

PRA2: Com realitzar la neteja i l'anàlisi de dades?

Àlex Franco Granell; Roger Esteban Fabró

Gener 2023

Descripció del dataset

Hem escollit investigar quins factors, patrons de conducta o situació ambiental influència més a la incidència de càncer. Per fer aquest anàlisi s'ha escollit tres datasets amb dades dels Estats Units que mostren la mortalitat per estat, i diversos paràmetres demogràfics. Concretament utilitzem els datasets de “Cancer Mortality & Incidence Rates: (Country LVL)”, les dades dels EUA de “Demographics & observation for pandemic escalation”, i un dataset d'usafacts.org que recull la població dels diversos estats del país. Els dos primers disponibles a kaggle.

Integració i selecció de dades