

# Metadatos

## noise\_pollution.csv



Los datos del archivo *noise\_pollution.csv* contienen la información correspondiente a la contaminación sonora la ciudad de Nueva York (dentro y fuera del área de Manhattan) y su proveniencia. Estos datos fueron extraídos del paper [SONYC Urban Sound Tagging \(SONYC-UST\): a multilabel dataset from an urban acoustic sensor network](#) publicado en la pagina de **Zenodo** el 14 de septiembre de 2020. Estos registros fueron tomados entre los años 2016 y 2019.

Variable	Tipo de dato	Tipo de variable	Descripción	Valores posibles
split	string	categorica	La división de datos para el análisis	train, test, validate
sensor_id	int	categorica	El ID del sensor del que proviene la grabación.	
audio_filename	string	cualitativa	El nombre del archivo de la grabación de audio	
annotator_id	int	categorica	El ID anónimo del anotador.	<b>positivo</b> = voluntario Zooniverse <b>negativo</b> = miembro del equipo SONYC <b>0</b> = verdad acordada equipo SONYC
borough	int	categorica	El distrito de la ciudad de Nueva York en el que se encuentra el sensor	<b>1</b> = Manhattan <b>3</b> = Brooklyn <b>4</b> = Queens
block	int	categorica	El bloque de la ciudad de Nueva York en el que se encuentra el sensor.	
latitude	float	númerica	La coordenada de latitud del bloque en el que se encuentra el sensor.	
longitude	float	númerica	La coordenada de longitud del bloque en el que se encuentra el sensor.	
year	int	nominal	El año del que proviene la grabación.	
week	int	nominal	La semana del año del que proviene la grabación.	Números del 1 al 52
day	int	nominal	El día de la semana del que proviene la grabación.	Numero de 0 al 6 con el lunes como inicio (es decir, 0=lunes).
hour	int	nominal	La hora del día de la grabación	Números del 0 al 23

<coarse_id>-<fine_id> _<fine_name> _presence	int	categorica	Indican la presencia de una clase de nivel fino. Se tiene una columna para cada clase.	<b>1</b> = presente <b>0</b> = no está presente <b>-1</b> = no se etiqueto
<coarse_id> _<coarse_name> _presence	int	categorica	Indican la presencia de una clase de nivel grueso. Se tiene una columna para cada clase.	<b>1</b> = presente <b>0</b> = no está presente <b>-1</b> = no se etiqueto
<coarse_id>-<fine_id> _<fine_name> _proximity	int	categorica	Indican la proximidad de una clase de nivel fino. Se tiene una columna par cada clase.	<b>1</b> = presente <b>0</b> = no está presente <b>-1</b> = no se etiqueto

## Clases

A continuación, se describen las clases que existen y a las cuales hace referencia las ultimas 3 filas del diccionario.

### 1. engine

- 1: small-sounding-engine
- 2: medium-sounding-engine
- 3: large-sounding-engine
- X: engine-of-uncertain-size

### 1. machinery-impact

- 1: rock-drill
- 2: jackhammer
- 3: hoe-ram
- 4: pile-driver
- X: other-unknown-impact-machinery

### 1. non-machinery-impact

- 1: non-machinery-impact

### 1. powered-saw

- 1: chainsaw
- 2: small-medium-rotating-saw
- 3: large-rotating-saw
- X: other-unknown-powered-saw

### 1. alert-signal

- 1: car-horn
- 2: car-alarm
- 3: siren
- 4: reverse-beeper
- X: other-unknown-alert-signal

### 1. music

- 1: stationary-music
- 2: mobile-music

- 3: ice-cream-truck
- X: music-from-uncertain-source

1. human-voice

- 1: person-or-small-group-talking
- 2: person-or-small-group-shouting
- 3: large-crowd
- 4: amplified-speech
- X: other-unknown-human-voice

1. dog

- 1: dog-barking-whining

## Consideraciones

La base de datos contiene 62.022 registros. No existen nulos ni duplicados.