Capstone Project
Natural Gas Price Prediction

Inspiration

Natural Gas has become a very precious resource to the world as it serves many purposes such as powering up different industries, heating our homes, and cooking food. Since the discovery of this natural resource, in the early decades of its use in America and other parts of the world, the prices were a bit high as there was less knowledge about the reserves, and due to the lack of advanced mining technologies, it was thought that it's going to end soon. This was the reason behind its high prices. As the technology developed and new mining techniques and machinery came along, the prices of Gas went down as it is said to have 100 more years of reserves. Another factor that plays a vital role in the prices of Natural Gas is the demand; as the demand increases in Winters, the prices go up a bit, and in summer, the prices come down as there is less demand at that time(*Natural Gas Prices Forecast & Predictions for 2022, 2023, 2025-2030*, n.d.).

The current global situation where Russia is invading Ukraine has impacted the prices of Natural Gas a lot, as the EU and United Kingdom have cut off their Natural Gas and Crude Oil demand from Russia, creating a resource scarcity. This has increased the Gas prices drastically; this spontaneous change in circumstances is exciting as this is something that was not predicted. We can take the Covid pandemic as an example here, where the world saw a halt in almost every sector. Still, it proved beneficial for the Natural Gas industry as it was the only one to keep its value and the reason behind it was the demand. Most of the world was home-bound, which increased the demand(*Natural Gas Price Forecast*, n.d.). As we see and hear nowadays that the world is turning towards renewable energy and sustainable energy with low carbon emissions to protect the planet, Natural Gas is an advantage in this regard as it has low carbon emissions compared to other fossil fuels.

This project aims to identify different data points from the data we collect for the United States. With the help of Machine Learning and Visualization techniques, we can predict the Gas Prices in different states of America.

Project Design

The whole idea is to develop an end-to-end web application that could provide us the visual information based on the analysis and predictions from the data for the Natural Gas prices in Petroleum Administration for Defense Districts (PADDs). The whole application is based on the Flask Web Framework, which runs on Python; Flask is a Model, View, Template (MVT)

framework, a microservice framework mainly used for the development of Data Analytics and Machine Learning based web applications. It has the power of Python, and support for Python packages makes it more flexible for integration with Machine Learning models and different visualizations. In terms of the prediction part of the project, we developed a Supervised Machine Learning model, which helps us predict the Natural Gas prices as it was trained on the historical data from which it finds the trends and patterns to predict the outcome with more accuracy. The last part was to visualize the data for the Natural Gas prices over time; for that, we had to develop a Data Visualization dashboard for which we used Tableau. Tableau is a potent tool for Data Analytics and Business Analytics as it provides much flexibility when visualizing data in a meaningful way. So, the goal was to develop a dashboard that could imagine the following:

- A plot of Natural Gas Prices over time
- A plot of Natural Gas prices and consumption over time
- A story of Natural Gas production over time

The dashboard for tableau is also deployed at this link.

Data Sources

The data we used for the whole project was from the Petroleum Administration for Defense Districts (PADDs), as they had many data from the past several years.

Hypothesis

From the development point of view of this project, we can hypothesize the following:

H1: A baseline web application provides us with the Natural Gas Price Prediction based on the historical data and visualizes the trends with the help of a dashboard.

The whole purpose of the hypothesis is to outline the desired outcome in a meaningful manner to help us focus on the primary goal.

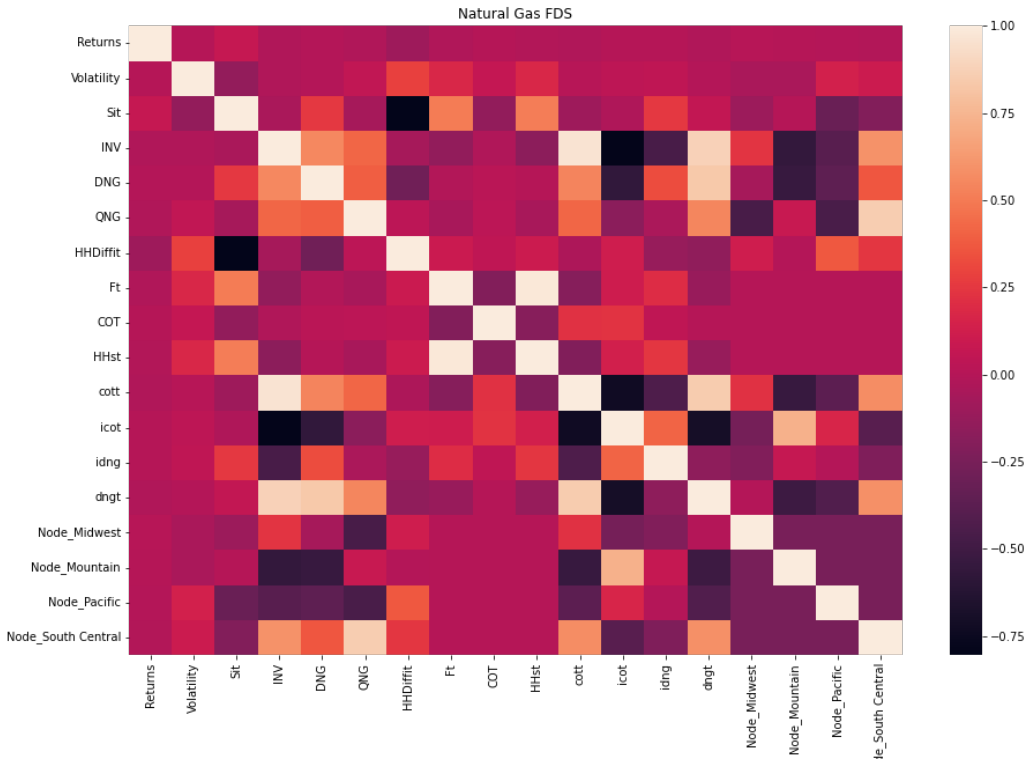
Cleaning Data/EDA

The data we got from PADDs had a lot of missing data and data that was not useful for our project. We applied some data preprocessing or data cleaning techniques to deal with it. Data Cleaning is a crucial part of any Data Analysis project as the outcomes or predictions are based on the data we feed into the model. To get the correct output from a Machine Learning model, it is essential to feed the correct data. With the help of the tools and libraries provided by Python, such as Pandas, we can deal with any garbage data or missing values in the data and make it more useable and

filtered so only the required things can go into our model. So, with the help of Pandas, we dealt with the missing data and removed some of the data that was not required for our project. Exploratory Data Analysis is also a technique that helps us identify any data points that could be useful for a problem. It is like the initial analysis, which provides insights into the data; this technique is also proper for identifying features that could be helpful for the Machine Learning model.

Machine Learning

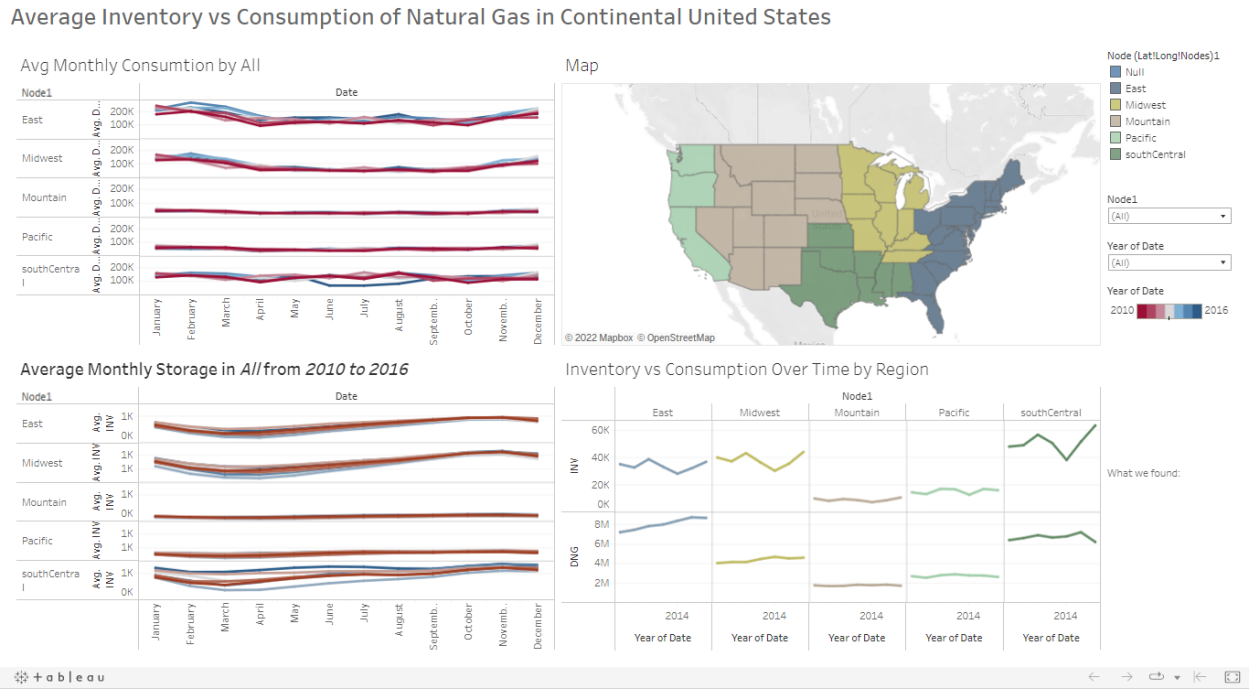
After performing the data preprocessing and Exploratory Data Analysis on the data, we could identify different features and elements from the data as the data was transformed into a cleaner shape which included other parameters such as Volatility, DNG, PNG, etc. The preprocessed data were used for the Linear Regression model training from which we are predicting the Natural Gas Prices. The model was trained on the 80-20 split, i.e., 80% for the training data and 20% for the test data. We can see the model's prediction correlation matrix in the following figure:



The developed Machine Learning model is integrated into the Flask web application, where we can display the predictions.

Data Visualization Dashboard

We also developed a Tableau dashboard that displays the necessary visualizations to identify any patterns or insights from it. The dashboard shows the Average Monthly Consumption, a map for the Continental United States which shows the different regions by colors, Average Monthly Storage in all areas from 2010 to 2016, and a comparison of Inventory with consumption over the period region-wise. The following figure shows the Dashboard visualizations:



Flask Web Application Interface

Conclusion

The overall aim of this project was to predict the Natural Gas prices across the Continental United States with the help of available data. We developed a Machine Learning model to predict the Natural Gas Prices. A Flask-based web application was designed to integrate the model prediction and the visualization from the Tableau dashboard. The most important part of this project was the Data Cleaning and Exploratory Data Analysis phase, where we dealt with the missing data and preprocessing data to make it more usable for the machine learning model. The overall purpose of this project is well achieved, and it can be expanded more in the future.

Limitations

There are several limitations we faced while working on this project. The major problem was acquiring the data, as there is minimal data available for getting it from open-source websites. The project is a very baseline product that aims to solve the problem of Natural Gas price prediction, which could help the governing bodies to prepare for different situations. The implemented Machine Learning model is a Linear Regression model, which is the most basic one and provides the predictions to a certain level. So, it needs to be extended to a more advanced level.

Future Work

In prospect, we can extend this project to a more advanced level by adding more features to the dashboard visualizations; we can add more improvements to the overall web application design and the Machine Learning model by adding more data and training it over it. We can also test and play around with other models and perform a performance comparison among them to see which model is best for our use case. There are many possibilities for improving this project or building upon it.

References

Natural gas price forecast: Prices hit 14-year highs on supply concern. (n.d.). Retrieved June 11, 2022, from https://capital.com/natural-gas-price-forecast

Natural Gas Prices Forecast & Predictions for 2022, 2023, 2025-2030. (n.d.). PrimeXBT. Retrieved June 11, 2022, from https://primexbt.com/for-traders/natural-gas-price-prediction-forecast/