Capstone Project Ideas (Module 1.2)
Phai Phongthiengtham
Springboard Data Science Intensive Course
July 10, 2017

Project 1: How do successful firms post their job vacancies?

Potential Clients: Human resource department and job board websites
Data:
1. Job postings from CareerBuilder (all over the US from 2012-2016)
2. Compustat Database (Data on firms)

Broad Idea: A company's success comes from each person's individual contributions and unique talents. Putting the right people in the right jobs is very important and this starts with a firm posting vacancies. In this project, I intend to explore what features of job posts can predict a firm's success (for example, profitability, Tobin's Q, stock market movement). To narrow the question down, I intend to focus on management positions.

Possible factors (very preliminary):
1. Length of a post
2. Explicit employment benefit packages

Table 1: Word Cloud from Management Positions

Project 2: Software Learning Recommendation Product

<u>Potential Clients</u>: Schools and job board websites
<u>Data</u>:
   1. Job postings from CareerBuilder (all over the US from 2012-2016)
<u>Idea</u>: In this project, I intend to create an algorithm that recommends which up-to-date software(s) a person should know given education, location and job prospect. To do so, I plan to use job postings from all over the US from 2012-2016.

Project 3: Predicting Video Game Sales and Ratings using Reddit Database

<u>Potential Clients</u>: Game companies
<u>Idea</u>: In this project, I intend to use Reddit database to build a model that can predict video game sales and ratings. Sentimental analysis of Reddit's comments would be the focus of the project.
<u>Data</u>:
1. Video Game Sales and Ratings
   - https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings

Data Sample (main dependent variables: *"Global Sales"* and *"Critic Score"*)

Name	Year of Release	Genre	Global Sales	Critic Score
Eagle Flight	2016	Adventure	0.07	72
Naruto Shippuden: Ultimate Ninja Storm 4	2016	Fighting	1.14	79
DiRT Rally	2016	Racing	0.37	85
Here They Lie	2016	Adventure	0.02	62
Call of Duty: Infinite Warfare	2016	Shooter	2.42	78
LEGO Marvel's Avengers	2016	Action	0.24	69
One Piece: Burning Blood	2016	Fighting	0.1	62
Skylanders Imaginators	2016	Platform	0.14	80
Rocket League	2016	Sports	0.37	85
Skylanders Imaginators	2016	Platform	0.1	78

2. Reddit Database: Reddit offers open source API service. The first link provides an overview. The second link is an application for API. The final one is python module *praw* that allows access to Reddit's API.
   - https://www.reddit.com/dev/api
   - https://github.com/reddit/reddit/wiki/OAuth2
   - https://pypi.python.org/pypi/praw