Laboratorio 3

Presentado por: Javier Ricardo Valderrama Gonzalez - 98001, Julián David Cifuentes Gonzalez - 100765

Objetivo del Laboratorio

El objetivo principal de este laboratorio es experimentar con las actividades relevantes que propone IBM para la fase de "comprensión de los datos" (data understanding) dentro de la metodología CRISP-DM (Cross Industry Standard Process for Data Mining). Esta fase incluye la recolección inicial de datos, su descripción, exploración y verificación de calidad.

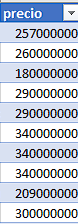
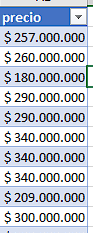
Introducción

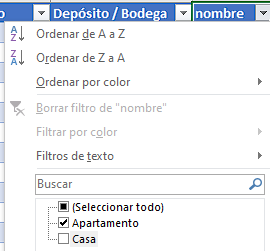
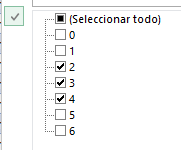
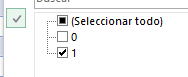
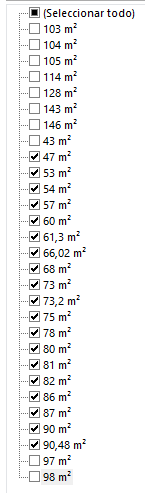
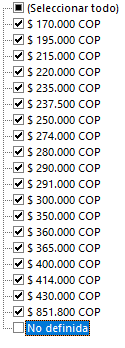
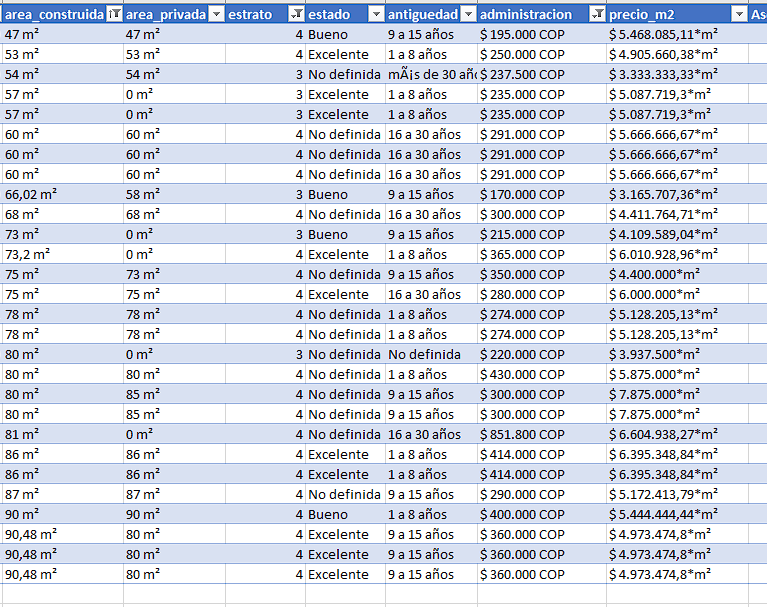
En el contexto actual de análisis de datos y big data, la limpieza de datos se ha convertido en una tarea crucial para garantizar la precisión y relevancia de los resultados obtenidos. Este trabajo se centra en la limpieza y análisis de un conjunto de datos que contiene información sobre apartamentos. El objetivo es identificar aquellos que cumplen con ciertos criterios específicos, que son de particular interés para un análisis detallado.

Nos centraremos en los apartamentos que sean de estrato 2 al 4 que cuenten con balcón parqueadero para visitantes y tengan estudios aparte de ello que sean mayores o iguales a 45 m2 y menores a 95 m2 de área construida.

Desarrollo

1. abrir la información en excel
2. seccionar los datosa en columnas ya que están delimitados por comas (csv)
3. seleccionar los datos y pasarlos a formato tabla
4. arreglamos los enunciado con tildes y con la letra ñ, como
   1. porteria / recepcion
   2. baños
   3. salon comunal
   4. balcón, etc..
5. seleccionamos las columnas area\_construida y area\_privada y reemplazar el carácter “mÂ²” por “m²”
6. en la columna antigüedad hacemos el campo del carácter “Ã±” por la ñ
7. en la columna de administración precio y precio\_m2 arreglamos el símbolo de pesos ($) quitando “Â” o ya sea añadiendolo como en el caso de precio

1. Una vez terminando filtramos la información solo por los apartamentos según el contexto del problema que planteamos.
   1. Filtramos solo los apartamentos.  
      
   2. Ahora filtramos los estratos del 2 al 4.  
      
   3. Filtramos que tenga Parqueadero para visitantes, Balcón y Estudio.  
      
   4. y por último filtramos los apartamentos que cumplan con los metros cuadrados solicitados que en este caso son entre 45 a 95 m2.  
      
   5. de aquí en adelante podemos hacer filtros un poco más sofisticados como el tema de la administración el descartar los apartamentos que sí tengan un valor de administración definida.  
      
   6. Con esto tenemos un total de 28 apartamentos encontrados donde también podremos filtrar por ubicación precio y más tipos de filtros que sean necesarios para el contexto en el que se trate el trabajo.  
      

Conclusión.

Esta práctica de laboratorio brinda la oportunidad de aplicar conceptos y técnicas de limpieza y preparación de datos dentro de un estudio de caso específico, siguiendo la metodología CRISP-DM de IBM. La ejecución correcta de estas actividades es crucial para garantizar la calidad y confiabilidad del análisis posterior, lo cual es importante para todos los proyectos de big data y minería de datos.