Data Science Final Project

By: Lucas Gral, Ben Paulson, and Thomas Benzshawel

For our project proposal, we propose predicting whether a game review on steam is positive or negative through sentiment analysis. We will be using this model to get a more specific steam review from a "recommend or don't recommend" score. ie, a percentage of positive from sentiment analysis. This will then lead us to predict the average score from critic reviews. Since more often than not, the steam percentage review is not representative of the critic review (due to limitations of the steam review system).

Our data set is a kaggle dataset of reviews from 565K games. The total number of reviews is 6.4 million, each with review text, a review score, and review votes.

	app_id	app_name	review_text	review_score	review_votes
0	10	Counter-Strike	Ruined my life.	1	0
1	10	Counter-Strike	This will be more of a "my experience with th...	1	1
2	10	Counter-Strike	This game saved my virginity.	1	0
3	10	Counter-Strike	• Do you like original games? • Do you like ga...	1	0
4	10	Counter-Strike	Easy to learn, hard to master.	1	1
5	10	Counter-Strike	No r8 revolver, 10/10 will play again.	1	1
6	10	Counter-Strike	Still better than Call of Duty: Ghosts...	1	1
7	10	Counter-Strike	cant buy skins, cases, keys, stickers - gaben ...	1	1
8	10	Counter-Strike	Counter-Strike: Ok, after 9 years of unlimited...	1	1
9	10	Counter-Strike	Every server is spanish or french. I can now f...	1	0

Our main objective while analyzing our dataset is to identify words that signal sentiment for negative and positive reviews. After analyzing or dataset, we plan to create a model to predict the sentiment of a game review beyond simply good or bad.