

Introducción

En el presente trabajo analizaremos datos de Terrenos de la Ciudad de Buenos Aires, que se encuentran a la venta. En base a sus características buscaremos predecir el valor a partir de ellas y veremos también de que manera impacta el entorno en él.

Además analizaremos distintos factores que afectan al valor de un terreno en CABA, y, determinar la posibilidad de predecir en un precio desconocido en base a dichas características.

Datasets

Todos los datasets fueron obtenidos de la página “Buenos Aires Data” (<https://data.buenosaires.gob.ar/>).

Datasets principales:

- Precio de Terrenos 2014
- Precio de Terrenos 2015
- Precio de Terrenos 2016
- Precio de Terrenos 2017
- Precio de Terrenos 2018

Se concatenaron los datasets enumerados para obtener el principal, correspondiente al período 2014-2018.

Datasets adicionales:

- Establecimientos educativos
- Espacios verdes públicos
- Oferta gastronómica
- Espacios culturales
- Delitos
- Estaciones de Bicicletas
- Bocas de acceso al subterráneo.

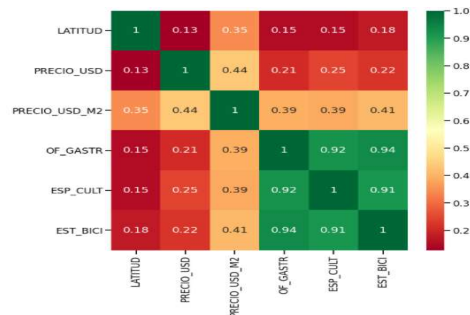
Al agregar esta información, relacionamos cada muestra de valor de terreno con su entorno inmediato, y nos permite analizar el valor del terreno más allá de las consideraciones generales.

Análisis Exploratorio de datos

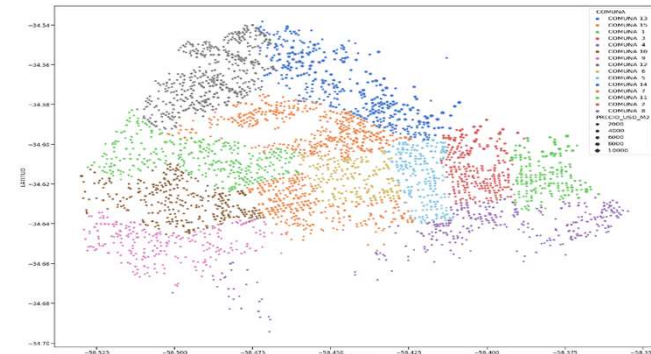
Para llevar adelante el objetivo de este informe, se realizó un relevamiento sobre el dataset armado.

En esta revisión, se buscó analizar:

Relación de las features con el precio del metro cuadrado en dólares



Distribución de la muestra



Resultados

COMPARACIÓN DE MODELOS DE REGRESIÓN

	Modelo	R2	MSE	RMSE	MAE
0	LR	0.349623	1.172242e+06	1082.701094	737.261813
1	RR	0.342877	1.184401e+06	1088.302043	737.025688
2	SVR	0.435441	1.017563e+06	1008.743219	623.955106
3	KNNR	0.522206	8.611771e+05	927.996285	547.958489

Conclusiones

- Es destacable la correlación lineal que presenta el valor del terreno (usd/m2) con Estaciones de bicicletas, Oferta gastronómica y Espacios culturales. Esto demuestra la incidencia positiva sobre el valor del terreno, cuando en su entorno geográfico existen espacios sociales y recreativos a donde la gente puede concurrir, en conjunto con la accesibilidad y movilidad.
- A pesar de haber seleccionado al modelo **KNN Regression** como el de mejor rendimiento para lograr la predicción de valor, consideramos que la precisión aún se puede mejorar, y el error por consiguiente reducirse. Para lograr este cometido, consideramos que sería oportuno el agregado de nuevas features que aporten información, como por el ejemplo el valor de FOT. Otra posibilidad sería la suma de nuevas muestras al entrenamiento y al testeo.

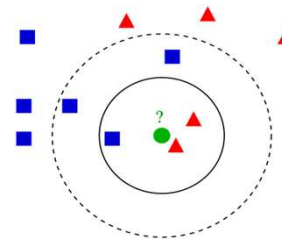
Método

KNN Regression

Es algoritmo de regresión donde durante el entrenamiento se determinan los K vecinos más cercanos por distancia euclídea (distancia par a par).

El yi a predecir se determina por la interpolación de los y en los vecinos.

Los pesos w indican cómo se interpolara cada K vecino : uniforme o por distancia.



$$d(x_a, x_b) = \sqrt{(x_{a1} - x_{b1})^2 + (x_{a2} - x_{b2})^2 + \dots + (x_{ap} - x_{bp})^2}$$