Hey, What's Up?
Gas Prices: Analyzing the Influences of U.S. Gas Price Trends
Our Team

Karen Wang
Phillips Academy '23
Princeton, NJ

Brian Ling
Lynbrook High School '23
San Jose, California

Jennifer Li
Westlake High School '23
Texas

Christine Lam
Hong Kong International School '24
Hong Kong
Objectives
Background

- Highest gas prices in U.S. history
- High Fluctuation in prices
  - Russia-Ukraine conflict
  - COVID-19
  - Lack of refinery space
Research Questions

What factors affect gas prices?

Why are gas prices so high?
Data + EDA
Explanatory Variables

**Economics**
- Imports, Exports, Production, Consumption, Poverty, Income

**Energy**
- Coal, Gas, Renewable Energy, Natural Gas, CO2 emissions

**Weather**
- Seasons

**Google Search**
- Oil, Gas
Datasets (Sources)

American Petroleum Institute
U.S. Energy Information Administration
The World Bank
National Broadcasting Company
Google Search Trends
Dataset

- Longitudinal data - 35 variables
- Each month from Jan 2000 - June 2022
- Descriptive data from previously mentioned sources
Response Variable: Gas Price

Average retail gas price of premium and regular gas in the U.S. in USD per month from the U.S. Energy Information Administration
Economics

Production: Total operating oil rigs, wells drilled, gas barrels produced, gas demanded and consumed

Trade: Oil imports and exports to the U.S. by Country, net imports and exports

Money: Income, Poverty %

Important Observations:

- Canada imports the most gas to the U.S.
- Consumption relatively consistent
- Production + Exports same behavior
- Drop in Imports
All Country Imports to USA over Month (in Thousands)

Month

Number of Barrels (thousand)

- Canada
- Colombia
- Ecuador
- Iraq
- Mexico
- Nigeria
- Russia
- Saudi Arabia
- South Korea
- UK
Oil Consumption, Production, Imports, and Exports over Year (in millions of Barrels)

- **Consumption** (red line)
- **Exports** (green line)
- **Imports** (teal line)
- **Production** (purple line)

**Y-axis:** Number of Barrels (millions)

**X-axis:** Year

The graph shows the trend of oil consumption, production, imports, and exports over several years, with consumption remaining relatively stable until around 2015, after which it shows a steady increase. Exports start at a lower level and show a significant increase over time. Imports show a pattern of fluctuation with a notable rise in the latter years. Production, indicated by the purple line, also exhibits a rising trend, especially from 2015 onwards.
Energy

Energy Consumed: Coal, Natural Gas, Gas, Renewable Energy

CO₂ Emissions by Sector: Agriculture, Commercial, Electricity, Industry, Residential, Transportation

Important Observations:

- Steady increase of renewable energy consumed
- Drop in coal consumed
- Overall decline of gas consumed, mirroring gas price drops in 2008 and 2020
- Drop in CO₂ emissions for Electricity and Transportation
Google Search

**Search Variable:** Popularity value from 0-100 for “Gas” and “Oil”

**Important Observations:**

- “Gas” and “Oil” generally mirror each other
- Spike in searches when sudden drop and rise in gas price
Analysis
Sentiment Analysis
LASSO and Regression
Obama seeks joint approach with Russia on Iran nuclear bomb

Oil rig workers hit with one-two punch of coronavirus and plummeting oil prices
Rise in Price

Fall in Price

Analysis
Initial Factors

Factors after LASSO + Back Selection

| Coefficients: | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------|----------|------------|---------|---------|
| (Intercept)   | -4.577e+00 | 8.926e-01 | -5.128  | 5.79e-07 *** |
| net.imports   | 1.213e-01  | 1.761e-02 | 6.886   | 4.42e-11 *** |
| income        | 8.871e-05  | 8.229e-06 | 10.781  | < 2e-16 *** |
| operation     | 7.529e-04  | 5.219e-05 | 14.425  | < 2e-16 *** |
| canImp        | 7.984e-06  | 2.036e-06 | 3.922   | 0.00013 *** |
| searchgas     | 2.741e-02  | 1.795e-03 | 15.266  | < 2e-16 *** |
| searchoil     | -1.099e-02 | 1.558e-03 | -7.055  | 1.61e-11 *** |
| renewal       | 1.929e-16  | 3.399e-17 | 5.675   | 3.76e-08 *** |
| Commercial    | -7.971e-09 | 1.708e-09 | -4.666  | 4.95e-06 *** |
| Electricity   | 3.552e-09  | 2.435e-10 | 14.589  | < 2e-16 *** |
| Transportation| -4.118e-09 | 3.360e-10 | -12.256 | < 2e-16 *** |
| seaspring     | 2.724e-02  | 4.423e-02 | 0.616   | 0.538488 |
| seasonsummer  | 5.633e-02  | 4.441e-02 | 1.269   | 0.205769 |
| seasonwinter  | -1.221e-01 | 4.602e-02 | -2.653  | 0.008466 ** |

---

Coefficients:

| Coefficients: | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------|----------|------------|---------|---------|
| (Intercept)   | -2.382e+01 | 8.889e+00 | -2.680  | 0.00789 ** |
| net.imports   | 3.812e-03  | 1.169e-01 | 3.262   | 0.00127 ** |
| income        | 8.593e-05  | 1.873e-05 | 4.592   | 7.16e-06 *** |
| poverty       | 1.600e-01  | 8.270e-02 | 1.942   | 0.05339 |
| wells         | -3.815e-05 | 1.551e-04 | -0.536  | 0.59237 |
| operation     | 3.682e-04  | 2.529e-04 | 1.456   | 0.14683 |
| musImp        | 5.255e-03  | 2.851e-03 | 1.852   | 0.060194 |
| colImp        | -3.346e-06 | 7.817e-06 | -0.428  | 0.66003 |
| maxImp        | -2.494e-04 | 1.502e-04 | -1.600  | 0.11087 |
| cliImp        | -1.217e-05 | 7.660e-06 | -1.588  | 0.11353 |
| aquImp        | -4.553e-05 | 9.335e-05 | 0.488   | 0.62617 |
| regImp        | 1.697e-06  | 4.009e-06 | 0.423   | 0.67255 |
| aquonImp      | -9.154e-06 | 1.131e-05 | -0.809  | 0.41922 |
| conImp        | 6.761e-06  | 2.455e-06 | 2.754   | 0.007864 ** |
| treeImp       | 3.332e-06  | 3.616e-06 | 0.921   | 0.35780 |
| catImp        | 2.512e-05  | 2.014e-05 | 1.247   | 0.21352 |
| searchgas     | 2.564e-02  | 2.079e-03 | 12.384  | < 2e-16 *** |
| searchoil     | -8.299e-03 | 1.780e-03 | -4.611  | 6.58e-06 *** |
| metaImp       | -1.002e-16 | 4.044e-17 | -2.523  | 0.01229 * |
| potImp        | -1.818e-05 | 1.258e-05 | -1.445  | 0.14906 |
| renewab       | 6.449e-17  | 1.842e-16 | 0.353   | 0.72472 |
| Agrivulture   | 3.911e-09  | 5.171e-09 | 0.756   | 0.45918 |
| Commercial    | -2.185e-09 | 7.102e-09 | -0.304  | 0.76155 |
| Electricity   | 1.751e-09  | 9.706e-10 | 1.804   | 0.07246 . |
| Industry      | 6.868e-09  | 1.658e-09 | 3.912   | 0.00012 *** |
| Residential   | 3.930e-09  | 3.575e-09 | 1.099   | 0.27282 |
| Transportation| -2.415e-18 | 2.428e-09 | -0.099  | 0.92985 |
| gasim         | -9.110e-06 | 6.485e-06 | -1.404  | 0.88840 |
| nuclear       | 5.453e+00  | 7.129e+00 | 0.765   | 0.44507 |
| hydroelectric | 1.052e+01  | 5.548e+00 | 1.897   | 0.05909 |
| geothermal    | 2.812e+02  | 9.331e+01 | 3.014   | 0.00286 . |
| wind          | 3.153e+01  | 1.184e+01 | 2.664   | 0.00827 ** |
| seaspring     | 6.488e-03  | 4.765e-02 | 0.136   | 0.89193 |
| seasonsummer  | 6.326e-02  | 4.586e-02 | 1.379   | 0.16912 |
| seasonwinter  | -1.235e-01 | 5.043e-02 | -2.449  | 0.01508 * |
LASSO Assumption Check

Residuals vs Fitted

Normal Q-Q

Im(allowprice ~ mexlmp + ecuadorlmp + canlmp + sklmp + coal + natgas + pet ...)

Im(allowprice ~ mexlmp + ecuadorlmp + canlmp + sklmp + coal + natgas + pet ...)
Conclusion
Findings

**INCOME:**
- Median annual income of US residents

**IMPORTS:**
- US petroleum net imports per month

**RENEWAB:**
- Metric tons of renewable energy usage in the US per month

**SEARCHGAS:**
- Units popularity in “gas” as a google search term

Conclusion
Limitations

- Dataset
- Time
Takeaways:

Gas Prices are generally lower during Winter months.

Importing more Gas into the U.S generally raises prices.

CO2 emissions of the Commercial, Electricity, and Transportation industries affect gas prices.

Important to note other factors (politics, current events, etc.)
Thank You!

Any Questions?

Special Thanks and Acknowledgements to:

*Our Professor:* Linda Zhao,

*Our Teaching Fellow:* Edward Zhang,

*Our Advisor:* Jeff Cai,

*The Wharton Data Science Summer Academy* and finally…

*The WiDS Organizational Committee!*
References


<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td>Time in year, month, day</td>
</tr>
<tr>
<td>allgasprice</td>
<td>Avg retail gas price of premium and regular gas in the U.S. in U.S. dollars per month</td>
</tr>
<tr>
<td>premgasprice</td>
<td>Premium retail gas price in the U.S. in U.S. dollars per month</td>
</tr>
<tr>
<td>reggasprice</td>
<td>Regular retail gas price in the U.S. in U.S. dollars per month</td>
</tr>
<tr>
<td>importacqcost</td>
<td>Gas price purchased by U.S. refiners in U.S. dollars per month</td>
</tr>
<tr>
<td>oilimport</td>
<td>U.S. import of crude oil in thousands of barrels per month</td>
</tr>
<tr>
<td>oildemand</td>
<td>U.S. total product demand in millions of barrels per day</td>
</tr>
<tr>
<td>consumption</td>
<td>U.S. petroleum consumption in millions of barrels per day</td>
</tr>
<tr>
<td>production</td>
<td>U.S. petroleum production in millions of barrels per day</td>
</tr>
</tbody>
</table>
### Dataset Factors & Explanation Cont.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>exports</td>
<td>U.S. petroleum exports in millions of barrels per day</td>
</tr>
<tr>
<td>net.imports</td>
<td>U.S. petroleum net.imports in millions of barrels per day</td>
</tr>
<tr>
<td>income</td>
<td>Median income per household in U.S. dollars per year</td>
</tr>
<tr>
<td>poverty</td>
<td>Poverty percentage in the U.S. per year</td>
</tr>
<tr>
<td>wells</td>
<td>Total crude oil wells drilled per month</td>
</tr>
<tr>
<td>operation</td>
<td>Total crude oil rotary rigs in operation per month</td>
</tr>
<tr>
<td>(country)Imp</td>
<td>U.S. crude oil imported from specific countries in thousands of barrels per month</td>
</tr>
<tr>
<td>searchgas</td>
<td>Google gas search result popularity value (0-100 where 100 is peak popularity) per month</td>
</tr>
<tr>
<td>searchoil</td>
<td>Google gas search result popularity value (0-100 where 100 is peak popularity) per month</td>
</tr>
</tbody>
</table>
## Dataset Factors & Explanation Cont.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>coal/natgas/petr/renewab</td>
<td>Total Consumption of Coal, Natural Gas, Petroleum, and Renewable Gas (in Metric Tons)</td>
</tr>
<tr>
<td>us_emission</td>
<td>Percentage that US emissions of CO2 makes up of global emissions of CO2</td>
</tr>
<tr>
<td>nuclear/hydroelectric/geothermal/solar/wind</td>
<td>Percentage that Each Type of Renewable Energy makes up of total Renewable Energy Consumption</td>
</tr>
<tr>
<td>Agriculture/Commercial/Electricity/Industry/Residential/Transportation</td>
<td>Emissions of CO2 for Each Sector (in Metric Tons of CO2 Equivalent)</td>
</tr>
<tr>
<td>imports</td>
<td>U.S. petroleum imports in millions of barrels per day</td>
</tr>
</tbody>
</table>
Can we predict future gas prices?

Yes

Using factors found with multiple regression

But...

Can't account for unpredictable factors (current events, politics) and potential lurking confounding variables (inflation)
Random Forest

0.572% mean testing error
Gas Supply Chain