The main goal of this initiative is to classify potential customers into 4 adoption categories to predict when they will purchase our new e-Reader. Each potential customer should be able to be considered an innovator (first week), early adopter (second or third week), early majority (fourth through eighth week), and late majority (after two months). With this information, we should hopefully be able to maximize our marketing efficiency by targeting marketing to each group of customers at the time(s) when they are most likely to respond. This initiative should have various organizational benefits. This project will attract new customers, encourage upgrades for existing customers, and connect them to our other services, leading to business growth in the short and long term. This initiative will be determined to be a success based on a couple of measurements. First, we can compare e-Reader sales of new generation vs previous generations. If more customers purchase the new e-Reader, it might be possible to say that the initiative was a success. Although the business has grown since the last e-Reader was released, we can control for that growth and identify whether the targeted marketing had an impact on overall sales. Another way to measure success of this initiative is to analyze individual customer’s predicted buying time vs actual buying time to see if targeted marketing was effective. The predicted buying time should have been based on past data which means it was the predicted time before applying any marketing effort. If many customers end up purchasing the e-Reader earlier than predicted, it might be safe to say that the marketing efforts paid off.

The Data Analytics Lifecycle is a key guide and process that we will follow throughout the project. In the discovery phase, we will analyze the project needs and frame the problem in its business context. In this case the Null Hypothesis is that all customers can be expected to purchase the new e-Reader at the same time. On the other hand, the Alternative Hypothesis is that some customers can be expected to purchase the new e-Reader at different times than others. In the data prep phase, we will turn the data that we have into the data that we need. This includes asking questions such as what data might predict buying time? Do we have this information? How can we get it? In the model planning phase, we will try to determine what method is best for this analysis. This could be a t-test, linear regression, ANOVA, or something else. In this case, the best method might be a chi-squared test, if most of the data ends up being categorical in nature. The next phase, model building, is when we will use the training data to build model and the testing data to evaluate it. Since we have all of the variables that we think might be useful, this phase is the process of working through each one to build the best model to predict customer buying time. The next step is to communicate the results. The model’s results and finding will need to be presented to all key decision makers as well as the marketing department who will be deploying strategies based on the model. This leads directly to the operationalize phase, where we will collaborate with marketing to create a plan which will maximize marketing efficiency of the e-Reader. The marketing efforts could include targeting the innovator group very early by offering a pre-order bundle. Additionally, the early majority might need more incentives such as a free trial or additional service offerings.

Following the data analytics life cycle will lead to the best possible results of this project when all is said and done. Each phase has benefits that will help our results be more predictable, reliable, and secure at the end of the day. The discovery phase ensures that there is a problem that needs to and can be solved and that we will use the optimal resources. The data prep phase ensures that the data we have is the best data to represent our population of potential customers, increasing reliability. The model planning phase ensures the statistical analysis can provide information that might be able to solve the problem. The model building phase optimizes performance and quality of the statistical analysis. When we communicate our results, it improves clarity of message, aligns objectives across departments, secures information and business practices. In the operationalize phase, we put our ideas into practice and watch the benefits come to life.

The data that is available at this time is a database of our current and potential customers that includes basic info, browsing & purchasing history. This is quality data because it covers a wide variety of customers over time, large sample. On top of that, the trends that we identify will hopefully be relatively apparent and stable over time. This means that there should be particular variables that clearly affect each customer’s buying time. There are also limitations to the data as it currently stands. The largest limitation is that there are likely many currently unavailable variables that could be more predictive than what we have. Potential variables I have considered that might be more predictive are home address, occupation, family size, and additional services that they use on our website. This leads to the questions of how can we get that information from our customers, and if it is available through the internet, can we obtain it legally?