

Introduction

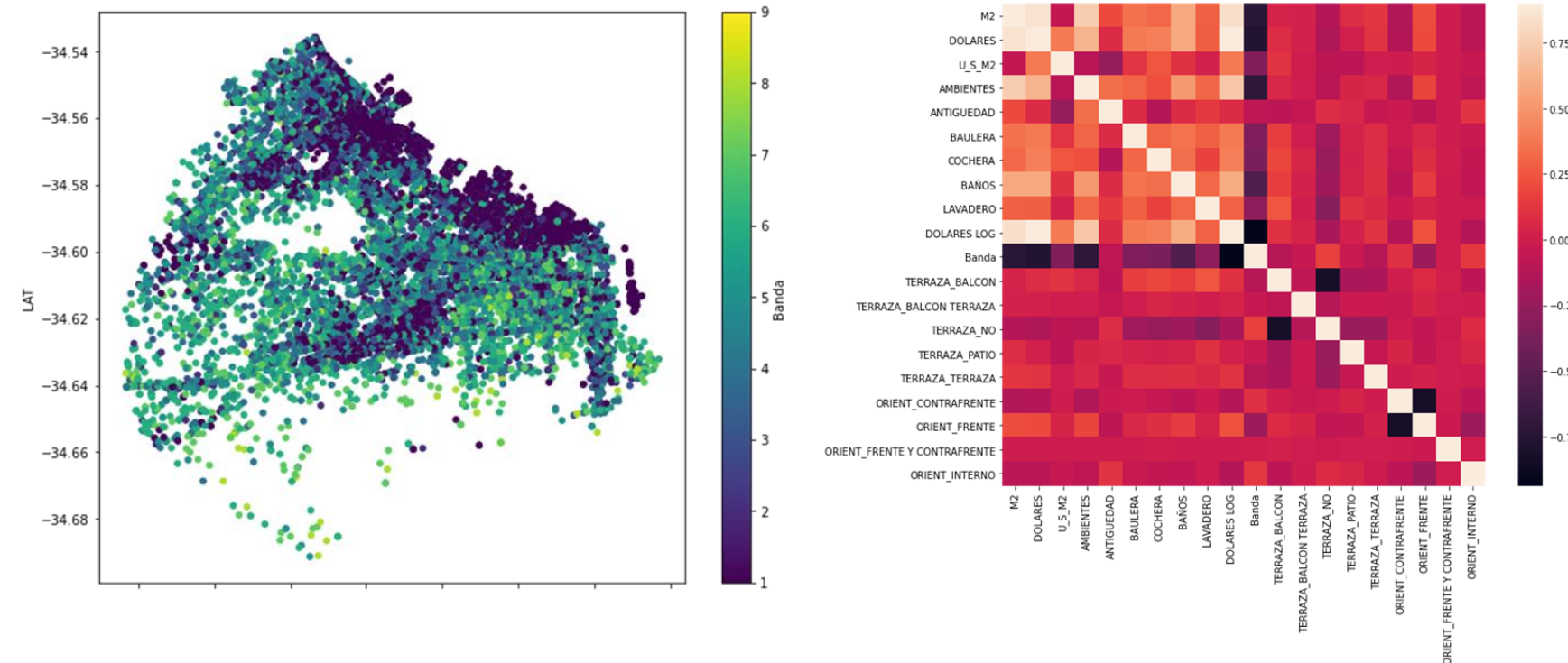
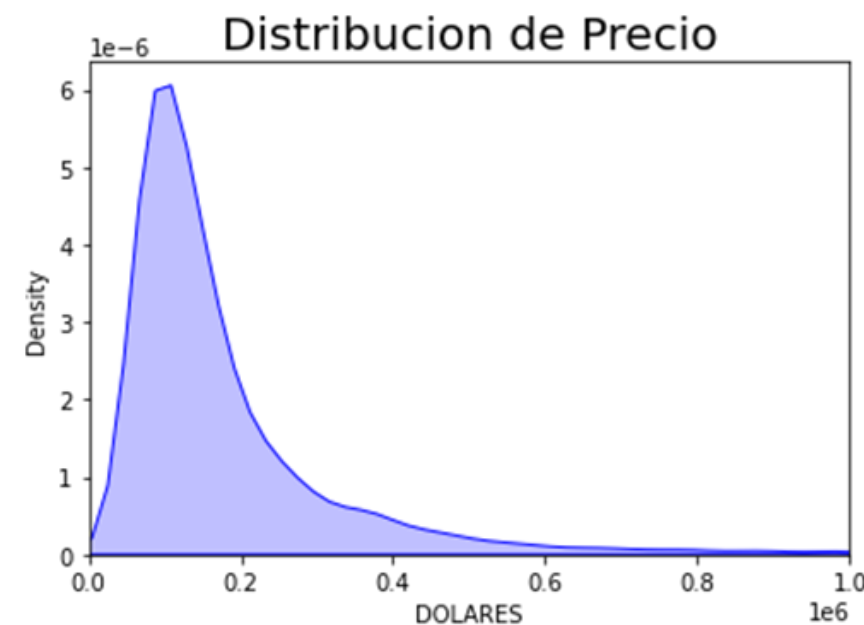
Este proyecto consta de un análisis del mercado inmobiliario en CABA para identificar la correlación entre las diferentes variables disponibles en el dataset y el precio de las propiedades, con el objetivo de estudiar la viabilidad del desarrollo de un modelo que sirva como cotizador de departamentos.

Datasets

El dataset utilizado se encuentra disponible en la página del gobierno de la ciudad, en el apartado perteneciente al Ministerio de Desarrollo Urbano y Transporte, Subsecretaría de Planeamiento, Dirección General de Diagnóstico territorial y Proyección Urbana.

Analysis Exploratorio de Datos

Llevamos a cabo un análisis exploratorio de los datos para así poder comprenderlos con mayor profundidad y completar la limpieza efectuada.



La zona donde se encuentra es determinante a la hora de establecer el precio de una propiedad. Los barrios que cuentan con departamentos en venta de mayor valor son Puerto Madero, Retiro, Recoleta, Palermo y Belgrano

Del heatmap vemos que se destacan las correlaciones entre el precio y la superficie y cantidad de ambientes.

Métodos

El objetivo era desarrollar un modelo de aprendizaje que sea capaz de cotizar un departamento en función de sus características. Para ello trabajamos con los siguientes modelos:

1. Lineal Regressor
2. Support Vectors Regressor (SVR)
3. KNN Regressor
4. Ridge Regressor

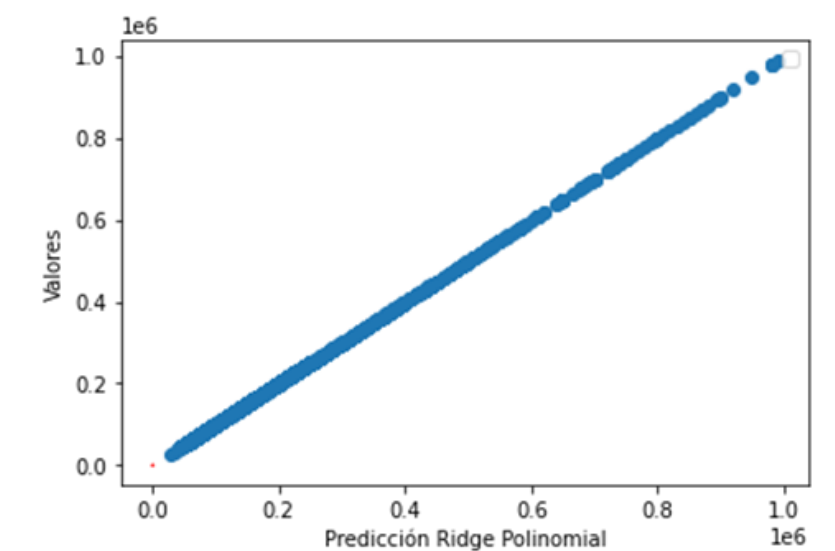
También transformamos nuestra base de datos mediante la herramienta polynomial features para encontrar relaciones no lineales entre las variables.

Resultados

Evaluamos el desempeño en función de el R2 (variabilidad explicada, valor ideal 1), el error cuadrático medio y el error absoluto medio.

	Model	R2	MSE	MAE
0	Regresión lineal	-8.979961e+07	1.411551e+18	1.146365e+09
1	SVR	-4.035431e-01	2.206215e+10	1.290024e+05
2	KNN	6.746505e-01	5.114136e+09	3.974467e+04
3	Ridge	-9.599846e+00	1.666179e+11	3.955118e+05
4	RIDGE Poly	9.999996e-01	5.850210e+03	2.440957e+01
5	KNN Poly	6.746505e-01	5.114136e+09	3.974467e+04
6	RIDGE Poly	9.999997e-01	5.225982e+03	1.676600e+01

El modelo predilecto fue el Ridge Regressor Poynomial que devolvió predicciones de elevada precisión.



Conclusiones

Si bien el modelo demostró gran desempeño, cuenta con sus limitaciones dadas por la antigüedad de la información manipulada (2013) y por la dificultad de seguir el rastro la dinámica del mercado y de tener en consideración variables estéticas no cuantificables.