

## Entregable 2 - Ejercicio 5:

### Criterios de exclusión de ejemplos

1. Se eliminan registros que pertenezcan a 'Suburb', quedándonos con los 30 suburbios que aparecen con mayor frecuencia.

### Características seleccionadas

#### Características categóricas

1. Type: tipo de propiedad. 3 valores posibles.
2. Suburb: suburbio donde está ubicada la propiedad. 30 valores posibles.

Todas las características categóricas fueron codificadas con un método OneHotEncoding.

#### Características numéricas

1. Rooms: cantidad de habitaciones.
2. Distance: distancia al centro de la ciudad.
3. Price: precio de compra/venta de la propiedad.
4. Postcode: código postal.
5. Bedroom2: cantidad de habitaciones con cama.
6. Bathroom: cantidad de baños.
7. Car: cantidad de lugares para estacionar autos.
8. Landsize: tamaño del terreno.
9. BuildingArea: metros cuadrados construidos.
10. YearBuilt: año de construcción.
11. Latitude: latitud
12. Longitude: longitud
13. environment\_price: se agrega el precio promedio diario de publicaciones de la plataforma AirBnB en el mismo código postal. [[Link al repositorio con datos externos](#)].

#### Transformaciones:

1. Todas las características numéricas fueron estandarizadas usando MinMaxScaler.
2. Las columnas 'YearBuilt' y 'BuildingArea' fueron imputadas utilizando el algoritmo IterativeImputer, con el estimador KNeighborsRegressor.

3. Se aplicó el método PCA para reducir la dimensionalidad y se calculó la varianza de las primeras 20 componentes.

## Datos aumentados

1. Se agrega la primera columnas obtenidas a través del método de PCA, aplicado sobre el conjunto de datos totalmente procesado.