

Criterios de exclusión/selección de ejemplos:

Luego de haber realizado un análisis exploratorio, el Df quedo conformado por las variables que se consideró afectarían más al precio de las propiedades, para ello tuvimos en cuenta la correlación entre las distintas variables y Price.

Otras variables como "Adress" o "Council Area" fueron excluidas por que se consideró que aportarían información brindada por otras si consideradas.

Características seleccionadas:

Características categóricas:

- **Type:** tipo de propiedad. Puede tomar 3 valores posibles.
- **Regionname:** Nombre de la Región donde se encuentra la propiedad.
- **Suburb:** Nombre del barrio donde se encuentra la propiedad. Esta variable fue seleccionada en un principio porque serviría para comparar semejanzas en el precio de un mismo barrio, pero como puede tomar 312 valores únicos es que finalmente se decidió excluirla.

Las características categóricas menos Suburb fueron codificadas con el método One Hot Encoding.

Características numéricas:

- **Rooms:** Cantidad de habitaciones.
- **Price:** Precio de venta de la propiedad,
- **Postcode:** Código Postal del conjunto de datos de Melbourne, esta fue seleccionada porque serviría para poder realizar la unión con el Df de Airbnb.
- **Bathroom:** Cantidad de baños.
- **Car:** Cantidad de cocheras.
- **Landsize:** Área ocupada por el terreno.
- **Building Area:** Área construida de la propiedad.
- **YeartBuilt:** Año de construcción de la propiedad.
- **Zipcode:** Código Postal de el conjunto de datos de Airbnb, esta variable se seleccione para poder unir los dos Df.
- **airbnb price mean:** precio promedio diario de las publicaciones del conjunto de Airbnb.
- **airbnb record count:** Cantidad de publicaciones con el mismo código postal.
- **airbnb weekly price mean:** Precio promedio semanal del conjunto de Airbnb.
- **airbnb monthly price mean:** Precio promedio mensual del conjunto de Airbnb

Las características numéricas Rooms, Bathroom y Car fueron codificadas con el método One Hot Encoding.

Transformaciones:

1. Todo el Dataframe fue escalado con el método MinMaxScaler entre (-1,1).
2. Se imputaron tanto variables categóricas como numéricas con el método KNeighborsRegressor.

Datos aumentados:

Se agregaron las 2 primeras columnas obtenidas a través del método de PCA, aplicado sobre el conjunto de datos totalmente escalado y con los valores faltantes imputados.