**Title**: Health Insurance Claims Analysis

**By**: Silvio Zabala

**Project Overview**:

This project aims to analyze and visualize patterns and uncover trends in health insurance claims data by analyzing multiple variables.

**Objectives**:

This project will focus on the following goals:

1. Explore and Analyze Claims Data
2. Uncover Claims Key Trends
3. Highlight Claims Data Outliers
4. Use of Interactive Visualizations

**Datasets**:

* The datasets are publicly available health insurance claims.
* The datasets contain over 100 unique records.
* First data set is sourced from Kaggle
  + Key variables include:
    - age
    - sex
    - bmi
    - children
    - smoker
    - region
    - charges
    - insurance claim
  + Website: <https://www.kaggle.com/datasets/mirichoi0218/insurance/data>
* Second datasets are sourced from Centers for Medicare & Medicaid Services (CMS). – They are committed to increasing transparency in the Health Insurance Exchanges with downloadable Public Use Files (PUFs) for research and the public.
  + Key variables include:
    - Claims received
    - Claims denied
    - Appeals
    - Issuer
    - State
  + Website: <https://www.cms.gov/marketplace/resources/data/public-use-files>

**Research Questions**:

1. How do claim charges vary by demographic factors (age, sex, BMI, smoking status, and region)?
2. Which states or regions have the highest claim charges and denial rates, and how do they compare across regions?
3. What is the impact of smoking status and BMI on claim charges or denial rates?
4. Which states or regions have the highest claim charges and denial rates, and what demographic factors contributes to this?
5. Which state or region has the highest rate of appeals and appeals overturned?

**Tools and Technologies**:

* Programing Languages: Python
* Data Cleaning: Pandas, NumPy
* Visualization Libraries: Plotly, Seaborn
  + Example Visualizations: Bar charts, line charts and geographical maps.
* Database: SQLite
* JavaScript Libraries: D3.js(v7), Plotly.js
  + Styling: Bootstrap CSS

**Timeline**:

1. Week 1:
   1. Complete project proposal
   2. Source data and create a database
   3. Clean and process data for exploratory analysis.
   4. Draft the visualizations and interactive dashboard
2. Week 2:
   1. Polished visualizations and interactive dashboard
   2. Complete README.md file.
   3. Create presentation

**References**:

* **Libraries and Frameworks:** Python, Matplotlib, Plotly, SQLite