This blog post is dedicated to people who started to study DataScience and got lost at some point. You know, like the moment when you already some know something but because the knowledge has so many disciplines included you starting (or already lost ) the notion of how it is all connected in the solid background.

We will try to make parallels between principal DataScience instruments and something that we all well know, so it will be easier to understand where you are now and where to go. So let's begin.

Let's suppose that you are a professional driver who literally can drive anything.

And every day you are getting some strange objective where to go and what to do:

- Go to Vancouver to pick up children from school

- It is snow in Mapple ridge; you need to clean it up

- Today is Nascar racing in Paris, go, go go…

And you have a garage full of different cars: White beetle, school bus, uncle Frank's Maserati, tank, and much much more.

Every day, after receiving your mission, you need to choose the right vehicle that can be used to accomplish this mission, and you must do it. I think you will agree that it will be bizarre and probably even not possible to clean the snow using Maserati.

As soon as we start to make parallels – our objective as a driver is smth that we should do as DataScientist, while the cars are models that can be used for it. Not any model can be used for any purpose. And even for the same objective, we can have different ways to solve it, using various means.

Now let's dive a little deeper into our imagination:

Besides choosing the car, we definitely should know how to drive. Driving skills are essential; otherwise, we risk ending up somewhere in a roadblock instead of winning the Nascar race.

This driving experience is our coding skills. We should know some primary programming language to make sure that we can drive to the destination and get knowledge of additional libraries and tools that help us achieve better performance.

What's next?

Engine!!! Different cars can have various engines! Or even the same car can have different types of engines. Maybe our old red sedan has an internal combustion engine, or it is not old, and it has an electrical motor? Or perhaps it is a diesel engine? The only thing that all engines have in common – that is the thing that drives our car forward.

The same functions have our algorithms inside our model. So it helps to solve the model and find optimal solutions. Different algorithms can be used for the same model (like various engines inside the same car).

And last not least:

As soon as we start driving our car towards the gloveful purpose, we always should consider the physical world and laws of physics around us. It comes mainly in 2 forms:

1. We should always consider the road conditions: Is the road wet or dry. Do I need to put on winter tires when it is +40 C around? How should I drive during strong wind conditions? Should I take Maserati when it is snowy outside?

I would make a parallel of these conditions with statistics that we should always consider when we decide which model to use(or which car to take)

1. There are also laws of physics that we don't see all the time, but they exist, and all our movement depends on them. For instance, if you know how the internal combustion engine functionality: it is based on many, many, many MANY explosions that constantly happen inside our engines. They give the energy for the car to move further.

I would call such "unseen" physics – Mathematical laws that exist. Still, we might don't even understand that they drive everything.

Just for fun: as small explosions power our engine that moves our cars, linear algebra and Mathematical laws are the ideas behind algorithms(Gradient descent, for instance) that resolve our models.

As you can see so far, everything inside the DataScience world is interconnected in the same manner as in cars, and this general understanding can help visualize how it happens.

Just a quick sum up where we have come so far:

Objective – What is your objective as a data scientist? (predict the number, make classification, etc.)

Driver – DataScientics

Car – your model that does all the work.

Engine – Algorithm inside the model

Laws of Physics – Statistics and Linear Algebra.

Why do we need to understand all these parallels?

1. The most obvious reason is that when you are not in the Data Science world and consider switching, it all looks like alien technologies. At the same time, it is not so difficult to understand in real life. You need to look at it like smth more familiar to you.

1. Secondly, it allows choosing the best strategy to study the materials.

There are tons of different online resources, websites, educational videos, and just bloggers who will tell you exactly what to do. But is it efficient? How important is the topic they discuss? All these questions can be tackled(at least in some form) when you can imagine how all connected.

For example, should I spend two weeks learning how gradient descent works and not knowing how to code well enough? (Translated version: If I am a driver, should I understand the detailed structure of the car engine and know everything about it when I even don't know how to drive)

1. You will be lost many times during your journey into the DataScience. Having some kind of easy-to-understand map will allow you to stop, realize what you are learning and why, and keep moving, or maybe stop focusing on these topics.

While there are much more topics and concepts in Data Science, they all come with time and experience. Making a mental map allows making this path much more straightforward.

How to structure your study based on this "map" you can find in my blog post: