Capstone Project 1 Proposal

- ## What is the problem

With the development of the network, there are increasingly number of people choose to purchase on-line. Given the shortage of information online, customers are always struggling with issues such as the size, quantity, colors, and etc. Thus, overview of others' reviews can help us get a quick impression of the products.

However, many issues remained related to current review platform. 1) How can we indicate whether the review is a positive sentiment or negative based on Recommended, Rating or positive feedback factors? 2) Can we predict the positive or negative sentiment of unlabeled reviews? 3) Which words are most popular? 4) Which features are the most important to customer, color, price, and/or size?

- ## Potential Clients

Appropriate customer review analysis can both enhance the efficiency of online shopping and help the retailers to improve their online e-commerce and align their market strategy.

- ## Data Acquiring

I will use Women's E-Commerce Clothing Reviews on Kaggles Datasets. This dataset includes 23486 rows and 10 feature variables. Each row corresponds to a customer review and includes the variables: Clothing ID, Age, Title, Review Text, Rating, Recommended IND, Positive Feedback Count, Division Name, Department Name, and Class Name.

- **How to solve this problem**

    - Using data waggling skill to clean the dataset
    - Building a correlation heatmap to indicate the relationship between each factors
    - Using Natural Language Processing technology to analyze the customer reviews

- **What to deliver**

  I will use Python 3.7 in the Jupyter Notebook Environment to perform the coding. I will create a slide deck presentation and prepare a final report.