Project:
Submit 3 Ideas for Your
First Capstone Project

Jiacheng(Dylan) Qian
Data Science Career Track
Springboard

I looked over the Internet and tried to find some datasets that are educational and practical for data science beginner. There are many projects are recommended by Analytic Vidhya, a data science website for both beginners and professionals. From the recommendations for beginning learners, I limited my choice to intermediate level and advanced level in order to challenge myself. Additionally, I chose 3 topics that interested me the most. Here are the topics name and their short description.

1. **Black Friday Data Set**

   This data set comprises of sales transactions captured at a retail store. It's a classic data set to explore your feature engineering skills and day to day understanding from your shopping experience. It's a regression problem. The data set has 550069 rows and 12 columns.

2. **Yelp Data Set**

   This data set is a part of round 8 of The Yelp Dataset Challenge. It comprises of nearly 200,000 images, provided in 3 json files of ~2GB. These images provide information about local businesses in 10 cities across 4 countries. You are required to find insights from data using cultural trends, seasonal trends, infer categories, text mining, social graph mining etc.

3. **Chicago Crime Data Set**

   The ability of handle large data sets is expected of every data scientist these days. Companies no longer prefer to work on samples, they use full data. This data set would provide you much needed hands on experience of handling large data sets on your local machines. The problem is easy, but data management is the key! This data set has 6M observations. It's a multi-classification problem.

I will take more time to look into the dataset and narrow down my choices to one of these datasets. Also, some of these datasets are designed for people to practice machine learning. I will try to focus on exploratory data analysis for capstone project 1, and maybe use very basic machine learning at the end of this project.