**Noise Pollution Dictionary**

**Columns**

*split*

The data split. (*train*, *validate, test*)

*sensor\_id*

The ID of the sensor the recording is from.

*audio\_filename*

The filename of the audio recording

*annotator\_id*

The anonymous ID of the annotator. If this value is positive, it is a citizen science volunteer from the Zooniverse platform. If it is negative, it is a SONYC team member. If it is 0, then it is the ground truth agreed-upon by the SONYC team.

*year*

The year the recording is from.

*week*

The week of the year the recording is from.

*day*

The day of the week the recording is from, with Monday as the start (i.e. 0=Monday).

*hour*

The hour of the day the recording is from

*borough*  
The NYC borough in which the sensor is located (1=Manhattan, 3=Brooklyn, 4=Queens). This corresponds to the first digit in the 10-digit NYC parcel number system known as Borough, Block, Lot (BBL).

*block*

The NYC block in which the sensor is located. This corresponds to digits 2—6 digit in the 10-digit NYC parcel number system known as Borough, Block, Lot (BBL).

*latitude*

The latitude coordinate of the **block** in which the sensor is located.

*longitude*

The longitude coordinate of the **block** in which the sensor is located.

*<coarse\_id>-<fine\_id>\_<fine\_name>\_presence*

Columns of this form indicate the presence of fine-level class. 1 if present, 0 if not present. If -1, then the class was not labeled in this annotation because the annotation was performed by a SONYC team member who only annotated one coarse group of classes at a time when annotating the verified subset.

*<coarse\_id>\_<coarse\_name>\_presence*

Columns of this form indicate the presence of a coarse-level class. 1 if present, 0 if not present. If -1, then the class was not labeled in this annotation because the annotation was performed by a SONYC team member who only annotated one coarse group of classes at a time when annotating the verified subset. These columns are computed from the fine-level class presence columns and are presented here for convenience when training on only coarse-level classes.

*<coarse\_id>-<fine\_id>\_<fine\_name>\_proximity*

Columns of this form indicate the proximity of a fine-level class. After indicating the presence of a fine-level class, citizen science annotators were asked to indicate the proximity of the sound event to the sensor. Only the citizen science volunteers performed this task, and therefore this data is not included in the verified annotations. This column may take on one of the following four values: (near, far, notsure, -1). If -1, then the proximity was not annotated because either the annotation was not performed by a citizen science volunteer, or the citizen science volunteer did not indicate the presence of the class.