5 questions a Data Scientist should ask their Customer

Paul F. Roysdon, Ph.D.

I. Questions

Data Scientists typically struggle with defining their problem statement by not asking their customer the correct questions. Below, I have provided 5 questions that should help *refine the problem*, and *define a successful outcome*.

1. Do you have data?
This might seem like an obvious question, but you would be surprised how often I am asked to apply "Data Science" to a problem, when the customer does not have data.

2. What type of data do you have?
This question should refine the problem and provide an indicator of the types of algorithms you might apply. For example: photos will not likely require natural language translation deep learning algorithms, unless the photos contain text. This question should also have several sub-questions, e.g.

- **Where will your data come from?**
- **What scales apply to your dataset?**
- **What is the quality of your data?**

These questions will further refine the additional steps you might need to perform your analysis.

3. What exactly do you want to learn from your data?
This question should define the scope of the problem and the desired outcome. Possibly the only metric the customer is interested in is a histogram of their data. It is up to you to find the *science* in the problem. For example a histogram might be informative, but the mean and variance of the data will provide measures for predicting future events, given the data. Similarly, the histogram might not be the Gaussian distribution that your customer expected, thus presenting you the opportunity to inform the customer of unique characteristics of their data. This simple additional insight will not only establish confidence, but will likely open new opportunities.

4. What are your requirements?
A good customer will always give you more requirements than are necessary. This is the time to *collect and prioritize* those requirements. This question should also have several sub-questions, e.g.

- **What Key Performance Indicators (KPIs) will you use?**
- **Which statistical analysis techniques do you want to apply?**
- **What data visualizations would you like to see?**
- **Who are the final users of the results?**

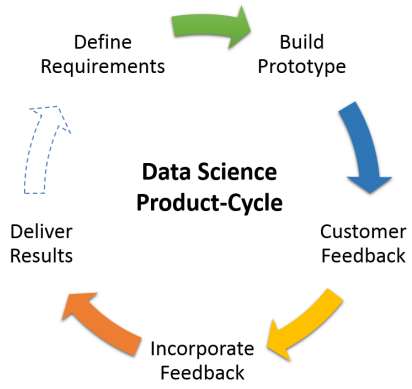
If your customer provides 10 requirements, but you know that you can only achieve 7 in the time that they provided, ask them what their top 5 requirements are, and surprise them by completing 7. This is not a hard rule, but rather a suggestion. Customers will always provide additional requirements.

5. What is your time-line?
It is very important to set expectations up front. This is also the time to inform them of the Data Science work-cycle. Customers (almost) always like to have a tight feedback loop, e.g. work on the problem, build and present a prototype for input by the customer, make the necessary modifications and deliver the result. Keeping your customer informed and involved will build trust and establish credibility.

Finally, I recommend asking your customer: **What else do I need to know?** At the end each of your meetings, your customer will almost always have (in the words of Steve Jobs) "just one more thing". Pay attention and stay engaged.

II. The DS Feedback Loop

The Data Science "Product-Cycle" is an important feedback loop, or cycle, that closely resembles the "Agile" cycle used in software development, see the figure below.



This loop should be performed in a timely manner. If there are several weeks required between steps, it is incumbent upon you to keep your customer informed of your progress, and possibly ask for input (if necessary). If you are unable to deliver a final product with all of the requirements, this is an opportunity to (possibly) start a revision of the project, given both your priorities and those of your customer.

Remember, take pride in your work and do not settle for mediocrity – you are a Data Scientist! If you follow these simple guidelines, you will have a greater chance for success.