Step 1: Frame the problem

Before you can solve a problem, you have to define the problem.

You'll often get ambiguous inputs from the people who have problems. You'll have to develop the intuition to translate scarce inputs into actionable outputs - and to ask the questions that nobody else is asking.

Say you're solving a problem for the VP of sales of your company. You should start by understanding their goals and the underlying 'why' behind their data questions. Before you can start thinking of solutions, you'll want to work with them to clearly define the problem.

To define the problem, you need to ask the right questions. For example,

1. Who are the customers?
2. Why are they buying our product?
3. How do we predict if a customer is going to buy our product?
4. What is different from segments who are performing well and those that are performing below expectations?
5. How much money will we lose if we don't actively sell the product to these groups?

You need as much context as possible for your numbers to become insights.

In response to your questions, the VP of Sales might reveal that they want to understand why certain segments of customers bought less than expected. Their end goal might be to determine whether to continue to invest in these segments, or deprioritize them. You'll want to tailor your analysis to that problem, and unearth insights that can support either conclusion.

It's important that at the end of this stage, you have all of the information and context you need to solve this problem.

Step 2: Collect the raw data needed for your problem

Once you've defined the problem, you'll need data to give you the insight needed to develop a solution. This part of the process involves thinking through what data you'll need, and finding ways to get that data, whether it's

querying internal databases or purchasing external datasets.

You might find out that your company stores all their sales data in a customer relationship management (CRM) software platform. You can export the CRM data in a CSV file for further analysis.

Step 3: Process the data for analysis

After you've collected all the raw data, you'll need to process it before you can do any analysis. Oftentimes, data can be messy, especially if it hasn't been well-maintained. You'll see errors that will corrupt your analysis: values set to null though they are actually zero, duplicate values, and missing values. It's up to you to go through and check your data and make sure you'll get accurate insights.

You'll want to check for the following common errors:

1. Missing values
2. Corrupted values
3. Timezone differences
4. Date range errors, such as data registered from before sales started

You'll need to look through aggregates of your file rows and columns, and sample some test values to see if your values make sense. If you detect something that doesn't make sense, you'll need to remove that data or replace it with a default value. You'll need to use your intuition here: if a customer doesn't have an initial contact date, does it make sense to say there was *no* initial contact date? Or do you have to hunt down the VP of Sales and ask if anybody has data on the customer's missing initial contact dates?

Once you're done working with those questions and cleaning your data, you'll be ready for exploratory data analysis (EDA).

Step 4: Explore the data

When your data is clean, you should start playing with it.

The difficulty here isn't coming up with ideas to test, it's coming up with ideas that are likely to turn into useful insight. You'll have a fixed deadline for your data science project (your VP of Sales is probably waiting on your analysis), so you'll have to prioritize your questions.

You should look for interesting patterns that explain why sales are reduced for the segment of the populations you've identified as the problem. You may notice they're not very active on social media, with few having Twitter or

Facebook accounts. You may also notice that most people in this segment are older than your general audience. At this point, you can now begin to trace these patterns to analyze the data more deeply.

Step 5: Perform in-depth analysis

This step of the process is where you will need to apply your statistical, mathematical and technological knowledge, and leverage all the data science tools at your disposal to crunch the data and find every insight you can.

In this example, you may have to create a predictive model that compares your under-performing group with your average customer. You may find that age and social media activity are significant factors in predicting who will buy the product.

If you asked enough of the right questions while framing your problem, you might realize that the company has been concentrating heavily on social media marketing efforts, with messaging aimed at younger audiences. You would also know that certain demographics prefer being reached by telephone rather than by social media.

You will begin to see the way the product has been marketed is significantly affecting sales. However, maybe this under-performing group isn't a lost cause: a change in tactics from social media marketing to more in-person interactions could change everything for the better. This is something you'll have to flag to your VP of Sales.

You can now combine all of those qualitative insights with data from your quantitative analysis to craft a story that moves people to action.

Step 6: Communicate results of the analysis

It's important that the VP of Sales understands why the insights you've uncovered are important. Ultimately, you've been called upon to create a solution throughout the data science process. Your ability to properly communicate your results will define the difference between action and inaction on your proposals.

You need to craft a compelling story that ties in your data with their knowledge. You start by explaining the reasons behind the under-performance of the older demographic. And alongside answers your VP of Sales gave you and the insights you've uncovered from the data, you can move to concrete solutions that address the problem. One solution would be to shift some resources from social media to personal calls.

You tie it all together into a narrative that solves the problem for your VP of Sales: she now has clarity for how she can reclaim sales and hit her objectives.

She is ready to act on your proposals.