US RIG COUNT ANALYSIS

Allowing international rig counts to drive domestic investments.

US RIG
COUNT
ANALYSIS

Contents of the Report

Introduction

The oil industry has a large affect on the standard of living of most Americans. It's peaks and troughs affect everything from gas prices at the pumps, energy bills and the job market.

Oil rigs are the life blood of the oil and gas industry and is a marker in revealing its health. The motivation for this analysis was to analyze global supply and demand data, Brent crude price and rig counts to predict US rig counts.

These predictions can be used by executives to determine where resources should be placed or pulled related to future rig activity.

Data Exploration

```
In [2]: 1 file = "../Data/Capstone_Data.csv"

In [3]: 1 oil_df = pd.read_csv(file)

In [4]: 1 oil_df.head()
```

Out[4]:

	Year	Quarter	Canada_D	Europe_D	Japan_D	US_D	China_D	Soviet_D	Asia_D	Other_D	Total_World_D	Canada_S	Mexico_S	North_Sea_S	Other_S
0	2019	Q3	2.57	15.44	3.43	20.88	14.37	5.58	13.74	26.24	102.25	5.47	1.93	2.96	4.64
1	2019	Q2	2.32	14.95	3.39	20.63	14.65	5.19	14.11	25.77	101.00	5.47	1.91	2.96	4.59
2	2019	Q1	2.31	14.82	4.06	20.55	14.46	5.15	13.95	25.22	100.49	5.43	1.91	2.96	4.85
3	2018	Q4	2.58	14.93	3.89	20.75	14.10	5.36	13.82	25.29	100.73	5.62	1.95	2.95	4.89

Data Gathering

Finding a single dataset with enough relevant information on US rig was a challenge. We combined the following 3 datasets:

- Global Oil Supply/Demand
- Brent Oil Prices
- Baker Hughes International Rig Count

Data Preparation

We quickly discovered the quarter column of our data was a string and not an integer.

This required using pandas get_dummies method to one hot encode the quarter data correctly.

Feature Exploration & Correlation

Due to our desire to create a more usable model for the end user, we wanted to try and eliminate some of the columns that appeared to be less highly correlated with our US Rig Count target

Data Exploration

In [8]:
```
#One Hot Encoding
dummies = pd.get_dummies(oil_df.Quarter)
oil_encode_df = pd.concat([oil_df, dummies], axis='columns')
oil_encode_df.head()
```

Out[8]:

rent_Crude_Price	Latin_America_Rigs	Europe_Rigs	Africa_Rigs	Middle_East_Rigs	Asia_Pacific_Rigs	Total_Intl_Rigs	Canada_Rigs	US_Rigs	Q1	Q2	Q3	Q4
61.95	195	190	114	422	224	1144	132	920	0	0	1	0
69.04	186	159	122	412	230	1109	83	989	0	1	0	0
63.10	188	92	116	398	235	1030	186	1046	1	0	0	0

Data Gathering

Finding a single dataset with enough relevant information on US rig was a challenge. We combined the following 3 datasets:

- Global Oil Supply/Demand
- Brent Oil Prices
- Baker Hughes International Rig Count

Data Preparation

We quickly discovered the quarter column of our data was a string and not an integer.

This required using pandas get_dummies method to one hot encode the quarter data correctly.

Feature Exploration & Correlation

Due to our desire to create a more usable model for the end user, we wanted to try and eliminate some of the columns that appeared to be less highly correlated with our US Rig Count target

Data Exploration

Correlations

In [13]:
```
1 df_scaled.corr()
```

Out[13]:

	Total_Intl_Rigs	Asia_Pacific_Rigs	Brent_Crude_Price	Latin_America_Rigs	US_Rigs
Total_Intl_Rigs	1.000000	0.861696	0.879123	0.720542	0.721919
Asia_Pacific_Rigs	0.861696	1.000000	0.837156	0.741667	0.759682
Brent_Crude_Price	0.879123	0.837156	1.000000	0.730043	0.829238
Latin_America_Rigs	0.720542	0.741667	0.730043	1.000000	0.862802
US_Rigs	0.721919	0.759682	0.829238	0.862802	1.000000

Data Gathering

Finding a single dataset with enough relevant information on US rig was a challenge. We combined the following 3 datasets:

- Global Oil Supply/Demand
- Brent Oil Prices
- Baker Hughes International Rig Count

Data Preparation

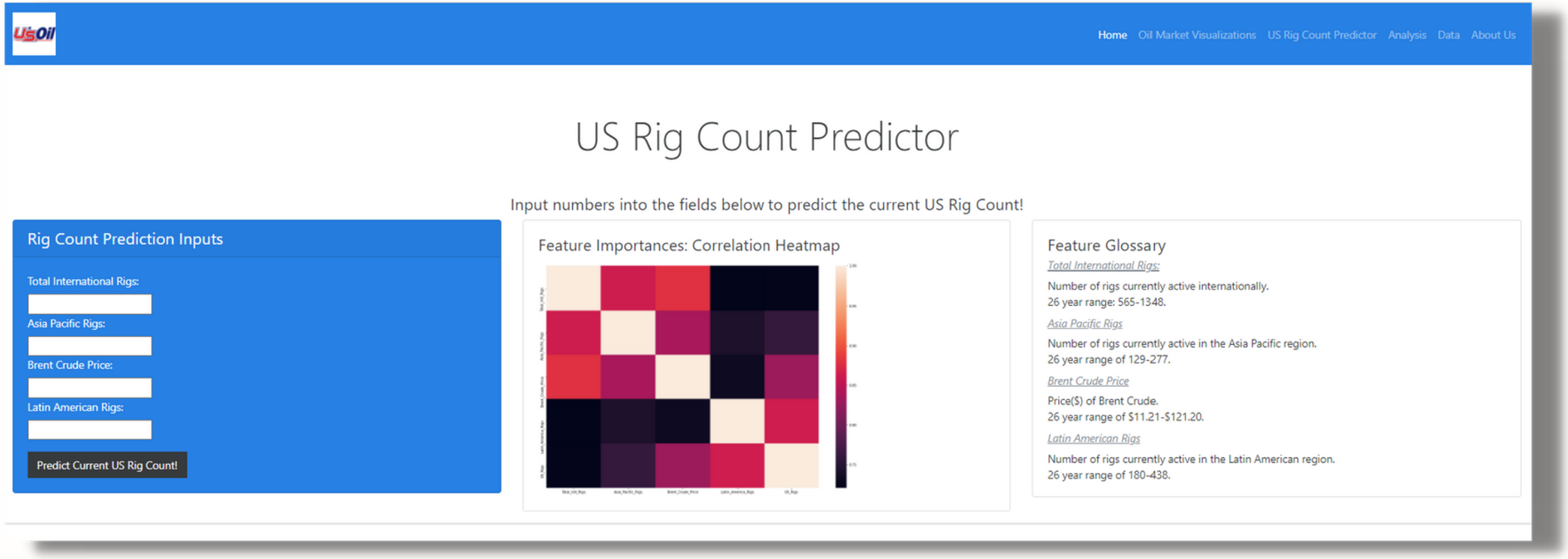
We quickly discovered the quarter column of our data was a string and not an integer.

This required using pandas get_dummies method to one hot encode the quarter data correctly.

Feature Exploration & Correlation

Due to our desire to create a more usable model for the end user, we wanted to eliminate some of the columns that appeared to be less highly correlated with our US rig count target.

US Rig Count Predictor



US Rig Count Predictor

Ultimately the best model proved to be the random forest model. Although it seemed that the model was overfitting the training data with a r-squared correlation value of .97 and a mean squared error of just 74, the random forest actually proved to be quite predictive for the testing data as well. Ultimately, with a r-squared correlation value of .89 and a mean squared error of 25225, the random forest model was the model we decided to choose to conduct our final analysis.

Rig Count Prediction Inputs

Takes various data points such as Asia Pacific rig count, Brent crude price and international rig count and outputs a prediction for US rig count.

Correlation Heatmap

A visual display of correlations between various features of the dataset.

Feature Glossary

Provides context for the rig count prediction inputs. The glossary also displays the lowest and highest values for the category.

US Rig Count Predictor

Ultimately the best model overall, proved to be the random forest model. Although it seemed that the model was overfitting the training data with a r-squared correlation value of .97 and a mean squared error of just 74, the random forest actually proved to be quite predictive for the testing data as well. Ultimately, with a r-squared correlation value of .89 and a mean squared error of 25225, the random forest model was the model we decided to choose to conduct our final analysis.

Rig Count Prediction Inputs

Takes various data points such as Asia Pacific rig count, Brent crude price and international rig count and outputs a prediction for US rig count.

Correlation Heatmap

A visual display of correlations between various features of the dataset.

Feature Glossary

Provides context for the rig count prediction inputs. The glossary also displays the lowest and highest values for the category.

US Rig Count Predictor

Ultimately the best model overall, proved to be the random forest model. Although it seemed that the model was overfitting the training data with a r-squared correlation value of .97 and a mean squared error of just 74, the random forest actually proved to be quite predictive for the testing data as well. Ultimately, with a r-squared correlation value of .89 and a mean squared error of 25225, the random forest model was the model we decided to choose to conduct our final analysis.

Rig Count Prediction Inputs

Takes various data points such as Asia Pacific rig count, Brent crude price and international rig count and outputs a prediction for US rig count.

Correlation Heatmap

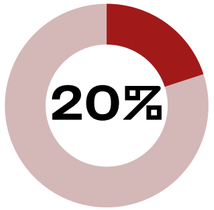
A visual display of correlations between various features of the dataset.

Feature Glossary

Provides context for the rig count prediction inputs. The glossary also displays the lowest and highest values for the category.

Visualizations

Our visualizations anchored on three elements – supply and demand, rig count and crude price analysis. With these elements at the focus we are able to overlay global and regional factors to glean greater insights.



The percentage of rigs located in the US.

Supply and Demand Analysis

Our demand and supply dashboard analyzes these metrics from a global perspective.

Users can explore:

- OPEC supply by year
- Brent crude price against global supply and demand

Brent Crude Price Analysis

Uncovers the relationship between crude price and regional rig count. The flagship visualization can be filtered by year, quarter and various averages.

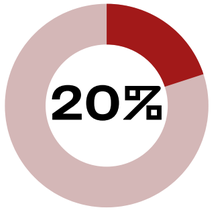
Regional Map with Rig Count

Quickly see who are the major players with regard to global average rig count.

The US Rig Count heatmap shows which year and quarter the US had the most active rigs.

Visualizations

Our visualizations anchored on three elements – supply and demand, rig count and crude price analysis. With these elements at the focus we are able to overlay global and regional factors to glean greater insights.



The percentage of rigs located in the US.

Supply and Demand Analysis

Our Demand and Supply dashboard analyzes these metrics from a global perspective.

Users can explore:

- OPEC supply by year
- Brent crude price against global supply and demand

Brent Crude Price Analysis

Uncovers the relationship between crude price and regional rig count. The flagship visualization can be filtered by year, quarter and various averages.

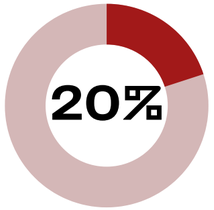
Regional Map with Rig Count

Quickly see who are the major players with regard to global average rig count.

The US Rig Count heatmap shows which year and quarter the US had the most active rigs.

Visualizations

Our visualizations anchored on three elements – supply and demand, rig count and crude price analysis. With these elements at the focus we are able to overlay global and regional factors to glean greater insights.



The percentage of rigs located in the US.

Supply and Demand Analysis

Our Demand and Supply dashboard analyzes these metrics from a global perspective.

Users can explore:

- OPEC supply by year
- Brent crude price against global supply and demand

Brent Crude Price Analysis

Uncovers the relationship between crude price and regional rig count. The flagship visualization can be filtered by year, quarter and various averages.

Regional Map with Rig Count

Quickly see who are the major players with regard to global average rig count.

The US Rig Count heatmap shows which year and quarter the US had the most active rigs.

Conclusion

The US oil and gas industry is very dynamic

The lion share of oil and gas production comes from shale formations that require a large input from fixed assets. The burden of these fixed assets is they force the industry to rapidly adjust every time there is a change in crude oil prices. Therefore , the leading indicator for oil field service companies is the number of drilling rigs.

In terms of production gauges rig counts are paramount and a necessity for adequate future planning and development.