Thermo Fisher

Thermo Fisher Scientific Inc is a scientific technology producer, providing software, reagents, and instruments, and other tools to researchers. Based in Waltham, Massachusetts, ThermoFisher has a market capitalization of 220 billion USD and over 130,000 employees as of late 2021.

Project Description

ThermoFisher wants HDAG to develop tools to analyze their competitors. They want to understand who their competitors are hiring, where they are advertising, how they are marketing themselves, and how their product portfolios are changing. To accomplish this goal, the project will have three phases:

1. **Initial scraping.** ThermoFisher wants us to initially look at a broad range of questions, and recognizes that some of these questions are more or less feasible to answer given the limitations of what data is accessible. This initial step will seek to gather as much relevant information as possible in order to understand which of their questions are answerable given the information we can access.
2. **Specification and deep data collection.** Once we understand what data is accessible and which companies the most data is available for, we will launch a deep dive into these competitor companies and collect all the data that we can on them.
3. **Analysis.** Once we have the desired dataset, we will analyze the data to determine which features correlate with company success or imply the introduction of new products, as well as differences between different competitors. Then, we will produce a qualitative and quantitative report on our findings and present it to ThermoFisher.

Internal Partner: Clayton Brown, Senior Strategy Manager

Datasets: TBD. You will be employing a variety of web scrapers in order to generate the dataset. Tools like this are a potential starting point.

Preferred Coding Langauges: Python

Specific Skills

1. NLP: experience with NPL is highly preferred as analyzing scraped text is likely to be crucial to this case.
2. Scraping: use of web scraping tools is a significant portion of the work of this case
3. Research: you will need to identify scrapers to obtain datasets
4. Machine Learning/feature engineering: once the dataset is generated, there is likely to be a high-dimension dataset without many features with clear links to the dependent variables we're looking at. Using

Expected technical difficulty: Intermediate