

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

Project 1: Predicting Video Game Sales and Ratings using Reddit Database

Potential Clients: Game companies

Idea: In this project, I intend to use Reddit database to build a model based on sentimental analysis that can predict video game sales and ratings.

Data:

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>

Project 2: Understand Crimes in Chicago Area through the Lens of Data Science

Idea: In this project, I intend use publicly available data on crimes in Chicago to predicts crimes and whether an arrest was made. Potential features include time of the day, distance to the closest police station, buildings nearby (such as library, school) and whether there had been a crime in proximity before.

Data:

1. Crimes in Chicago (2001-present)

- <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>

Data Sample

Case Number	Date	Block	Primary Type	Description	Arrest	Location
HV445884	8/24/12 17:00	053XX S MARSHFIELD AVE	CRIMINAL DAMAGE	TO VEHICLE	FALSE	(41.796703956, - 87.665873032)
HV444536	8/23/12 15:30	033XX N MARSHFIELD AVE	THEFT	FROM BUILDING	FALSE	(41.942257928, - 87.669951491)
HV444338	8/6/12 9:00	009XX S WABASH AVE	THEFT	FROM BUILDING	FALSE	(41.870334861, - 87.625899331)
HV445887	8/25/12 1:00	009XX W BELMONT AVE	THEFT	FROM BUILDING	FALSE	(41.939918899, - 87.653369734)
HV444538	8/24/12 9:00	035XX W 57TH ST	THEFT	\$500 AND UNDER	FALSE	(41.789749572, - 87.711789147)

2. GIS and Geospatial Data in Chicago Area (for example, police station, school)

- <https://www.cityofchicago.org/city/en/depts/doit/provdrs/gis.html>

