

# Metadatos

## air\_quality.csv



Los datos del archivo *air\_quality.csv* contienen la información correspondiente a la calidad del aire en la ciudad de Nueva York. Estos datos fueron extraídos de la página [NYC Open Data](#) donde fueron provisto por el Department of Health and Mental Hygiene (DOHMH). Esta base de datos fue subida inicialmente el 24 de octubre de 2014 y actualizada por ultima vez el 10 de abril de 2024. Contiene los datos de las medidas de calidad del aire desde el año 2005 al 2022.

Variable	Tipo de dato	Tipo de variable	Descripción
UniqueID	int	categórica	Identificador de registro único
indicatorID	int	categórica	Identificador del tipo de valor medido en el tiempo y el espacio.
Name	string	categórica	Nombre del indicador
Measure	string	categórica	Cómo se mide el indicador
Measure info	string	categórica	Información (como unidades) sobre la medida
Geo Type Name	string	categórica	Tipo de conjunto geográfico
Geo Join ID	float	categórica	Identificador del área geográfica del barrio, utilizado para unir a archivos de geografía cartográfica para realizar mapas temáticos
Geo Place Name	string	categórica	Nombre de barrio
Time Period	string	nominal	Descripción del tiempo al que se aplican los datos
Start_Date	string	fecha	Valor de fecha para el inicio del período de tiempo
Data Value	float	cuantitativa	El valor real de los datos para este indicador, medida, lugar y hora
Message	float	cualitativa	Notas que se aplican al valor de los datos

# Consideraciones

La base de datos contiene 18.025 registros. Solo se encuentran valores nulos en una columna y no existen duplicados para todas las columnas.

Variable	Observacion	Acciones
Measure	Se tienen datos que corresponden a unidades y no a tipo de medida (ej. Number per km2)	Es posible que esta inconsistencia sea pasada por alto para el análisis
Start_Date	<ul style="list-style-type: none"><li>Solo se tienen valores para Enero y Febrero y los días 1, 6 y 12 de los meses.</li><li>Existen 202 en los cuales el formato no corresponde a %d/%m/%Y como los demas.</li></ul>	Se tendra en cuenta esta diferencia al momento de pasar estos datos a tipo <i>datetime</i>
message	Esta columna contiene solo nulos	Se procederá a eliminar esta columna del DataFrame