Brandon Croarkin and Lou Zhang

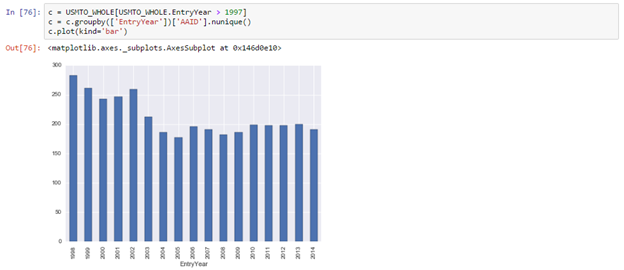
7/13/2016

**Springboard Data Story – Prediction USMTO Submissions**

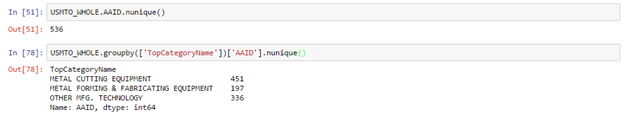
The US Manufacturing Orders (USMTO) survey collects data from US-based manufacturing companies and compiles orders data to capture a monthly picture of the US manufacturing market. USMTO is a free survey and currently collects data from around 180 different companies every month. It is the duty of the analysts at AMT to ensure that current participating companies submit their orders data and to reach out to as many manufacturing companies as possible for recruitment in free time. Due to the other various duties of analysts, we do not always have the time and resources devoted to selectively reach out to companies that are in danger of not submitting. The following visualizations use USMTO data to display participation trends.

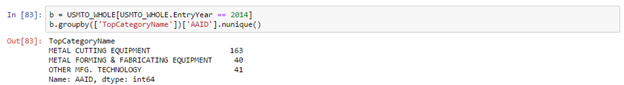
**The Companies**

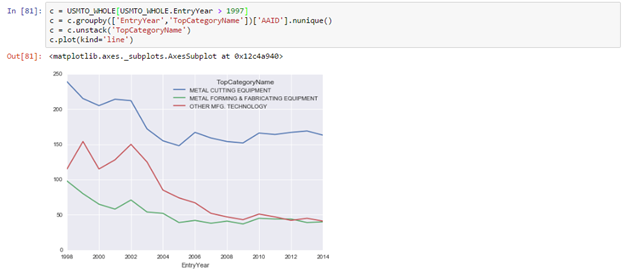
Over the years, the composition of USMTO companies has shifted dramatically. Our dataset spans the beginning of the survey in 1998 to 2014. During this time the number of companies participating has shifted from 283 to 191. That number is even lower today, around 180.



In total, 536 companies have participated in the survey across the three categories. The largest category has consistently been Metal Cutting Equipment. The largest drop in participation has been seen in Other Mfg. Technology. You can also see from this that companies participate across categories.





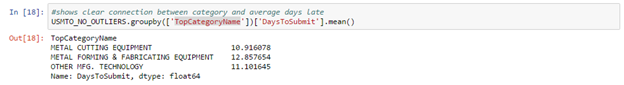


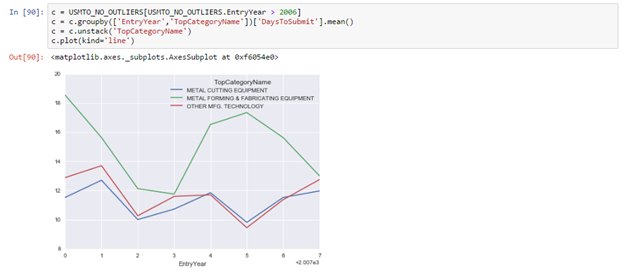
The goal of this project is to assess the likelihood of a participant to submit so our analysts can focus their attention on the companies that are most likely to not submit and drop from the survey. We are assuming that companies that take longer to submit are more likely to drop from the survey, but this is a hypothesis that we would need to investigate further.

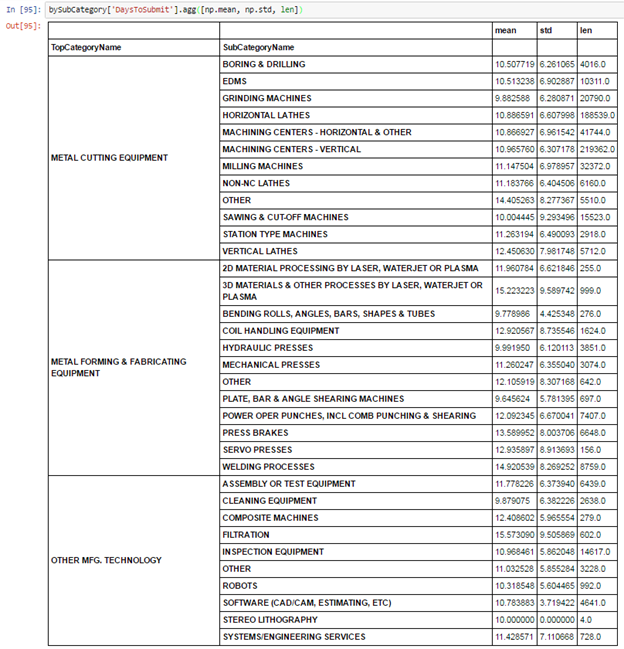
This would also give us more time to recruiting new companies to the program so the survey can get back to the numbers it once had. Higher participation would result in a better, more accurate survey.

**The Categories**

USMTO consists of three parent categories (Metal Cutting, Metal Forming, and Other Manufacturing Equipment) under which fall numerous Sub-Categories and specific machine codes. Each of these categories presents different administrative problems due to various factors. Our Metal Cutting category has the best market coverage, capturing nearly 80% of the total market. Metal Forming and Other Manufacturing however, capture much lower shares of the markets. Consequently, participants are more likely to submit their Metal Cutting orders as opposed to the other categories since the reports they get out of it are more valuable. Metal Forming is dominated by two main companies that dropped from the program. This has resulted in our Metal Forming category having much worse coverage than our other categories. Participants are now less likely to submit into this category as the survey results that come out of it are less useful.





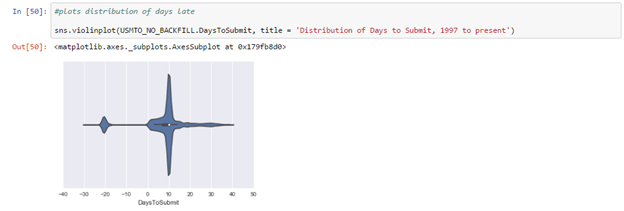


**Important Dates**

There are many dates that are very important in the submission patterns of USMTO. Data is technically due on the 10th of the following month. In order to be eligible to view USMTO in Advance, you need to submit your data by the 15th. USMTO in Advance is an app that provides an insight into the month’s orders before the official reports are released. It does so by comparing all company orders that have been submitted by the 15th for the report month and comparing it to the sum of those same company’s orders from the month before. Reports are processed on the first Monday of the month after the month after the order month. For example, May reports are processed on the first Monday of July. More personalized calls and emails are usually sent to delinquent participants during the week prior to this, with the larger customers typically being the customers that we focus on.

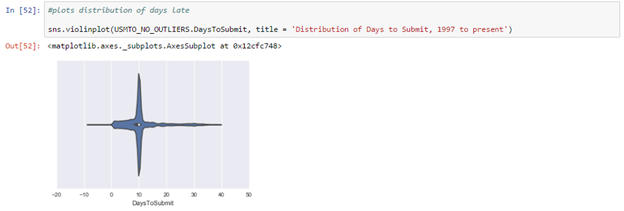


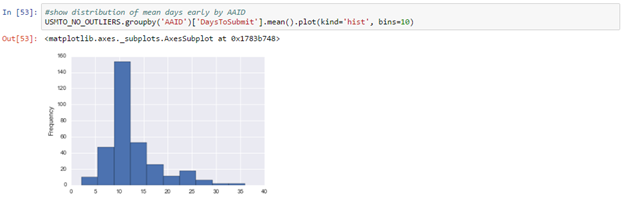
Initially plotting the Days to Submit in a Violin Plot gives us this output. With a quick glance we can see that there is a grouping a little in front of 0 showing that most people submit around this time range. However, the outliers in the graph make it unreadable. These outliers are from companies that backfill data. Companies are only able to view data for periods that they submit data for. So, a company that enters the survey in 2016 who wants to see data from 2006 would need to submit their company data from 2006. This results in these massive outliers. In order to make our analyses relevant to companies that are active in the current period and remove outliers we need to remove these backfills.



Now, we can much more clearly see some of the patterns in data submission. We see that starting at 0, there is a jump in data submission with a spike right before 10 days as companies scramble in trying not to be “late”. Automatic emails are sent during these dates reminding all companies to submit. After this the automated emails stop. Now, we can see a steady trickle of companies in with a minor drop after the 15th when USMTO in Advance is no longer available, and a minor increase between the 25th and 35th where the calls and emails to participants are being made. What is up with the spike around -20 though? After investigating this we discovered that there was a group of companies that were submitting data for a month before the month was over. It also turned out that a lot of these submissions were companies submitting 0. Submitting 0 still counts as submitting for the month. While some of these may have been data entry errors, we found a few companies that were regularly submitting early, we believe with the intent to just view the reports without submitting any of their company data. Many of these companies have a fear that their company’s confidential data may be leaked.

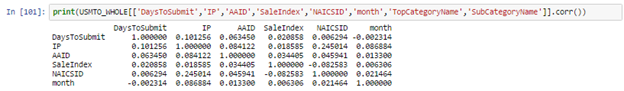
After adjusting for this, we can finally make a graph that shows the trends of submissions through the month for the appropriate range. This data can also be visualized with a histogram, which is shown below the violin plot.





**Next Steps**

Running all these tests and visualizations on our data set has shown us many interesting trends. However, there also do not appear to be many variables that have strong correlations with Days To Submit.



We have recently discovered a new data source that we are looking to add to our models, USMTO report usage. Currently, we don’t track how often a company logs into USMTO, but we do capture how often they run reports. Within USMTO there are numerous reports that you can run to see orders filtered and broken out by different criteria. We hypothesize that companies that run reports more often are more likely to submit data on time. Initial hypothesis testing has confirmed this theory and we believe it can be used to improve our models. The setback is that our databases are currently only holding 90 days of this data at a time. Our initials models and data testing have not used 2016 data, so this data would not currently apply. It also means that we don’t have historical data to test these theories on. In conclusion, our data story reveals significant insights into the data, and has allowed us to pursue further options to delve deeper into the data and come up with new ways to slice and dice our project in a meaningful way.