1. **What is the problem you want to solve? Why is it an interesting problem?**

The problem here is one of trying to estimate the insurance premium that should be charged based on various factors of an individual. Insurance being one of the most important industry in business, a lot can be learned by studying the various factors of an individual that goes into charging the specific premium. Ability to predict the insurance amount would enable companies to optimally price to provide the best customer satisfaction as well as generate revenue and profit for the company.

1. **What data are you going to use to solve this problem? How will you acquire this data?**

The data is a CSV containing various features of an individual such as age, sex, bmi, #children, whether the individual is a smoker, geographic location.

The data is available at Kaggle for download and use.

3. **In brief, outline your approach to solving this problem. You might not know everything in advance, and this approach may change later. This might include information like:**

**a. Is this a supervised or unsupervised problem?**

This is a supervised problem wherein a dataset is trained to predict a specific outcome.

b. **If supervised, is it a classification or regression problem?**

This is a linear regression problem as it attempts to generate a continuous dollar value that represents the insurance premium.

c**. What are you trying to predict?**

We are trying to predict the insurance premium for an individual given the multiple features listed above.

d**. What will you use as predictors?**

We will use a test and train data set to compare and predict the accuracy of the insurance premium.

e**. Will you try a more “traditional” machine learning approach, a deep learning approach, or both?**

This Capstone project attempts a more “traditional” machine learning approach with an emphasis on developing a machine learning framework that adheres to the all stages of machine learning staring with data exploration, wrangling, cleansing, training and verification and final deployment of the model.

1. **What will be your final deliverable? Will it be an application deployed as a web service with an API or a more robust web/mobile app.**

The final deliverable is a highly accurate machine learning model/algorithm able to predict insurance premiums given a variety of input features.

It is expected that the application would be deployed as a web-service eventually.

1. **What computational resources would you need at a minimum to do this project? You may not have a very clear sense now but work with your mentor to come to an estimate. In real industrial applications, you’ll often be called upon to provide resource estimates at the beginning of a project. a. Processing power (CPU) b. Memory c. Specialized hardware such as GPUs**

A 1.9 GHz 16GB memory on a 64-bit operating system on x-64 based processor is expected to suffice the computational needs for the project.