

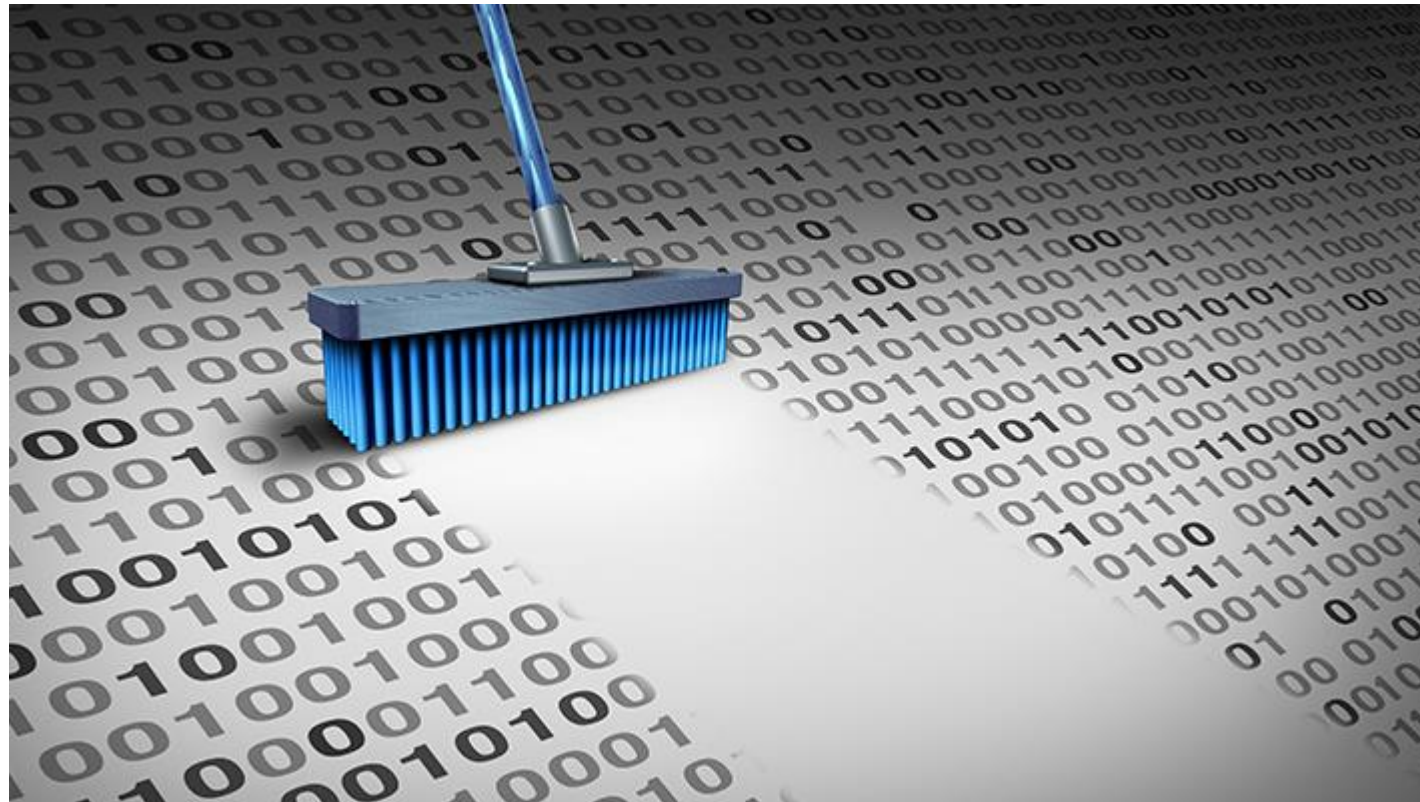


Limpieza de
datos:
Imputación de
datos en
blanco

¿Cómo tratar la ausencia de datos?



En la vida real, los datos no siempre están completos. Esto puede deberse a distintos factores, sin embargo, es necesario saber cómo tratar este fenómeno



¿Cómo tratar la ausencia de datos?

Tipo	Definición
Datos no imputables	Pertenecen a datos censurados y no se lograron capturar, o no es posible identificar
Datos imputables	No pueden ser identificados explícitamente pero que pueden ser subsanados

¿Cómo tratar la ausencia de datos?



Datos imputables

¿Por qué estos datos están en blanco?

Tipo	Definición
Totalmente al azar	Cualquier valor tiene la probabilidad de ser valor en blanco
Datos al azar	Se podría explicar que sean faltantes debido a algunas características
Totalmente no al azar	Se debe a variables no controladas

¿Cómo imputarlos?



Métodos

Si pueden ser imputados, existen diversos métodos:

- Imputación por medidas de tendencia central
- Imputación múltiple
- Imputación por algoritmos de machine learning

Column 0	age	years_seniority	income	parking_space	attending_party	entree	pets	emergency_contact
Tony	48	27	<input type="text"/>	1	5	shrimp	<input type="text"/>	Pepper
Donald	67	25	86	10	2	beef	<input type="text"/>	Jane
Henry	69	21	95	6	1	chicken	62	Janet
Janet	62	21	110	3	1	beef	<input type="text"/>	Henry
Nick	<input type="text"/>	17	<input type="text"/>	4	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Bruce	37	14	63	<input type="text"/>	1	veggie	<input type="text"/>	NA
Steve	83	<input type="text"/>	77	7	1	chicken	<input type="text"/>	n/a
Clint	27	9	118	9	<input type="text"/>	shrimp	3	None
Wanda	19	7	52	2	2	shrimp	<input type="text"/>	empty
Natasha	26	4	162	5	3	<input type="text"/>	<input type="text"/>	-
Carol	<input type="text"/>	3	127	11	1	veggie	1	****
Mandy	44	2	68	8	1	chicken	<input type="text"/>	null