Laboratorio 3

Presentado por: Javier Ricardo Valderrama González - 98001, Julián David Cifuentes González - 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.

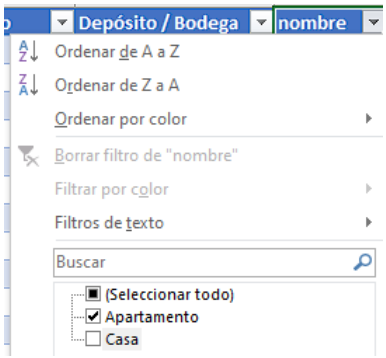
Desarrollo

1. abrir la información en Excel
2. seccionar los datos en columnas ya que están delimitados por comas (csv)
3. seleccionar los datos y pasarlos a formato tabla
4. arreglamos los enunciados con tildes y con la letra ñ, como
    a. portería / recepción
    b. baños
    c. salón comunal
    d. 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

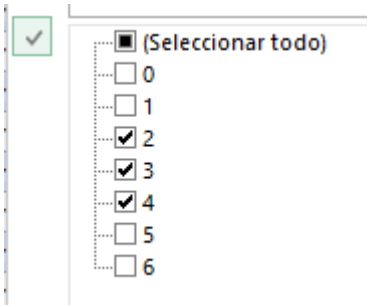
precio	precio
257000000	$ 257.000.000
260000000	$ 260.000.000
180000000	$ 180.000.000
290000000	$ 290.000.000
290000000	$ 290.000.000
340000000	$ 340.000.000
340000000	$ 340.000.000
340000000	$ 340.000.000
209000000	$ 209.000.000
300000000	$ 300.000.000

8. Una vez terminando filtramos la información solo por los apartamentos según el contexto del problema que planteamos.

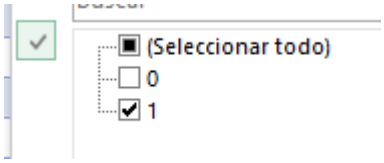
    a. Filtramos solo los apartamentos.



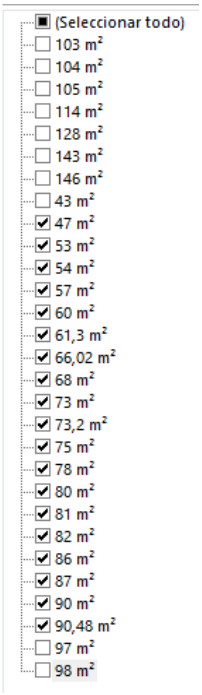
    b. Ahora filtramos los estratos del 2 al 4.



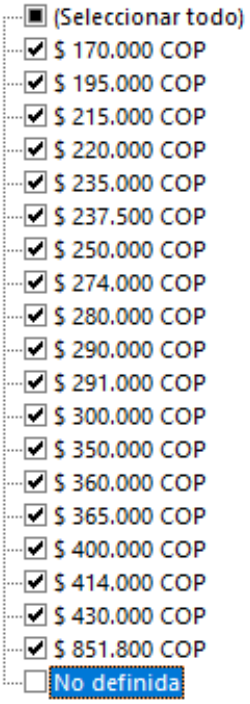
    c. Filtramos que tenga Parqueadero para visitantes, Balcón y Estudio.



d. y por último filtramos los apartamentos que cumplan con los metros cuadrados solicitados que en este caso son entre 45 a 95 m2.



e. 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.



f. 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.

area_construida	area_privada	estrato	estado	antiguedad	administracion	precio_m2	As
47 m²	47 m²	4	Bueno	9 a 15 años	$ 195.000 COP	$ 5.468.085,11*m²	
53 m²	53 m²	4	Excelente	1 a 8 años	$ 250.000 COP	$ 4.905.660,38*m²	
54 m²	54 m²	3	No definida	mÃ¡s de 30 añ	$ 237.500 COP	$ 3.333.333,33*m²	
57 m²	0 m²	3	Excelente	1 a 8 años	$ 235.000 COP	$ 5.087.719,3*m²	
57 m²	0 m²	3	Excelente	1 a 8 años	$ 235.000 COP	$ 5.087.719,3*m²	
60 m²	60 m²	4	No definida	16 a 30 años	$ 291.000 COP	$ 5.666.666,67*m²	
60 m²	60 m²	4	No definida	16 a 30 años	$ 291.000 COP	$ 5.666.666,67*m²	
60 m²	60 m²	4	No definida	16 a 30 años	$ 291.000 COP	$ 5.666.666,67*m²	
66,02 m²	58 m²	3	Bueno	9 a 15 años	$ 170.000 COP	$ 3.165.707,36*m²	
68 m²	68 m²	4	No definida	16 a 30 años	$ 300.000 COP	$ 4.411.764,71*m²	
73 m²	0 m²	3	Bueno	9 a 15 años	$ 215.000 COP	$ 4.109.589,04*m²	
73,2 m²	0 m²	4	Excelente	1 a 8 años	$ 365.000 COP	$ 6.010.928,96*m²	
75 m²	73 m²	4	No definida	9 a 15 años	$ 350.000 COP	$ 4.400.000*m²	
75 m²	75 m²	4	Excelente	16 a 30 años	$ 280.000 COP	$ 6.000.000*m²	
78 m²	78 m²	4	No definida	1 a 8 años	$ 274.000 COP	$ 5.128.205,13*m²	
78 m²	78 m²	4	No definida	1 a 8 años	$ 274.000 COP	$ 5.128.205,13*m²	
80 m²	0 m²	3	No definida	No definida	$ 220.000 COP	$ 3.937.500*m²	
80 m²	80 m²	4	No definida	1 a 8 años	$ 430.000 COP	$ 5.875.000*m²	
80 m²	85 m²	4	No definida	9 a 15 años	$ 300.000 COP	$ 7.875.000*m²	
80 m²	85 m²	4	No definida	9 a 15 años	$ 300.000 COP	$ 7.875.000*m²	
81 m²	0 m²	4	No definida	16 a 30 años	$ 851.800 COP	$ 6.604.938,27*m²	
86 m²	86 m²	4	Excelente	1 a 8 años	$ 414.000 COP	$ 6.395.348,84*m²	
86 m²	86 m²	4	Excelente	1 a 8 años	$ 414.000 COP	$ 6.395.348,84*m²	
87 m²	87 m²	4	No definida	9 a 15 años	$ 290.000 COP	$ 5.172.413,79*m²	
90 m²	90 m²	4	Bueno	1 a 8 años	$ 400.000 COP	$ 5.444.444,44*m²	
90,48 m²	80 m²	4	Excelente	9 a 15 años	$ 360.000 COP	$ 4.973.474,8*m²	
90,48 m²	80 m²	4	Excelente	9 a 15 años	$ 360.000 COP	$ 4.973.474,8*m²	
90,48 m²	80 m²	4	Excelente	9 a 15 años	$ 360.000 COP	$ 4.973.474,8*m²	

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.