**Data Science Practicum Project Proposal**

1. Lydia Swatek
   * Phone: 303-910-6364
   * Email: lswatek@regis.edu
2. Title: Inflation Rate Forecasting for High Consumption Food Commodities
3. Problem Statement:
   * Due to the high level of volatility in the market, especially for food, it is imperative that organizations have more agility and accuracy in their inflation forecasting. Food items are one of the commodities that are broken out when analyzing the consumer price index (CPI) and are extremely volatile. Based off of this information, organizations have to forecast their food pricing separately from the rest of their goods or services which increases the amount of time an organization needs to forecast pricing for their business. To try and reduce this lift, the US CPI averages will be compared to food CPIs to see if there is an easier and faster way to forecast food price based off of the national CPI.
4. Data Science Tasks
   * Data visualization which will be used to understand past trends and future forecasting
   * Supervised machine learning will be used to see how close a variety of real world data sets come to meeting the consumer index report
5. Data
   * The data being used for this practicum will be values from the consumer index report, as well as, 2-3 real world purchasing datasets for a variety of food commodities. While only 2-3 datasets will be used for this practicum the datasets will have a large amount of spend over multiple years when possible.
6. Analysis
   * The data will be analyze in multiple ways such as KNN analysis, text analysis and data visualization using Python and Tableau.
7. Challenges:
   * Understanding which world events are contributing to spikes in inflation rates vs what might be coincidental may be challenging. To mitigate this each world event will be analyzed to make sure that it is a contributing factor, however, coincidental events will be broken out into a separate output to insure we are not removing the data all together because some events may appear coincidental but are not.
8. Timeline:

|  |  |
| --- | --- |
| Week | Deliverable |
| Week 2 | Project Proposal and Data Collection |
| Week 3 | Data Cleansing and EDA |
| Week 4 | Machine Learning Analysis |
| Week 5 | Machine Learning Analysis |
| Week 6 | Data Analysis |
| Week 7 | Data Visualization |
| Week 8 | Presentation |

1. GitHub Repository:
   * <https://github.com/lswatek/Data-Science-Practicum-.git>