**Rachael Phillips**

**May 28, 2021**

Fuzzy deduplication techniques demonstrated using two jupyter notebooks.

Some Python modules used:

import recordlinkage

from sklearn.feature\_extraction.text import TfidfVectorizer

import sparse\_dot\_topn.sparse\_dot\_topn as ct

from scipy.sparse import csr\_matrix

from ftfy import fix\_text

from pyphonetics import Soundex

import jellyfish

Repository:

<https://github.com/rachaelph/Code>

Jupyter notebooks:

<https://github.com/rachaelph/Code/blob/main/Data%20Deduplication.ipynb>

<https://github.com/rachaelph/Code/blob/main/Fuzzy%20Matching%20Techniques.ipynb>