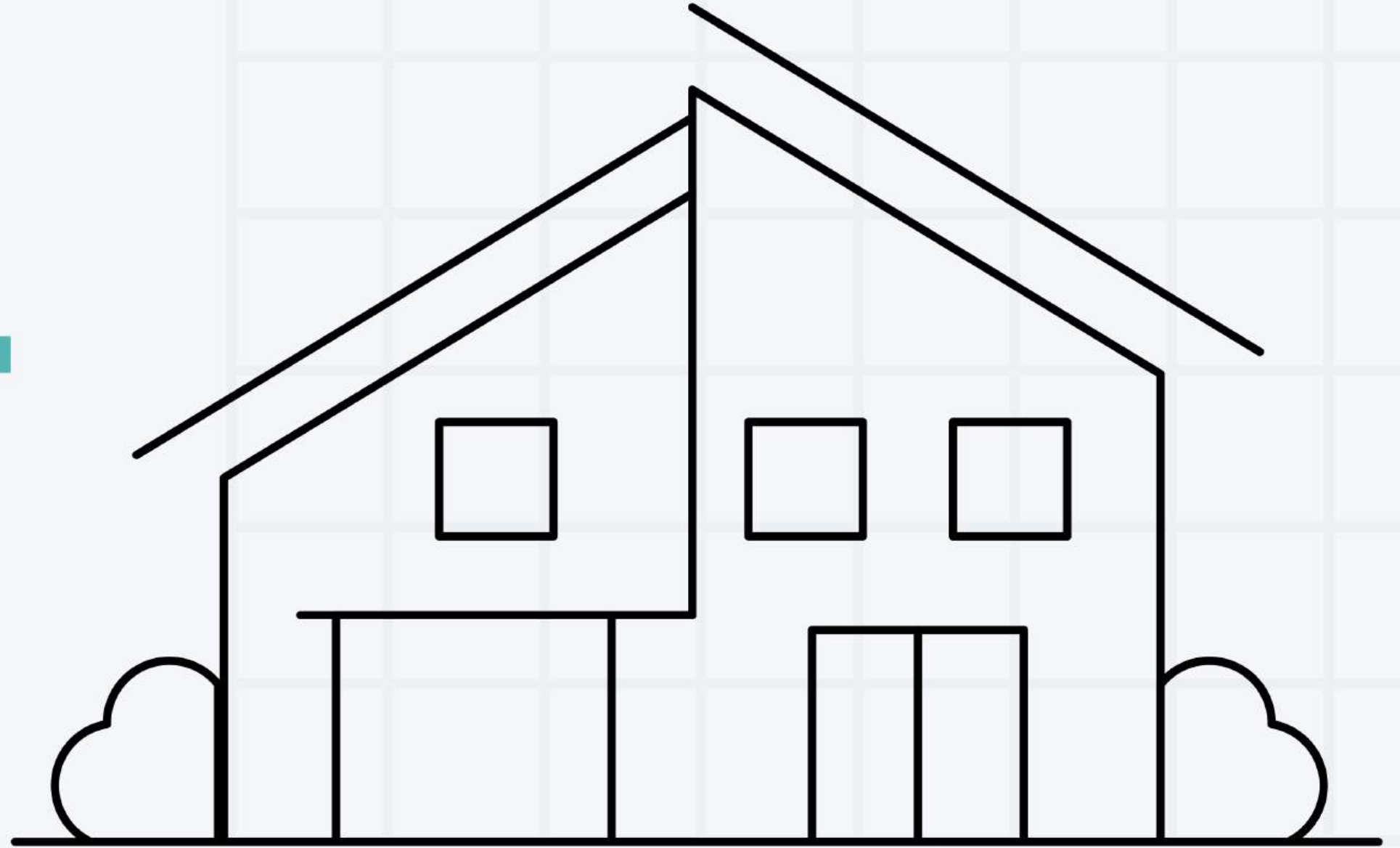


# Real Estate

---



Fernando Sanz-Extremuera, Gael Moreira, Laura Nieto.



```
NAL_XML', false);  
are("5.2", PHP_VERSION, ">")) {  
or greater is required!!!");  
loaded("pcre")) {  
Info requires the pcre extension to php in order  
y.");  
PP_ROOT.'/includes/autoloader.inc.php';  
uration  
APP_ROOT.'/config.php';  
'PSI_CONFIG_FILE') || !defined('PSI_DEBUG'))  
w Template("/templates/html/error_config.html  
->fetch();  
out javascript  
: strt
```



# Datos

PRECIOS	HAB	BAÑOS
TERRENO	ID	FECHA VENTA
FECHA CONSTRUC.	PIES2 ÚTILES	VISITAS
ESPACIO SOTANO	DATOS 15 VECINOS	ETC





# 01

Exploración básica.

# 02

Limpieza de datos ya revisados.



# 03

Testeo de todos los metodos de regresión y  
seleccion del mas efectivo.

# 04

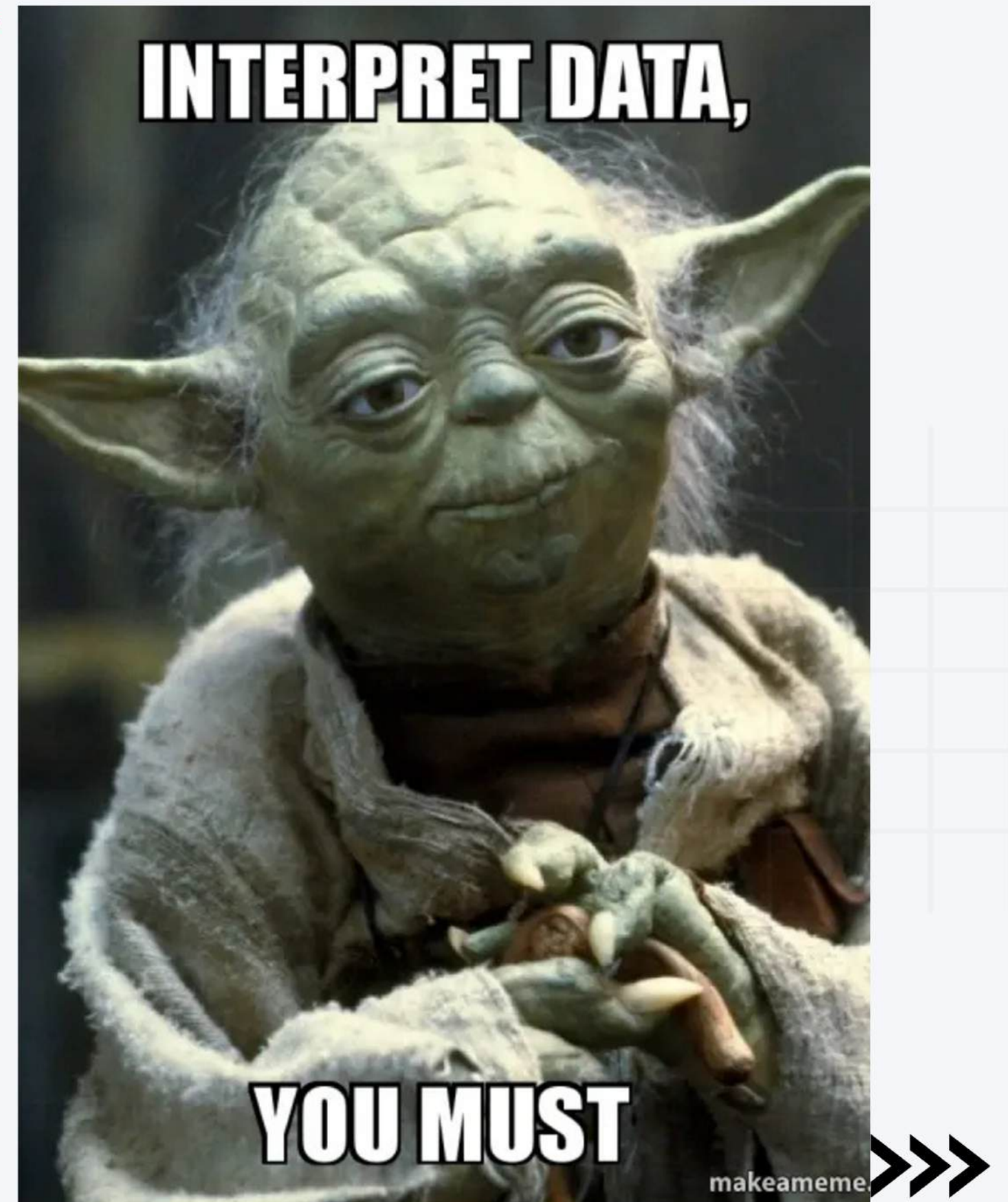
Representacion de los datos finales





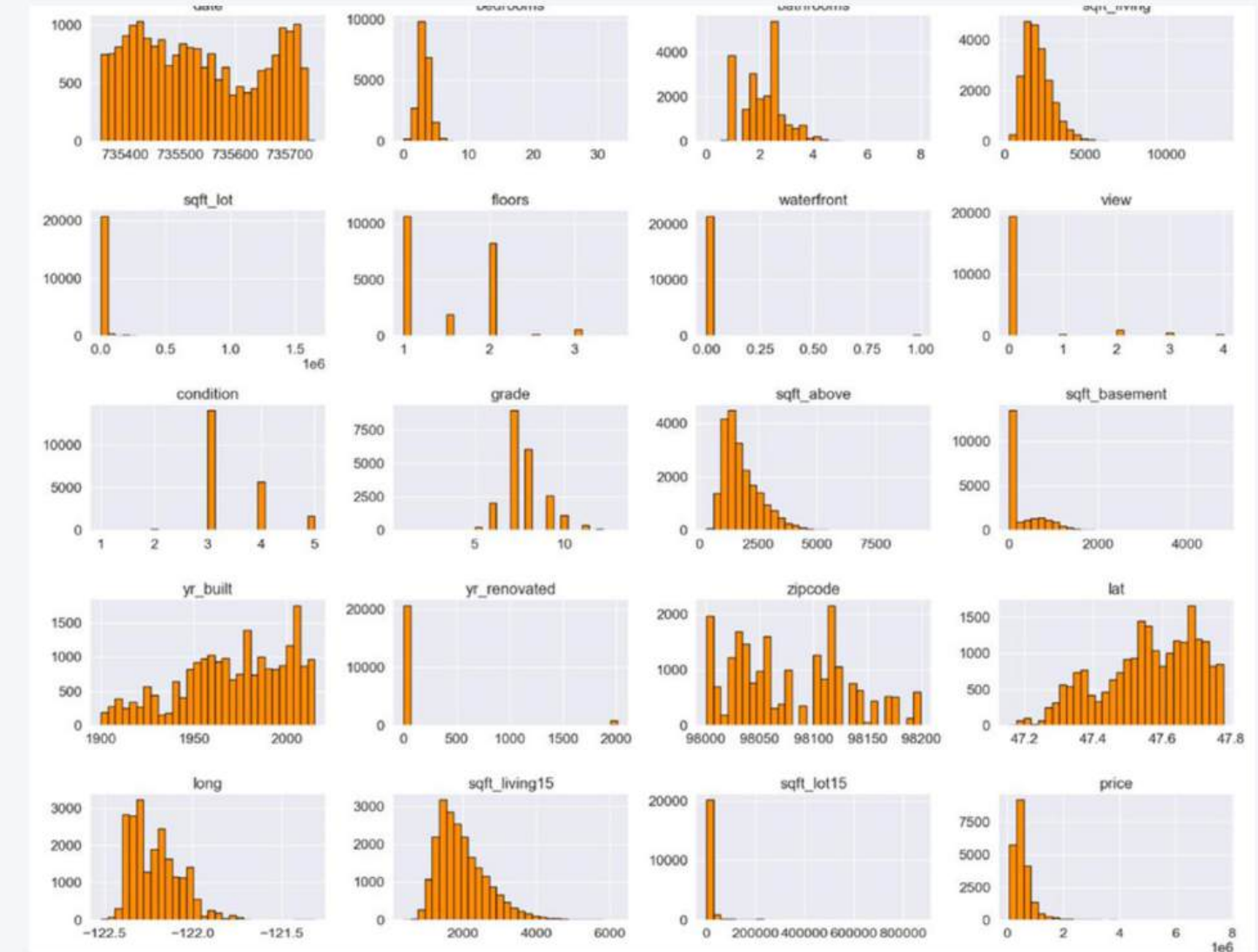
# Limpieza

- Id como Indice
- 33 habitaciones
- 0 hab y 0 baños
- Fecha a ordinal

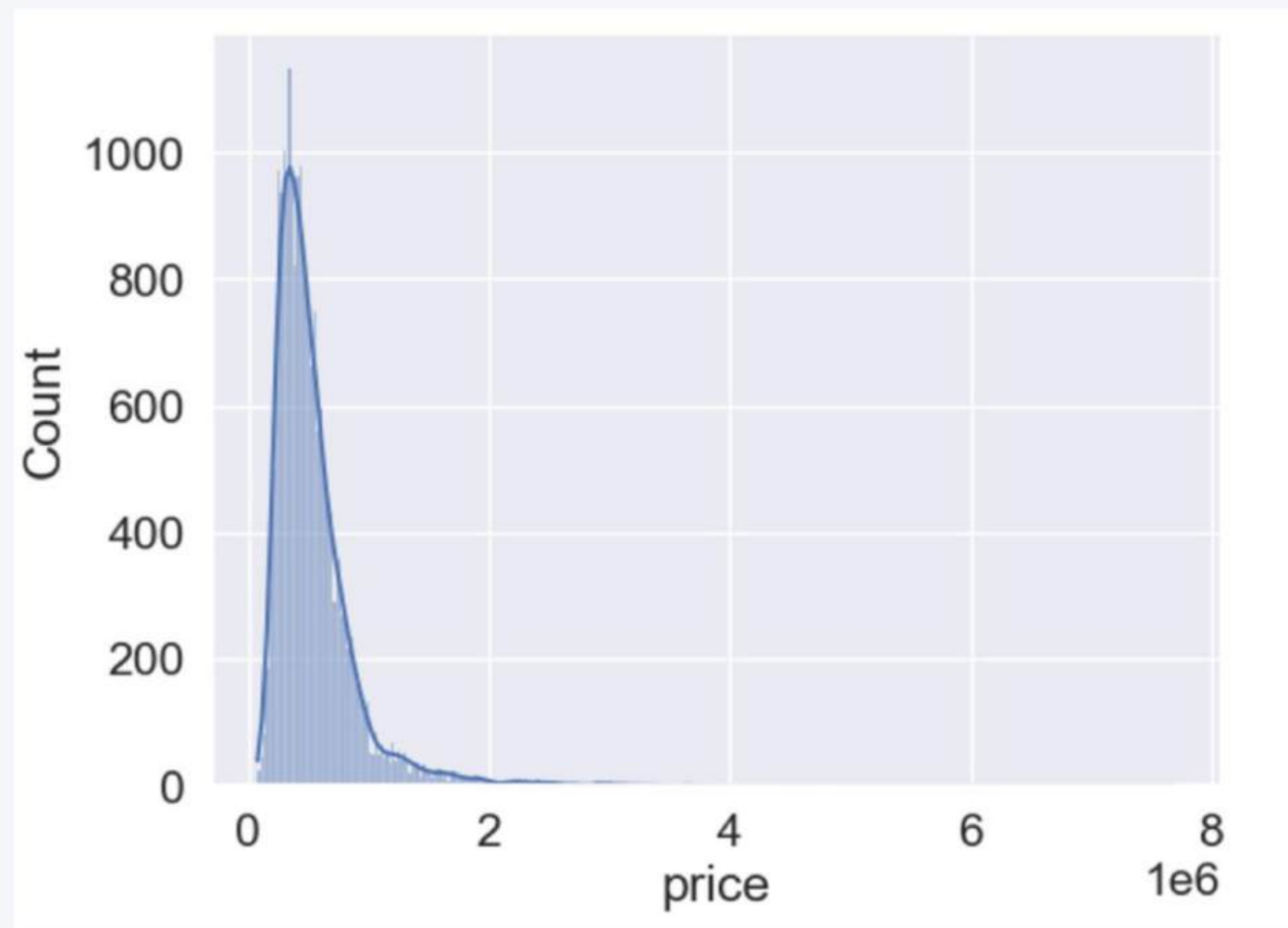


# Chequeo

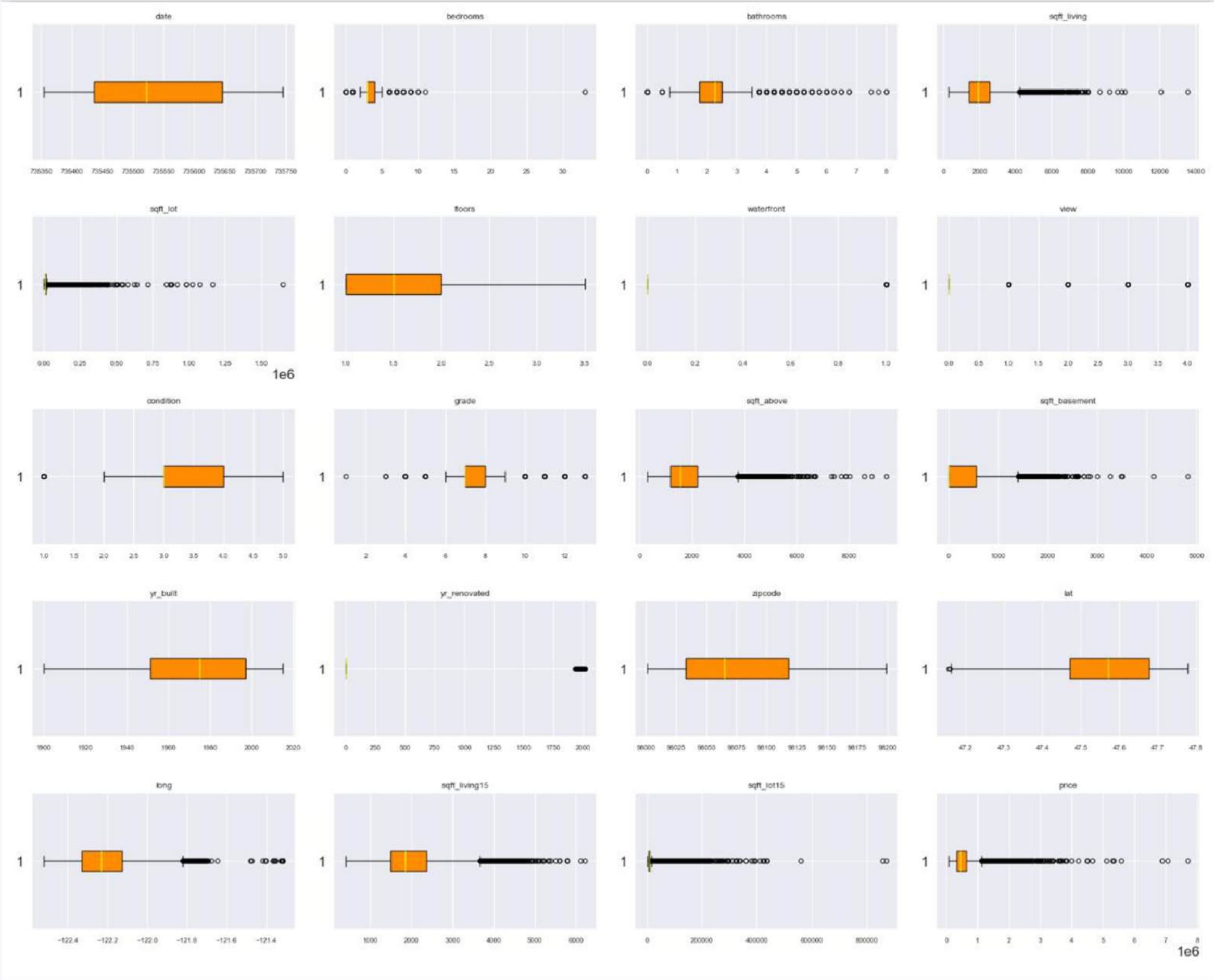
Chequeo de distribuciones



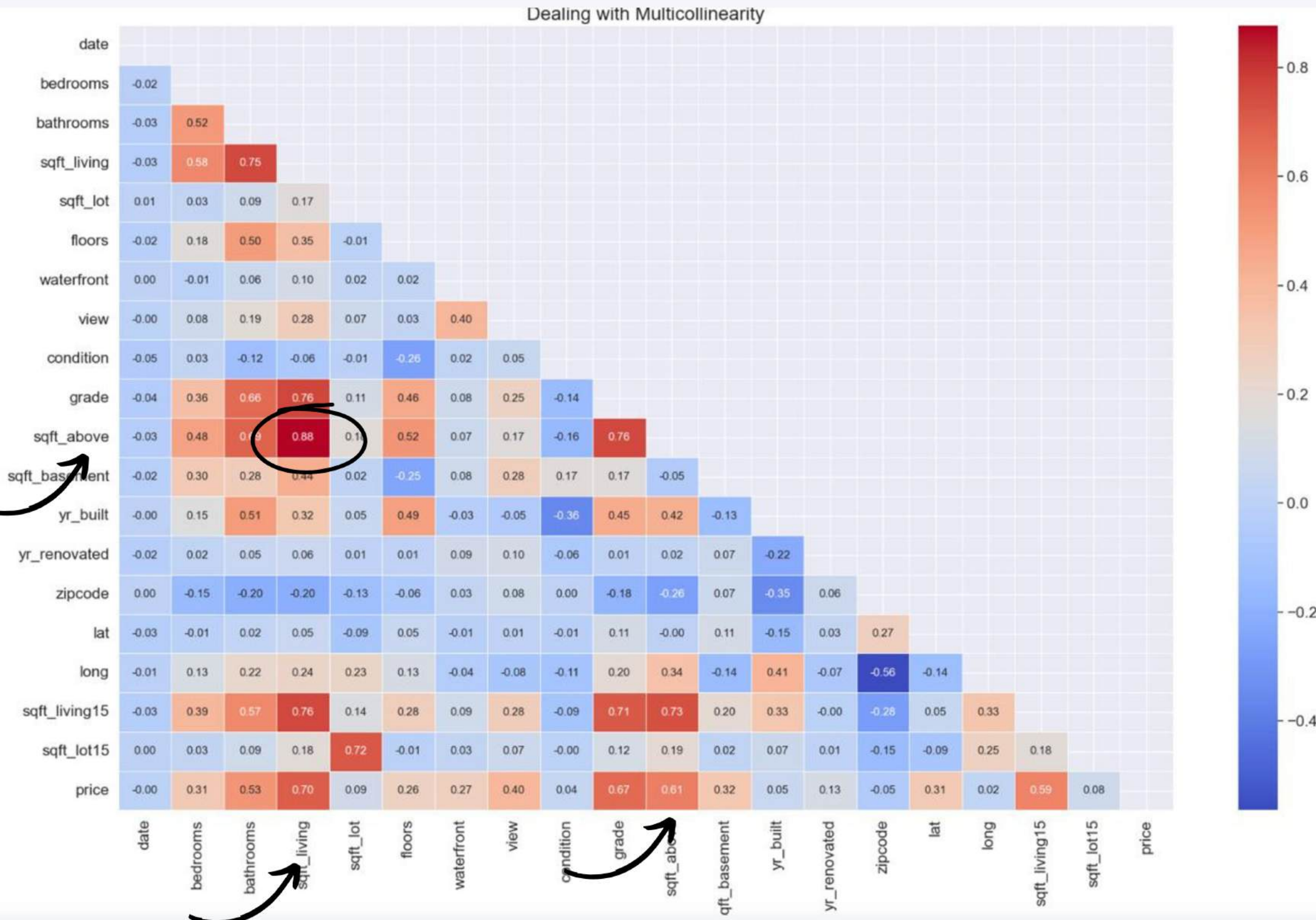
Distribución del precio





[illegible]

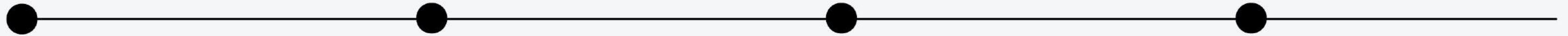
# Hacemos un Heatmap de Correlaciones



Eliminamos sqft\_above  
por relación con  
sqft\_living



# Método Fer



## Modelo Base

Tan solo eliminando la feature que genera multicolinealidad y probamos bastantes modelos de predicción

## Modelos intermedios

Eliminamos outliers muy claros  
Modificamos features:

- sqft living neighbors
- sqft lot neighbors

## Modelos fallidos

Probamos a eliminar features  
Probamos a modificar features:

- Juntar año de construcción y renovación

## Modelo final

Volvemos al modelo con mejores métricas





# Sopa de Modelos

XGBoost RF Regression

Decision Tree Regression

Ridge Regression

**XGBoost Regression**

Elastic Net Regression

Neural Network Regression

Polynomial Regression (degree=2)

Lasso Regression

**Random Forest Regression**

K-Nearest Neighbors Regression

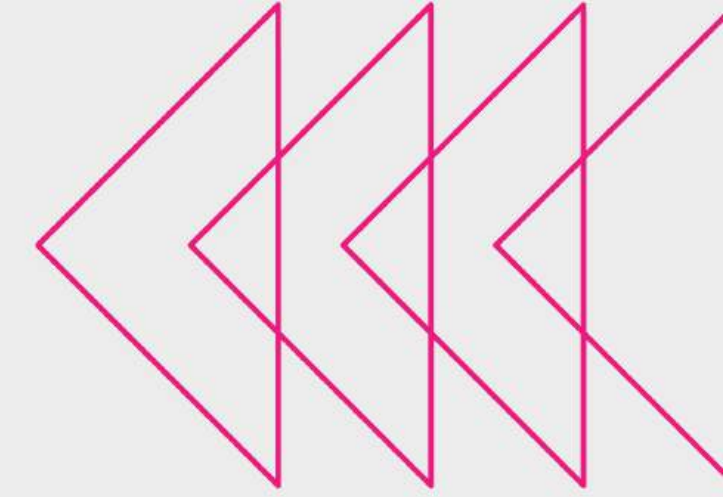
Linear Regression

Bayesian Regression

**Gradient Boosting Regression**

Support Vector Regression

# Base Model



**Gradient  
Boosting**  
**85,7**

**XGBoost**  
**87,4**

**Random  
Forest**  
**85,4**

2

1

3



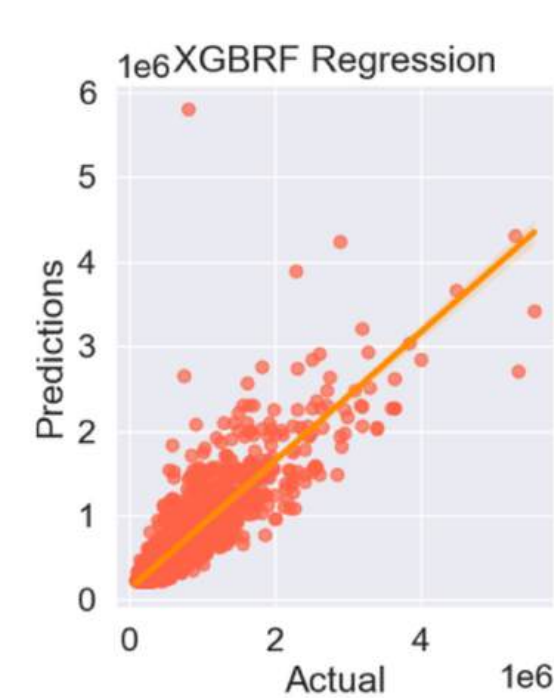
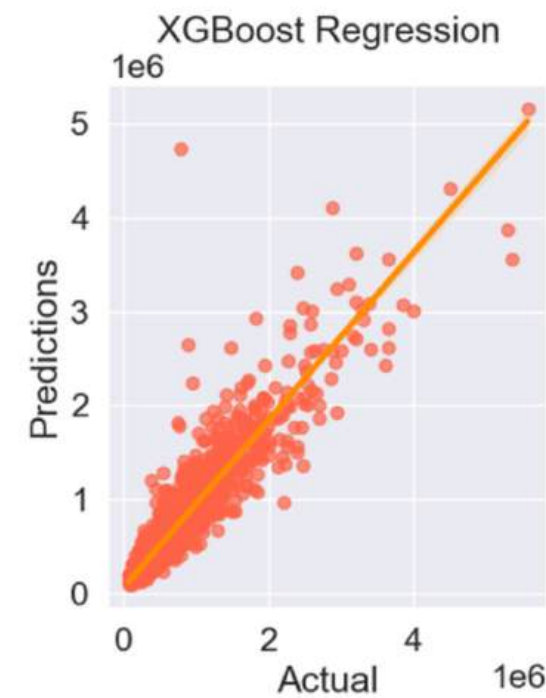
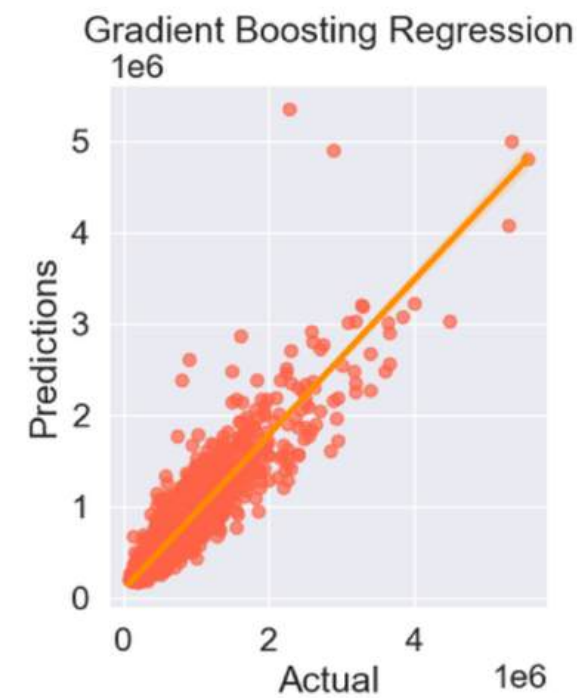
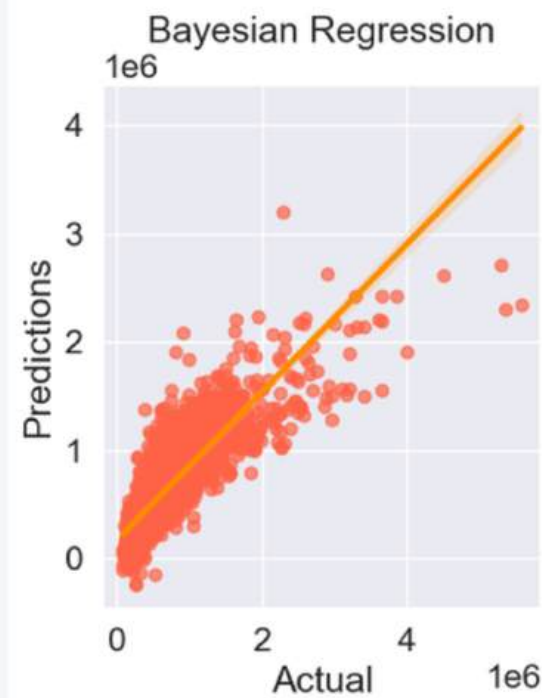
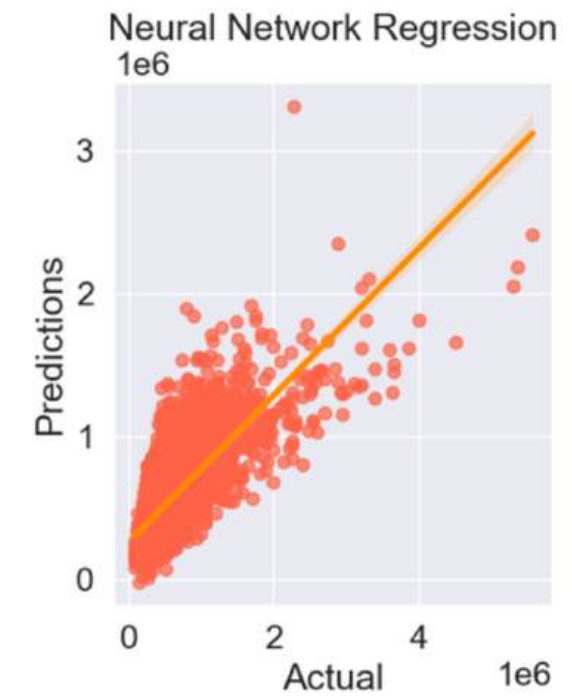
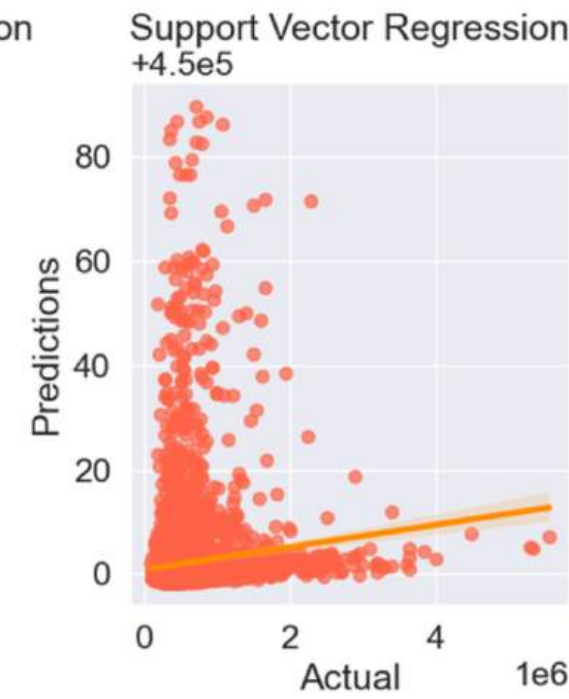
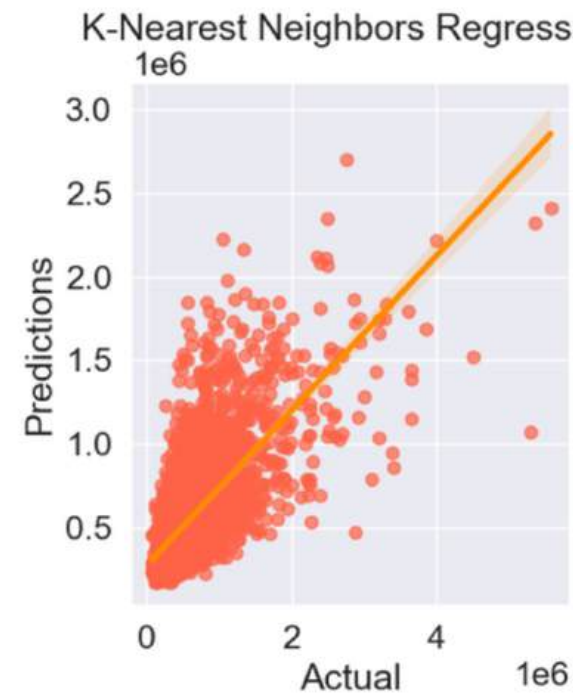
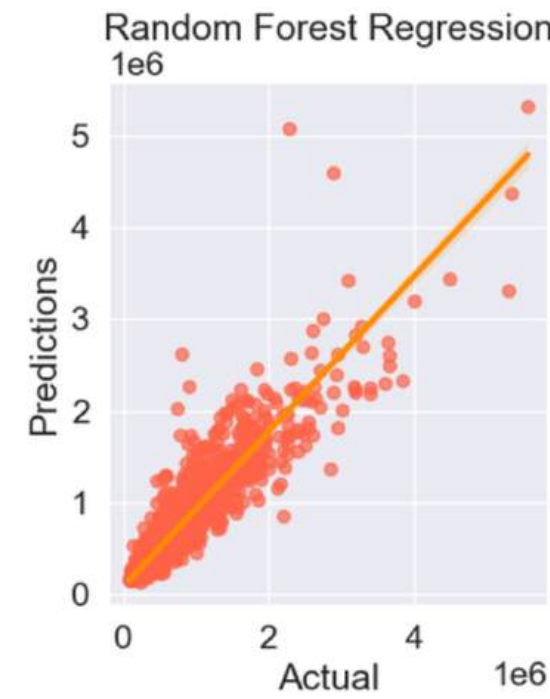
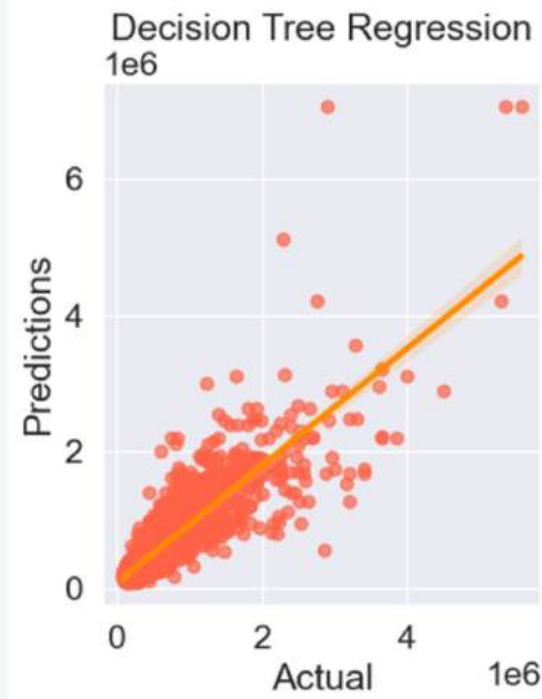
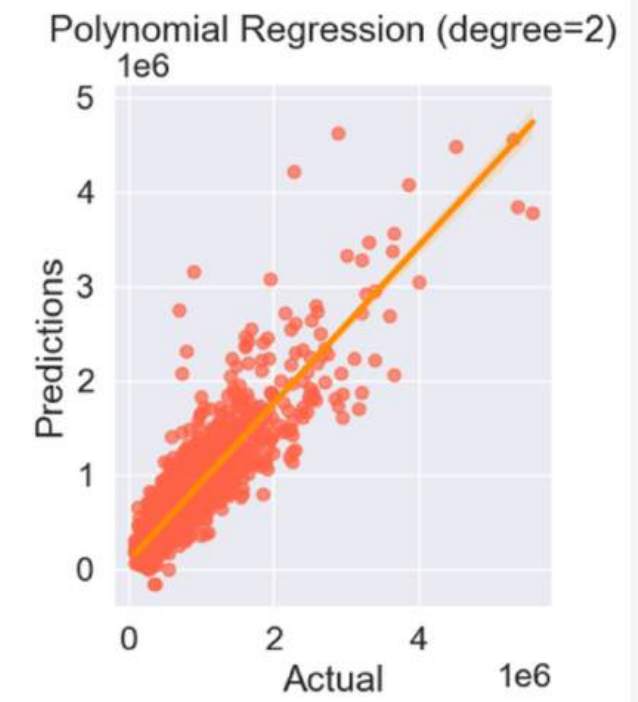
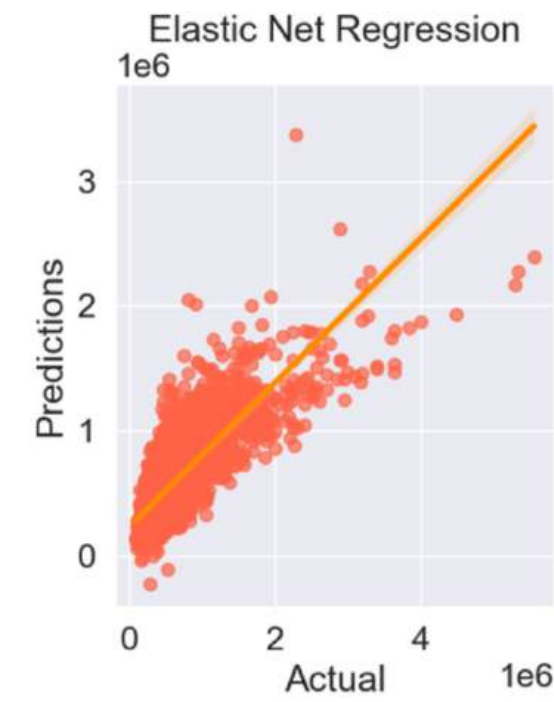
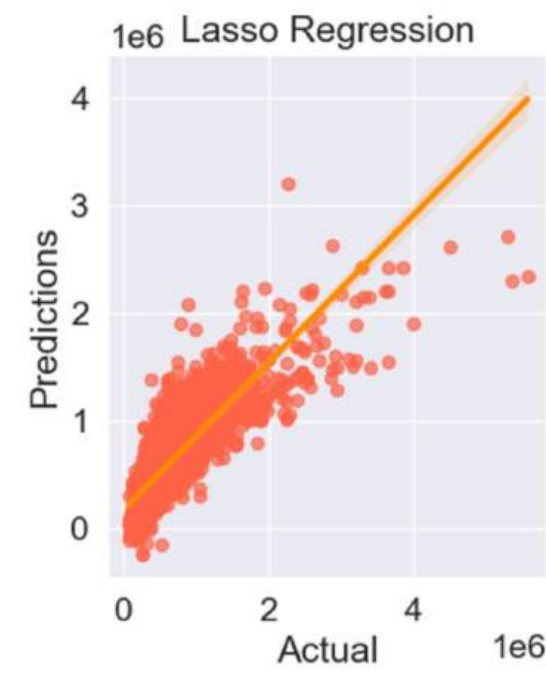
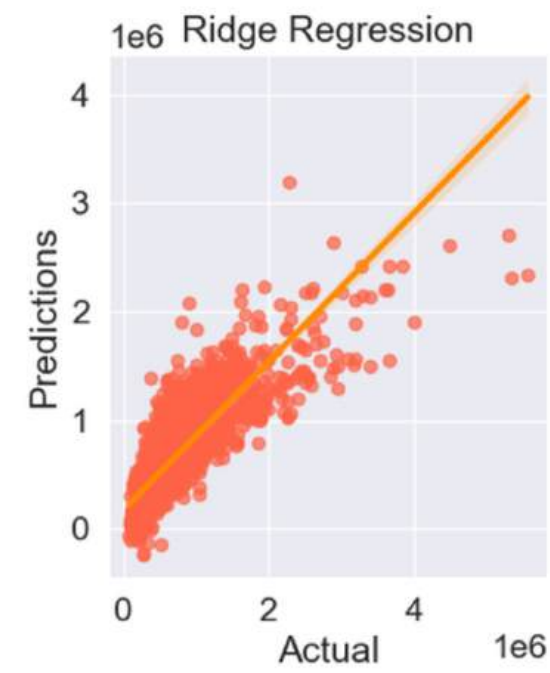
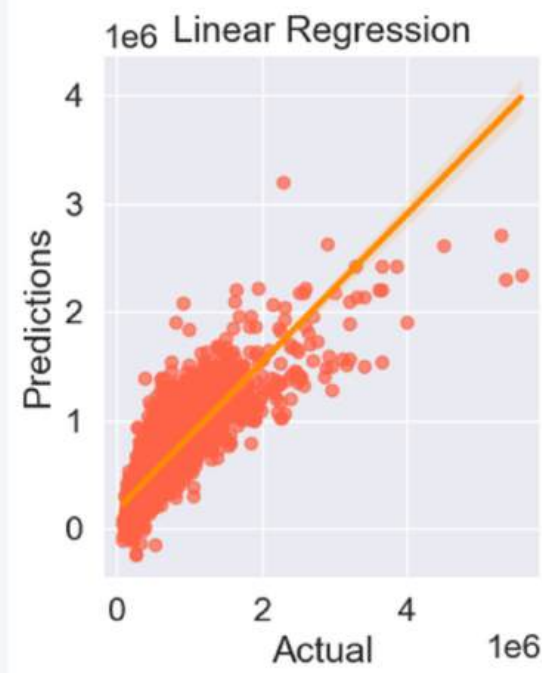


# Regression

Model	R <sup>2</sup>	RMSE	MSE	MAE
Linear Regression	0.7005	207924.99	4.32e+10	127181.99
Ridge Regression	0.7004	207957.99	4.32e+10	127177.60
Lasso Regression	0.7005	207926.16	4.32e+10	127181.60
Elastic Net Regression	0.6115	236829.63	5.60e+10	148877.54
Polynomial Regression (degree=2)	0.8058	167448.28	2.80e+10	102545.23
Decision Tree Regression	0.7290	197795.45	3.91e+10	102487.99
Random Forest Regression	0.8540	145173.51	2.10e+10	73486.95
K-Nearest Neighbors Regression	0.4814	273611.07	7.48e+10	165269.83
Support Vector Regression	-0.0636	391845.82	1.53e+11	226854.95
Neural Network Regression	0.5751	247678.24	6.13e+10	163202.95
Bayesian Regression	0.7004	207958.31	4.32e+10	127177.56
Gradient Boosting Regression	0.8575	143420.99	2.05e+10	80282.63
<b>XGBoost Regression</b>	<b>0.8742</b>	<b>134739.01</b>	<b>1.81e+10</b>	<b>69674.45</b>
XGBoost RF Regression	0.7674	183242.03	3.35e+10	99885.87

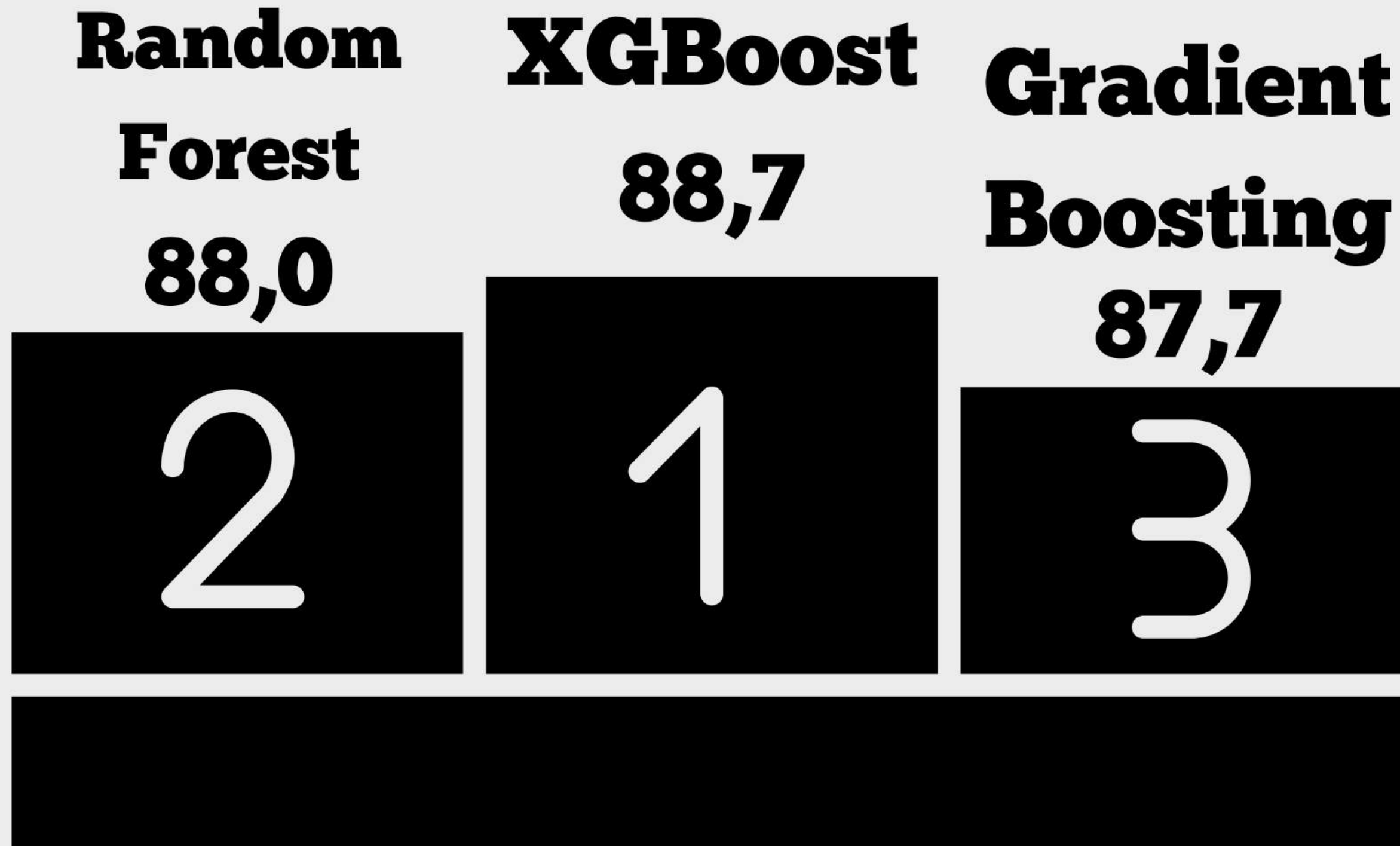








# **Modelo Final**



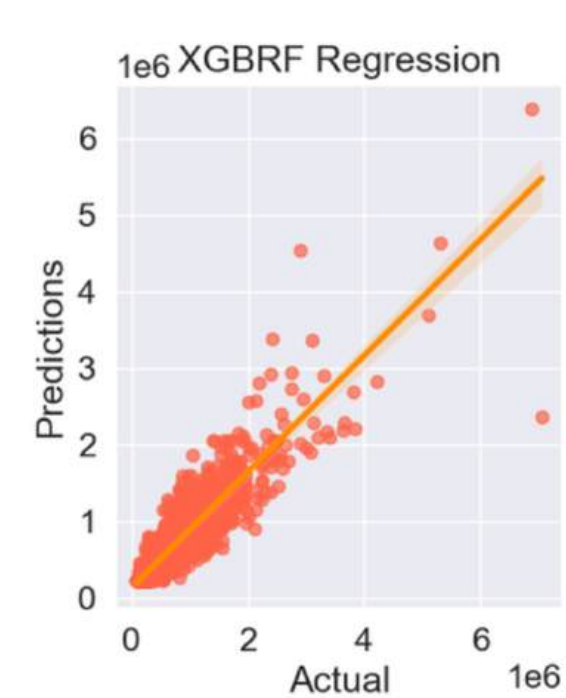
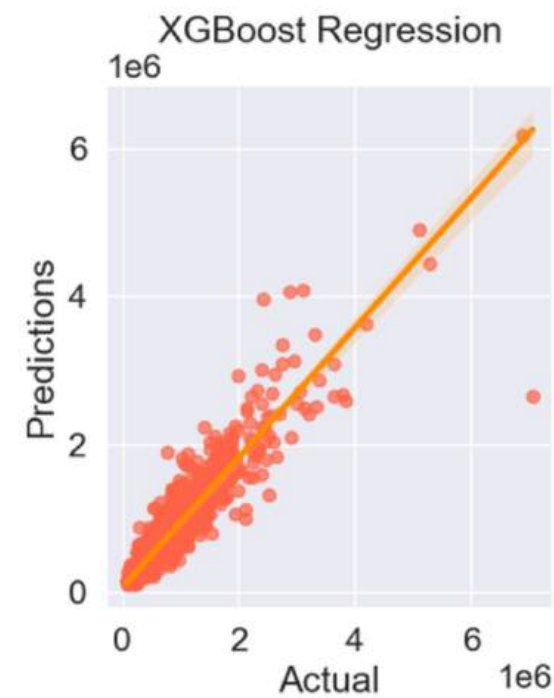
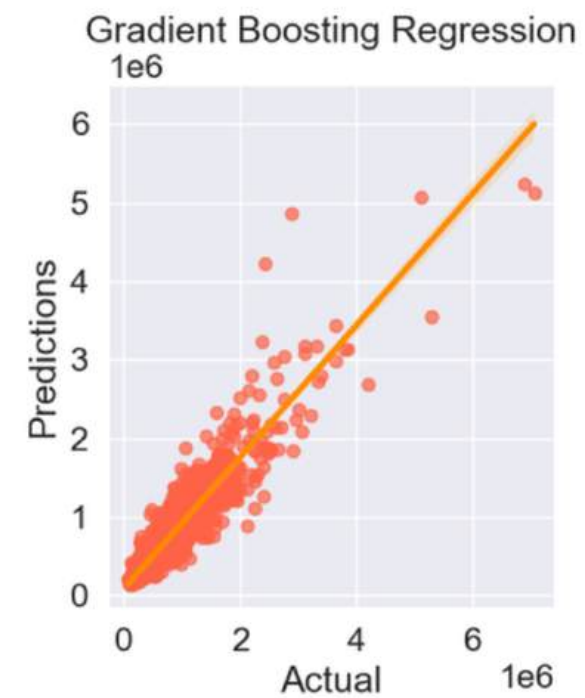
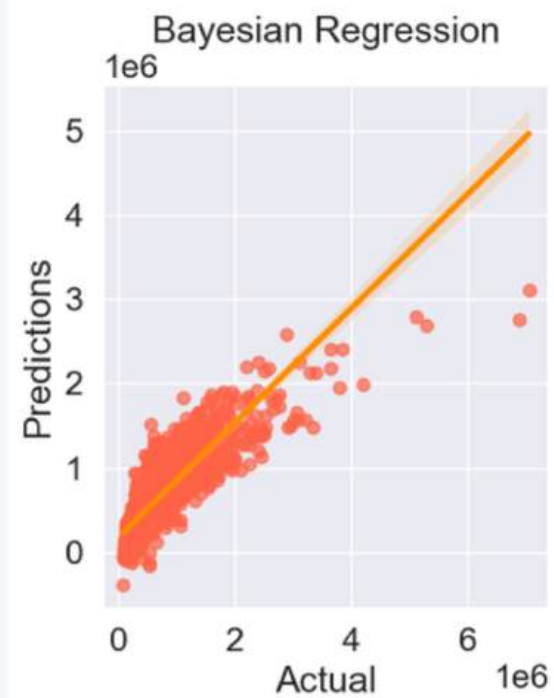
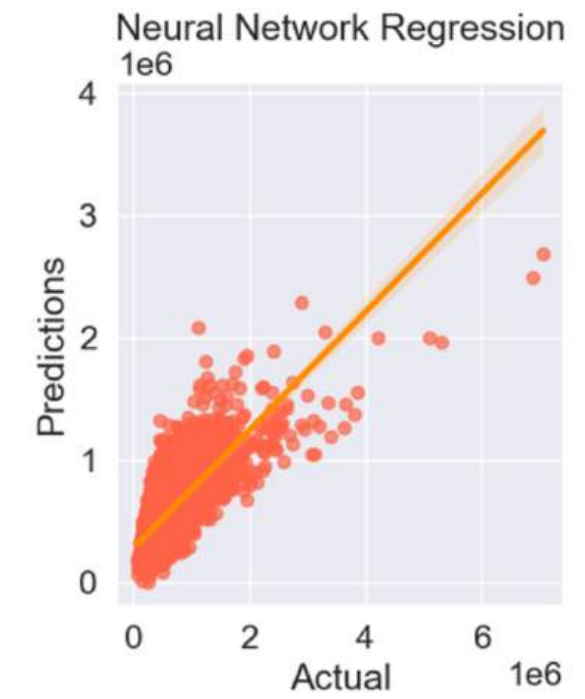
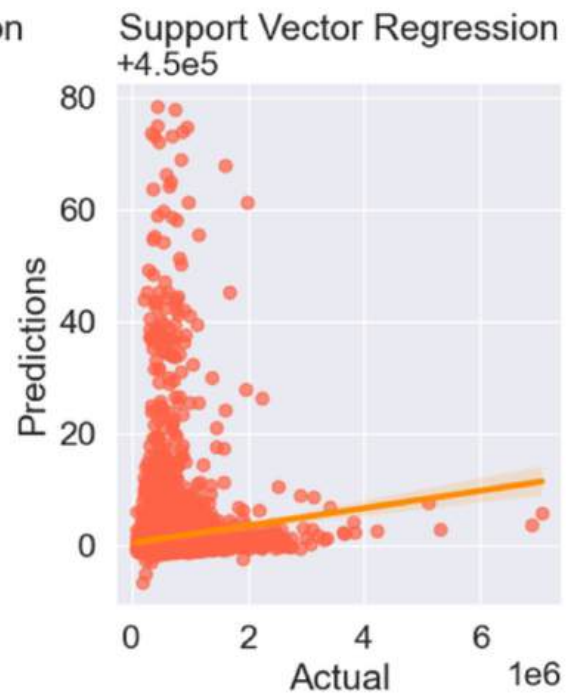
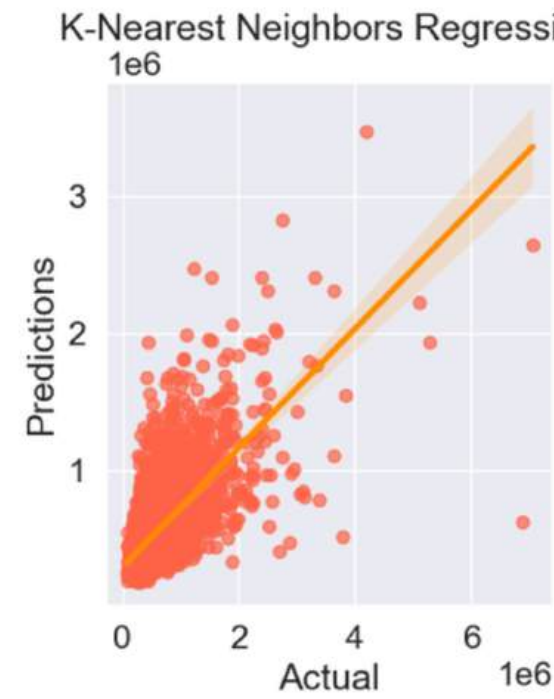
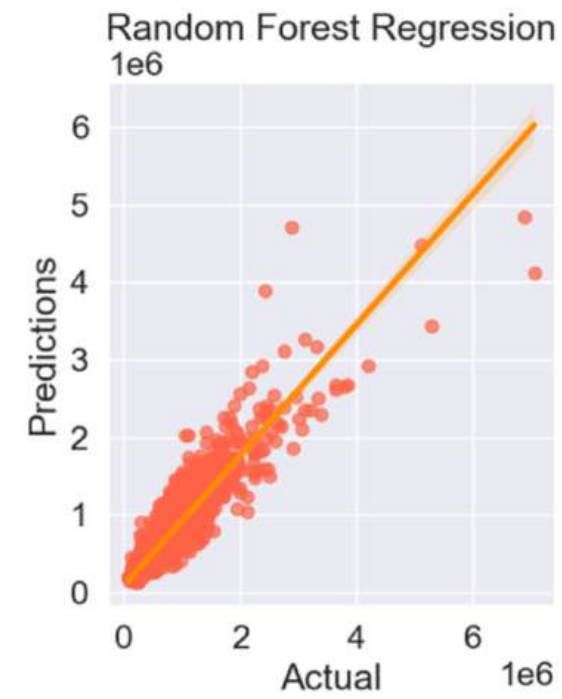
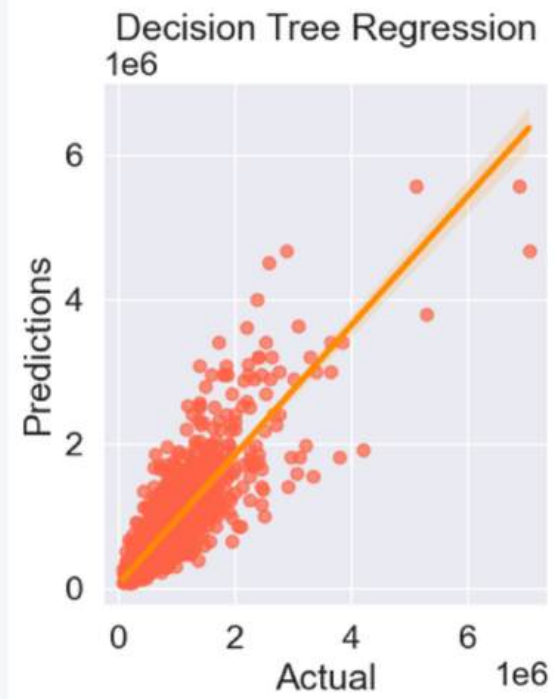
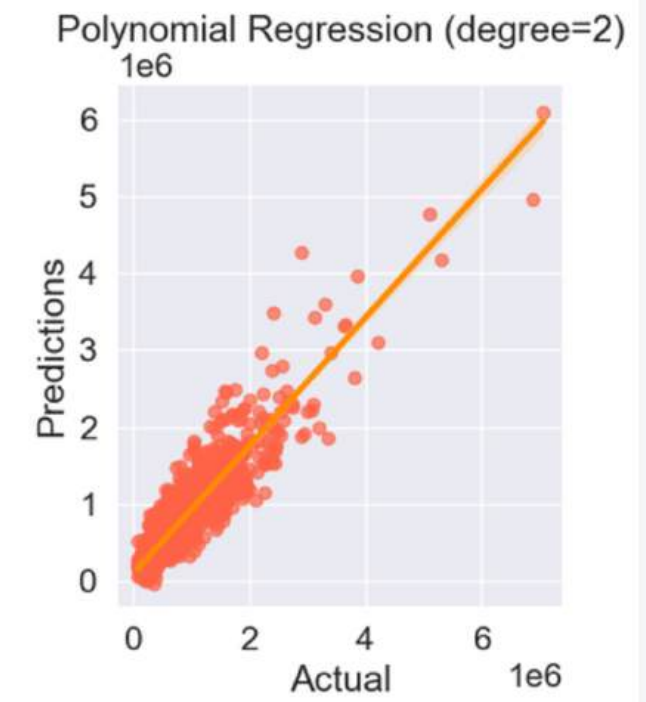
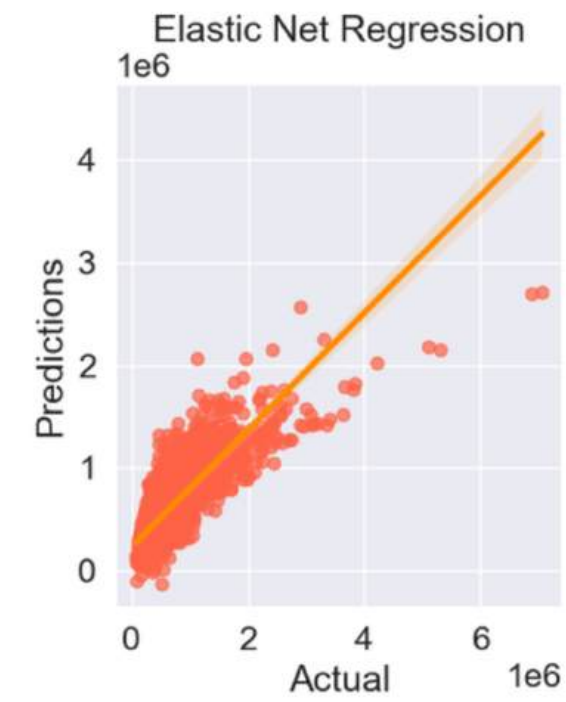
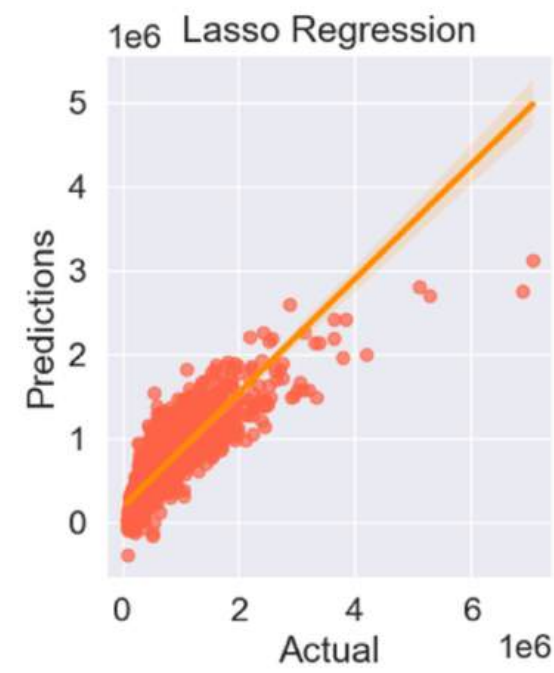
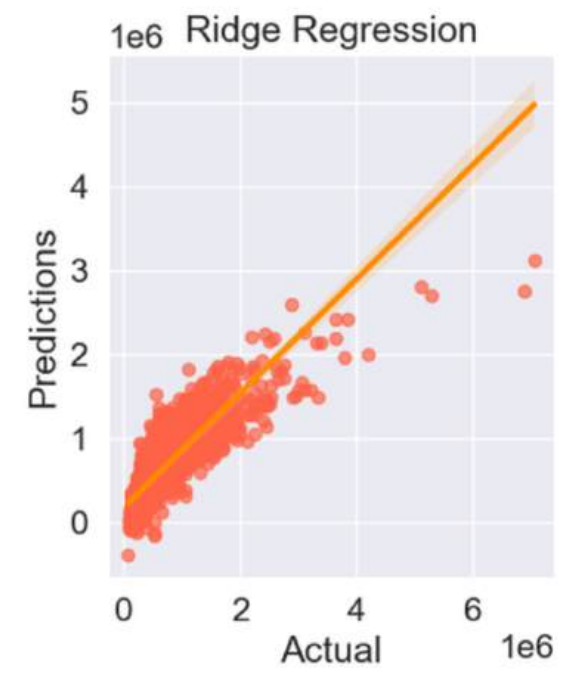
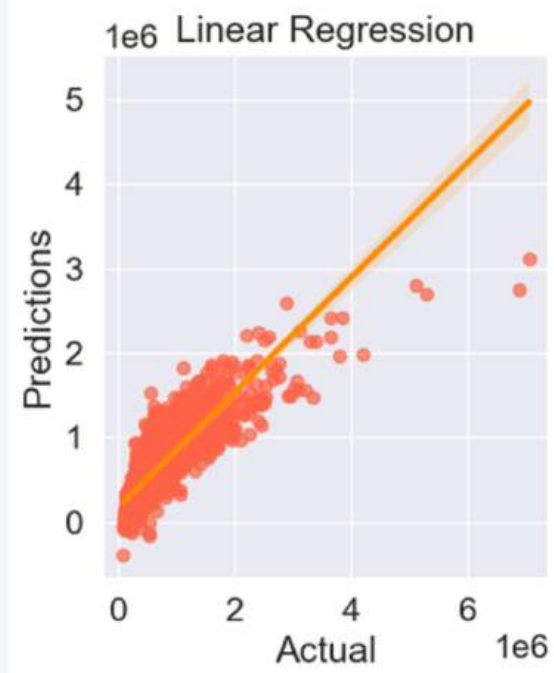
# Regresion

- Creamos `lot_diff_neighbors` y restamos los pies cuadrados del terreno con los pies cuadrados de la media de los 15 vecinos
- Creamos `sqft_diff_neighbors`. y restamos los pies cuadrados de la casa con los pies cuadrados de la media de los 15 vecinos.

Model	R <sup>2</sup>	RMSE	MSE	MAE
Linear Regression	0.7034	206688.50	4.27e+10	126068.94
Ridge Regression	0.7034	206712.84	4.27e+10	126042.28
Lasso Regression	0.7034	206689.29	4.27e+10	126067.88
Elastic Net Regression	0.6187	234376.29	5.49e+10	145328.31
Polynomial Regression (degree=2)	0.8352	154097.73	2.37e+10	100342.39
Decision Tree Regression	0.7387	194012.45	3.76e+10	104091.64
Random Forest Regression	0.8800	131469.00	1.72e+10	69690.92
K-Nearest Neighbors Regression	0.4458	282560.25	7.98e+10	166188.14
Support Vector Regression	-0.0620	391139.20	1.52e+11	224906.23
Neural Network Regression	0.5159	264076.50	6.97e+10	168159.08
Bayesian Regression	0.7034	206712.61	4.27e+10	126042.52
Gradient Boosting Regression	0.8787	132185.32	1.74e+10	75400.25
<b>XGBoost Regression</b>	<b>0.8870</b>	<b>127566.00</b>	<b>1.62e+10</b>	<b>66518.98</b>
XGBoost RF Regression	0.8147	163402.32	2.67e+10	92873.72









# Problemas principales







# Problemas principales



- Usar Funciones
- Correlacionados con la target
- Buscar Modelos
- Año de renovación == 0, cambiar por año de construcción
- Normalización con Log Transform
- Baños





**Gracias!**





# Regresion

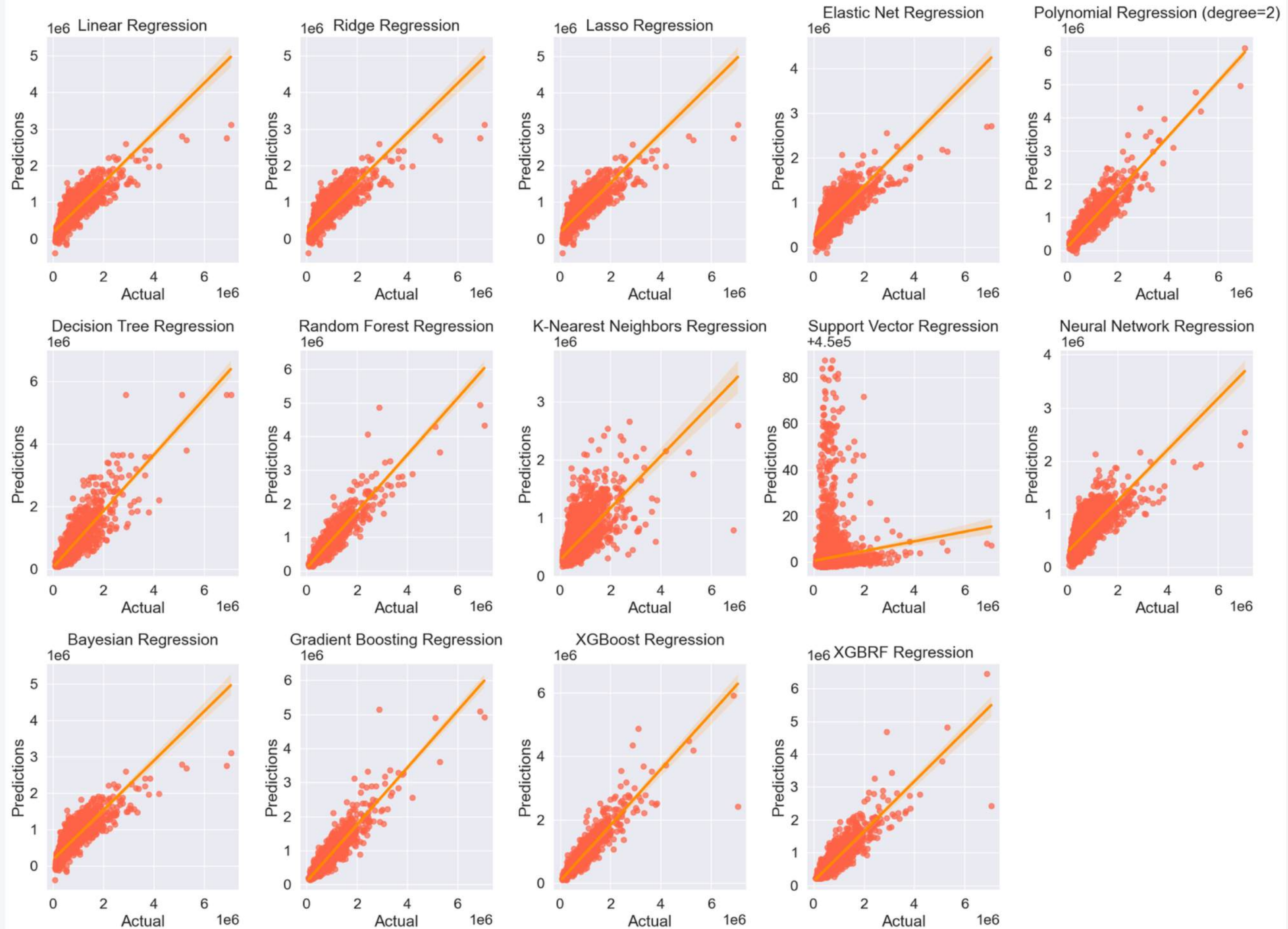
Eliminando:

- 33 habitaciones
- 0 habitaciones y 0 baños

Model	R²	RMSE	MSE	MAE
Linear Regression	0.7034	206688.50	4.27e+10	126068.94
Ridge Regression	0.7034	206712.84	4.27e+10	126042.28
Lasso Regression	0.7034	206689.29	4.27e+10	126067.88
Elastic Net Regression	0.6187	234376.32	5.49e+10	145328.34
Polynomial Regression (degree=2)	0.8351	154147.94	2.37e+10	100477.55
Decision Tree Regression	0.7435	192232.19	3.69e+10	101483.09
Random Forest Regression	0.8809	130993.04	1.71e+10	69446.66
K-Nearest Neighbors Regression	0.4638	277924.38	7.72e+10	160126.65
Support Vector Regression	-0.0620	391138.89	1.52e+11	224906.08
Neural Network Regression	0.4994	268546.82	7.21e+10	169098.55
Bayesian Regression	0.7034	206712.61	4.27e+10	126042.52
Gradient Boosting Regression	0.8743	134584.61	1.81e+10	75994.56
<b>XGBoost Regression</b>	<b>0.8822</b>	<b>130265.68</b>	<b>1.69e+10</b>	<b>66080.33</b>
XGBoost RF Regression	0.8180	161918.88	2.62e+10	92051.23









# Escalado min-max

