Project 4 - Proposal:

For this project, I am using data to explore relationships between scalar data and the sites that collect this data.

The beauty industry estimated global value is 571 billion US dollars in 2023 with an expected growth rate annually is 3.8% (Statista, 2023) and is a highly competitive landscape for keeping the top positions as well as for new innovative products. Part of this massive product landscape is the use of botulinum toxin for aesthetic appeal – better known as Botox – which is an injectable treatment for different areas of the face like frown lines, crow's feet, etc. Botox is made by Allergan/Abbie who see annual sales of about 3.5 billion US dollars for aesthetic indications and 3.5 billion US dollars for therapeutic indications. For this project, I will focus on the aesthetic indication of frown lines.

The data for this project will come from two sources: 1) physician evaluations of subjects over time after an injection of botulinum toxin and 2) what locations are people looking to get treated as Google Trends data.

The first data source details the evaluation of subjects' frown lines on 4-point scale of Severe, Moderate, Mild or None to determine if the subject responded to treatment. Subjects were given an injection (Drug or Placebo) and evaluated every 4 weeks up to 36 weeks. These data were collected at different locations throughout the United States and Canada during 2018 and 2019.

The second data source from Google Trends shows people who are interested in getting treated for different metro-areas in the United States and Canada from 2019 through 2022.

The business questions being addressed are:

- What is the relationship of location to ratings? Do some locations have better performance than others?
- To predict which locations produce the best outcomes
- To predict if areas where people have a higher interest in being treated have better outcomes

The fields in the first dataset are:

Study – Study Identifier, 1 or 2

SiteID – Location ID, there are about 60 locations in each study

Country - US or Canada

State – State in Country

City – City in Country

Postal Code – local postal code

ID – Subject Identifier

Age – Age in years

Sex – Gender as F for Female and M for Male

Race – Subject Race

Treat – Treated or not, 1 is treated and 2 is placebo

Visit – Study Visit

ScaleScore – Severity score where 0 is None, 1 is Mild, 2 is Moderate and 3 is Severe

Had_response_YN – Subject had a response compared to baseline

Fields in the second dataset:

Country – US for United States or CA for Canada

Region_MetroArea – name of local metro area

Score – Google Trends score out of 100, how many searches were done for ‘Botox near me’ relative to the total number of searches done on Google.

In order to merge these two data sources together, I have another data source which is zip codes and city names mapped to metro areas. I am still working on how this merge will happen, but the end result is to have a single dataset.