With this article I want to start a series of articles showing how to use some of the important skills of a Data Scientist. To start, in this article I will be talking about web scraping and how I’ve built a Python web scraper using Beautiful Soup, a Python library that make it easy scraping information from a website.

My motivation for these articles came some months ago, when I decided to build a personal project on a subjected that I was passionate about and where I could use some of the important skills for a Data Scientist. Just when I was thinking about it I heard about a tour where the American rock band [Evanescence](https://en.wikipedia.org/wiki/Evanescence) would join forces with the Dutch symphonic metal band [Within Temptation](https://en.wikipedia.org/wiki/Within_Temptation), April this year. Since I love music and I found amazing the idea of putting together these two bands, seemed to me a natural choice to explore it in my project.

The two bands lead by the most influential and creative women in the rock world, Amy Lee from Evanescence and Sharon den Adel from Within Temptation, named the tour **WORLD COLLIDE**. [According to Amy Lee](https://www.blabbermouth.net/news/evanescence-and-within-temptation-to-join-forces-for-worlds-collide-2020-european-tour/), although they live in very distant countries and have some differences in their music, they also share a lot of similarities, so the name reflects well the embracing of both.

Based on this some questions started popping up: How data can show us how similar and different the bands are? What are the expectations of the fans while tour approaches and what are their reactions during the tour? And this is nice because it shows other things that are important for a Data Scientist: be curious, ask questions, look for answers in your data, and if there is no data enough, go chase it!

In a broader way, my goal developing this project is to answer these and other questions. I want to analyze and compare these bands using data retrieved through web scraping, using Spotify and Twitter APIs, and by applying important data scientist’s skills as cleaning, wrangling, visualization, as well as doing some NLP.

Here I’ll start retrieving the data I need. We will start by collecting all lyrics from both bands. I’ll show how to build a Python web scraper using Beautiful soup, applying some simple string methods as well as other tools to obtain the lyrics. You will also notice that you don’t need to be an HTML expert to do this (I’m certainly not one), just a bit of observation.

So, let’s start web scraping….