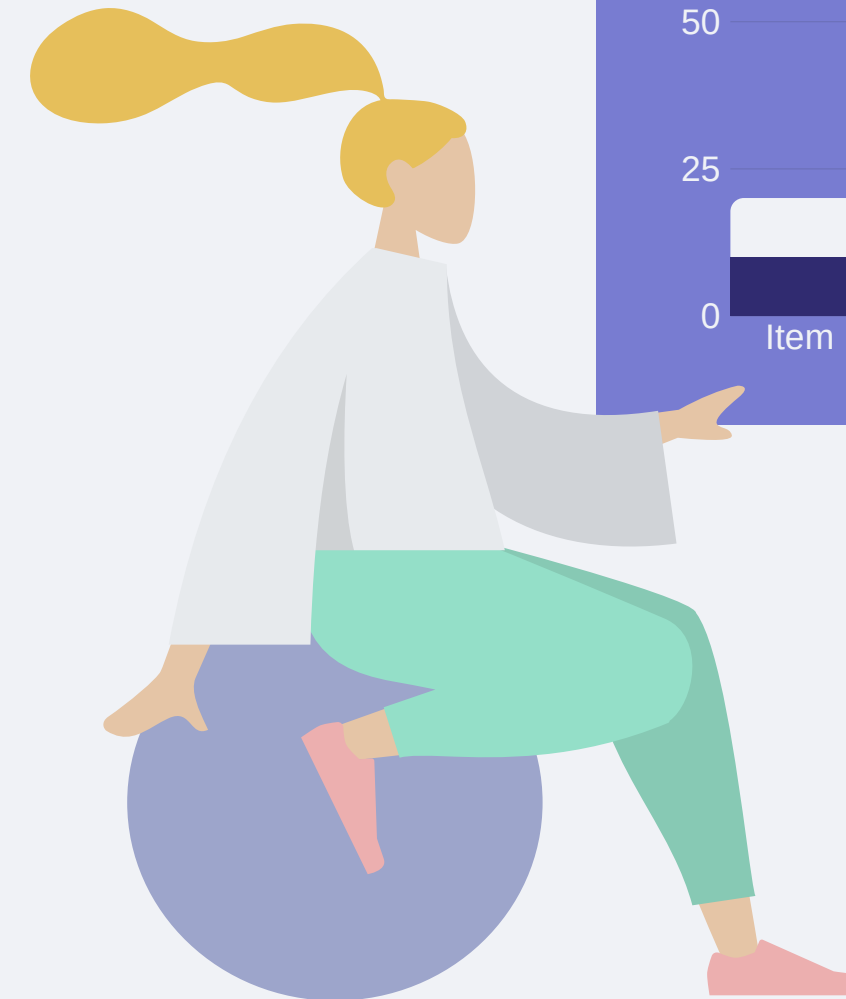
Variables utilizadas



0	id	13	rooms
1	ad_type	14	bedrooms
2	start_date	15	bathrooms
3	end_date	16	surface_total
4	created_on	17	surface_covered
5	latitude	18	price
6	longitude	19	currency
7	country	20	price_period
8	province	21	title
9	city	22	description
10	neighbourhood	23	property_type
11	15	24	operation_type
12	16		



# ANÁLISIS EXPLORATORIO DE LOS DATOS



# Modificaciones del Dataset

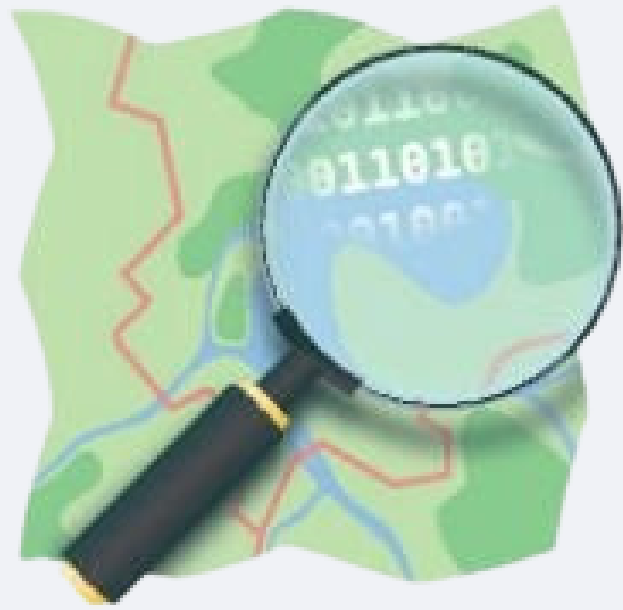


**PASO 1: ELIMINACIÓN DE COLUMNAS**

**PASO 2: ELIMINACIÓN DE FILAS**

**PASO 3: REEMPLAZO LOS NULOS POR LA FRECUENCIA**

**PASO 4: VALORES DE LA COLUMNA CITY USANDO UN GEOLOCALIZADOR**



# OpenStreetMap

Utiliza datos de OpenStreetMap para encontrar ubicaciones en tierra por latitud y longitud ([reverse geocoding](#)).

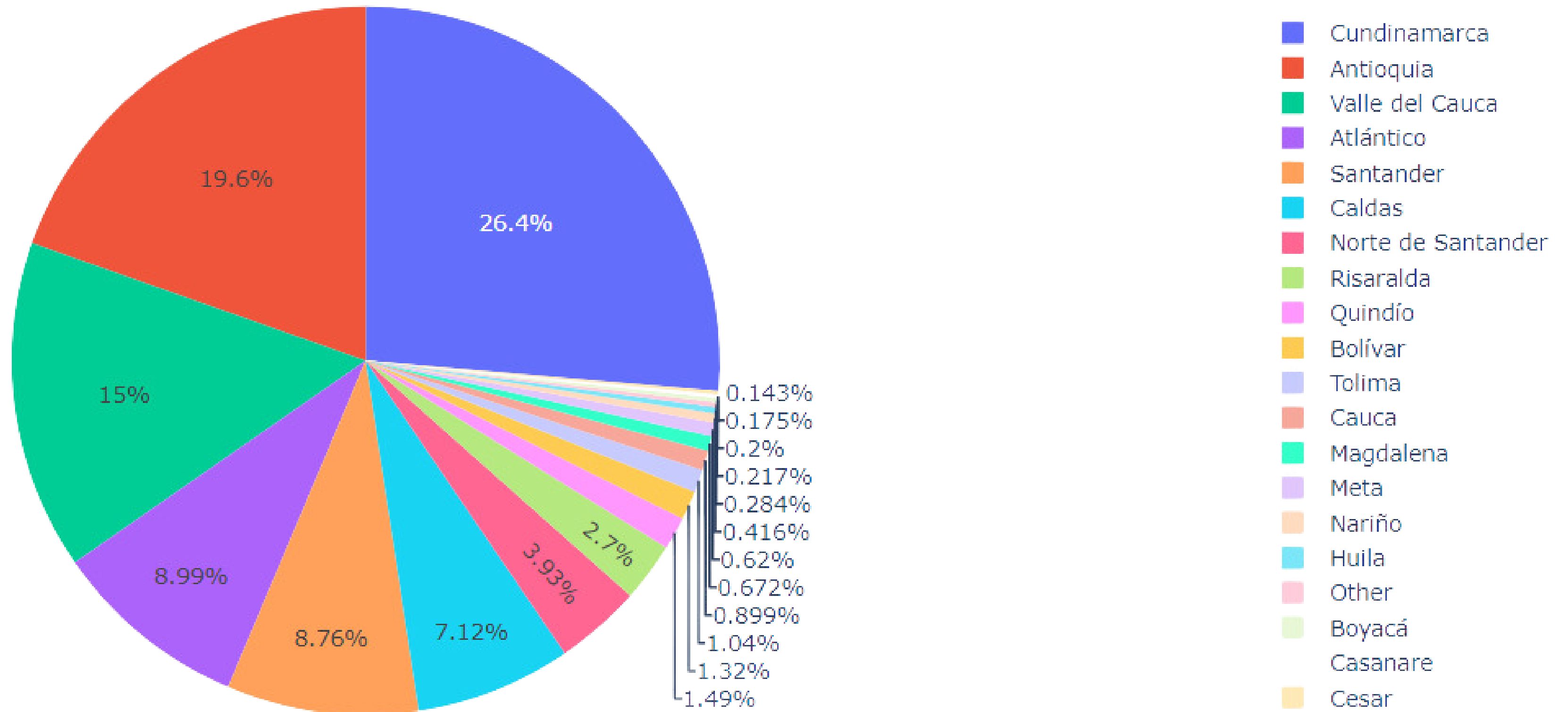
## Geolocalizador: Nominatim



# VISUALIZACIÓN

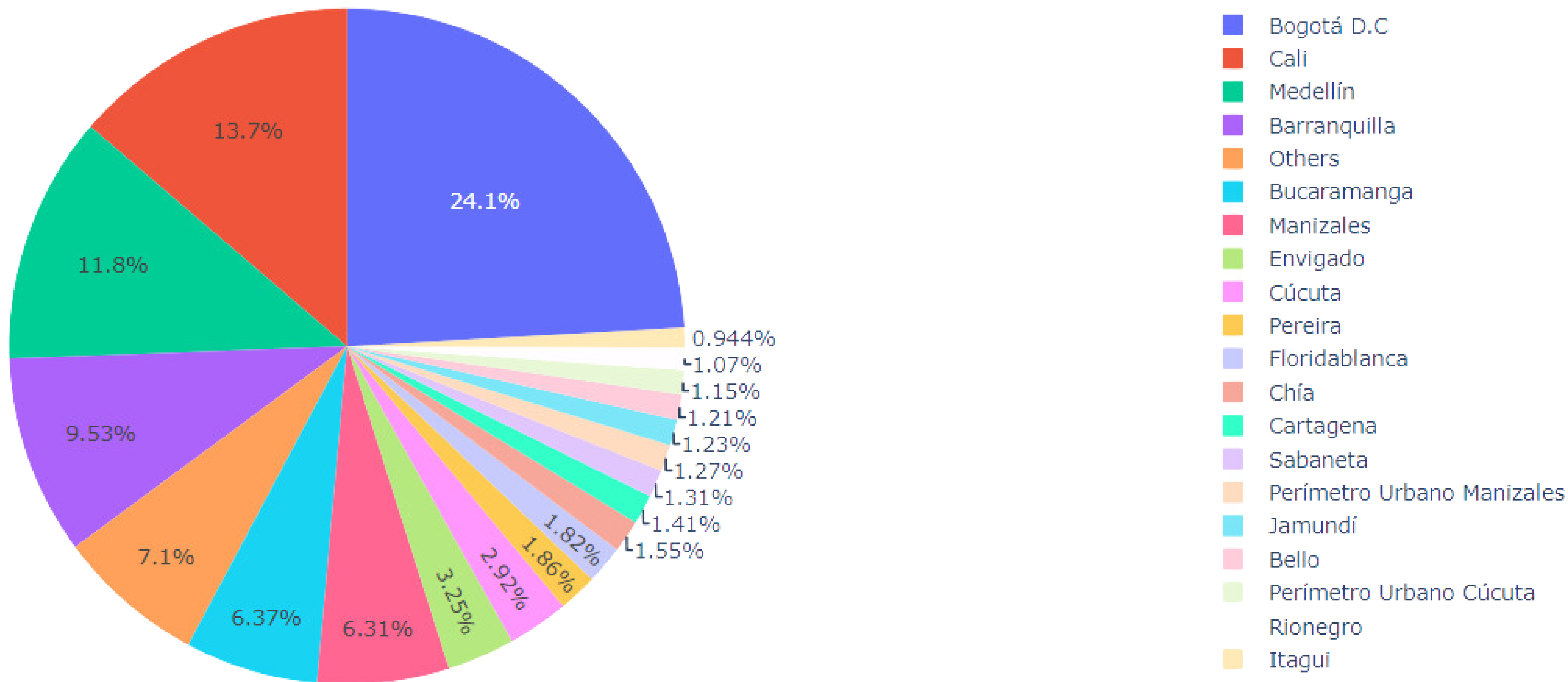


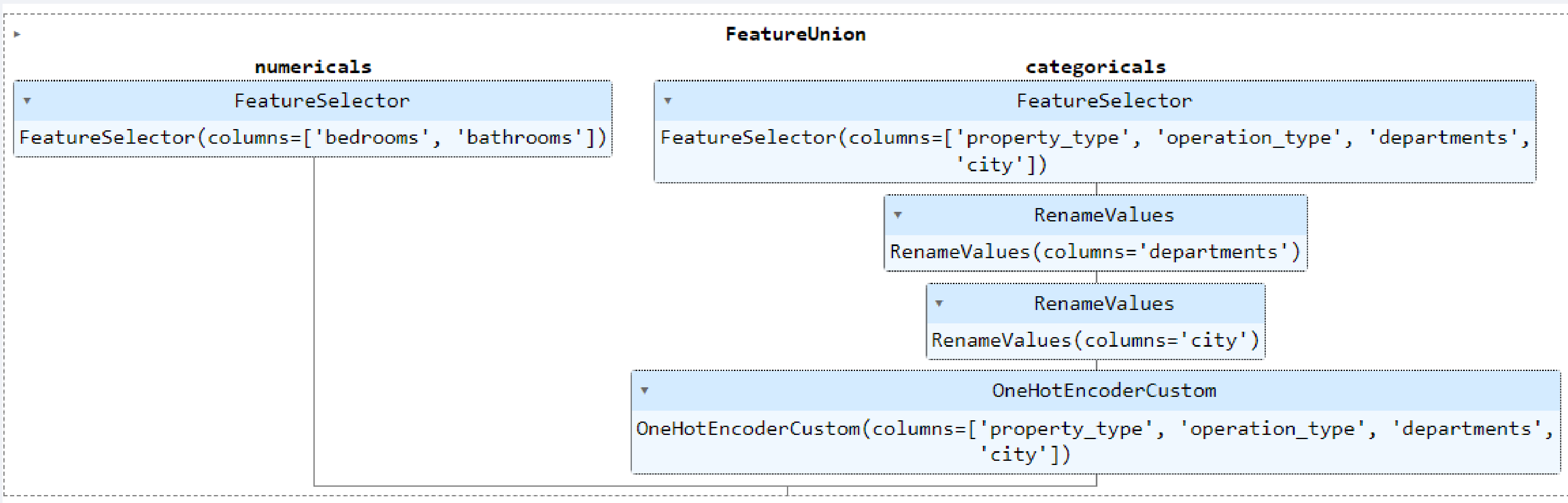
# DEPARTAMENTOS





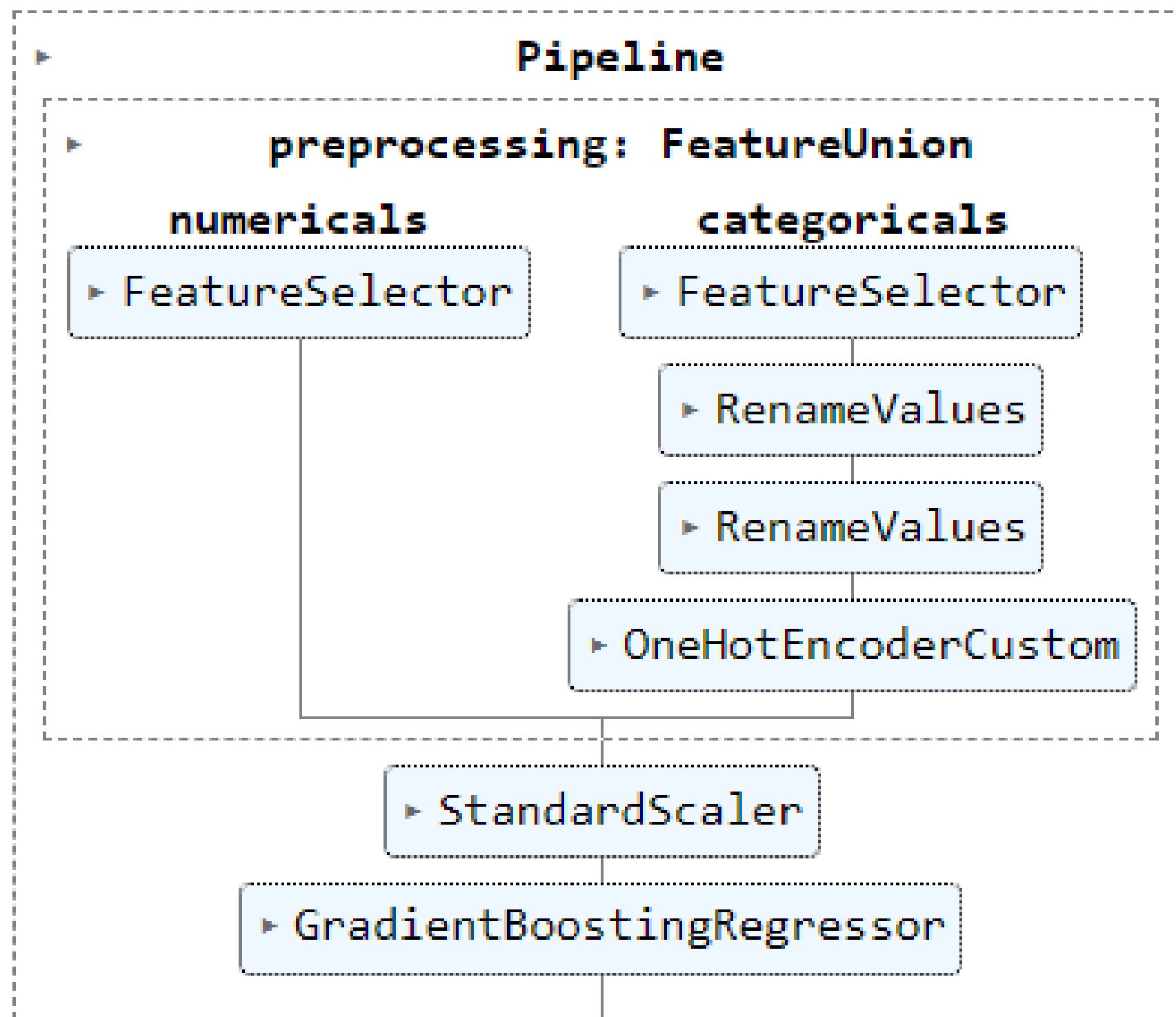
# CIUDADES





# PIPELINE: PREPARACIÓN

# PIPELINE GENERAL



1. Regresión Lineal
2. Regresión Ridge
3. Regresión Lasso
4. Regresión Random Forest
5. Regresión Gradient Boosting

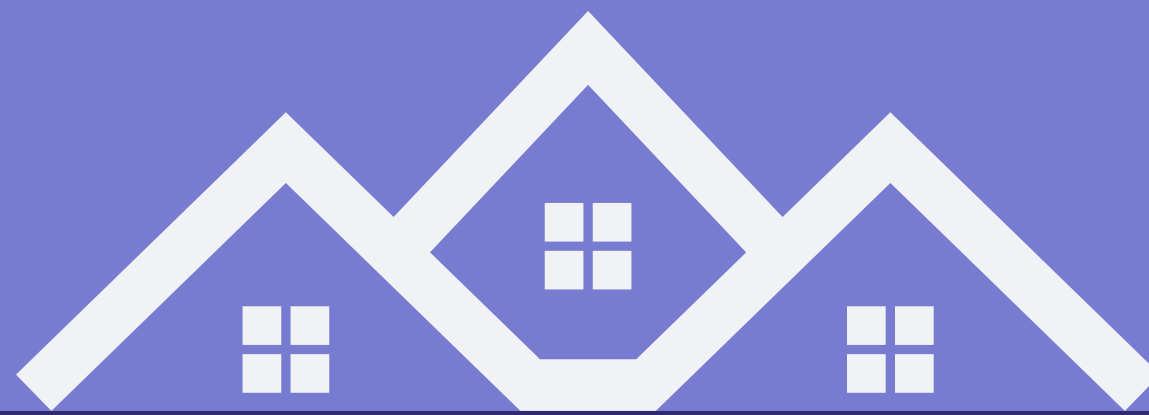
# RESULTADOS DE LOS MODELOS



# COMPARACIÓN DE LOS MODELOS

	MODELO	TIPO DE RESPUESTA	TIPO DE RELACIÓN	REGULARIZACIÓN
	LINEAL	Numérica	Lineal	No
	RIDGE	Numérica	Lineal	<b>L2</b> , penaliza los coeficientes grandes
	LASSO	Numérica	Lineal	<b>L1</b> , penaliza los coeficientes pequeños
	RANDOM FOREST	Numérica o categórica	No lineal o compleja	No
	GRADIENT BOOSTING	Numérica o categórica	No lineal o compleja	Sí, usa el aprendizaje por gradiente para optimizar una función de pérdida

# Referencias



- Platzi. (s.f.). ¿Qué es regresión lineal? [Blog]. Recuperado de <https://platzi.com/blog/que-es-regresion-lineal/>
- Interactive Chaos. (s.f.). Regresión Ridge. Recuperado de <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/regresion-ridge>
- Interactive Chaos. (s.f.). Regresión Lasso. Recuperado de <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/regresion-lasso>
- IBM. (s.f.). ¿Qué es Random Forest? [Página web]. Recuperado de <https://www.ibm.com/mx-es/topics/random-forest#:~:text=El%20random%20forest%20es%20un,problemas%20de%20clasificaci%C3%B3n%20y%20regresi%C3%B3n>.
- DataScience.eu. (s.f.). Gradient Boosting: What You Need to Know. Recuperado de <https://datascience.eu/machine-learning/gradient-boosting-what-you-need-to-know/>