**Tableau - Demo**

Thanks Cheng for the demo.

I’ll cover tableau visualization in simple two steps, assuming we have tweets collected and processed. Step one is to generate charts that show twitter trend, the trend with sentiment breakdown, overall sentiment summary, top retweeted tweets, and top favorite tweets.

Second step is to put the charts into Dashboard and StoryLine in order to summarize multiple visualizations. The visual on the left bottom is to show overview and the two visuals on the right bottom are for popular tweets for deeper analysis.

**We can think some use cases of these visualizations.** For example, if there is a tweets pattern that always peaks at evening on weekdays and is quite on weekends. It is better for a product marketer to launch marketing programs before evening on weekdays when people tweets about the product.

Another example is that looking at popular tweets to see if there is any particular tweets require attention. In the tweets we collected for Amazon Echo, there are some tweets about Amazon Echo is being hacked and turned into a spy device. These tweets were retweeted many times. Product marketing team might take actions to ease their concerns.

**Project Goals and Outcomes**

Overall, we achieved the goals that we initially set in the planning phase. We built a repeatable process for data collection, storage and analysis of tweets. By changing a search keyword, we can repeat the process for other products.

We did sentiment analysis of the products that we chose. We verified that simple sentiment analysis can uncover insights and bring values for marketers.

One missing piece is serving layer in our data architecture. Since we didn’t implement SQL or NoSQL database, the architecture doesn’t have capability for querying.

**Architecture – Production**

Our project focused on building prototype for new product sentiment analysis. In order to bring our prototype to production, we need to consider scale along with other factors. The diagram describes the data architecture for production.

**Architecture – Pilot vs. Production**

Instead of using the diagram, let me use a simplified chart to explain how to transform from pilot to production component by component.

* First, move batch data collection with simple python script to a combination of batch data collection / real-time data collection. Kafka and Apache storm are candidates for the implementation of real-time data collection. With that we can collect/analyze tweets real-time.
* Second, move data storage from EBS to HDFS and integrate data table schema to Hive so that we can build serving layer.
* Third, along with the change in storage, we need data processing for distributed computing. Pyspark is a good candidate since it allows us to reuse some of our current python codes.
* Fourth, better natural language analysis module such as NLTK should uncover more insights.
* Lastly, we want to improve tableau visualization to support real time updates instead of manual data connections so that we can automate the process.

**Getting to Production**

In addition to the changes in architecture that I have covered, changes for usability and features are also important. One idea is to build user interface GUI to allow custom key word search. Another improvement is to make tableau dashboard to provide prescriptive actions. For example, if negative tweets exceed 50% of total tweet volume in one hour, the dashboard send alert message so that marketers can take some actions.

With these changes, we believe our new product sentiment analysis become more robust and ready for scale and commercial use.

Thanks for listening to our presentation.