

Analytics starts by asking the right questions

Prepared by Sam Heidari, Data Scientist, CPA Alberta

Data analysis requires quite a bit of thinking. Often when you've completed a good data analysis, you've spent more time thinking than doing. The thinking begins before you even look at a dataset, and it is well worth devoting careful thought to your questions. This point cannot be over-emphasized, as many of the "fatal" pitfalls of a data analysis can be avoided by expending the mental energy to get your question right.

So what is a good business question? A good question should target **solving a particular business issue**, or parts of it at least. It may arise from many sources, including how to better attract customers, how to price a product or service, or how to find errors and fraud. These questions should be very **specific** and must be **potentially answerable by data analytics**. There is no point asking an unfocused question or a question that is not answerable.

For example, questions that accountants and auditors may be interested in include:

1. Are employees circumventing internal controls over payments?
2. Are there any suspicious travel and/or entertainment expenses?
3. How can we increase the amount of add-on sales of additional goods to our customers?
4. Are our customers paying us in a timely manner?
5. How can we predict the allowance of loan losses for our bank loans?
6. How can we find transactions that are risky in terms of accounting issues?
7. How often do we have transactions above \$100,000?
8. Who authorizes cheques above \$10,000?
9. How can accounting errors be identified?

But first, let's delve a little deeper into the six types of questions that might be asked in the process of analytics:

1. Descriptive:

Descriptive questions, along with exploratory questions, are a great place to start any analytics process. They will provide great insight to your data while laying the foundation for more advanced questions, if they are needed. Descriptive questions will try to summarize characteristics of a set of data. For example: What is the average income of Canadian households? What is our company's most successful service by revenue?

2. Exploratory:

Exploratory questions will ask if there are any patterns present in the data. For example: Is Project A revenue increasing over the past two years? Has the transportation cost decreased over the last year? Is there an observed year-over-year decrease in carbon emissions after new regulations?

3. Inferential:

Here we will hypothesize that a situation is “true” and try to find evidence from the data to support or reject our hypothesis. It will normally start by proposing the hypothesis based on our observations, assumptions, or experiences. Then we put that to test. We normally test our hypothesis on a smaller sample size, and if it holds, we assume it has a good chance of being true for the larger population. These types of questions are very helpful when the actual data is very large and we only have access to a subset of it. Election exit polls are a common example of this. A good hypothesis to test in this case would be “Candidate A wins the election by a margin of X percentage points”.

A more common example would be in marketing. A sample hypothesis would be “Spending \$1,000 in Google ads improves sales by 20 per cent,” or “Spending \$1,000 in Google ads improves sales more than spending the same amount on billboards.”

Since the inferential questions are typically answered on a subset of the larger dataset, it is imperative to get the sampling right. It is very common, even among veteran analysts, to miss this opportunity and come to the wrong conclusion.

4. Predictive:

These questions will ask what set of input parameters may result in a particular outcome. These don't need to be an exact cause-and-effect relationship. Often times, a simple correlation between the input parameters and outcomes is enough. We are all familiar with commodity price prediction. Another question might be, “How much would my next year's sales/costs/revenues be?”

5. Causal:

A causal question asks if changing one parameter will change another—basically, trying to go beyond mere correlations in the dataset and focus on the cause and effect. This is a very common form of question in manufacturing or operation root-cause analysis, and medical research, among other fields. They can be difficult to answer without a controlled sampling and a good dataset. For example, “Does eating more vegetables improve the immune system against a particular viral infection?” “Does Drug A cure a disease?”

6. Mechanistic:

While most other types of question ask about causes and correlations, this question type will ask how a particular outcome may emerge.

Before we start our exercise, consider that one can ask many different questions types about the same issue. For example, with the question about YoY carbon emission decrease in the exploratory analysis above, we can ask an inferential question by defining a hypothesis for it and then trying to find significant support for that in the data. One possible hypothesis for that question might be: After new regulation, year-over-year carbon emission has decreased. Now we can search our data and see if we can support this hypothesis.

One way of asking the same question in a predictive manner would be: What would the trend of carbon emission five years be after implementing the new emission laws? A causal question could be asked after we test the provisions of the new regulation in a controlled environment, and make some observations. For example, in a few specific manufacturing facilities where every other parameter is kept the same before and after, we can ask: Does the new regulation cause a reduction in carbon emission? But which one should we ask? Well, not surprisingly, that depends on what problem are we trying to solve and how we want to approach it. The following exercise should help us to get a feeling for it.

Works Referenced

- [1] R. D. Peng and E. Matsui, *The Art of Data Science*, Leanpub, 2018.
- [2] J. V. Richardson, R. A. Teeter and K. L. Terrell, *Data Analytics For Accounting*-Second Edition, McGraw-Hill Education, 2020.

Group exercise

Now let's put our thinking hats on and test what we learned:

Consider a particular business problem that you or your company have experienced, such as recent loss of revenue, lower sales numbers, or difficulty attracting young employees. Let's call it "Situation A". Then think about the data you have to run the analysis on and potentially find a solution to Situation A.

You should form a few questions to guide the process of analysis. Write down a few good questions to better understand your data in the context of Situation A and potentially explore trends that may exist in your data.

Ask yourself if you should spend time and resources to solve this issue at all:

- a) What if we ignore Situation A? What would be the cost to our business?
- b) Can we ask a predictive or inferential question about the potential impact of the existing situation?
- c) Is it specific enough? Is it answerable?

Let's say you have decided to tackle the problem, and you need to do some root-cause analysis to come up with a solution:

- d) What do you need to know to enable a root cause analysis of Situation A?
- e) What are the mechanics of Situation A, or the correlation between it and other parameters?

Now let's look into the future! With your great analytical work, you have found the solution to Situation A and decided to implement it:

- f) How do you monitor the actual impact?
- g) Did our solution cause the outcome?

What questions would you ask? Make sure they are very specific!