Step 1 is defining the problem. Put simply, what question(s) are you trying to answer? Once you understand that you need to then think about what data is available to you to answer the question:

- ▶ Is the data directly related to the question?
- ▶ If it is not, can you create a proxy relationship to be able to link it?
- ▶ Is the data you need even available within the enterprise or elsewhere?

As part of step 1, you also need to specify the inputs and outputs of the model you are going to build as these may change as you change and tweak the model. Finally, dont forget the most commonly forgotten piece of any new initiative. Determine, up front, how you are going to measure the results. - What measure of accuracy are you going to use? Is that level of accuracy good enough for the business? - How will you benchmark the results? - What criteria are you going to use to determine success or failure?

Step 2

Step 2 is more rote and technical process the data. Collect the data (more is always better in this analysts mind but more does not always mean easier or better results). In general, more recent data is better and the data need to be consistent. Dont skimp on cleaning the data. While this may end up taking the most time, it is critical and erroneous data will create erroneous results.

Step 2

Transforming the data is also worth the time and effort to improve the modeling process including such things as: - Converting non-numerical data to numeric (or vice versa) - Standardizing thing such as coding, definitions, costs, combining variables, etc. Predixion gives you tools, built into the software, to help you analyze, clean, and classify your data on the front end.

Step 3

Step 3 is running the initial model. Part of step 3 is to split the dataset into a test dataset and a validation dataset. If you really want to be able to test the accuracy of the model once it has been built, you need to do this. The software will walk you through this and will do this for you by default, holding back 30values. This is also the step whereby you will choose the method or methods by which you want to build the model and process the data.

Step 3

As you become more familiar with predictive modeling and with
your own data you will find that certain types of data and certain
types of analyses or problems do lend themselves more or less to
certain types of modeling. But if you are just starting out, you can
use the software to guide you in the choice of a model or simply
choose to run all the models against your data. Once done, run
the model and move on to step 4.

Step 4 is to evaluate the initial results. Are the results acceptable? Are they what you were expecting to see? Do you understand the results? Do they answer the question you are trying to answer? If the answer is yes, then move on to the next step. If the answer is no, then consider the following: - Try using different algorithms/models - Try using different data elements or repackaging what you have - Consider collecting more or different data - Consider redefining the problem, changing the question and the means to an answer as you better understand your data and your environment

- Part of the learning process may be to try and not boil the ocean with your models.
- Think about setting up the model to run on a number of scenarios starting from the simple and most straightforward and then progressing to more and more complex.

Step 5

Step 5 is to select the final model.

- ▶ Dont be afraid to try a number of different models and then when you are satisfied with the results, choose the best one.
- ▶ We will talk about means of assessing the accuracy of your model in a bit.
- ▶ For now, choose the final model and consider whether you want to rerun the entire dataset against the selected model and re-examine the results

Step 6

Step 6 is to test the final model. This is another one of those things that often seems not to get done. It is important to test the final model and the only way to do so is to take the selected model and run it against a second, unrelated dataset (e.g. - the validation dataset or the portion of the dataset that was held back for this purpose) and assess the results.

Do not tweak or change the model in any way at this point as it will invalidate any comparison to the initial results. If the results are similar and you are satisfied with them you can move on to the final step. If you are not, then go back (to step 3) to reassessing the model and the data, make any necessary or desired changes and try re-running the model again.

Step 7 is to apply the model and run the prediction. There are actually two parts to this. One is done if you want to refine the model then you can use the output from the model to determine next steps and potential intervention or changes. Continue to test the model as much or as often as needed. But when you are satisfied, please do two things: 1) Run the necessary measures to test the final accuracy of the model (see our next discussion points) 2) Take the output from the model and turn it back into language and output for the business that answers the initial question you set out to answer and makes it useful/usable for them