**Capstone Project proposal**

Finalize one Capstone idea based on the feedback you got from mentor(s) and peers on your Section 1 submission, and based on your newly acquired understanding of the tools and data wrangling. Submit a project proposal - a short (1-2 page) document that answers the following questions:

What is the problem you want to solve?

1. regression model for predicting medical claim frauds.
2. regression model for predicting people at risk of being obese or any chronic disease.

What data are you going to use for this? How will you acquire this data?

Exploratory analyses followed by regression model for predicting the main culprit factors in patients’ readmission. The goal of this project is comparing two approaches for solving the problem through utilizing the same regression model/s. The first approach is non-selective which include as many involved factors as possible in the regression model. This model can be used by data scientists with no prior subject matter experience. The second approach is selective which is depending on prior research to determine the main physical, environment, and socio-economic factors and then incorporate them in different regression models to identify the best one for the given situation.

The data set is 10,000 records of virtual patients.

In brief, outline your approach to solving this problem (knowing that this might change later).

The approach will show the following:

1. Information about the data set to get familiarized with the collected variables of the given problem.
2. A-B test for exploratory analyses.
3. Statistics for testing the significance of the differences those have been shown in A-B test.
4. Testing regression model.
5. Applying the model on different problems (the 3rd and 4th capstones).
6. 2nd capstone idea: regression model for predicting medical claim frauds.
7. 3rd capstone idea: regression model for predicting people at risk of being obese.
8. The reason for utilizing the same regression model on different problems is to examine the transferability of the model between different cases or scenarios.

What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

The deliverables will be the following:

1. The documentation processes. It will be word document that explain the steps of the project, the utilized tools, the results, and the future suggestions.
2. List of visualizations those are produced by GUI application (such as Tableau) and those produced by python.
3. Python scripts of the project including data importing, wrangling, sub-setting, visualization, modelling, and future proposal for model development.

The proposal will be part of a GitHub repository for the proposed project. All code and further documentation will be added to this repository.

Once the mentor has approved the proposal, the GitHub repository URL will be shared on the community and ask for feedback.