The Poetry Project

Analyzing data is fun. However, as data analysts most of the times we clean, describe and visualize data that has more to do with our clients than with ourselves.
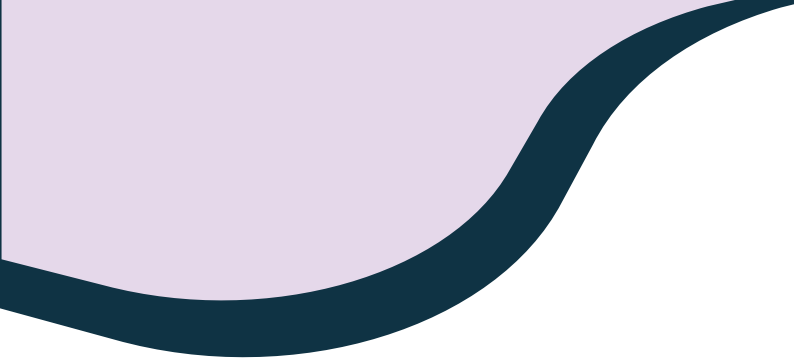
I wanted to change that.

I did an analysis of my personal poems which I wrote from beginning of August 2021 until November 2021. Not only did I write down the words of my poems but I also recorded some data around the day that I was writing it.

- How was the weather?
- Did I eat healthy?
- Did I listen to music? Which genre?
- Did I read a book? Which category of book?
- How long did I meditate?

In my analysis I was looking for correlations between weather, my eating, listening, reading and meditation habits and the wordcount of my poems that day. I was excited by the question:

What makes more creative on some days than on others?

I conducted exploratory data analysis, regression analysis, cluster analysis and time series analysis in Python with Jupyter Notebook.

To be honest, there were some limitations to my data. I had only spent two months on collecting the data and I didn't write a poem every day nor did I read or listen to music for the whole time. This created a lot of NULL values in my data. But nevertheless, it was fun analyzing the data and I learned a lot from it.

I learned that

- the wordcount of my poems is most highly correlated to how long I meditate,
- it is not trivial to bring data you spontaneously collected in an Excel spreadsheet into the right format for pandas,
- and that a lot of NULL values in a short period of time makes it hard to stationarize data for time series analysis.

I had most fun with

- exploring textblob for the first time
- looking at the wordcloud for my poems (you find it in Script 01)
- producing and looking at all the graphs in my exploratory data analysis (Script 02)
- seeing my productivity displayed in a time series analysis (Script 05)

Have fun with exploring my project!