**Title**: Health Insurance Claims Analysis

**By**: Silvio Zabala

**Project Overview**:

This project aims to analyze and uncover trends in health insurance claims 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. What are the trends of health insurance claims denials by region?
2. Are there any correlations with demographic claim charges and claim denials?
3. What demographics correlate with claims received?

**Tools and Technologies**:

* Programing Languages: Python
* Visualization Libraries:

**Timeline**:

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

**References**:

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