NETS 213: Final Project questionnaire

\* Required

Top of Form

**How many teammates are in your group?**\*

 3

**May we have your permission to feature your project on the class website?**\*

We will use it to advertise the class and as an example to next year's course.

* +   ***Yes***
  +   No

**Name 1**\*

 Matt Labarre

**PennKey username 1**\*

mlabarre

**Team member 1 - Can we list your name listed alongside your project?**\*

FERPA guidelines require that we get your explicit permission. It's OK to say no!

* +   ***Yes***
  +   No

**Name 2**\*

Jeremy Laskin

**PennKey username 2**\*

 jlaskin

**Team member 2 - Can we list your name listed alongside your project?**\*

FERPA guidelines require that we get your explicit permission. It's OK to say no!

* +   ***Yes***
  +   No

**Name 3**

 Chris Holt

**PennKey username 3**

holtc

**Team member 3 - Can we list your name listed alongside your project?**

FERPA guidelines require that we get your explicit permission. It's OK to say no!

* +   ***Yes***
  +   No

**Name 4**



**PennKey username 4**



**Team member 4 - Can we list your name listed alongside your project?**

FERPA guidelines require that we get your explicit permission. It's OK to say no!

* +   Yes
  +   No

Basic Project Info

**Name of your project**\*

Crowdsourcing for Cash Flow



**Give a one sentence description of your project.**\*

Examine the correlation between sentiment expressed in tweets about a certain company and that company’s performance in the stock market during that time period.

Please use the name of the project in your description.



**URL to the logo for your project**\*

Create a PNG file, and save it in your github repo. Include the full path to your logo here (prefix with[https://github.com/](https://www.google.com/url?q=https%3A%2F%2Fgithub.com%2F&sa=D&sntz=1&usg=AFQjCNHReqsuKT6C86HcgL4TbSevF24rxQ))



**What problem does it solve?**\*

This project does not specifically solve a problem, but rather tackles the challenge of understanding stock market movement, and attempting to correlate it to Twitter activity. Instead of solving a problem, this concept provides a market research tool, and would allow retail investors (people like you and me) to conduct the same sentiment research conducted by analysts at large firms like Goldman Sachs and BlackRock.



**What similar projects exist?**\*

 There are some sentiment analysis tools available currently. These include StockFluence, The Stock Sonar, Sentdex, and SNTMNT, all of which perform sentiment analysis on Twitter or other news feeds.

**What type of project is it?**\*

* +   Human computation algorithm
  +   ***Social science experiment with the crowd***
  +   A tool for crowdsourcing
  +   A business idea that uses crowdsourcing
  +   Other: 

**What was the main focus of your team's effort**\*

* +   Engineering a complex system
  +   ***Conducting an in depth analysis of data***
  +   Something in between

**How does your project work?**\*

Our project begins by writing a Twitter-scraping script that retrieves tweets that contain #AAPL or @Apple. Once these tweets are received, we create a sentiment analysis crowdsourcing job on CrowdFlower. This job asks the crowd to evaluate the sentiment expressed in the tweet from the options of: negative, neutral, positive, or “This tweet is not about the company Apple.” Once this is complete, we retrieve stock market data for AAPL over the desired time span, and perform advanced analytics to attempt to correlate the sentiment data with the market data. This involves bucketing the sentiment data bucketed by 30 minute time intervals, calculating the z-scores of these buckets, and plotting these values and the similarly bucketed stock market positions, and comparing the data. We also calculated day-by-day correlation values to see how the sentiments compared to the movement of the stock price on a given day.



**URL to the flow diagram for your project**\*

Create a PNG file, and save it in your github repo. Include the full path to your diagram here (prefix with[https://github.com/](https://www.google.com/url?q=https%3A%2F%2Fgithub.com%2F&sa=D&sntz=1&usg=AFQjCNHReqsuKT6C86HcgL4TbSevF24rxQ)).



**Provide a link to your final presentation video**\*

Give the full path to your Vimeo video.



The Crowd

**Who are the members of your crowd?**\*

 CrowdFlower participants

**For your final project, did you simulate the crowd or run a real experiment?**\*

* +   Simulated crowd
  +   ***Real crowd***

**If the crowd was simulated, how did you collect this set of data?**



**If the crowd was simulated, how would you change things to use a real crowd?**



**If the crowd was real, how did you recruit participants?**

We used CrowdFlower, which recruited participants for us



**How many unique participants did you have?**\*

257



Incentives

**What motivation does the crowd have for participating in your project?**\*

* +   ***Pay***
  +   Altruism
  +   Enjoyment
  +   Implicit work
  +   Reputation
  +   Other: 

**How do you incentivize the crowd to participate?**\*

We used CrowdFlower’s payment system to incentivize the crowd to participate. Our job was very large – we had 3,969 units, and we wanted at least 3 judgments per unit. Since we wanted to minimize cost, we initially set out to pay 2 cents per job, and each job contained 10 tweets to evaluate as either negative, neutral, or positive sentiment. However, this initial payment scheme was not a great enough incentive for the crowd, and we had very low participation rates to start. After modifying our payment plan several times, we ended up paying the crowd 5 cents to evaluate 15 tweets.



**Did you perform any analysis comparing different incentives?**\*

* +   ***Yes***
  +   No

**If you compared different incentives, what analysis did you perform?**

We did not perform a rigorous analysis, but we actively monitored how the crowd responded to different payment options. We had to vary our price per HIT and the number of tweets per HIT many times before we finally found an effective incentive plan.



**If you have a graph analyzing incentives, include the HTML for a Google graph here.**



**Caption for your graph (incentives)**



What the crowd gives you

**What does the crowd provide for you?**\*

The crowd provided us very simple data: whether the sentiment in the tweet towards Apple was positive, neutral, or negative. We wanted to keep this simple because we believed that adding options such as “very positive” and “very negative” would cause for too much subjectivity and variation amongst the crowd. We would have had to have had many more judgments per tweet in order to develop a consensus on the sentiment, and we did not believe that adding such options were necessary.

**Is this something that could be automated?**\*

* +   Yes
  +   ***No***

**If it could be automated, say how. If it is difficult or impossible to automate, say why.**\*

 Yes, we could automate this process, but that would involve creating a sentiment analysis platform, which is very difficult for vocabulary like that seen in stock articles. Additionally, we would still need to manually label large amounts of training data, which would have been impossible given the time limits. We could use twitter queries to presort positive and negative tweets for the stocks, but then there would be no crowd aspect to this particular project.

**Did you train a machine learning component from what the crowd gave you?**\*

* +   Yes
  +   ***No***

**If you trained a machine learning component, describe what you did.**



**Did you analyze the quality of the machine learning component?**

For instance, did you compare its quality against crowd workers using an n-fold cross validation?

* +   Yes
  +   ***No***

**If you have a graph analyzing a machine learning component, include the HTML for a Google graph here.**



**Caption for your graph (machine learning component)**



**Did you create a user interface for the crowd workers?**\*

Answer yes even if it's something simple like a HTML form on CrowdFlower.

* +   ***Yes***
  +   No

**If yes, please give the URL to a screenshot of the crowd-facing user interface.**

Save the screenshot as a PNG file, and put it in your github repo. Include the full path to your image (prefix with [https://github.com/](https://www.google.com/url?q=https%3A%2F%2Fgithub.com%2F&sa=D&sntz=1&usg=AFQjCNHReqsuKT6C86HcgL4TbSevF24rxQ)). You can include multiple screenshots, one per line.

 Screenshots in docs folder

**Describe your crowd-facing user interface.**

We kept our interface simple to make it as easy as possible for the crowd. We hoped that simple and conscise interfaces and instructions would heighten the accuracy of the crowd’s work. We also wrote robust definitions for what constitutes a positive, negative, and neutral tweets, again to heighten accuracy. 

Skills

**Do your crowd workers need specialized skills?**\*

* +   Yes
  +   ***No***

**What sort of skills do they need?**\*

 We targeted our job towards members of CrowdFlower in the U.S. only so that they were fluent in English and could understand the tweets well. This is the only skill necessary.

**Do the skills of individual workers vary widely?**\*

* +   Yes
  +   ***No***

**If skills vary widely, what factors cause one person to be better than another?**



**Did you analyze the skills of the crowd?**\*

* +   Yes
  +   ***No***

**If you analyzed skills, what analysis did you perform?**

How did you analyze their skills? What questions did you investigate? Did you look at the quality of their results? Did you analyze the time it took individuals to complete the task? What conclusions did you reach?



**Do you have a Google graph analyzing skills?**

* +   Yes
  +   ***No***

**If you have a graph analyzing skills, include the HTML for a Google graph here.**



**Caption for your graph (skills)**



Quality Control

**Is the quality of what the crowd gives you a concern?**\*

* +   ***Yes***
  +   No

**How do you ensure the quality of the crowd provides?**\*

Quality was of significant concern. Since each individual task of analyzing sentiment in a tweet is quick and simple, there was concern that workers would hastily evaluate as many tweets as possible to receive more money. This concern was exacerbated by the fact that we did not pay workers much, and low paying, high volume jobs can lead to poor quality from the crowd. In order to ensure high quality results, we created 104 test questions, and deployed our job in quiz mode, where the workers had to answer a minimum of 5 quiz questions with at least 70% accuracy. Additionally, each HIT had one test question in it.

This quality control method proved successful. CrowdFlower’s Contributor analytics file reports a “trust\_overall” score for each individual worker. The average trust was 0.85. However, this number is slightly skewed, because some workers who provided zero judgments were given trusts of 1. After filtering out these workers, we still received a high average trust of 0.79. Additionally, we calculated a weighted-trust metric, where the trust\_overall was multiplied by the number of judgments that the worker made, allowing us to calculate an average trust-per-judgment value. This value was 0.83. All of these metrics are very close in value, which points to a fairly consistent level of quality across workers. Thus, we can conclude that our quality control mechanism was successful, and maintained a high level of quality throughout the job.

If quality if a concern, then what did you do for quality control? If it is not a concern, then what about the design of your system obviates the need for explicit QC? This answer should be substantial (several paragraphs long).



**Did you analyze the quality of what you got back?**\*

For instance, did you compare the quality of results against a gold standard? Did you compare different QC strategies?

* +   ***Yes***
  +   No

**What analysis did you perform on quality?**

We did not preform highly robust quality control analytics, because a simple review of the contributors file suggested that the quality was high. The metrics that were calculated are discussed above, namely, an average trust\_overall quantity, an average trust\_overall for workers that contributed more than 0 judgments, and a weighted trust metric that factored in exactly how many judgments each worker contributed, giving us an average trust-per-judgment value. All of these values were high, at around 0.8.

**Do you have a Google graph analyzing quality?**

* +   Yes
  +   No

**If you have a graph analyzing quality, include the HTML for a Google graph here.**



**Caption for your graph (quality)**



Aggregation

**How do you aggregate the results from the crowd?**\*

We conducted robust aggregation work on the results from the crowd. First, we read through the CrowdFlower results, and for each tweet, we assigned a negative sentiment a value of -1, a neutral sentiment a value of 0, and a positive sentiment a value of +1. Then, we bucketed each tweet by the timestamp of when they were tweeted – the buckets were 30 minute intervals. For each bucket, we summed up the scores of the tweets in the bucket, assigning an overall score per bucket. Additionally, we only considered the tweets that were tweeted during stock market hours (9:30am – 4pm, Mon-Fri). Once the scores for each bucket were determined, we calculated the average bucket score, and the standard deviation of the bucket scores, allowing us to calculate z-scores for each bucket. This normalized the data, accounting for the varying amount of tweets per bucket.



**Did you analyze the aggregated results?**\*

* +   ***Yes***
  +   No

**What analysis did you perform on the aggregated results?**\*

The analysis performed on the aggregated results is described above.

What questions did you investigate? Did you compare aggregated responses against individual responses? What conclusions did you reach?



**Do you have a Google graph analyzing the aggregated results?**

* +   Yes
  +   ***No***

**If you have a graph analyzing the aggregated results, include the HTML for a Google graph here.**



**Caption for your graph (aggregation)**



**Did you create a user interface for the end users to see the aggregated results?**\*

* +   Yes
  +   ***No***

**If yes, please give the URL to a screenshot of the user interface for the end user.**

Save the screenshot as a PNG file, and put it in your github repo. Include the full path to your image (prefix with [https://github.com/](https://www.google.com/url?q=https%3A%2F%2Fgithub.com%2F&sa=D&sntz=1&usg=AFQjCNHReqsuKT6C86HcgL4TbSevF24rxQ)). You can include multiple screenshots, one per line.



**Describe what your end user sees in this interface.**

This can be a short caption for the screenshot. Alternately, if you put a lot of effort into the interface design, you can give a longer explanation of what you did.



Scaling Up

**What is the scale of the problem that you are trying to solve?**\*

Ultimately, the problem that is being tackled is discovering the correlation between tweets about a company and its performance in the stock market. In this project, we simply looked at standard tweets from standard Twitter users; however, this concept can be scaled greatly to look at specific types of tweets, and who is tweeting them. For example, our project did not ask our crowd to read articles linked in tweets, and judge the sentiment of such articles. Additionally, we did not weight tweet sentiment for users that were news sources or prominent figures in the financial industry (Carl Icahn, Warren Buffet, etc.). We simply looked at standard Twitter users who were primarily tweeting about personal complaints (or positive experiences) with Apple, which, understandably, is not very tightly correlated to stock market movement. Scaling this project to include more companies and collecting a more specialized set of tweets can make for an interesting study on the correlation of tweets to stock market movement on a large scale.



**Would your project benefit if you could get contributions from thousands of people?**\*

* +   Yes
  +   ***No***

**If it would benefit from a huge crowd, how would it benefit?**



**What challenges would scaling to a large crowd introduce?**



**Did you perform an analysis about how to scale up your project?**

For instance, a cost analysis?

* +   Yes
  +   No

**What analysis did you perform on the scaling up?**

What questions did you investigate? What conclusions did you reach?



**Do you have a Google graph analyzing scaling?**

* +   Yes
  +   No

**If you have a graph analyzing scaling, include the HTML for a Google graph here.**



**Caption for your graph (scaling up)**



Project Analysis

**Did your project work?**\*

Our project did work, even though the exact results did not necessarily point to tight correlation. We were able to utilize crowdsourcing to get sentiment analysis on nearly 4000 tweets about Apple, and then write programs to synthesize this data, and correlate it to Apple’s performance in the stock market. The success of this project was not to be measured by whether or not the correlation existed, but rather if a valid analysis could be conducted, specifically, via crowdsourcing. In our case, we conducted a very valid analysis – no aspect of our project contains skewed or inaccurate data or analytical methods. Therefore, even though our results did not point to significant correlations, we have a very good idea as to how we can repeat the study and obtain a more interesting outcome.

How do you know? Analyze some results, discuss some positive outcomes of your project.



**Do you have a Google graph analyzing your project?**

* +   Yes
  +   No

**If you have a graph analyzing your project, include the HTML for a Google graph here.**



**Caption for your Google graph (project analysis)**



**What were the biggest challenges that you had to deal with?**\*

One of the biggest challenges that we had to deal with was restructuring our project slightly when we ran into significant issues with CrowdFlower. Initially, we wanted to post three jobs on CrowdFlower to study three different companies (Apple, Microsoft, and Google), and, since these companies are competitors, conduct further analysis on how tweets about one company affected the stock market performance of another. Unfortunately, our CrowdFlower subscriptions expired after Apple’s job ran, and, after exploring the possibility of positing the other two jobs on MTurk, it was clear that this was not a viable option, and would taint the quality of our data. Therefore, we switched gears slightly, and decided to just conduct a more robust analysis on Apple’s tweets and market performance. Additionally, we depended on the twitter search queries to get a representative section of tweets about the stocks, but our queries may not have been robust or varied enough to get enough tweets about the stocks.

**Where there major changes between what you originally proposed and your final product?**\*

* +   Yes
  +   ***No***

**If so, what changed between your original plan and your final product?**

The change that we made to our project was not major, we were just unfortunately unable to repeat the analysis on several companies, and had to stick to simply one.

**What are some limitations of your product?**\*

We do not believe that our project consists of many sources of error. The only possible source of error would be inaccurate sentiment analysis from the crowd, but we implemented a strong quality control method that returned highly trusted results, according to CrowdFlower. The analytics that we performed on the CrowdFlower data was very standard, and did not introduce new sources of error. However, we would have liked to have either conducted this study on a longer time scale, or on multiple companies, to have obtained more data and thus validate our results further. Additionally, we could improve the twitter search queries to get more relevant results, as well as collecting a much longer duration of tweets. Furthermore, we could try correlating current stock prices to past twitter sentiment, with the idea that it takes some time for the sentiment to effect market prices due to trading delays.

If yours is an engineering-heavy project, what would you need to overcome in order to scale (cost/incentives/QC…)? If yours was a scientific study, what are some sources of error that may have been introduced by your method.



**Did your results deviate from what you would expect from previous work or what you learned in the class?**\*

* +   Yes
  +   ***No***

**If your results deviated, why might that be?**



Other info (optional)

**Is there anything else you'd like to say about your project?**

If you have additional information about your project that didn't fit into the above questions, put it here.



|  |
| --- |
| *Never submit passwords through Google Forms.* |

Bottom of Form

Powered by

[Google Forms](https://www.google.com/forms/about/?utm_source=product&utm_medium=forms_logo&utm_campaign=forms)

This content is neither created nor endorsed by Google.   
[Report Abuse](https://docs.google.com/forms/d/1SndUhVcdVr1DI10tEML6gKxMrh4tBz5LZ7hVbOaBeqQ/reportabuse?source=https://docs.google.com/forms/d/1SndUhVcdVr1DI10tEML6gKxMrh4tBz5LZ7hVbOaBeqQ/viewform?usp%3Dsend_form) - [Terms of Service](http://www.google.com/accounts/TOS) - [Additional Terms](http://www.google.com/google-d-s/terms.html)

Screen reader support enabled.