* What is the problem you want to solve?

Capstone Idea 2 – Data Cleaning : Health Data for Big Cities

The primary objective of this project is to apply Data Cleaning techniques so that the clean dataset could be used for meaningful purposes for further analysis. If the Dataset is not clean, it may turn out that the data set you’re analyzing isn’t really suitable for what you’re trying to do,

Here we are using a Dataset that illustrates health status of 26 of the US’ largest and most urban cities.

* Who is your client, and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

Primary audience is Govt Health Department in respective cities who could utilize the finding for disease control and prevention.

* What data are you using? How will you acquire the data?

This dataset illustrates health status of 26 of the nation’s largest and most urban cities as captured by 34 health (and six demographics-related) indicators. These indicators represent some of the leading causes of morbidity and mortality in the United States and leading priorities of national, state, and local health agencies. Public health data were captured in nine overarching categories: HIV/AIDS, cancer, nutrition/physical activity/obesity, food safety, infectious disease, maternal and child health, tobacco, injury/violence, and behavioral health/substance abuse.

The Dataset is available in data.world:

<https://data.world/health/big-cities-health>

* Briefly outline how you'll solve this problem.

Using Python skill, will clean the dataset to meet the following objectives:

* Be spread over multiple files.
* Have a lot of nuance, and many possible angles to take
* Require a good amount of research to understand.
* Be as “real-world” as possible

.

* What are your deliverables? Typically, this includes code, a paper, or a slide deck.

Deliverables will include – i) Code (via GitHub) ii) Document