**INFORMATION SEEKING**

Dataset#1: **U.S. Chronic Disease Indicators (CDI)**

|  |  |  |
| --- | --- | --- |
| 1 | Data Citation | U.S. Chronic Disease Indicators (CDI) :Centres for Disease Control and Prevention, August 20, 2018, Retrieved from: <https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi> |
| 2 | License/Terms of use | This dataset is intended for public access and use.  License information: <https://opendefinition.org/licenses/odc-odbl/> |
| 3 | Why is the data interesting | This data is interesting as it provides very detailed information about the various chronic diseases impacting adults- both males and females. It covers a range of diseases resulting from alcoholism to arthritis, cancers, cardiovascular diseases, kidney disorders, disabilities etc to name a few. This data can be filtered as per all the states present in the USA. Also, the dataset has categorization based on age, race, ethnicity, crude prevalence etc. This can aid us to analyse the pattern of chronic diseases across the USA and their correlation with the filters discussed above. |
| 4 | Potential data-users/ decision makers | The state/national health department, government/ international health organizations, medical researchers, hospitals, data analysts in the medical field. |
| 5 | Three questions the data might help to answer | 1. Most common chronic health conditions among adults. This may help assess the contributing factors and cure. 2. Categorization of diseases based on age and race. This might help establish if certain diseases are specific to a particular group. 3. Prevalence of diseases in specific states of interest in the USA. |

Dataset#2: **Air Quality Measures on the National Environmental Health Tracking Network**

|  |  |  |
| --- | --- | --- |
| 1 | Data Citation | Air Quality Measures on the National Environmental Health Tracking Network: Centres for Disease Control and Prevention, August 20, 2018, Retrieved from: <https://catalog.data.gov/dataset/air-quality-measures-on-the-national-environmental-health-tracking-network> |
| 2 | License/Terms of use | This dataset is intended for public access and use.  License information: <https://opendefinition.org/licenses/odc-odbl/> |
| 3 | Why is the data interesting | This data is interesting because it provides the measure of air quality based on the number of days that have the maximum average ozone concentration over the National Ambient Air Quality Standard. This data is available over a period of 14 years across all the states and counties in the USA. This data has been collected from 4000 stations across the country over the years and is helpful for determining the extent of air pollution in every state. |
| 4 | Potential data-users/ decision makers | The Environmental Protection Agency, world organizations interested in air pollution, researchers (areas of air quality, pollutants, ozone depletion), government. |
| 5 | Three questions the data might help to answer | 1. Ozone layer depletion patterns over the past decade. 2. Which states/counties have the poorest/best air quality? And how do they compare with the National Ambient Air Quality Standard? This can help us figure out the contributing factors for the good/bad air quality. 3. Is there an increase/decrease in the air pollutants/ ozone depletion since 1999? |

Dataset#3: **Behavioral Risk Factor Data: Tobacco Use**

|  |  |  |
| --- | --- | --- |
| 1 | Data Citation | Behavioural Risk Factor Data: Tobacco Use (2011 to present) :Centres for Disease Control and Prevention, October 14, 2016, Retrieved from: <https://catalog.data.gov/dataset/behavioral-risk-factor-data-tobacco-use-2011-to-present-e0ad1> |
| 2 | License/Terms of use | This dataset is intended for public access and use.  License information: <https://opendefinition.org/licenses/odc-odbl/> |
| 3 | Why is the data interesting | This data is interesting because it provides information on tobacco use from 2011-2016 across all the states in the USA based on various categories such as Cigarette use, e-cigarette use, smokeless tobacco etc. It also gives data about the people from the sample who have stopped using tobacco. This data has categorization based on gender/overall population, race and age range. This can help analyse the cigarette smoking status based on demographics. |
| 4 | Potential data-users/ decision makers | The State Tobacco Activities Tracking and Evaluation (STATE) System, researchers in the areas of tobacco use and effects, oral cancers etc. |
| 5 | Three questions the data might help to answer | 1. Which is the most common type of smoking (smokeless/ cigarettes/e-smoking) among the population across different states? This can be used to analyse if a correlation exists between the smoking type and the associated diseases/death rates. 2. Which age group/race engages the most in tobacco use? 3. What percentage of the sample has been successful with quit attempts? |