Thanks David.

**Here is our progress on the project so far.**

We came up with data architecture based on assessment of demands of the processing and analysis. I will talk about this a little bit more detail on the next slide.

We also tested twitter API and collected tweets relevant for the two products we selected.

We tried some basic sentiment analysis which was to count positive words and negative words in a tweet to tell whether tweet is positive or negative. Objective was to understand mechanics of sentiment analysis.

**We envision simple data architecture.**

From the left, we collect tweets utilizing tweepy and then ingest the data into HDFS running on AWS instance. Then use pyspark to process the data and use NLTK for sentiment analysis. Lastly, visualize insights we gained from the analysis by using tableau.

**There are five key tasks to tackle in the remaining weeks**

First, we need to collect larger data. Since we have already figured out basics of twitter APIs, our focus is putting data out of twitter APIs without damaging twitter credential.

We also need to define analytics metrics for tweets analysis. We are interested in tweet volume, trend, sentiment analysis, words association, influencer and so forth. But we may focus a few of them instead of analyzing all of them.

If we have time left, we might explore better options to load the data into HDFS than simply use “HDFS put” command.

Once the data is stored in HDFS, we analyze it with pyspark and sentiment analysis modules.

Lastly, visualize the insights we gained from the analysis by using tabluea.

ma