**Assignment 2 Data and Models**

**Chapter 2 Data:**

**Data should guide the process of identifying data sets that are most relevant to the real-world problem in Assignment 1:**

**Please answer the following question about Data:**

* **Who**is the data set about?
* **Who** were sampled in this data set?
* **Who** were over sampled or under sampled? Are they representative of the main characters in Assignment 1?
* Is there any identifiable information or is there any risk of disclose identifiable information? This is fundamentally about the sampling issue, and anonymity.
* **What** events, activities, behaviors, and observations etc. are recorded by the data set?
* Does the data set record the targeted events, activities, behaviors, etc. in Assignment 1? This is fundamentally about the variables.
* **When**did the event, activity, behavior, and observation, etc. take place? **When** were the data collected?
* Is it longitudinal or cross-sectional?
* Are they real time data?
* How old or fresh are the data? To what extent generalization can be made across time to inform Assignment 1? This is fundamentally about the temporal structure of the data set, and the external validity of the data set across time.
* **Where** did the event, activity, behavior, and observation, etc. take place? **Where** were the data collected if the information is available? What does the geographical coverage of the data set look like?
* Does the data set contain geographical information (GIS)? Is this a local, regional, national, or global data set?
* To what extent generalization can be made across settings to inform Assignment 1? This is fundamentally about geographic variables in the data set, and the external validity of the data set across settings.
* **Why**did the event, activity, behavior, or observation etc. take place? **Why** were the data collected?
* **How:** If you would like, you can add a dimension of how. How did it happen? Sometimes, the answer to how can be covered by what, when and where.

**Chapter 3 The Scientist and AI:**

The objective of Chapter 3 is to select or create the most relevant ML algorithms given chapter 1 and Chapter 2. The data sets should be properly aligned so that they can be used in ML. Accordingly, Chapter 3 describes a story about data scientists (the students) and AI together help the people/community find the solution. The main story line is about iterative experimentation that data scientists and AI conduct collaboratively along the journey of “rescuing” people in need.

**“Chapter 3 the Scientist and AI” is a story about the data scientist and AI’ selecting the most relevant ML algorithms given Assignment 1 and Data:**

* **Who:** In this story, there are three main characters: 1) the people/the community who needs help, 2) the data scientist (that is you), and 3) AI.

**Please answer the following questions:**

* How much does the data scientist understand Assignment 1 (domain) and data?
* **What** models and analysis did the data scientist and AI apply to fulfill the need of the people or the community?
* Can the data scientist estimate and select data for their goals from Assignment 1? Can they map data sets from Assignment 2 onto appropriate ML models?
* Can the data scientist connect Story 1 with ML models/stories about what a ML model can do? To perform good ML research, what in-depth knowledge and experience with ML algorithms and ML stories does a data scientist need?
* **When** has to do with the iterations, how much time did it take for experimentation? How efficient is the modeling/algorithm?
* Can the data scientist determine the acceptance level of the model (validation with accuracy and runtime performance) considering the targeted users?
* **Where** has to do with the learning environment. Where did this experiential learning process take place? For example, it was part of an online Deep Learning course.
* **Why** explains the modeling. Explain the ML model you are using?
* **How**: If you would like, you can add a dimension of how. How did it happen? Sometimes, the answer to how can be covered by what, when and where.