Applied Data Science Capstone Proposal

**Project Title:**

“An Exploration of Consumer Complaint Sentiment”

**Introduction:**

As part of its Federal responsibility to protect the American consumer, the Consumer Financial Protection Bureau receives more than 50,000 weekly complaints relating to financial goods and services (CFPB). The CFPB states that most of these complaints are responded to within 15 days. In addition to being a direct conduit to corporations, the bureau also securely shares complaints with other Federal and State agencies to facilitate supervision and enforcement activities, as well as to support market monitoring.

In our investigation, we will test whether CFPB complaints can serve as indicators of broader financial stress and consumer sentiment. Specifically, we will evaluate whether complaint trends hold predictive value for near-term societal shifts, and whether changes in complaint patterns align with major state or national policy changes, as the CFPB suggests.

**Project Outline:**

Using previous analysis (<https://mylestym.github.io/2025/08/27/Analyzing-Servicemember-Financial-Complaints/>) as a structural basis, we will analyze publicly available CFPB complaint data with an eye towards geospatial and cultural analysis with a goal to identify any potential trends. Additionally, if feasible, we will develop a model based on this analysis with the goal of predicting potential changes in consumer sentiment or other societal factors.

To accomplish this, we will collect a complaint dataset from CFPB encompassing a wide enough timespan as to glean meaningful insights. Where applicable, we may draw upon other datasets to provides additional contextual information and potentially identify deeper trends. For example, when looking closely at mortgage complaints, the most complained about product in our previous analysis (linked above), we may find use in conducting correlation analysis with the House Price Indexes dataset maintained by the Federal Housing Finance Agency. Perhaps the Department of Transportation database of auto sales provide additional insights when analyzing car loan complaint information.

**Process Outline:**

Data preparation — Using Python and necessary libraries (SciPy, Seaborn, TensorFlow, etc.), we will first clean the datasets, dropping missing data, standardizing date formats, categorical variables where necessary, and other processes necessary to prepare the data for analysis.

Exploratory Analysis — Here we will seek to understand the dataset from a thirty-thousand-foot perspective. In addition to standard EDA, we will seek to answer the following questions. What products, companies, and geographic regions receive the most complaints? What is the average response time by company, by complaint type? The findings in this phase will inform subsequent steps.

ML Analysis — Our Primary goal will be to analyze narrative as pointers to larger societal change. Speaking generally, we will utilize NLP to conduct sentiment analysis (BERT) to examine complaint narrative data to look more closely at the context under which the complaints were made. For example, we may seek to examine if and when instances of narratives containing reference to jobless increase before, during, or after economic downturn. These are simply examples of analysis that could potentially be conducted after drawing initial conclusions from EDA.

Findings — Our findings, to include visualizations will be presented in report format. Additionally, a public GitHub repository will hold Jupyter notebooks and Python files to include documentation outlining process and explaining code.

Works Cited

*About Us*. Consumer Financial Protection Bureau. (n.d.). https://www.consumerfinance.gov/about-us/