# Entry 3 - Frame the Problem

In Entry 2 I decided to follow the [ML Project Checklist](https://www.kdnuggets.com/2018/12/machine-learning-project-checklist.html) from [Hands on Machine Learning with Scikit-Learn & TensorFlow](https://www.amazon.com/Hands-Machine-Learning-Scikit-Learn-TensorFlow/dp/1491962291). The first step in that process is to frame the problem and look at the big picture.

## The Problem

Generally, when completing this step, I’ve already got a project that needs to be completed. It usually comes from business partners, like a product manager or executive. An example would be when Amazon first asked ‘How do we put similar products in front of customers that they might be interested in purchasing?’ (No, I don’t work for Amazon, nice guess though).

These problems are generally ill-defined, vague, and sometimes contradictory. Most of the time when the business partner comes with their question, I have to figure out what need or opportunity the partner is trying to address, collect requirments for the end product, and define how to measure the real-world success (or failure) of the project.

The ML Project Checklist focuses on these kinds of business oriented problems and so has suggestions like: ‘Define the objectives in business terms’ and ‘Is the performance measure aligned with the business objective.’ But there are also a good number of more generic suggestions.

## The Options

As these ‘diary entries’ are to practice small, specific parts of the process, there isn’t a ready-made question to answer. My question is only limited by the data that I can [get](http://archive.ics.uci.edu/ml/datasets.php), [find](https://earthdata.nasa.gov/), or [create](https://data.world/datasets/open-data).

While there are certainly datasets out there wherein [companies pose real-world questions](https://www.kaggle.com/datasets), I think I’m going to go another route.

Since I’m a nerd and a writer, the question(s) I want to answer revolve around a story I’ve been thinking about involving terraforming Mars.

## The Proposed Solution

### Domain Experts

This is one of the many areas where domain expertise comes into play. Because I’m an amature hobbiest space nerd, I know to ask quesitons like: ‘What are the biggest issues in terraforming Mars?’ Then come up with an answer like: ‘The two most obvious are liquid water and air.’

Domain experts are invaluable resources. They have information I’d never even guess at and readily supply context. For this initial step, they can help frame what the question should be. Take the credit card industry for example. Fraud can mean huge losses of both money and reputation. Domain experts can help define what fraud means - is it customers incorrectly challenging charges by legitimate businesses because they don’t want to pay or is it thieves using credit card numbers than don’t belong to them or is it people who apply for a card then never pay their first bill? Framing the problem should address things like this. Is the goal to tackle everything at once, or focus on a specific thing?

### Level of Detail

The more concrete and detailed the answer to ‘defining the problem’ is the better. Everyone involved in a project brings their own perspective to it. These varying perspectives mean that everyone expects different things from the results. This is the time, right at the beginning, to spell out what the results will be, how they’ll be used, what success/completion looks like, and how all that aligns with larger objectives.

A high level of detail, where everyone is on the same page, may or may not (probably not) help prevent [scope creep](https://en.wikipedia.org/wiki/Scope_creep), where everyone tries to sneek in their own pet project that can be even tangentally related to the actual goal. But by making sure everyone is starting from the same page, additions to the project can be quantified and the impact to the current plan more easily measured.

That said, the problem definition will change over the course of the project as background research is completed, requirements are collected, roadblocks are encountered, and the business environment evolves.

### Pick a Problem

This brings me back to: what is my current goal? And how will the information be used?

The story takes place in two timelines. Something close to present day, wherein the how and why of Earth’s ecosystem begins to be unraveled so that people understand how a planetary system works. The other is the future, where that information has been put to use turning Mars into a planet that people can live on without domes or space suits or other artificial means.

There are a lot of great minds working on figuring out how and why the Earth does what it does. But the Mars issue is mostly speculative. A solution really just has to pass the smell test (this is a science-*fiction* story afterall). Which brings me back to the two biggest hurdles to Martian occupation: liquid water and air.

Based on the information collected from rovers and satellites, we know there is frozen [water on Mars](https://en.wikipedia.org/wiki/Water_on_Mars). This means temperature is a factor in liquid water. On Earth, our atmosphere helps hold heat and [regulate temperature extremes](https://www.sciencedaily.com/terms/earth's_atmosphere.htm). I’m gonna take a leap here and conclude that air is the biggest hurddle to life.

Based on some old documentary I probably watched as a small child (yes instead of cartoons I watched the Discovery Channel back when it was the only channel playing documentaries), Mars had an atmosphere way back when, but didn’t have sufficient gravity to hold onto it. The Swedish Institue of Space Phsics published [a paper in 2017](https://www.sciencedaily.com/releases/2017/12/171207140513.htm) with basically the same findings. Gravity isn’t the only factor in play when it comes to having an atmosphere. A magnetic field is thought to help protect an atmosphere from the [solar wind](https://en.wikipedia.org/wiki/Solar_wind#Atmospheres). And if solar wind plays a part in stripping away atmosphere, then distance from the sun is probably also a factor.

### Define the Problem

Now I’m onto something. This is the kind of thing that can be turned into a data science problem. What factors can be altered? Magnetic field is [thought to be formed by](https://www.sciencefocus.com/space/do-all-planets-have-magnetic-fields/) a molten rock and metal core and a planet’s rotation. Good luck changing that. Distance from the sun feels pretty fixed. [Asteroid mining](https://en.wikipedia.org/wiki/Asteroid_mining) has been a thing in recent years, let’s say asteroids can be diverted and crashed into Mars in a controlled fashion, which would increase mass and thus gravity.

Now I have a concrete problem: **Holding all other factors constant, what mass is needed to retain an atmosphere on Mars?**

## The Failure

Confession time. Coming up with the problem didn’t go as smoothly as the above may imply.

I knew I wanted to do something around the Martian atmosphere. So my first problem statement had atmosphereic mass as the target. When I went to try to make a prediction, I would have had to guess and check planetary mass until the model came up with an acceptable atmospheric mass.

The key takeaway of this failure? Consider which factor will be the target when considering the problem question. What am I trying to predict - which variable is the one that changes? Yes, having an atmosphere is my end goal, but the factor I want to vary is mass. Afterall, there is a limited range of atmosphere within which life as we know it will survive. I can just plug those in as max and min values to get a range for change in mass.

Correctly defining the problem seems to be a common problem. I’ve seen it at multiple companies I’ve worked at, across different departments, and heard others comment on how challenging it is in my extracurricular research via a variety of mediums. It’s easy to underestimate the importance of this step.

## Next Up

Get the data.