**Understanding the Business: The Data Science Skill We Don’t Talk About Enough**

Why you need to master this skill, and how to go about it



Ada Bojana, Montenegro (Image by the author)

Every aspiring data scientist I talk to thinks their job starts when someone else gives them:

1. a dataset, and
2. a clearly defined metric to optimize for, e.g. accuracy

But it doesn’t.

It starts with a business problem you need to understand, frame, and solve. This is the key data science skill that separates senior from junior professionals.

And in this article, I’ll show you how you can train this data science skill, with a real-world example.

**The Starting Point of a Data Science Project**

In the real world, data science projects start from a business problem. They are born to move a key business metric (KPI).

The data scientist’s job is to translate a business problem into the **\*right\*** data science problem. Then solve it.

To translate a business problem into \*the right\* data science problem you do 2 things:

1. ask questions
2. explore the data to find clues.

There is nothing more frustrating than building a great data science solution, to the wrong business problem.

Let’s go throw an example.

**Example**

Imagine you are a data scientist at a top ride-sharing app company. And your product lead tells you:

👩‍💼: “We want to decrease user churn by 5% this quarter”

We say that a user churns when she decides to stop using our ride-sharing app.

There are different reasons behind user churn. For example:

* “Another ride-sharing app company (aka direct competitor) is offering better prices for that geo” (pricing problem)
* “Car waiting times are too long” (supply problem)
* “The Android version of the app is very slow” (client-app performance problem)

You build this list ↑ by asking the right questions to the rest of the team. You need to understand the user’s experience using the app, from HER point of view.

Typically there is no single reason behind churn, but a combination of a few of these. The question is: which one should you focus on?

This is when you pull out your great data science skills and EXPLORE THE DATA 🔎.

You explore the data to understand how plausible each of the above explanations is. The output from this analysis is a single hypothesis you should consider further.

Depending on the hypothesis, you will solve the data science problem differently.

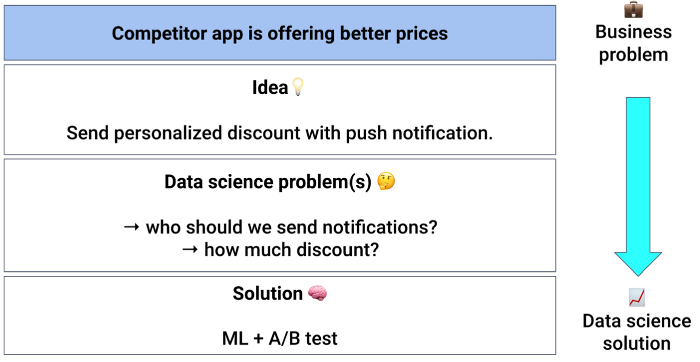
For example:

**Scenario 1: “A Competitor Is Offering Better Prices” (Pricing Problem)**

One solution would be to somehow detect/predict the segment of users who are likely to churn (possibly using an ML Model) and send personalized discounts via push notifications. To test your solution works, you will need to run an A/B test, so you will split a percentage of app users into 2 groups:

* The A group. No user in this group will receive any discount.
* The B group. Users from this group that the model thinks are likely to churn, will receive a price discount in their next trip.

You could add more groups (e.g. C, D, E…) to test different pricing points.

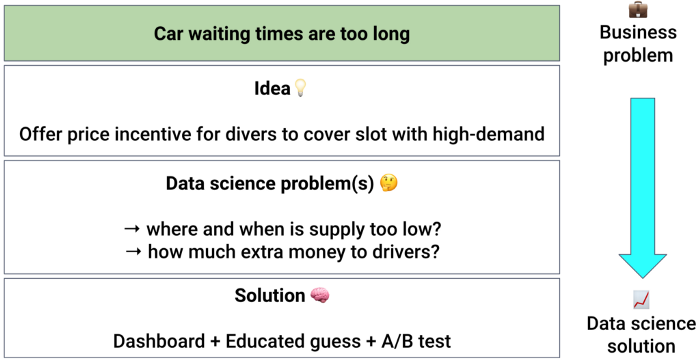


From pricing problem to a data science solution (Image by the author)

**Scenario 2: “Car Waiting Times Are Too Long” (Supply Problem)**

In this case, there is no pricing problem, but a lack of drivers to pick up clients. The problem is different, so the solution must also be different.

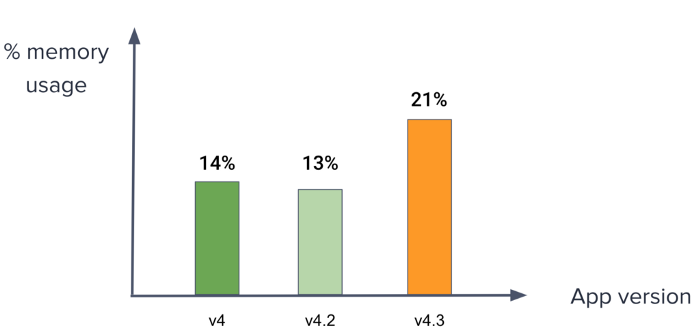
Something you can do is to identify the location and time where supply is too low and offer a price incentive for divers to cover these slots. This way you can balance better supply and demand, and reduce car waiting times.



From supply problem to a data science solution (Image by the author)

**Scenario 3: “The Android Version Of The App Is Very Slow” (App Performance Problem)**

Imagine you explore the data on memory consumption of the app, and find out that the latest version of the app consumes almost double the memory as the previous versions.



Break down app memory consumption by app version (Image by the author)

This is strange, so you go and ask the customer support team if they had received any complaints from users.

It turns out that most users do not contact support, but stop using the app, and use an alternative. However, there are still a few users who complained, and mentioned the new version of the app was not “very responsive”.

Bingo. You found an issue in the newest version of the app.

*How do you solve this?* Go to the frontend devs, show them the breakdown of use churn by app version, and convince them they should release a new version of the app with better performance.



From tech issue to a data science solution (Image by the author)

**To Sum Up**

* Translating business problems into \*the right” data science problem is the key data science skill that separates a senior from a junior data scientist.
* Ask the right questions, list possible solutions, and explore the data to narrow down the list to one.
* Solve this one data science problem