Sameer Rau

CSCI 4243W Writing Assignment #3

The Data Science market as of 2010 is worth roughly 100 billion dollars, a number that has increased by roughly ten percent each year within the past decade. Therefore, The Data Visualization Project falls under a division in analytics that is in need of an innovation that is able to look at a greater amount of data simultaneously that would otherwise fall to the wayside to other data sets or spreadsheets that may be deemed to be more important. Sure, there are currently existing statistical packaging platforms that do a reasonable job of getting the data analyzed at a respectable rate, however each tool comes with its own shortcomings. For example, Stata, SAS, and R requires an analyst to enter in statements or commands before the actual analysis itself is carried out. Moreover, all three have space constraints, so the size of the spreadsheet can not be overly large, and only one spreadsheet can be looked at any time (it is possible to merge the spreadsheet, but again the final spreadsheet can go over the space limits, and merging spreadsheets can lead to undesired consequences). With the market continuously growing at an exponential rate, the expectation is that similar innovations will be developed, however currently no such product has been formulated in the marketplace. The Data Visualization Project combines automated analysis with aesthetically pleasing graphical components that makes it a unique tool in a field that has several products that do similar forms of analysis in the same way.

The intended market that The Data Visualization Project aims to target include all analysts, from those working in smaller companies or startups to larger companies to educational purposes for students who may be tasked with working with larger data. The basic features of The Data Visualization Project are proposed to be free, so it would directly benefit those working in smaller companies where the funding may be needed to be allocated elsewhere, or students attending college that may not have the money to pay for an expensive tool. The main market driver for this innovation is the premise that the analysis is automated. As mentioned previously, in order to use one of the currently existing statistical packages, one has to have a working knowledge as to how to produce a graph or run a regression. Not only will the demand be high to have a data tool that looks at as much data as possible, but it will greatly benefit those who do not know a great deal of commands in any of the platforms. Most of the time spent learning new libraries and doing the analysis can instead be shifted towards drawing conclusions and making decisions as to what steps to take next.

There are several potential risks in bringing the innovation to a larger stage. One risk would be handling malicious input if one were to import a spreadsheet that has unusual values in a column that would not be expected, and another would be the format of the spreadsheet itself. Currently, for The Data Visualization Project to function as intended, the data has to be placed in a certain way; the columns of the variable names go at the top with all the corresponding values going below each respective column. However, if the data itself is not presented in such a way but the numbers are all over the place, the analysis may still go through and the results returned may not be of use or meaningful. In addition to any formatting issues that may occur, space limitations will still exist in how much data can be seen at any point. A database is being used to store multiple spreadsheets as a means of looking at as much data at one time to discover trends that may exist within spreadsheets. Unfortunately, importing too many datasets or file sizes that get way too large can crash the analyzing process, resulting in the product being unusable. Any one of these factors, or a combination of a couple may result in the individual or company utilizing the tool to abandon it in favor of another currently existing tool that does not have the same flaws.

The revenue potential for The Data Visualization Project is very high if marketed the correct way. In the past year, SAS racked in over three billion dollars worth of revenue, largely due to the greater familiarity that many companies have with the tool over R, even though the latter is free and does not required a renewed license to continue using over the course of several years. Therefore, the tool needs to be directly catered to all companies doing analysis big and small. One way to carry this out would be to purchase Google Adwords, a tool which allows a product to be advertised on the search engine when certain keywords are entered. A certain amount of money is paid to Google each time an individual clicks on the link, however in the long run the reach that Adwords would provide would lead to a greater number of people using the product, and therefore lead to a large amount of revenue. Another way of earning money would be to allow a company to advertise on the website. The Data Visualization Project is expected to operate on a web-based platform rather than a software application. Therefore, another company can increase their reach when various analysts visit the website, and at the same time money can be earned in order to maintain the Visualization tool. A final proposition of potential revenue would be the added benefits of a membership. If a certain sum of money is paid, the analyst can get an upgrade with more features at their disposal. Possible features could include importing spreadsheets of any format to including more graphics to support the data that has been analyzed. The membership could be charged on either a monthly or yearly basis, depending on how often the individual using The Data Visualization tool. Even though the plan is for the tool to earn potential revenue down the line, the preliminary focus is to establish a large number of users who utilize The Data Visualization tool, and therefore the core tools such as automated summary statistics and basic graphs will always be free.

The commercial opportunity that The Data Visualization Project provides benefits all fields of society. In the field of healthcare, a large amount of data is available regarding the effectiveness of several drugs that cure illnesses. Analyzing the results of the data quickly can lead to doctors definitively prescribing a certain therapy to their patient. In business and marketing, quickly finding trends in a large amount of data can help possibly describe target customers that may be interested in a certain product that a company sells. Similarly, handling a large amount of data relating to education can help determine what engages elementary and middle school students, and what makes certain students perform better academically than others. In sports, conclusions can be drawn as to what enhances athletic performance. In short, large amounts of data are coming in at a rapid rate in all fields, and having a tool such as The Data Visualization Project can help make decision making easier and help shape society for the better.

There are no major environmental or health issues associated with the project. As mentioned earlier, the tool aims to solve health issues by making it easier in the healthcare field to make decisions of how to cure illnesses by looking at correlations of a large sample size of data over time. It is appropriate for people of all ages (since there is no need to know how to do the actual analysis itself) with the only necessity being having a computer, an internet connection for which to import the files to the website, and a passion for finding patterns in large sets of data. Although it can be technically used for analyzing demographics and trends in crime, for example, there is a very small chance that someone would exploit the tool to create chaos. Therefore, there is no expectation currently that the product would need regulation from the government. The only warning to take away, as with all tools analyzing data, is to not make decisions and believe everything based on what the data tells alone. Data can be skewed and may not always be accurate, not to mention the fact that the data may not take into consideration all factors that go in to making a final decision.

Although it is expected that there will be other competitors in the industry seeking to do similar work, the product could potentially be used unethically so that one could steal the entire framework and market The Data Visualization tool as a new product in itself. Hacking the server in order to reap the benefits of obtaining a membership could be another possible way that the product could be used unethically. The tool could also be used as a means of introducing bias into the data being looked at. For example, if a company was analyzing the amount of hours that each person worked per week for a full year the spreadsheet could be edited beforehand to show a different conclusion, and the Data Visualization tool could be blamed for misinterpreting the data. Perhaps the biggest unethical decision that can be made is not giving credit for when the analysis has been done using The Data Visualization tool. Such unethical practices can not be easily solved, but basic integrity is expected when using the website. In summary, there are large expectations for The Data Visualization tool to make the lives of analysts everywhere easier and improve society for the better.