#66DaysOfData – Day 1: Asking Important Questions

Today I start my second round of Ken Jee's #66DaysOfData challenge. Since I had already started the #RoadToDataScience series (which I will continue), I decided to do another set, focused on this challenge.

Every day, I will be posting a small article about my project and progress. I am sure some days will be more interesting than others. I will go through the whole process of a project, from asking questions to building models, and the best part is, I have no idea what I will be able to accomplish. It might be an amazing predictive model, or a list of busted hypotheses. Either way, I will have fun, learn a lot, and share every single step day by day.

Where do you start with a Data Science project? In my experience, the best place to begin is by asking a simple question. Now remember, we have no idea if we will be able to answer it, that is the way of science, you ask questions, you look for data, you experiment with it. At any point, you can go back, change the question, create new hypothesis, jump back, and jump forward. At the end, you get something from a useful model, to a lot of lessons (you always get the lessons, and those are the most valuable discoveries).

Back to my project. Since I am a behavioral scientist and data scientist, I decided to mix both, and work on something that intrigues me, and at the same time, I believe could be unbelievably valuable (although again, this is just another hypothesis): **Can we predict job success from personality traits?**

This is a big question, and we want to be thorough, so let us unpack it to prevent confusion, this will also help refine this question as we move forward. There are two parts to this question, the first one is Job Success, what does it mean? In today's highly competitive tech, the average time a person stays in a particular position is 3 years, at this time, people leave for other, better paid or more challenging jobs. Let us set Job Success as a metric which states that a person stayed at least 3 years at a company. Notice that I am using a metric from the Tech industry, so I will have to bind my question to this industry to be more effective.

The second part is personality traits, what do we mean by this? In this case, my background in psychology comes to the rescue. There are several personality traits models, like MBTI (Myers – Briggs) and DiSC, which are well known, but I am a stickler for open science and good evidence,

because of this I will stick to the OCEAN/HEXACO model, which has set and descriptive and well-studied traits. It is also an open model and has been used for many years in research.

From all this, we can now rephrase our question to make it more specific: **Can we predict if a new employee will stay for at least 3 years at a Tech company by using OCEAN/HEXACO personality traits model?**

Here we go, our progress today was to define a clear question to start our investigation process and see where it leads us. Remember, we have no idea if this is possible, or even if we have the data we need or enough data to answer this question. But asking questions (and being ready to change them a few hundred times) is the beginning of any great Data Science project.

Next Time – Let us find some data

Jack Raifer Baruch

Follow me on Twitter: @JackRaifer

Follow me on LinkedIN: jackraifer

About the Road to Data Science - #66DaysOfData Series

Road to Data Science series began after I experienced the first round of Ken Jee´s #66DaysOfData challenge back in 2020. Since we are starting the second round of the challenge, I thought it would be a good idea to add small articles every day where I can comment my progress.

I will be sharing all the notebooks, articles and data I can on GitHub: https://github.com/jackraifer/66DaysOfData-Road-to-Data-Science

Please do understand I might have to withhold some information, including code, data, visualizations and/or models, because of confidentiality regards. But I will try to share as much as possible.

Want to follow the #66DaysOfDataChallenge?

Just follow Ken Jee on twitter **@KenJee_DS** and join the #66DaysOfData challenge.

You can also reach out to me at any time through **LinkedIN** or **Twitter.**