1. **Header**

World Inflation Data

Spring 2025 Data Science Project

Erica Honadel, Livia Earl, Abhyuday Srivatsa, Tony Zhang, Rohan Tadisetty

**Contributions:**

Everyone: Helped worked on the project ideas (A).

Livia Earl: She did the dataset curation and preprocessing (B), where the data was cleaned (deleting rows because they were missing not at random). Additionally, she did the data exploration and summary statistics (C) to get a better understanding of the dataset and the final tutorial report creation (G). The report includes finding helpful links and editing the Jupyter Notebook.

Erica Honadel: She did one of the statistical tests (ANOVA) (C). She also did the visualization (F) for ML 1 and ML 2 algorithms.

Rohan Tadisetty: He did the visualizations for each of the statistical tests (F) and the ML random forest model (ML 3) (D and E).

Abhyuday Srivatsa: He did one of the statistical tests (GDP: Pearson Correlation) (C) He also did the visualization (F) for ML 3 algorithm.

Tony Zhang: He did one of the statistical tests (Corruption Index: Linear Regression) (C) and the ML k-means clustering models (ML 1 and ML 2) (D and E).

1. **Introduction**

Our topic is exploring the Global Inflation Rates and how they relate to corruption perception index and GDP from 1960 to present. Our topic is exploring the Global Inflation Rates and how they relate to corruption perception index (CPI) and GDP from 1960 to present. [Inflation](https://www.google.com/url?q=https%3A%2F%2Fwww.pgpf.org%2Farticle%2Fwhat-is-inflation-and-why-does-it-matter%2F) is the rate at which prices for goods and services increase. It often occurs at a rate of 2% per year. Inflation affects many different facets of the economy. It is important to understand how rapid increase of these rates can lead to negative outcomes like increased interest on national debt. Inflation also reflects the state of the economy. For example, lower inflation rates can be seen when there is high unemployment and economic recession. On the other hand, higher inflation rates normally correlate to low unemployment. For these reasons and more, it is important for data scientists to examine past inflation rates to help better understand the future.

Some questions we want to answer: Were different regions affected differently during the Great Recession from 2007 to 2009? Is there a linear relationship between a country’s corruption perception index and its inflation? Is there a correlation between a country’s inflation rate and GDP? Which countries and subregions have similar inflation trends? Can we accurately predict the 2024 inflation and GDP? It is important to answer these to have a better understanding of the global economy and how countries and regions are similar and different. We can use these patterns to predict the future economy.

1. **Data Curation**

Global Inflation Data: <https://www.kaggle.com/datasets/fredericksalazar/global-inflation-rate-1960-present> This includes data from 1960 to 2023 with the inflation rate of each country for each year. It also includes the subregions.

Global Corruption Data: <https://www.kaggle.com/datasets/tr1gg3rtrash/global-corruption-index> This dataset has the corruption index for 180 countries from 2013 to 2021.

GDP: <https://www.kaggle.com/datasets/alejopaullier/-gdp-by-country-1999-2022>

This dataset includes 180 countries and their GDPs for each year from 1999 to 2022.

Inflation GDP Data: Merged Dataset from the ones above including inflation and GDP.

The data transformation and database set up is in the Data Curation section of the Jupyter Notebook.

1. **Exploratory data analysis**

All of our exploratory data analysis is included in the Exploratory Data Analysis section of the Jupyter Notebook.

1. **Primary Analysis**

The first two machine learning techniques are k-means clustering where we look at which countries and subregions have similar inflation trends. The reasoning is that k-means can group similar features into clusters that can help us determine where new inflations may be located. The model is in the Primary Analysis: ML 1 and Primary Analysis: ML 2 sections in the Jupyter Notebook.

The third machine learning techniques is a random forest to predict 2024 inflation rates and then test them against the actual 2024 data. The model is in the Primary Analysis: ML 3 section in the Jupyter Notebook.

1. **Visualization**

The visualizations can be found in sections Visualization: ML 1, Visualization: ML 2,

and Visualization: ML 3.

1. **Insights and Conclusions**

Inflation is a reflection of the state of the economy throughout history. Looking at inflation rates can highlight economic recessions and correlate to other economic measures like GDP. Through the data science pipeline, we have observed the role that inflation plays in the economy.

Each analysis has its own conclusions about the data. Overall, there is a small correlation between inflation rate and GDP but no correlation between CPI and inflation. Additionally, countries can have similar inflation trends but also be affected differently by the recession. Throughout the notebook there are links to help an uninformed reader better understand. Through our analysis, an informed reader may learn more about the global inflation rate and the general economic state of the world.