Capstone II: Initial project Ideas

(1) What makes people happy?

Problem Statement: Happiness is the best measure of psychological health. In a society, more happiness means less problems – less crimes, less health concerns, etc. This gives more political and economic stability to a country.

The main questions are (1) Is happiness entirely related to emotions or there are social or economic features involved? (2) Can we predict the happiness of people using economic features? (3) What can a government do to make its people happy?

Clients: The local governments or government of all countries that want to make people happy.

Data: I use the data from the link below:
https://knoema.com/atlas/topics
This website contains many features that includes Agriculture, crime statistics, Demographics, Economy, Education and happiness index. Several of these data would be gathered together by web scraping from the link. Additional data would be acquired from the link
https://www.kaggle.com/unsdsn/world-happiness

Approach: We will begin with data exploration. We will present a comparative study of several country on several economic features. In addition, we will build a model that predicts the happiness based on other socio-economic features.

(2) Finding a great college:

Problem Statement: High school student aspire to go the college/university that match with their expectation. Finding a best match is always a complicated question. People consider several factors, such as affordability, performance of college, location. Knowing all these features in a single package will help student select their best choice.

Clients: High school students looking for colleges.

Data: I will mainly use the data provided by the US government.
https://catalog.data.gov/dataset/college-scorecard

In addition, college affordability will be obtained from the link:

https://catalog.data.gov/dataset/college-affordability-and-transparency-list-explanation-form-201617/resource/d0be8edb-33df-4319-b1fb-7e02793d65c5

And find additional data from Kaggle:
https://www.kaggle.com/flyingwombat/us-news-and-world-reports-college-data
https://www.kaggle.com/wsj/college-salaries#degrees-that-pay-back.csv

Approach: Aside from data exploration, we use unsupervised machine learning to classify colleges based on affordability and their performances.

(3) Spending for science:

Problem: My client, national science foundation (NSF) spends millions of dollars every year to science research. NSF wants to understand whether or not the spending has been utilized for the best possible outcomes. The goal of this project is to explore the previous awards and make recommendation to the client so that helps in making strategies in research spending.

Data: We use the data in the link:

https://www.research.gov/research-portal/appmanager/base/desktop?_nfpb=true&_eventName=viewQuickSearchFormEvent_so_rsr

Approach: This project focusses in understanding how effectively the funding has been utilized? Who are more productive (small grants vs big grants)? Who produces high impact research?