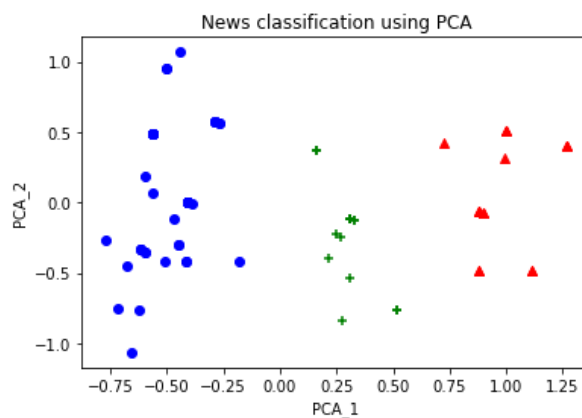# Team Renegade

## Preliminary Project Report – Data Science Bootcamp

o   The goal of the project is to investigate the stock movements of the market corresponding to a specific piece of information. The information that we work with decide the dataset. There are two types of information we want to focus on: the news and financial reports release.

- News/Posts: We collect the headlines from google news and transform them into vectors of words (strings). After that, we utilize a corpus of common words from the headlines and generate an embedding to represent these headlines as vectors of 0 and 1 (since vectors of numbers are easier to work with in machine learning models than vectors of words). In terms of labels, we will label the headline as positive, negative or neutral.
- Financial reports release: We collect the financial ratios from some selected websites for some companies and create dataframes containing all these ratios for every company. With these financial ratios, we will analyze the relationship between the ratios and the stock movements of the corresponding stock price according to the financial release.

o   The stakeholders of the project are banks, financial analysts and any person that look for materializing information release (most likely frequent traders) into stock movements.

o   Modeling approach:

- For the news/posts component, we will run PCA to detect the news clusters within the dataset. If we can find such clusters, we have evidence that the classification should be successful (the news can be classified). The preliminary result on a small dataset of news is shown below, which indicates that the news can be classifiable:



- For the financial release component, we will run a regression/neural network over the financial ratios and the prior stock price movements as the regressors and the stock price movements on a determined time window as the response. We will figure out how a certain financial ratio release will affect the stock price movements on that day of release or a small period of time after the release. Currently we do not have the preemptive analysis since gathering the dates of financial release and the stock price on those specific dates takes time.