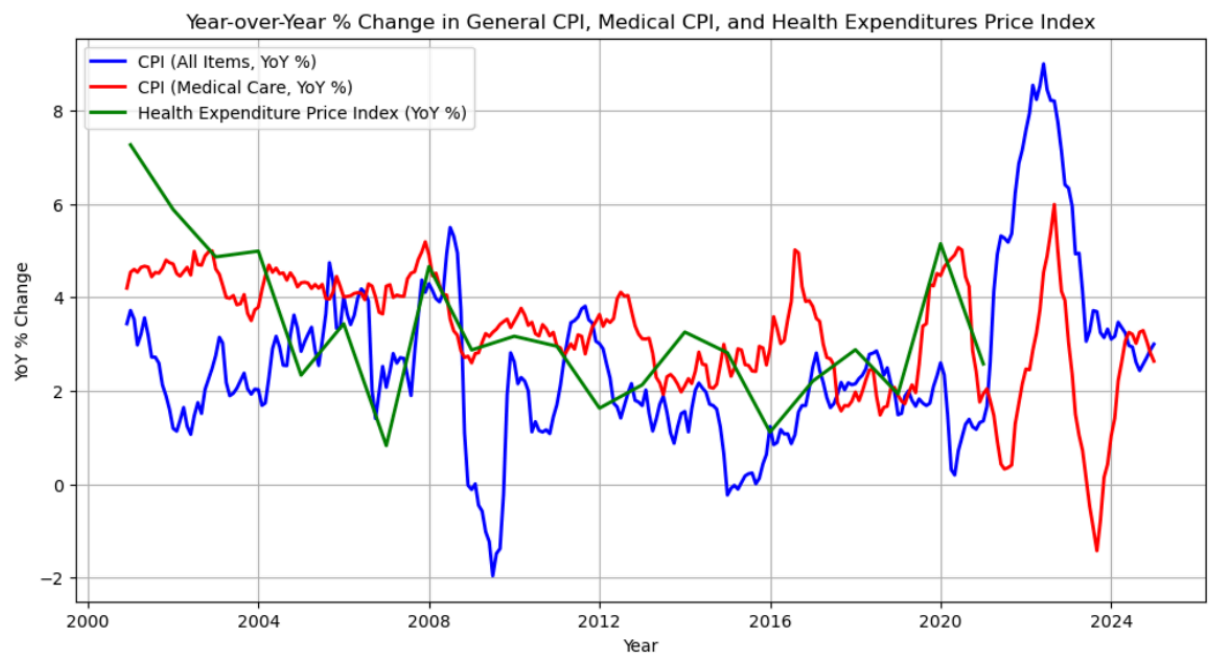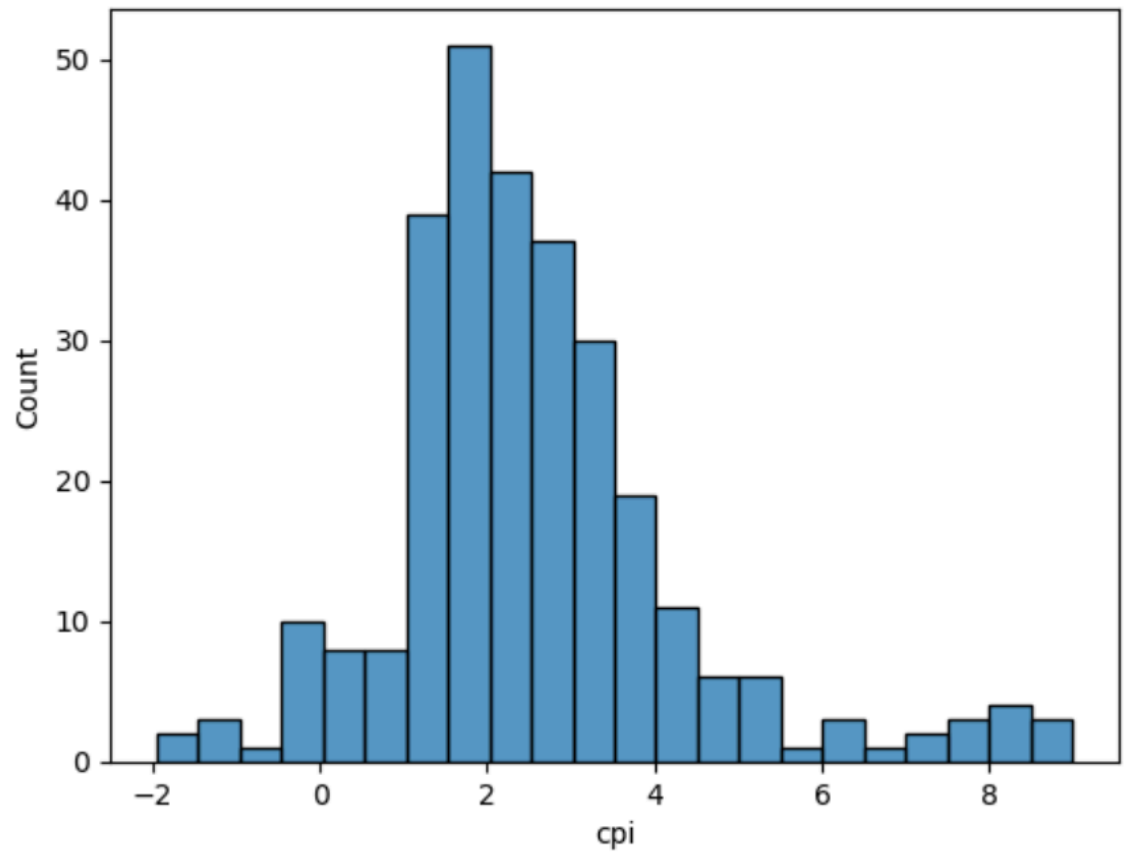# Project 2 DS 4002 Data Appendix

1. Dataset 1: [DS4002P2_RawData.xlsx]

- **1.1 Unit of Observation:** there are two different sections within this Excel file, the first of which shows the annual values of the Health Expenditure Index (HEI), and the second of which shows the monthly values of both the Consumer Price Index(CPI) and Medical Consumer Price Index(MCPI).
- **1.2 Variables Overview:** For the HEI section the variables are observation_date, which shows the year in which the HEI observation was made, and HLTHSCPIBLEND_PC1 which shows the HEI. For the CPI and MCPI section, the first variable is also observation_date, however, the date increases monthly instead of annually as there are more observations of CPI and MCPI. The two other variables in this section are CPIAUCSL_PC1 which shows the CPI and CPIMEDSL_PC1 which shows the MCPI
- **1.3 Variable Subsections :**
  - **[Variable Name]:** observation_date, contains the data at which the data for other variables has been recorded
  - **Data Type:** Date
  - **Notes/Observations:** In the two different sections of the data set this variable can either increase monthly or annually.
- **1.4 Variable Subsections :**
  - **[Variable Name]:** HLTHSCPIBLEND_PC1 contains data on the Health Expenditure Index, updated annually.
  - **Data Type:** float
  - **Notes/Observations:** None
- **1.5 Variable Subsections :**
  - **[Variable Name]:** CPIAUCSL_PC1, contains the data for the overall consumer price index (CPI) which is updated monthly.
  - **Data Type:** float
  - **Notes/Observations:** None
- **1.6 Variable Subsections :**
  - **[Variable Name]:** CPIMEDSL_PC1, contains the data for the medical consumer price index (MCPI), which is updated monthly
  - **Data Type:** float
  - **Notes/Observations:** None

- **1.4 Overall Dataset Descriptive Statistics:**
  - The annual section of the data set contains 21 rows, each showing the HEI for a single year, and the monthly section of the dataset contains 290 rows.

● **2.5 Figures and Visualizations:**

Distribution of total CPI % changes from a year ago (measured monthly)

Year-over-Year % Change in General CPI, Medical CPI, and Health Expenditures Price Index

2. Dataset 2: [Established.csv]

- **2.1 Unit of Observation:** Each row represents a single observation of the CPI, HEI, and MCPI for a specific date.
- **2.2 Variables Overview:** Each observation contains an observation_date, a cpi, and a mcpi value.
- **2.3 Variable Subsections :**
  - **[Variable Name]:** observation_date, contains the date for which an observation was recorded.
    - **Data Type:** Date
    - **Notes/Observations:** The date increases by month from 2000-12-31 to 2025-01-31
  - **[Variable Name]:** cpi, contains the monthly value of the consumer price index.
    - **Data Type:** Float
  - **[Variable Name]:** mcpi, contains the monthly value of the medical consumer price index.
    - **Data Type:** Float
  - **[Variable Name]:** hei, contains the annual value of the health expenditure index
    - **Data Type:** Float
    - **Notes/Observations:** Since the data on HEI only updates annually, the values recorded are repeated 12 times each in the table, in order to match the monthly recording of the other variables. Furthermore, since the HEI data only reaches from 2000-12-31 to 2021-01-31 there are multiple rows without an HEI value at the bottom of the table.

- **2.4 Overall Dataset Descriptive Statistics:**
  - The Dataset contains a total of 291 rows and 4 columns