The Impact of Gasoline Prices on U.S. Gasoline Consumption

Nikolas Lee-Bishop

The purpose of this report is to investigate the relationship between gasoline consumption and gasoline prices in the US. In the real world, economics is based on the principle of supply and demand. The basis of this law is that as prices go up, demand goes down. However, gasoline is an elastic good due to its lack of replacements, so I was curious to see if there was a noticeable difference over time.

The process started by looking for data that could be used. The two datasets that I decided to use were both from the US Energy Information Administration. The two datasets were chosen for their public source being trustworthy and long time period of data. There was no statement about how the data was collected that I’m aware of. The data was collected over years and was last updated at whatever date, by the government, for some purpose. The two datasets are ‘Product Supplied for Finished Motor Gasoline’ which I will reference as the consumption dataset and ‘U.S. Gasoline and Diesel Retail Prices’ which I will reference as the prices dataset.

The first dataset that I started cleaning was the consumption dataset. This dataset consisted of 7 columns: a date column, a total U.S. column, and 5 regions. Each row of data represented a month. The total U.S. column started in January 1945, and the 5 regions started at January 1981. To start with I removed the two header rows that contained the title and source key info. Next, I got rid of the 5 regions and only kept the ‘Date’ column and ‘U.S. Product Supplied of Finished Motor Gasoline (Thousand Barrels)’ which was renamed to ‘Gasoline\_Consumption(Millions of Gallons)’. The Date column was converted to a datetime data type and filtered to include data between 1/1/1994 and 12/31/2024. These dates were selected as they were the first and last year that both datasets had full data for. The next step was converting ‘Gasoline\_Consumption’ to millions of gallons by dividing by multiplying by 1000 and 42 and then dividing by a million. The conversion to millions of gallons was done since the prices dataset was in millions of gallons and to make readability easier.

Once data cleaning was done, the next step was testing basic summary statistics, kurtosis, and skew. The skew and kurtosis was IDK. While doing summary statistics, it was discovered that there was 372 months with a mean of 11,211 millions of gallons a month. The minimum was 7,390 millions of gallons a month and the max was 12,803 millions of gallons a month. The standard deviation was 820 millions of gallons a month.

Once there was a basic understanding of the data, the next step was visualization. The first thing plotted was a simple line graph of Monthly Gasoline Consumption. Then a 12 month rolling average was added to create a smoother line. After that a month-to-month percentage change chart was created to more easily visualize changes from a month-to-month perspective. From there a line chart that had 13 lines, 1 for each month and 1 for a total, was created. Using plotly you can choose which months to view to make it easier to read. Next, with the help of ChatGPT, I asked for help to create a chart that highlighted key economic events that made a change to gas consumption patterns. The final visualization was a heatmap that plotted every month of every year and used darker colors to represent higher monthly consumption.

The second dataset was the prices dataset. This dataset had 16 columns: a date column, and 15 different variations of the average retail price for various fuel tyes and grades including regular, midgrade, premium, and diesel fuels, separated by formulation (conventional, reformulated, and sulfur level). Each row of data represented a month. The two columns I kept were the ‘Date’ column and the ‘U.S. All Grades All Formulations Retail Gasoline Prices (Dollars per Gallon)’. Explain why I kept this column. The first date of data was September 1990 but All Grade All Formulations column didn’t start until April 1993. To start with the cleaning, the top 2 rows of the dataset were removed that contained the title of the chart and the source key info.