CO Open in Colab

COGS 108 – Final Project: Examining the Effects of the Pandemic on Gas Prices Across America

Here is the published video:
https://youtu.be/84L9PQCwoWk

Permissions

Place an X in the appropriate bracket below to specify if you would like your group's project to be made available to the public. (Note that student names will be included (but PIDs will be scraped from any groups who include their PIDs).

- [X] YES – make available
- [ ] NO – keep private

Overview

This project focuses on the factors that lead to gasoline prices fluctuating. Throughout our analysis, we take a look at how the pandemic and the economic status of each state affect gasoline prices. We believe that due to the lockdowns caused by the pandemic gas prices would drop. We also believe that higher economic states have higher gas prices. Our data proves our hypotheses and shows how gas prices are affected by these factors.

Names

- Ian Chin
- Jacqueline Gonzalez
- Jisu Kim

- Sumiah Ali

Research Question

Do the following factors have an effect on gas prices: start of pandemic, economic status of a state and the type of gas?

Background & Prior Work

Gasoline prices are always fluctuating. In the United States alone, we saw gas prices reach all time highs and we also saw them decline at times. There are several factors that may account for these changes, such as crude oil costs, taxes and refining costs and profits [2]. Even factors such as weather can cause changes in gas prices. The two factors that we looked into for this project were the 2020 pandemic and the economic status of each state to see how and if these factors result in gas prices changing.

The 2020 global pandemic caused many countries including the United States to issue a national emergency, which resulted in lockdowns. This in turn reduced economic activity across the country. Previous research on the impact that the Covid-19 pandemic had on gasoline prices showed that there was a sharp drop in March and April of 2020. This research showed that this drop was a result of low demand, increase in supply, and decreased storage space [1]. It also showed that the initial drop did not last long, because gas prices quickly rose back up, as the demand for it returned.

There is also previous research that reveals the relationship between the economic status of a state and their gas prices. This research indicates that gasoline prices vary across states and regions. They show that this could be due to a number of reasons including, "distance from supply, supply distributions, and retail competition and operating

costs" [3]. In this project, we take a look at three states in different economic brackets, which may allow us to analyze whether or not the states with a higher economic status also have higher gas prices. We would like to determine through our data and analysis how gas prices change based on these factors.

References:

[1] - From the barrel to the pump: the impact of the COVID-19 pandemic on prices for petroleum products
https://www.bls.gov/opub/mlr/2020/article/from-the-barrel-to-the-pump.htm

[2] - Gasoline explained, Factors affecting gasoline prices
https://www.eia.gov/energyexplained/gasoline/factors-affecting-gasoline-prices.php

[3] - Gasoline explained, Regional gasoline price differences
https://www.eia.gov/energyexplained/gasoline/regional-price-differences.php

Hypothesis

We believe that gas prices during the start of the pandemic, march 2020 the gas prices will be lower because there will be less people on the road due to quarantine. The states that are more populated will have higher prices. The gas prices will rise no matter the type of gas.

Dataset(s)

- Dataset Name: U.S. Gasoline and Diesel Retail Prices 1995-2021
- Link to the dataset:
  https://www.kaggle.com/datasets/mruanova/us-gasoline-and-diesel-retail-prices-19952021?resource=download

- Number of observations: 51

- Dataset Name: From the barrel to the pump:
  the impact of the COVID-19 pandemic on
  prices for petroleum products

- Link to the dataset:
  https://www.bls.gov/opub/mlr/2020/article/from-
  the-barrel-to-the-pump.htm
- Number of observations: 10

- Dataset Name: State Gas Price Averages

- Link to the dataset:
  https://gasprices.aaa.com/state-gas-price-
  averages/
- Number of observations: 50

- Dataset Name: Gross Domestic Product by
  State

- Link to the dataset:
  https://apps.bea.gov/iTable/iTable.cfm?
  reqid=70&step=30&isuri=1&year_end=-1&acrdn=1&
  industry&state=0&yearbegin=-1&unit_of_measure=
- Number of observations: 51

1-2 sentences describing each dataset.

The first data set describes the varying gas prices
based off time, the second dataset describes
pandemic and gas prices, and the third describes
gas prices of each state.

If you plan to use multiple datasets, add 1-2
sentences about how you plan to combine these
datasets.

The first dataset will be used to compare the
difference between the three states in pertaining to
their respective enconomic status. The second data
set will describe the relationship between the
pandemic and gas prices.

Setup

In [1]: ## YOUR CODE HERE

```
## YOUR CODE HERE
%matplotlib inline

import numpy as np
import pandas as pd
import math
import scipy.stats as stats
import warnings
warnings.filterwarnings('ignore')

# visualization
import matplotlib.pyplot as plt
import seaborn as sns

# add more setups later
```

Data Cleaning

Describe your data cleaning steps here.

In [2]:
```
## YOUR CODE HERE
## FEEL FREE TO ADD MULTIPLE CELLS PER SECTIO
gas_prices = pd.read_csv('https://raw.githubu

gas_prices['Date'] = pd.to_datetime( gas_pric

##only 2020 data

gas_prices = gas_prices[(gas_prices['Date'] >
gas_prices = gas_prices.drop(labels = ['A2',
gas_prices = gas_prices.rename({'A1':'All', '
gas_prices = gas_prices.reset_index()
gas_prices = gas_prices.drop(axis=1,labels =

## Pandemic Gas Prices

pandemic_gas_prices = pd.read_csv('https://ra
pandemic_gas_prices = pandemic_gas_prices.tra
pandemic_gas_prices.columns = ['Month','Chang
pandemic_gas_prices = pandemic_gas_prices[1:]
pandemic_gas_prices = pandemic_gas_prices.ast
pandemic_gas_prices = pandemic_gas_prices.set

##State Gas Prices

state_gas_prices = pd.read_csv('https://raw.g
state_gas_prices = state_gas_prices.rename({'
state_gas_prices['Regular'] = state_gas_price
state_gas_prices['Mid-Grade'] = state_gas_pri
state_gas_prices['Premium'] = state_gas_price
state_gas_prices['Diesel'] = state_gas_prices
state_gas_prices = state_gas_prices[state_gas
state_gas_prices = state_gas_prices.melt(id_v
state_gas_prices = state_gas_prices.astype('f
state_gas_prices = state_gas_prices.reset_ind

## State GDP
```

```
state_gdp = pd.read_csv('https://raw.githubus
state_gdp = state_gdp[state_gdp.GeoName.isin(
state_gdp = state_gdp.melt(id_vars=['GeoName'
```

Data Analysis & Results

Include cells that describe the steps in your data
analysis.

In [3]:
```
gas_prices.shape
```

Out[3]: (52, 6)

In [4]:
```
gas_prices.head()
```

Out[4]:

	Date	All	Regular	Mid-Grade	Premium	Diesel
0	2020-01-06	2.665	2.578	2.973	3.214	3.079
1	2020-01-13	2.657	2.570	2.964	3.209	3.064
2	2020-01-20	2.625	2.537	2.940	3.183	3.037
3	2020-01-27	2.595	2.506	2.911	3.163	3.010
4	2020-02-03	2.546	2.455	2.872	3.125	2.956

In [5]:
```
gas_prices.describe()
```

Out[5]:

	All	Regular	Mid-Grade	Premium	
count	52.000000	52.000000	52.000000	52.000000	5
mean	2.258250	2.167519	2.583000	2.834115	
std	0.193219	0.194744	0.186252	0.184935	
min	1.870000	1.773000	2.221000	2.478000	
25%	2.192500	2.101000	2.520000	2.772500	
50%	2.258000	2.167500	2.578500	2.827500	
75%	2.311000	2.222500	2.625250	2.873250	
max	2.665000	2.578000	2.973000	3.214000	

In [6]:

```
plt.rcParams['figure.figsize'] = (17, 7)
sns.set_theme(context='notebook',style='white
sns.lineplot(data=pd.melt(gas_prices,['Date']
plt.legend(title='Gas Type')
```

Out[6]:   `<matplotlib.legend.Legend at 0x7ff20b5ac490>`
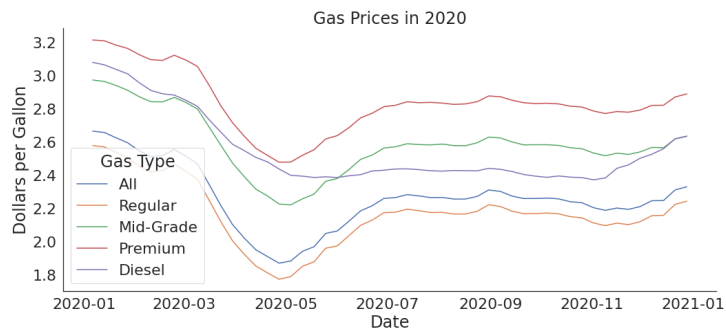
This shows the prices of various types of gas
throughout 2020. There is a notably sharp decrease
in price of all types of gas around March, which
reaches its minimum at May, then quickly increases
until July, where it remains steady for the rest of the
year.

In [7]:
```
pandemic_gas_prices.shape
```

Out[7]:   `(10, 1)`

In [8]:
```
pandemic_gas_prices.head()
```

Out[8]:

Change in Gas Price

Month	
January 2020	-1.6
February 2020	-3.4
March 2020	-10.5
April 2020	-20.6
May 2020	-3.5

In [9]:
```
pandemic_gas_prices.describe()
```

Out[9]:

Change in Gas Price

count	10.00000
mean	-3.67000
std	14.16671
min	-31.40000

25%	-8.75000
50%	-2.50000
75%	4.70000
max	14.40000

In [10]:
```
sns.lineplot(data=pandemic_gas_prices[0:-2],x
```

Out[10]:
```
[Text(0, 0.5, 'Percent Price Change per Mont
h'),
 Text(0.5, 1.0, 'Change in Gas Price in 202
0')]
```
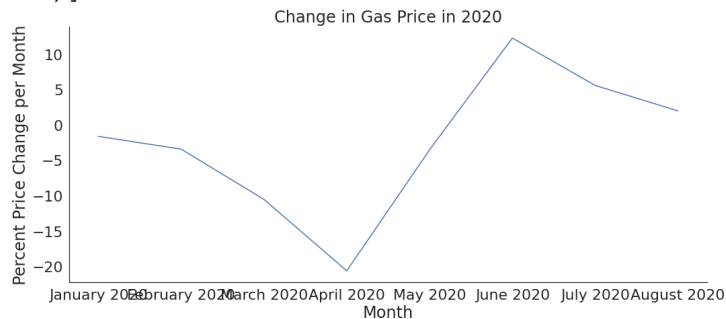

This graph uses a different data set to show gas prices similar to the previous graph, though this one uses percentage change. Like before, prices drop hard in March and April and don't come back up until June.

In [11]:
```
state_gdp.shape
```

Out[11]: (24, 3)

In [12]:
```
state_gdp.describe()
```

Out[12]:

	GDP
count	2.400000e+01
mean	1.185517e+06
std	1.447929e+06
min	1.229587e+05
25%	1.451938e+05
50%	2.358810e+05
75%	3.052559e+06
max	3.513348e+06

In [13]:
```
print(state_gdp)
```

```
      GeoName    Quarter          GDP
0      Alabama    2020:Q1     231792.8
1     Arkansas    2020:Q1     132472.4
2   California    2020:Q1    3083611.2
3      Alabama    2020:Q2     210576.3
4     Arkansas    2020:Q2     122958.7
5   California    2020:Q2    2801946.6
6      Alabama    2020:Q3     231581.2
7     Arkansas    2020:Q3     133036.8
8   California    2020:Q3    3042208.0
9      Alabama    2020:Q4     233635.5
10    Arkansas    2020:Q4     134534.9
11  California    2020:Q4    3100984.9
12     Alabama    2021:Q1     238126.4
13    Arkansas    2021:Q1     138770.9
14  California    2021:Q1    3207138.8
15     Alabama    2021:Q2     243786.8
16    Arkansas    2021:Q2     143004.5
17  California    2021:Q2    3321704.1
18     Alabama    2021:Q3     248991.2
19    Arkansas    2021:Q3     145923.6
20  California    2021:Q3    3384335.2
21     Alabama    2021:Q4     257465.4
22    Arkansas    2021:Q4     150482.7
23  California    2021:Q4    3513347.5
```

In [14]:
```
sns.lineplot(data=state_gdp,x='Quarter',y='GD
```

Out[14]:
```
[Text(0.5, 1.0, 'GDP in Various States in 202
0-2021')]
```
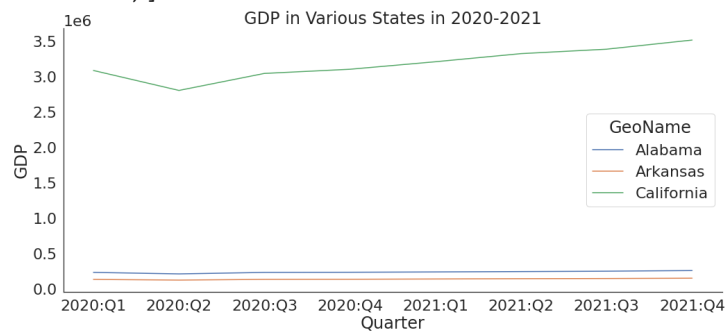

This shows the GDP of a few different states during 2020 and 2021. Like with gas prices, GDP fell somewhat in the second quarter of 2020. This is most noticable in a large state like California, but looking at the data shows that a decrease did occur, just not as significantly. Furthermore, the decreases here are nowhere near as extreme as with gas prices.

In [15]:
```
state_gas_prices.shape
```

Out[15]:
```
(12, 3)
```

In [16]:
```python
state_gas_prices.head()
```

Out[16]:

	State	Type	Price
0	Alabama	Regular	3.952
1	Arkansas	Regular	3.881
2	California	Regular	5.785
3	Alabama	Mid-Grade	4.290
4	Arkansas	Mid-Grade	4.183

In [17]:
```python
state_gas_prices.describe()
```

Out[17]:

	Price
count	12.000000
mean	5.032583
std	0.912046
min	3.881000
25%	4.263250
50%	4.949000
75%	5.833250
max	6.502000

In [18]:
```python
sns.barplot(data=state_gas_prices,x='Type',y=
```

Out[18]:
```
[Text(0, 0.5, 'Dollars per Gallon'),
 Text(0.5, 1.0, 'Gas Prices in Various State
 s')]
```
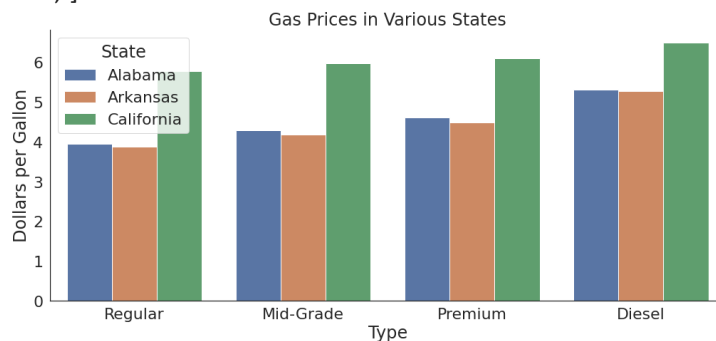

This graph shows the recent prices of different types of gas in the states of Alabama, Arkansas, and California. The main point to note here is that California, which has a higher population and GDP than the other states, also has consistently higher prices for all types of gas.