**Information Seeking Assignment**

**U.S. Department of Health and Human Services/ Centers for Disease Control and prevention**

**Name:** U.S. Chronic Disease Indicators (CDI)

**Data Citation:**

U.S. Department of Health and Human Services/ Centers for Disease Control and Prevention (2016). U.S. Chronic Disease Indicators [Dataset, data file]. retrieved from <https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi-e50c9/resource/61d71c08-f317-4af8-95b8-763f4bfadc4f> date accessed 9/8/17.

**License and terms of use:**

This dataset has an Open Database License (ODbl) <https://opendatacommons.org/licenses/odbl/summary/> which enable the its users to copy and distribute the date, produce works from this database, and modify or build upon from the database. However, it is important to make aware of this license to any other user with who the dataset is shared and its important to keep any notices from the original database, any changes to this dataset must also be adapted to the ODbl terms, also if you restrict the work of this dataset for personal/work related usage, it is important to redistribute a version without any restrictions made to it.

**Description:**

This data set published by the Centers of Disease Control and Prevention demonstrates 124 indicators that were developed by consensus that helped identify the prevalence of chronic diseases throughout the U.S. states, territories, and provinces. These indicators can help answer the underlying cause of why certain chronic disease are more prevalent in certain parts of the country than others. This dataset was created in June 2016 and most recently modified in July 2017, therefore it gives recent information about potential causes of Chronic Diseases on certain parts of the country. This dataset also provides information regarding race and ethnicity, which are variables that public health professionals take into consideration when studying the prevalence, patterns, and distribution of chronic diseases throughout a geographic region.

**Potential Users:**

* Public health professionals studying the prevalence rates of certain chronic disease throughout the United States.
* Doctors and scientist who want to do further research to determine how they can improve the chronic disease trends in these particular parts of the country.
* Journalists who may want to write about the current health state of individuals throughout the U.S. and how culture, food, and other aspects may have contributed to the development of these chronic diseases.

**Three Questions:**

* Based on these chronic disease indicators, what can we do to improve the health of the American people throughout the country?
* What patterns do we see in the prevalence of any particular chronic disease based on gender, race/ethnicity?
* How can we use this information to develop health programs that provide health education and screenings to community members in the regions with the highest prevalence rates?

**Students’ Academic Performance**

**Name:** Mining educational data to predict student’s academic performance using ensemble methods.

**Data citation:**

Amrieh E. A., Hamtini T., & Aljarah. (2016). Mining educational data to predict student’s academic performance using ensemble methods [database jornal]. Retrieved from <https://www.kaggle.com/aljarah/xAPI-Edu-Data> date accessed 9/8/2017.

**License and terms of use:**

The license and terms of use are provided by ShareAlike 4.0 International (CC BY-SA 4.0). <https://creativecommons.org/licenses/by-sa/4.0/>The license enables the users the ability to make copies and redistribute this dataset in any format. The terms also permits the remix, transformation, and build upon this data set for any purpose even if it’s for commercial purposes as long as credit is always given to the publisher if the dataset is modified or redistributed.

**Description:**

This particular dataset is focused on education and learning. The purpose of collecting this dataset was to collect data on various students throughout the world and determine their learning experience using a “learner activity tracker” called experience API (xAPI). The xAPI is part of a training learning architecture (TLA) which tracks the students monitor usage and adapts to its users learning pattern over time. For this dataset information about students from different countries in the Middle East, Latin America, and the USA were collected. This data includes students from different academic levels (elementary, middle, and high school). Information about the students also contains the students’ gender, class participation, subjects they studied and the level of involvement of their parents in their education.

**Potential Users:**

* UX/UI designers who are trying to develop an efficient and effective learning application that students from around the world could use in their academic studies.
* The education system who is trying to compare the effectiveness on online education using a learning application tool, as oppose to a traditional classroom setting.
* Teachers and parents who are collaborating together to provide the students the best learning opportunity and environment.
* Older students who may wish to attend school online as oppose to attending a traditional classroom.

**Three Questions:**

* How are test scores and other assignments grades compared between students who use this learning tool and students who do not use it?
* If the usage of this application tool is highly successful, could the education system possibly implement a full online education for students K-12?
* If the education system decides to implement a full online education system, how will this create different education approach affect the teachers’ job market as well as students who may not have easy access to a computer or Internet?

**Significant Earthquakes, 1965-2016**

**Name:** Significant Earthquakes 1965-2016. Date, time, and location of all earthquakes with magnitude of 5.5 or higher

**Data citation:**

US Geological Survey (2017). Significant Earthquakes, 1965-2016 Date, time, and location of all earthquakes with magnitude of 5.5 or higher [Dataset of earthquakes 5.5 or higher worldwide]. Retrieved from <https://www.kaggle.com/usgs/earthquake-database/data> date accessed 9/8/2017.

**License and terms of use:**

This dataset is a public domain licensed by CC0 1.0 Universal (CC0 1.0). <https://creativecommons.org/publicdomain/zero/1.0/>The individuals who worked on gathering this information have waived all of their rights to the work worldwide. Therefore this data can by copied, modified, and distributed without the need of asking for permission to do so.

**Description:**

The US Geological Survey has provided this dataset that was gathered by the National Earthquake Information Center (NEIC), which provides data on earthquakes incidents worldwide for the past fifty years. The information about the earthquakes incidences provided by the NEIC describes the date and time each earthquake occurred. It also provide the location of the earthquake incident (Latitude and Longitude) and most importantly it provides the magnitude of the earthquake. All the earthquakes provided in this dataset only include earthquakes of a 5.5 magnitude or higher. Therefore, we can assume that a lot more earthquakes have occurred throughout the past fifty years which are not included in this dataset because they were probably weaker than 5.5 in magnitude.

**Potential users:**

* Geologist and other scientist who may want to study the patterns of earthquakes incidences throughout the last fifty years.
* Scientists who are trying to understand how global warming and other human induced changes in our environment may be affecting the frequency of earthquake incidences.
* The citizens of certain parts of the world where the frequency of earthquakes is high and may want to know how they can be prepare for future earthquake incidences.

**Three questions:**

* Based on this dataset, is there any particular time period within these past fifty years where the frequencies of earthquakes have been the highest?
* How can this dataset help geologist and other scientist predict future earthquake incidences in particular parts of the world that have high earthquake frequencies?
* Is there something underneath the ground of high frequency earthquakes locations that may be influencing the frequent incidence of earthquakes? Could there be some human induced causes of why earthquakes are more frequent on certain parts of the world more than others?

**Word count**: 1,052/1240.