Project Ideas:

**Dataset 1: CMS Doctors and Clinicians Data:**

**66MB**

<https://data.cms.gov/provider-data/dataset/mj5m-pzi6#data-dictionary>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable Name | Variable Label | Description | Length | Values |
| NPI | Professional Identification | Unique clinician ID assigned by NPPES | 10 | string |
| PAC ID | Ind\_PAC\_ID | Unique individual clinician ID assigned by PECOS | 10 | string |
| Clinician Enrollment ID | Ind\_enrl\_ID | Unique ID for the clinician enrollment that is the source for the data in the observation | 15 | string |
| Provider Last Name | Provider Last Name | Individual clinician last name | 35 | string |
| Provider First Name | Provider First Name | Individual clinician first name | 25 | string |
| Provider Middle Name | Provider Middle Name | Individual clinician middle name | 25 | string |
| Suffix | suff | Individual clinician suffix | 10 | string |
| Gender | gndr | Individual clinician gender | 1 | M/F/U |
| Credential | Cred | Medical credential such as MD, DO, DPM, etc. | 3 | string |
| Medical school name | Med\_sch | Individual clinician’s medical school | 100 | string |
| Graduation year | Grd\_yr | Individual clinician’s medical school graduation year | 4 | numeric |
| Primary specialty | Pri\_spec | Primary medical specialty reported by the individual clinician in the selected enrollment | 60 | string |
| Secondary specialty 1 | Sec\_spec\_1 | First secondary medical specialty reported by the individual clinician in the selected enrollment | 60 | string |
| Secondary specialty 2 | Sec\_spec\_2 | Second secondary medical specialty reported by the individual clinician in the selected enrollment | 60 | string |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable Name | Variable Label | Description | Length | Values |
| Sec\_spec\_3 | Secondary specialty 3 | Third secondary medical specialty reported by the individual clinician in the selected enrollment | 60 | string |
| Sec\_spec\_4 | Secondary specialty 4 | Fourth secondary medical specialty reported by the individual clinician in the selected enrollment | 60 | string |
| Sec\_spec\_all | All secondary specialties | All secondary medical specialty reported by the individual clinician in the selected enrollment | 200 | string |
| Telehlth | Telehealth | Indicator for whether clinician offers telehealth services over video and/or audio | 1 | Y = Medicare fee-for-service claims indicate that clinician offers telehealth services |
| Facility Name | Facility Name | Legal organization name of the group practice that the individual clinician works with – will be blank if the address is not linked to a group | 70 | string |
| Org\_PAC\_ID | Group PAC ID | Unique group ID assigned by PECOS to the group that the individual clinician works with – will be blank if the address is not linked to a group | 10 | string |
| num\_org\_mem | Number of group members | Total number of individual clinicians affiliated with the group based on Group Practice PAC ID | 8 | numeric |
| adr\_ln\_1 | Line 1 Street Address | Group or individual's line 1 address | 55 | string |
| adr\_ln\_2 | Line 2 Street Address | Group or individual's line 2 address | 55 | string |
| ln\_2\_sprs | Marker of address line 2 suppression | Marker that the address as reported may be incomplete | 1 | Y |
| City/Town | City/Town | Group or individual's city | 30 | string |
| State | State | Group or individual's state | 2 | string |
| ZIP Code | ZIP Code | Group or individual's ZIP code (9 digits when available) | 15 | string |
| Telephone Number | Telephone Number | Phone number is listed only when there is a single phone number available for the address | 20 | string |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable Name | Variable Label | Description | Length | Values |
| ind\_assgn | Clinician accepts Medicare Assignment | Indicator for whether clinician accepts Medicare approved amount as payment in full | 1 | Y = Clinician accepts Medicare approved amount as payment in full, M = Clinician may accept Medicare Assignment |
| grp\_assgn | Group accepts Medicare Assignment | Indicator for whether group accepts Medicare approved amount as payment in full | 1 | Y = Group accepts Medicare approved amount as payment in full, M = Group may accept Medicare Assignment |
| adrs\_id | Address ID | Unique identifier for the practice location; offices within the same building, but varied by suite or floor, will have the same Address ID aside from the final two characters | 25 | string |

## **Dataset 2: Word Bank Data – NASDAQ**

The World Bank ([WB](https://data.nasdaq.com/databases/WB)) data feed contains data from hundreds of countries and regions around the world, from multiple categories such as finance, economy, energy, education, health, poverty, agriculture, employment, population, land use, foreign aid, climate change, government expenditures, literacy, mortality, and patents.

## WB Data (WB/DATA)

| **Columns** | **Data type** | **Description** | **Filter** | **Primary Key** |
| --- | --- | --- | --- | --- |
| series\_id | string | unique series identifier | ✅ | ✅ |
| country\_code | string | 3-letter country/region code | ✅ | ✅ |
| country\_name | string | name of country/region |  |  |
| year | integer | year of data point |  | ✅ |
| value | double | value of data point |  |  |

## WB Metadata (WB/METADATA)

| **Columns** | **Data type** | **Description** | **Filter** | **Primary Key** |
| --- | --- | --- | --- | --- |
| series\_id | string | unique series identifier | ✅ | ✅ |
| name | text | name of series |  |  |
| description | text | description of series |  |  |

**Dataset 3: NYC Taxi Trip Data 2023**

https://catalog.data.gov/dataset/2023-yellow-taxi-trip-data

These records are generated from the trip record submissions made by yellow taxi Technology Service Providers (TSPs). Each row represents a single trip in a yellow taxi. The trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off taxi zone locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts.