

Datos Vacios

Luis Correa, Valeria Márquez, Maily Silva

Resumen

Teniendo como base un listado de datos reflejado entre filas y columnas en la cual unas filas pueden estar completamente llenas con sus respectivos datos, como algunas no pueden estar con todos los datos sino con la mayoría de ellos o por lo menos con dato en específico. Respecto a esto debemos ubicar en que filas hay datos nulos y mirar que hacer con ellos.

Palabras Claves: datos, datos vacíos, estrategia, filtro, dropna().

1 Introducción

Un datos vacío es llamado a aquel que no tiene valor o que es nulo, esto puede ser debido a varias razones tales como falta de información, quizás a la falta de participación en una encuesta o a cualquier acontecimiento que no permitió asignarle un valor a dichos datos. La ausencia de estos datos pueden llegar a generar problemas en el análisis de estos. Por lo tanto existen diferentes métodos el cual podemos aplicar para darle respuesta a la pregunta de que voy a hacer con estos datos vacíos o nulos?, entre esos métodos o soluciones está la identificación de valores faltantes, impacto en el análisis, eliminación de datos faltantes y por ultimo la imputación de valores.

2 Objetivos

- Objetivo General
 - Definir una estrategia el cual nos permita saber que se hará con aquellos datos que son nulos o vacíos para así evitar que afecte negativamente al procesamiento de datos
- Objetivos Específicos
 1. Identificar los valores faltantes o datos vacíos dentro del conjunto de datos generados en filas y columnas.
 2. Implementar un método que defina que se realizará con los datos vacíos

3 Descripción de la actividad

Esta actividad se basa en qué podemos hacer con aquellos datos que son nulos, es decir, que no tienen ningún valor asignado. Con respecto a esto debemos aplicar una estrategia el cual nos permita definir e identificar como serán tratados estos datos.

4 Gráfica

- Conjunto de datos incluyendo las filas vacías

index	Unnamed: 0	Habitaciones	Baños	Estrato	Antigüedad	Piso N°	Administración	Precio m²	Parqueaderos	Estado	Tipo de apartamento	Precio	Área construida (m²)	Área privada (m²)
0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	240000000	NaN	NaN
1	1	1.0	2.0	4.0	1 a 8 años	3	No definida	\$ 4.032.258,06m²	NaN	NaN	NaN	250000000	62	62
2	2	1.0	2.0	6.0	9 a 15 años	NaN	No definida	\$ 6.428.571,43m²	2.0	NaN	NaN	675000000	105	0
3	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	647307563	NaN	NaN
4	4	1.0	1.0	5.0	NaN	3	\$ 167.000 COP	\$ 4.700.000m²	2.0	NaN	NaN	235000000	50	0
5	5	1.0	1.0	4.0	NaN	NaN	No definida	\$ 6.187.500m²	1.0	Excelente	NaN	198000000	32	0
6	6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	561960000	NaN	NaN
7	7	2.0	2.0	6.0	menor a 1 año	Otro	\$ 180.000 COP	\$ 7.421.875m²	1.0	Excelente	NaN	475000000	64	64
8	8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	631862400	NaN	NaN
9	9	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	Sobre planos	NaN	297126000	NaN	NaN
10	10	4.0	5.0	6.0	1 a 8 años	15	\$ 650.000 COP	\$ 6.162.780,77m²	2.0	Excelente	NaN	1060000000	172	172
11	11	3.0	2.0	5.0	9 a 15 años	2	No definida	\$ 3.394.127,17m²	NaN	Bueno	NaN	420000000	125,97	115
12	12	2.0	2.0	3.0	9 a 15 años	5	\$ 40.000 COP	\$ 4.818.181,82m²	1.0	Excelente	Duplex	265000000	55	55
13	13	3.0	3.0	4.0	9 a 15 años	1	\$ 244.000 COP	\$ 4.487.099,59m²	1.0	Excelente	Duplex	360000000	80,23	80,23
14	14	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	347300000	NaN	NaN
15	15	2.0	2.0	6.0	1 a 8 años	Otro	No definida	\$ 4.857.142,86m²	1.0	NaN	NaN	340000000	70	0
16	16	3.0	2.0	5.0	9 a 15 años	2	\$ 60.000 COP	\$ 4.551.282,05m²	1.0	NaN	NaN	355000000	78	0
17	17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1417743028	NaN	NaN
18	18	3.0	3.0	6.0	NaN	NaN	No definida	\$ 6.461.538,46m²	NaN	NaN	NaN	840000000	130	130
19	19	3.0	2.0	5.0	16 a 30 años	4	No definida	\$ 3.698.630,14m²	1.0	Bueno	NaN	270000000	73	0
20	20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	543532000	NaN	NaN
21	21	2.0	2.0	4.0	NaN	NaN	\$ 230.000 COP	\$ 5.333.333,33m²	1.0	NaN	NaN	320000000	60	0
22	22	3.0	3.0	4.0	16 a 30 años	7	\$ 311.000 COP	\$ 3.197.674,42m²	1.0	NaN	NaN	275000000	86	86
22	22	3.0	3.0	4.0	16 a 30 años	7	\$ 311.000 COP	\$ 3.197.674,42m²	1.0	NaN	NaN	275000000	86	86
23	23	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	295210000	NaN	NaN
24	24	2.0	2.0	4.0	16 a 30 años	11	No definida	\$ 4.081.632,65m²	NaN	NaN	NaN	200000000	49	0
25	25	3.0	2.0	5.0	1 a 8 años	Otro	\$ 281.000 COP	\$ 6.798.801,57m²	1.0	Excelente	NaN	590000000	86,78	87
26	26	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	En construcción	NaN	488100000	NaN	NaN
27	27	5.0	5.0	4.0	más de 30 años	Otro	\$ 843.000 COP	\$ 2.188.552,19m²	1.0	Excelente	NaN	650000000	297	297
28	28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	214626000	NaN	NaN
29	29	NaN	NaN	6.0	NaN	NaN	NaN	NaN	NaN	Sobre planos	NaN	553675000	NaN	NaN
30	30	3.0	2.0	5.0	1 a 8 años	Otro	\$ 450.000 COP	\$ 5.800.000m²	2.0	Excelente	PentHouse	638000000	110	0
31	31	3.0	3.0	5.0	9 a 15 años	NaN	\$ 320.000 COP	\$ 4.702.380,95m²	1.0	NaN	NaN	395000000	84	84
32	32	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	156520000	NaN	NaN
33	33	4.0	5.0	6.0	9 a 15 años	7	\$ 900.000 COP	\$ 4.280.155,64m²	2.0	Excelente	Duplex	1100000000	257	257
34	34	1.0	1.0	5.0	1 a 8 años	9	\$ 65.500 COP	\$ 6.000.000m²	NaN	Excelente	Loft	180000000	30	30
35	35	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	825894200	NaN	NaN
36	36	3.0	4.0	5.0	más de 30 años	3	\$ 607.000 COP	\$ 4.170.212,77m²	1.0	Excelente	Duplex	980000000	235	235
37	37	3.0	5.0	6.0	9 a 15 años	Otro	\$ 1.050.000 COP	\$ 7.400.000m²	3.0	NaN	NaN	1850000000	250	250
38	38	3.0	2.0	3.0	1 a 8 años	NaN	\$ 163.000 COP	\$ 4.200.000m²	NaN	NaN	NaN	210000000	50	0
39	39	2.0	2.0	4.0	16 a 30 años	Otro	\$ 165.000 COP	\$ 4.907.407,41m²	1.0	Excelente	NaN	265000000	54	0
40	40	4.0	3.0	4.0	16 a 30 años	4	No definida	\$ 3.465.116,28m²	1.0	NaN	NaN	298000000	86	86
41	41	3.0	2.0	5.0	1 a 8 años	Otro	\$ 218.000 COP	\$ 4.666.666,67m²	1.0	Bueno	NaN	350000000	75	75

- Conjunto de datos pero SIN incluir las filas vacías

index	Unnamed: 0	Habitaciones	Baños	Estrato	Antigüedad	Piso N°	Administración	Precio m²	Parquesaderos	Estado	Tipo de apartamento	Precio	Área construida (m²)	Área privada (m²)
1	1	1.0	2.0	4.0	1 a 8 años	3	No definida	\$ 4.032.258,06*m²	NaN	NaN	NaN	250000000	62	62
2	2	1.0	2.0	6.0	9 a 15 años	NaN	No definida	\$ 6.428.571,43*m²	2.0	NaN	NaN	675000000	105	0
4	4	1.0	1.0	5.0	NaN	3	\$ 167.000 COP	\$ 4.700.000*m²	2.0	NaN	NaN	235000000	50	0
5	5	1.0	1.0	4.0	NaN	NaN	No definida	\$ 6.187.500*m²	1.0	Excelente	NaN	198000000	32	0
7	7	2.0	2.0	6.0	menor a 1 año	Otro	\$ 180.000 COP	\$ 7.421.875*m²	1.0	Excelente	NaN	475000000	64	64
10	10	4.0	5.0	6.0	1 a 8 años	15	\$ 650.000 COP	\$ 6.162.790,71*m²	2.0	Excelente	NaN	1060000000	172	172
11	11	3.0	2.0	5.0	9 a 15 años	2	No definida	\$ 3.334.127,17*m²	NaN	Bueno	NaN	420000000	125,97	115
12	12	2.0	2.0	3.0	9 a 15 años	5	\$ 40.000 COP	\$ 4.818.181,82*m²	1.0	Excelente	Duplex	265000000	55	55
13	13	3.0	3.0	4.0	9 a 15 años	1	\$ 244.000 COP	\$ 4.487.099,59*m²	1.0	Excelente	Duplex	360000000	80,23	80,23
15	15	2.0	2.0	6.0	1 a 8 años	Otro	No definida	\$ 4.857.142,86*m²	1.0	NaN	NaN	340000000	70	0
16	16	3.0	2.0	5.0	9 a 15 años	2	\$ 60.000 COP	\$ 4.551.282,05*m²	1.0	NaN	NaN	355000000	78	0
18	18	3.0	3.0	6.0	NaN	NaN	No definida	\$ 6.461.538,46*m²	NaN	NaN	NaN	840000000	130	130
19	19	3.0	2.0	5.0	16 a 30 años	4	No definida	\$ 3.698.630,14*m²	1.0	Bueno	NaN	270000000	73	0
21	21	2.0	2.0	4.0	NaN	NaN	\$ 230.000 COP	\$ 5.333.333,33*m²	1.0	NaN	NaN	320000000	60	0
22	22	3.0	3.0	4.0	16 a 30 años	7	\$ 311.000 COP	\$ 3.197.674,42*m²	1.0	NaN	NaN	275000000	86	86
24	24	2.0	2.0	4.0	16 a 30 años	11	No definida	\$ 4.081.632,65*m²	NaN	NaN	NaN	200000000	49	0
25	25	3.0	2.0	5.0	1 a 8 años	Otro	\$ 281.000 COP	\$ 6.798.801,57*m²	1.0	Excelente	NaN	590000000	86,78	87
27	27	5.0	5.0	4.0	más de 30 años	Otro	\$ 843.000 COP	\$ 2.188.552,19*m²	1.0	Excelente	NaN	650000000	297	297
30	30	3.0	2.0	5.0	1 a 8 años	Otro	\$ 450.000 COP	\$ 5.800.000*m²	2.0	Excelente	PentHouse	638000000	110	0
31	31	3.0	3.0	5.0	9 a 15 años	NaN	\$ 320.000 COP	\$ 4.702.380,95*m²	1.0	NaN	NaN	395000000	84	84
33	33	4.0	5.0	6.0	9 a 15 años	7	\$ 900.000 COP	\$ 4.280.155,64*m²	2.0	Excelente	Duplex	1100000000	257	257
34	34	1.0	1.0	5.0	1 a 8 años	9	\$ 65.500 COP	\$ 6.000.000*m²	NaN	Excelente	Loft	180000000	30	30
34	34	1.0	1.0	5.0	1 a 8 años	9	\$ 65.500 COP	\$ 6.000.000*m²	NaN	Excelente	Loft	180000000	30	30
36	36	3.0	4.0	5.0	más de 30 años	3	\$ 607.000 COP	\$ 4.170.212,77*m²	1.0	Excelente	Duplex	980000000	235	235
37	37	3.0	5.0	6.0	9 a 15 años	Otro	\$ 1.050.000 COP	\$ 7.400.000*m²	3.0	NaN	NaN	1850000000	250	250
38	38	3.0	2.0	3.0	1 a 8 años	NaN	\$ 163.000 COP	\$ 4.200.000*m²	NaN	NaN	NaN	210000000	50	0
39	39	2.0	2.0	4.0	16 a 30 años	Otro	\$ 165.000 COP	\$ 4.907.407,41*m²	1.0	Excelente	NaN	265000000	54	0
40	40	4.0	3.0	4.0	16 a 30 años	4	No definida	\$ 3.465.116,28*m²	1.0	NaN	NaN	298000000	86	86
41	41	3.0	2.0	5.0	1 a 8 años	Otro	\$ 218.000 COP	\$ 4.666.666,67*m²	1.0	Bueno	NaN	350000000	75	75
42	42	2.0	2.0	6.0	NaN	NaN	\$ 320.000 COP	\$ 6.216.216,22*m²	1.0	NaN	NaN	480000000	74	0
44	44	1.0	2.0	4.0	1 a 8 años	3	No definida	\$ 4.032.258,06*m²	NaN	NaN	NaN	250000000	62	62
45	45	1.0	2.0	6.0	9 a 15 años	NaN	No definida	\$ 6.428.571,43*m²	2.0	NaN	NaN	675000000	105	0
47	47	1.0	1.0	5.0	NaN	3	\$ 167.000 COP	\$ 4.700.000*m²	2.0	NaN	NaN	235000000	50	0
48	48	1.0	1.0	4.0	NaN	NaN	No definida	\$ 6.187.500*m²	1.0	Excelente	NaN	198000000	32	0
50	50	3.0	2.0	5.0	1 a 8 años	Otro	\$ 281.000 COP	\$ 6.798.801,57*m²	1.0	Excelente	NaN	590000000	86,78	87
53	53	3.0	2.0	3.0	1 a 8 años	NaN	\$ 163.000 COP	\$ 4.200.000*m²	NaN	NaN	NaN	210000000	50	0
54	54	3.0	3.0	4.0	9 a 15 años	1	\$ 244.000 COP	\$ 4.487.099,59*m²	1.0	Excelente	Duplex	360000000	80,23	80,23
55	55	4.0	3.0	4.0	16 a 30 años	4	No definida	\$ 3.465.116,28*m²	1.0	NaN	NaN	298000000	86	86
56	56	5.0	5.0	4.0	más de 30 años	Otro	\$ 843.000 COP	\$ 2.188.552,19*m²	1.0	Excelente	NaN	650000000	297	297
58	58	2.0	2.0	6.0	1 a 8 años	Otro	No definida	\$ 4.857.142,86*m²	1.0	NaN	NaN	340000000	70	0
59	59	4.0	5.0	6.0	9 a 15 años	7	\$ 900.000 COP	\$ 4.280.155,64*m²	2.0	Excelente	Duplex	1100000000	257	257
61	61	2.0	2.0	4.0	16 a 30 años	11	No definida	\$ 4.081.632,65*m²	NaN	NaN	NaN	200000000	49	0
62	62	2.0	2.0	6.0	menor a 1 año	Otro	\$ 180.000 COP	\$ 7.421.875*m²	1.0	Excelente	NaN	475000000	64	64

5 Análisis de las gráficas

- Gráficas que incluyen los datos vacíos (1 y 2 son continuas)

En estas dos gráficas podemos ver una pequeña parte de lo que es el conjunto donde están todos los datos, y podemos observar que también hay datos vacíos o nulos. Para ser más específicos, hay 430 filas y 14 columnas en total, pero debido a su extensión solo proporcionamos una pequeña parte de dicho conjunto.

A estos datos vacíos les aplicamos el método de eliminación, ya que en su mayoría tenían más del 50% de datos nulos. Optamos por este método debido a que en estos datos no encontramos manera de intuir entre dos columnas para darle valor a otra, es decir, estimar un valor para una columna basándose en la información de otras dos columnas.

- Gráficas que NO incluyen datos vacíos (3 y 4 son continuas)

Por otro lado, en estas dos gráficas continuas podemos observar una parte del conjunto de datos pero sin aquellos datos vacíos o nulos. De tal manera nos queda un total de 274 filas con valores asignados y 14 columnas, es decir, se eliminaron 156 filas que tenían menos del 50% de datos no nulos.

6 Conclusión

El manejo de datos vacíos o nulos es importante en el análisis de datos, ya que su existencia puede influir de manera significativa en la calidad de los resultados y las conclusiones que se derivan de los modelos y análisis. Los valores nulos pueden surgir por diversas razones, como errores en la recolección de datos, registros incompletos o problemas técnicos en la base de datos.

Es fundamental identificar estos datos nulos, ya que dejarlos sin procesar puede llevar a errores. Los métodos para procesar los datos nulos pueden ser eliminarlos, imputarlos o mantenerlos si son relevantes para el análisis. La decisión sobre cómo manejar los datos vacíos depende del contexto del análisis, el tipo de datos y el impacto que estos valores puedan tener en el modelo o los resultados.