**Data Mining for the Office of Science and Engineering Laboratories**

**Data Story**

The Food and Drug Administration (FDA) was curious about the progress of their Office of Science and Engineering over time. To help them understand their office's productivity, I analyzed their publications data to see when, under which division and program, and by whom the publications were made. By measuring collaboration among authors in divisions and programs, I aimed to assess productivity.

In my analysis, I explored the impact of the number of authors per publication-on-publication counts. Through exploratory data analysis (EDA) plots, I found that the number of authors did not significantly affect publication counts within divisions. This suggests that the work quality of a division alone does not drive publication numbers; rather, it is influenced by the number of people working in each division. Therefore, collaboration does not appear to be a defining factor in division productivity.

Plot 2 – publication counts per program for the top 10 authors: confirms collaboration which could be a factor for more publications

**A graph of a number of authors

Description automatically generated with medium confidence**

Anova test – significance of average number of authors per publication in each Program

A number and a result

Description automatically generated with medium confidence

However, when analyzing the data for programs, an analysis of variance (ANOVA) test revealed a significant difference in publication counts based on the number of authors per publication. This indicates that collaboration within each program does impact program productivity. As a result, we can identify the most successful programs based on this analysis.

Now that we know the number of authors makes a difference, let's explore who the authors are and identify the most successful ones in this office. The data product I created, an interactive bar plot, can help us visualize this information.

The data analysis conducted revealed intriguing insights regarding author collaboration and its impact on publication counts across different programs. While collaboration did not emerge as a measure of success at the division level, it played a significant role in driving productivity within each program. This underscores the importance of fostering collaboration for program success. Notably, the DBP division demonstrated exceptional productivity, emerging as the most productive division in the analysis. This achievement highlights their outstanding performance. Furthermore, the analysis identified the top three most collaborative authors: Phillips KS, Nagaraja S, and Scully CG. These authors showcased their expertise by collaborating with multiple programs, contributing to a diverse range of research areas. In terms of program productivity, the Cardiovascular, Medical Imaging and Diagnostics, and AI and ML programs under the OSEL division stood out as the most productive. Their accomplishments reflect their dedication and commitment to advancing their respective fields. It is also worth noting that the COVID-19 pandemic had a detrimental impact on productivity across all programs, authors, and divisions. However, as the situation improved, programs resumed their publishing activities, and the productivity within the OSEL division remained consistent.