EXPLORACIÓN PROPERATTI DATASET

1. Importamos bibliotecas
2. Importamos archivo e inspeccionamos dataset
3. Renombramos columnas para evitar confusiones cuando trabajemos con los datos creando un diccionario (recordemos que se pueden duplicar values pero no keys en este objeto)
4. Elimino columnas en donde gran parte de sus valores son nulos identificando cuales son esos campos antes de proceder a eliminar los registros duplicados. Elimino las siguientes columnas:
   1. ID
   2. Operación, era la misma para todos los casos
   3. País, todo corresponde a Argentina
   4. URL
   5. Imagen
   6. Expensas, pocos registros poblados
   7. Piso, abundan los nulos
5. Se detectaron 5251 registros duplicados luego de acotar el dataframe a 20 columnas
6. Eliminamos 5251 registros (representa 4,3%) un y nos queda un dataframe de 115.969 registros
7. Evaluación de columnas por separado, intentaremos completar valores nulos/no existentes por medio de expresiones regulares a partir del título y la descripción, o a través de otros métodos. A su vez, debemos chequear si existen outliers para poder analizarlos. Completamos los barrios que estaban vacios que se daba solo en Tigre con 23 casos
8. Controlamos columnas de superficies para detectar los casos con errores o outliers
9. Completamos el campo superficie\_total\_m2 con el valor del campo precio\_por\_m2\_cubierto en caso de que el primero sea nulo
10. Vamos a hacer la verificación de las columnas superficie cubierta vs superficie total y en caso que superficie cubierta > total, entonces pisaremos esa columna
11. Vamos a hacer la verificación de las columnas ambientes detectando también cantidad de outliers. Boxplot de ambientes para ver gaficamente los outliers
12. Verificación de tipo de cambio en ese momento
13. Elimino los 3 registros en donde la moneda no es dolar o pesos argentinos
14. Ambientes (Rooms), caculo media y desvío std para saber cantidad de outliers. Son 709 outliers y además el 60% de los registros son nulos. Buscamos en los campos description y title aquellos números o palabras que hacen referencia a ambientes, cuartos, habitaciones y generamos expresiones regulares. Tambien se encontraron registros de hoteles los cuales vamos a identificar con expresión regular y los vamos a eliminar del dataframe. Para poder tener la cantidad de ambientes le sumamos 1 a todos aquellos valores que encontramos. Se consiguieron completar mas de 30.000 registros que antes estaban nulos a partir de las regex mencionadas anteriormente
15. Genero dummies para conocer características y amenities y luego veo esto a través de pivot\_table para saber los porcentajes por zona geográfica
16. Gráficos para ver valores M2 por tipo de propiedad y barrio de CABA llegando la misma conclusión que Wendy
17. Analsis de columna relacionada a precios, corregimos localidades y barrios en donde los valores son erróneos (Villa Ballester, Aldo Bonzi, Boedo, San Cristobal)
18. Completamos precio en pesos y dólares a partir de precio M2 y superficie total M2
19. Realizamos histograma de precios USD M2 para ver la distribución de los valores del dataframe