

DATA DOCUMENTATION

Data Source:

The Office of National Coordinator for Health Information technology (ONC)'s Health IT have numerous health information technology data which are accessible through its open-source REST APIs. "<https://www.healthit.gov/data/api>" is the website that hosts the information and documentation regarding accessing of these APIs to query the data. It has medical data predominantly part of public health information. The datasets utilized for this project work "Electronic Prescribing Adoption and Use by County" and "Office-based Health Care Provider Database". The former has the information of adoption of electronic prescribing practices in each county of all states in the USA. The information is from the year 2014. The second dataset contains the additional data about the number of providers, primary care providers, physicians, primary care physicians, nurse practitioners, primary care nurse practitioners, physician assistants, and primary care physician assistant. The primary care subset refers to the health professional having expertise in adolescent medicine, pediatrics, family practice, general practice, geriatrics, internal medicine, obstetrics, or gynecology. Majorly it contained the county information which included the state or region name, state FIPS, county name, county FIPS, combined FIPS. FIPS or Federal Information Processing Standard is a special code to denote the region. There are many levels of code for state level and county level, and the combined code uses elements from both to create a unique code for the county in the whole country.

Accessing the Data:

The above-mentioned URL is used to access the webpage from python environment. The program then scrapes the dataset list from the page where there are about 23 different datasets. Then, web crawling is carried out to retrieve the URL extension of the required dataset. The "requests" library then fetches that website. That website in the same domain has details about the dataset, and detailed documentation of the data dictionary explaining the variables and its characteristics. This webpage contains the required API website to the data, which is obtained through crawling. Then it is called to retrieve the dataset in the runtime. It is a list of dictionary where each dictionary refers to a county. In the second dataset, there are additional information at national and state level which are left out. There is a last dictionary in both the dataset for "None Region" which was removed. Since, this is a static dataset, caching is not required.

Data Summary:

The first data set had 8 variables for each instances.

- 1) Region - The region name for the data which includes all 50 states and the District of Columbia.
- 2) Region_code - The abbreviation code for the region name. This also includes all 50 states and the District of Columbia.
- 3) Period - Time period of the data collected.
- 4) County_name - The county name corresponding to the specific line of data.
- 5) State_FIPS - State Federal Information Processing Standard for the region.
- 6) County_FIPS - County Federal Information Processing Standard for the county name.
- 7) FIPS - Federal Information Processing Standard which combines region and county.
- 8) pct_phys_e_Rx_ehr - Percent of Physicians Actively Using an EHR to e-Prescribe via the Surescripts Network.

The second data 14 variables for each instances.

- 1) Region - The region name for the data which includes all 50 states and the District of Columbia.
- 2) Region_code - The abbreviation code for the region name. This also includes all 50 states and the District of Columbia.
- 3) Period - Time period of the data collected.
- 4) State_FIPS - State Federal Information Processing Standard for the region.
- 5) County_FIPS - County Federal Information Processing Standard for the county name.
- 6) FIPS - Federal Information Processing Standard which combines region and county.
- 7) All_providers - Total count of Office-based Physicians, Nurse Practitioners, and Physician Assistants
- 8) All_primary_care_providers - Total count of Office-based Primary Care Physicians, Nurse Practitioners, and Physician Assistants
- 9) All_physicians - Total count of Office-based Physicians
- 10) All_primary_care_physicians - Total count of Office-based Primary Care Physicians
- 11) All_nurse_practitioners - Total count of Office-based Nurse Practitioners
- 12) All_primary_care_nurse_practitioners - Total count of Office-based Primary Care Nurse Practitioners
- 13) All_physician_assistants - Total count of Office-based Physician Assistants
- 14) All_primary_care_physician_assistants - Total count of Office-based Primary Care Physician Assistants

The second data set was combined with the first by matching the FIPS of each counties in the datasets. Providers, physicians, nurse, assistant count were taken while their primary care counterparts' count was left out.