**Información relevante**

​Para este proyecto se realizará un modelo que se aplique para el mercado de bienes raíces de Colombia, para ello se utilizaran las columnas con mayor correlación con la columna target la cual fue creada a partir de la media del precio de las distintas propiedades otorgándole 1 a las propiedades caras y 0 a las baratas

**Descripción del problema**

​Para predecir si el precio de un inmueble es barato o caro se usará un modelo de aprendizaje supervisado random forest classifier el cual funciona bien si ajustar los hiperparametros y con problemas de clasificación y reduce el riego de overfiting

**Entrega:**

​\*Un archivo model.ipynb

\*Un archivo dianece03.csv

​**Métrica a utilizar**

​Como método de evaluación del desempeño del modelo, se utilizará la métrica de Exhaustividad (Recall) para las propiedades caras, a partir de la matriz de confusión (Confusion Matrix).​$$ Recall=\frac{TP}{TP+FN}$$​Donde TP son los verdaderos positivos y FN los falsos negativos.

Adicionalmente, se incluye la Accuracy como métrica de control.​

**Archivos provistos**

​Se proveen los archivos dentro del archivo comprimido 'properties\_colombia.zip':

* 'properties\_colombia\_train.csv': Contiene 197549 registros y 26 dimensiones, el cual incluye la información **numérica** del precio.
* 'propiedades\_colombia\_test.csv': Contiene 65850 registros y 25 dimensiones, el cual no incluye la información del precio.​

**Descripción de las dimensiones**

* id - Identificador del aviso. No es único: si el aviso es actualizado por la inmobiliaria (nueva versión del aviso) se crea un nuevo registro con la misma id pero distintas fechas: de alta y de baja.
* ad\_type - Tipo de aviso (Propiedad, Desarrollo/Proyecto).
* start\_date - Fecha de alta del aviso.
* end\_date - Fecha de baja del aviso.
* created\_on - Fecha de alta de la primera versión del aviso.
* lat - Latitud.
* lon - Longitud.
* l1 - Nivel administrativo 1: país.
* l2 - Nivel administrativo 2: usualmente provincia.
* l3 - Nivel administrativo 3: usualmente ciudad.
* l4 - Nivel administrativo 4: usualmente barrio.
* l5 - Nivel administrativo 5.
* l6 - Nivel administrativo 6.
* rooms - Cantidad de ambientes.
* bedrooms - Cantidad de dormitorios (útil en el resto de los países).
* bathrooms - Cantidad de baños.
* surface\_total - Superficie total en m².
* surface\_covered - Superficie cubierta en m².
* price - Precio publicado en el anuncio.
* currency - Moneda del precio publicado.
* price\_period - Periodo del precio (Diario, Semanal, Mensual)
* title - Título del anuncio.
* description - Descripción del anuncio.
* property\_type - Tipo de propiedad (Casa, Departamento, PH).
* operation\_type - Tipo de operación (Venta).
* geometry - Puntos geométricos formados por las coordenadas latitud y longitud.​