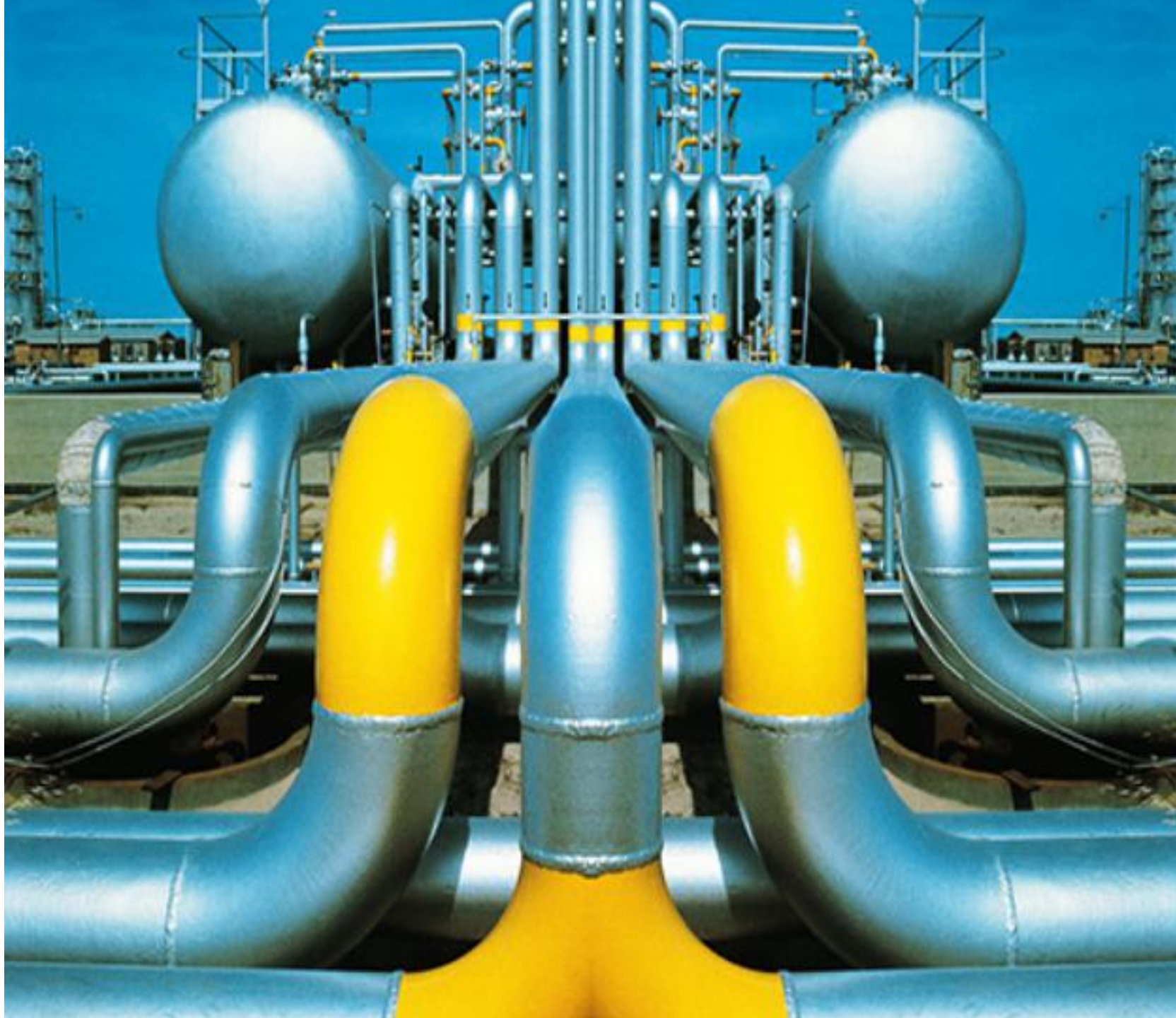
US Natural Gas Prices - Future Performances

by

Angus Ogubuike



Introduction

- Natural gas consists primarily of methane
- It is created using one of two mechanisms: biogenic and thermogenic.
- It is a major source of power generation, especially for heating and cooling systems in the U.S.
- In 2015, Natural gas was the nation's second-largest source of energy in the United States, after petroleum.
- The natural gas industry includes exploration for, production, processing, transportation, storage, and marketing of natural gas and natural gas liquids. The exploration for and production of natural gas and petroleum form a single industry, and many wells produce both oil and gas.
- The most commonly quoted producer price for natural gas is the Louisiana-based Henry Hub price, which is futures-traded on NYMEX.

Problem Statement

“Determine the major factors that influence Natural Gas Prices. Use these factors to predict natural gas prices future performance”

Data for the study

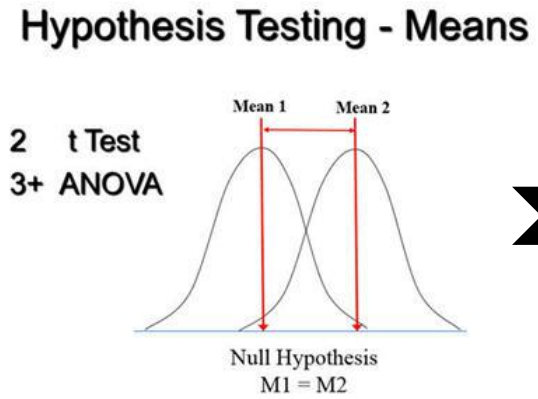
Data Set:

Natural Gas Data Set from the United States Energy information Administration:

https://www.eia.gov/dnav/ng/ng_pri_fut_s1_d.htm

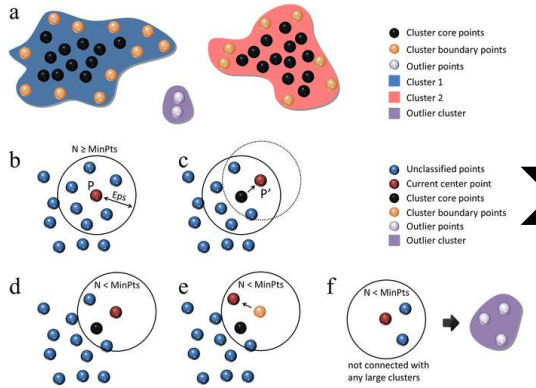
INDEPENDENT VARIABLES	DERIVED VARIABLES	DEPENDENT VARIABLE
- Average Coal Price	- Weather Status	- Gas Price
- Oil Price	- Gas Price Status	
- Gross Gas Production		
- Total Gas Consumption		

Approach to solving the problem



Hypothesis Testing

Data Wrangling



Anomaly detection

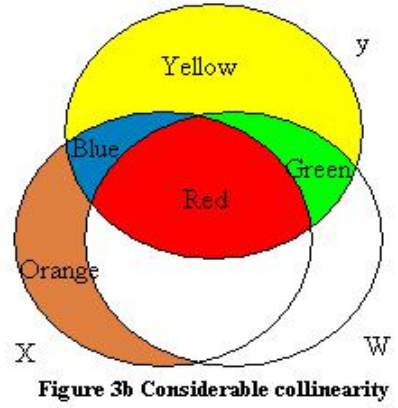
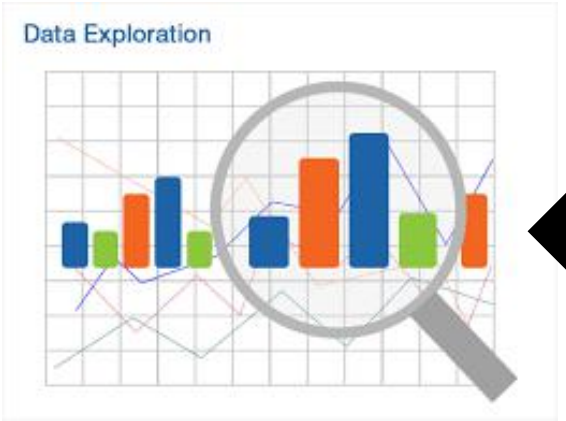


Figure 3b Considerable collinearity

Quantifying Multicollinearity

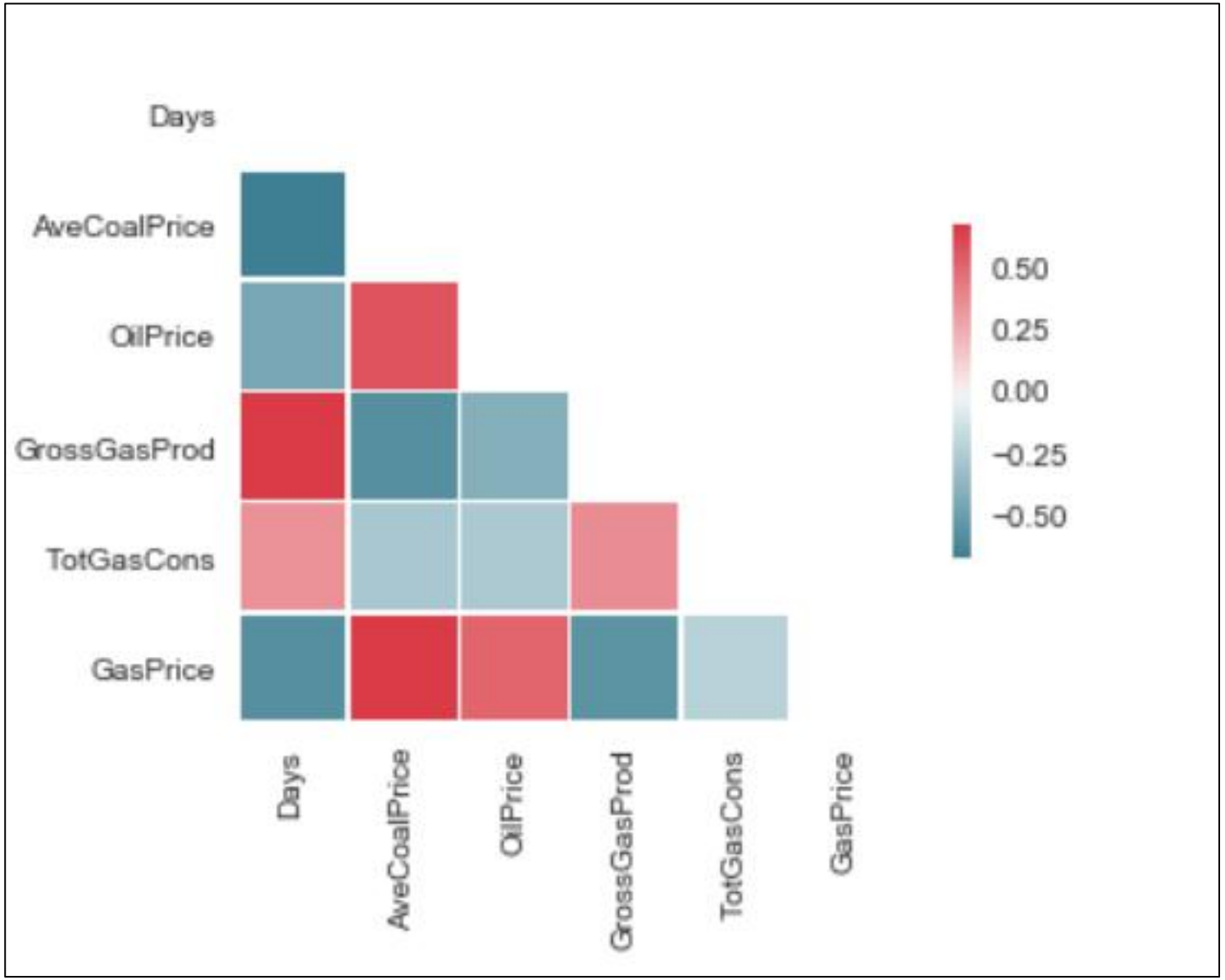
Data Analysis



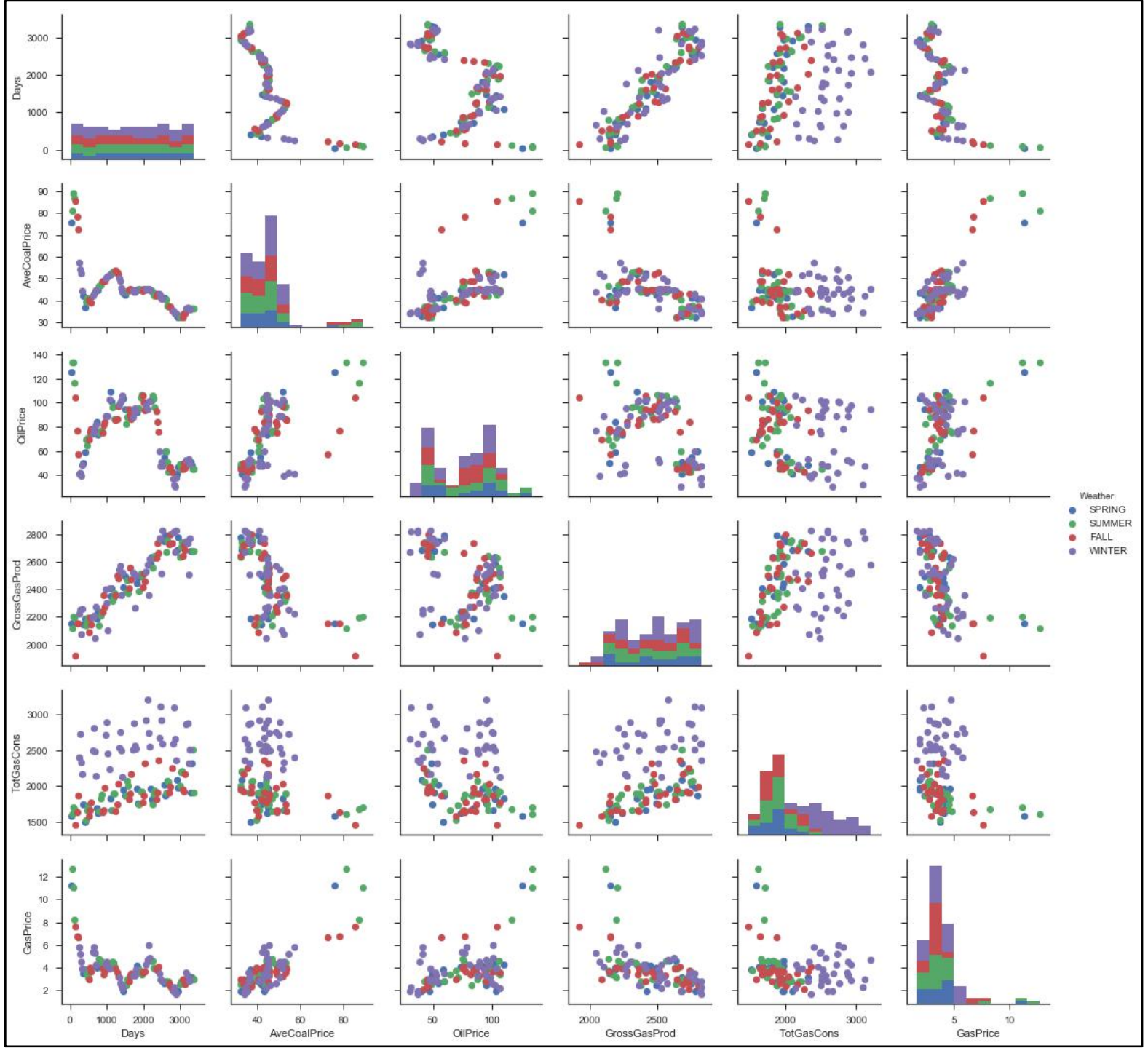
Data Exploration

Accuracy Boosting/ Machine Learning

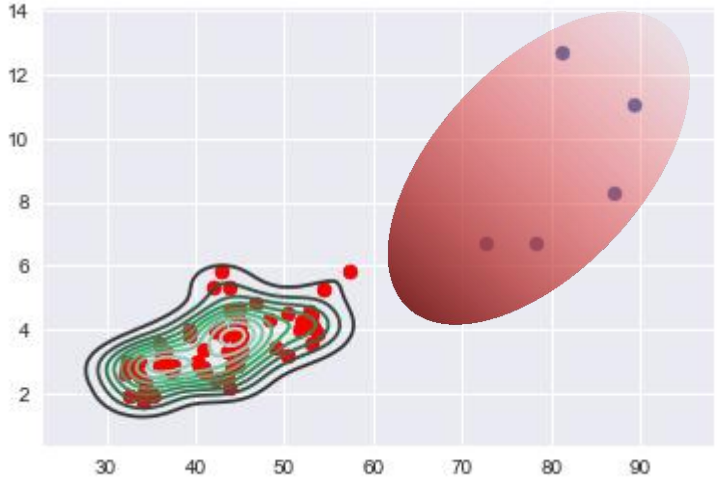
Correlation Matrix



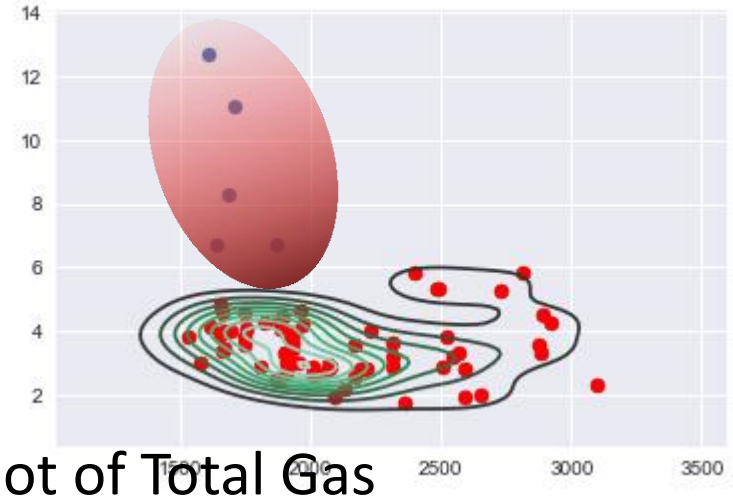
Scatter Matrix of All Variables Included in the Study Data-Set



Anomaly Detection



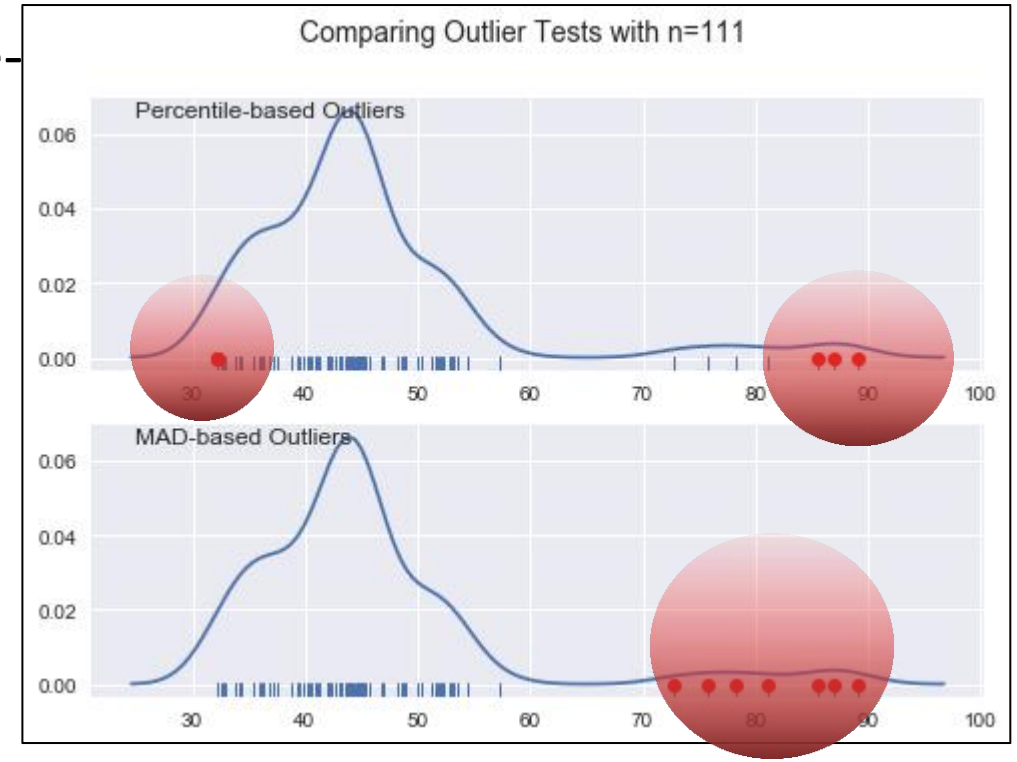
kde Plot of Average Coal Price vs Gas Price



kde Plot of Total Gas Consumption vs Gas Price

Median-absolute-deviation (MAD) based outlier detection

Percentile based outlier detection



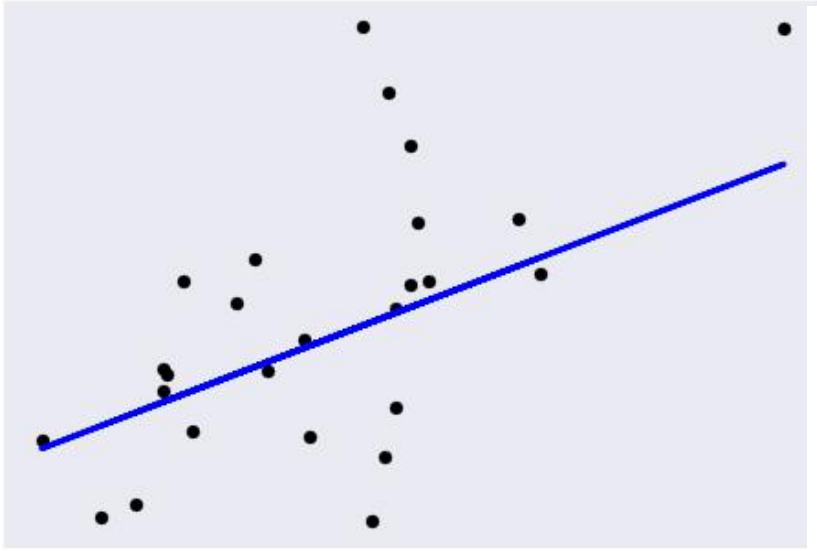
- **Anomalies were removed from subsequent analysis**

Outliers

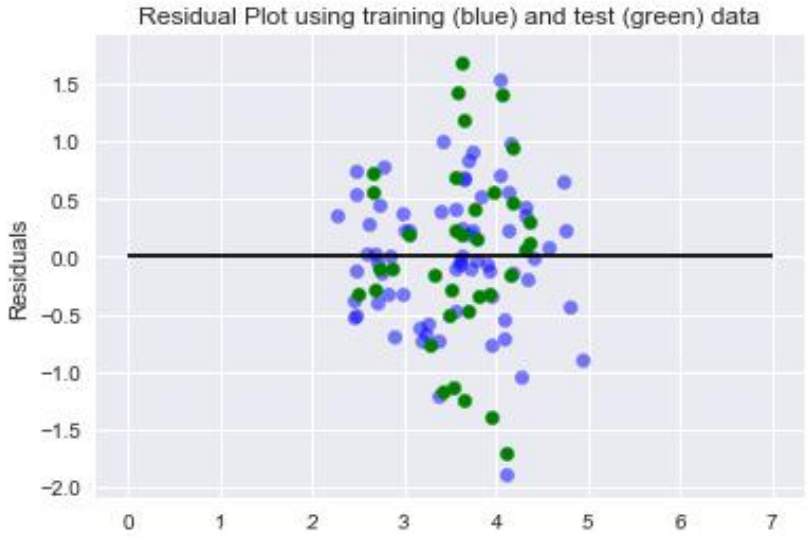
Linear Regression

PROCEDURE

- Regress Gas Price vs Each Independent Variable
- Regress Gas Price vs All Independent Variables



Gas Price vs Average Coal Price



Residual Plot for all variables regression

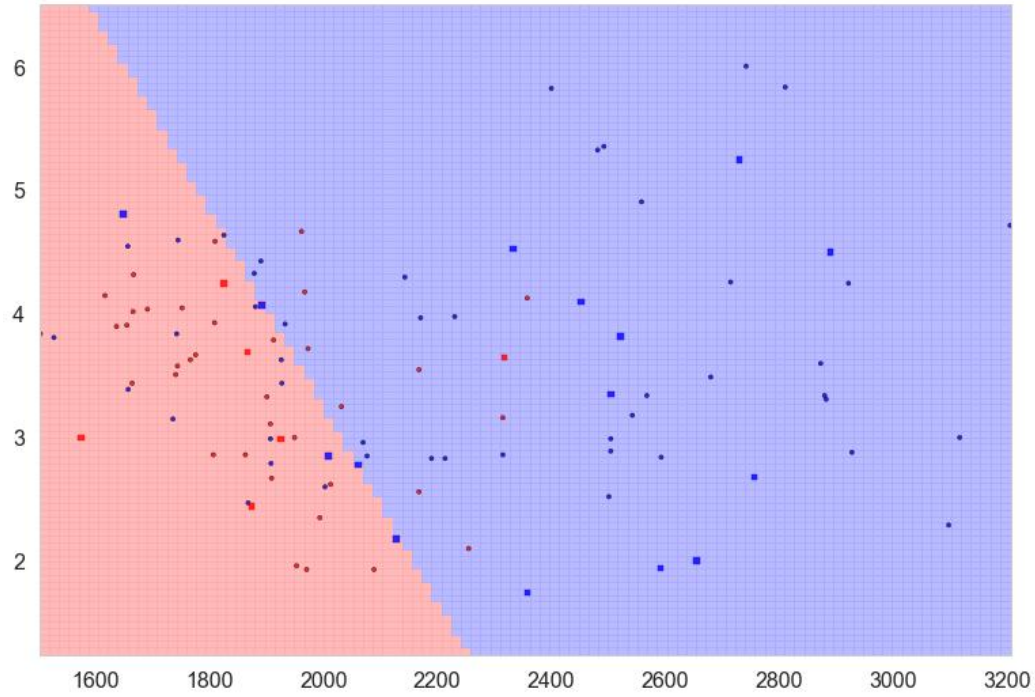
Gas Price vs	Score	MSE	Cross Val Score
Average Coal Price	0.33	0.75	0.60
Oil Price	0.10	0.62	
Gas Production	0.19	0.75	
Gas Consumption	-0.04	0.58	
All Variables	0.47	0.44	

Logistic Regression

PROCEDURE

- Use Gas Price Status as dependent variable
- HIGH Gas Price: >$3.00

  LOW Gas Price: <$3.00



Classification of Gas Price wrt average Coal Price

Gas Price vs	Accuracy Score (Training)	Accuracy Score (Test)
Average Coal Price	0.87	0.81
Oil Price	0.88	0.81
Gas Production	0.82	0.76
Gas Consumption	0.84	0.86
Days	0.88	0.81

Model Tuning

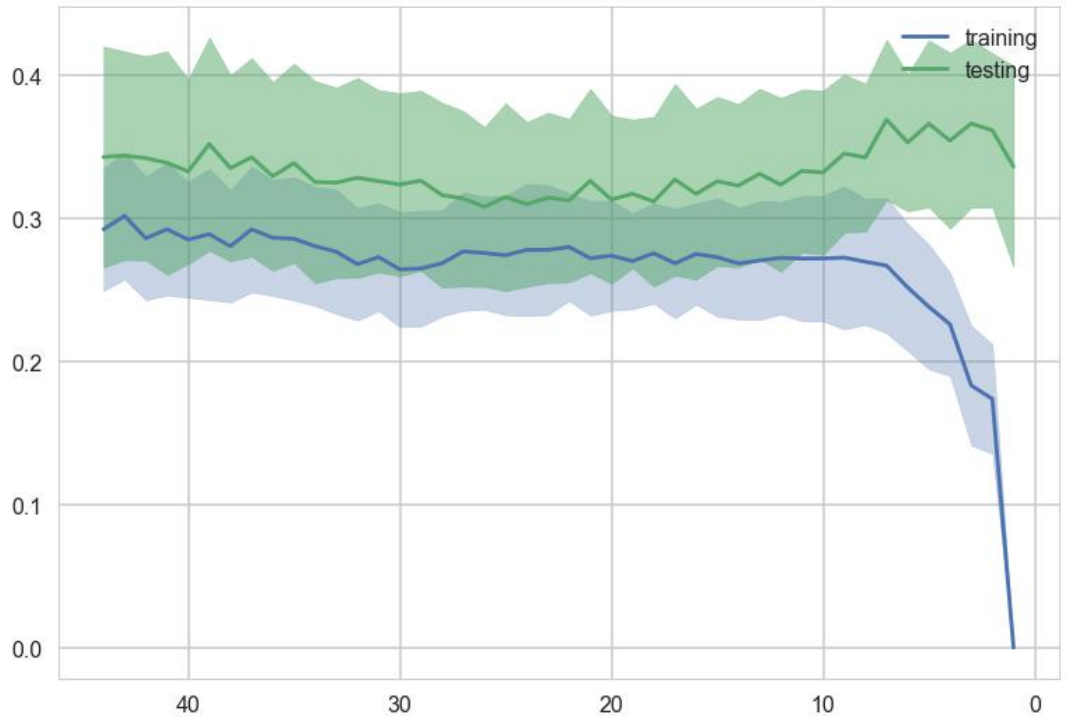
PROCEDURE

- Cs = [0.001, 0.1, 1, 10, 100]
  - max_score = 0.76
  - best_C = 0.001

- Accuracy score = 77%
- This suggests that the risk that a portion of the sample is misclassified is 23%

Gas Price vs	Logistic Regression	Grid CV Search
All variables	0.77	0.77 (CV=5)

Principal Components Analysis



Error against complexity (k), and cross-validation

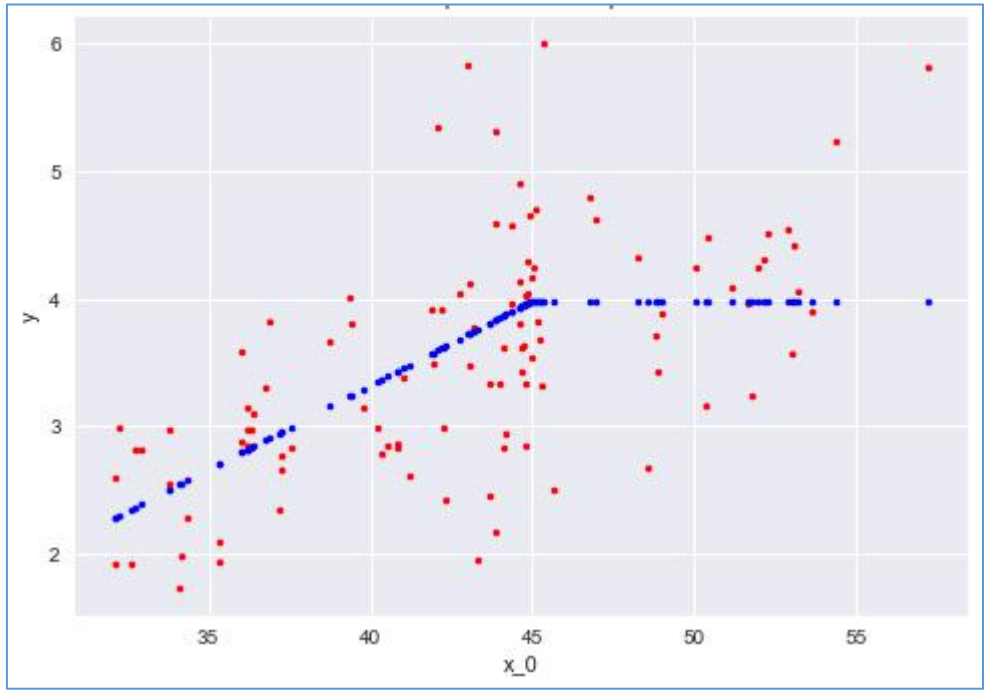
- Sum of explained variance ratio of PCA = 0.99925003031
  - pc1 = 80.9% of variance
  - pc2 = 18.8% of variance
  - pc3 =0.2% of varance
- Only two components are enough to explain the variation in the model.

Prediction

- Using Ridge Regression: Coef of Ridge =

['AveCoalPrice', 'OilPrice', 'GrossGasProd', 'TotGasCons']:

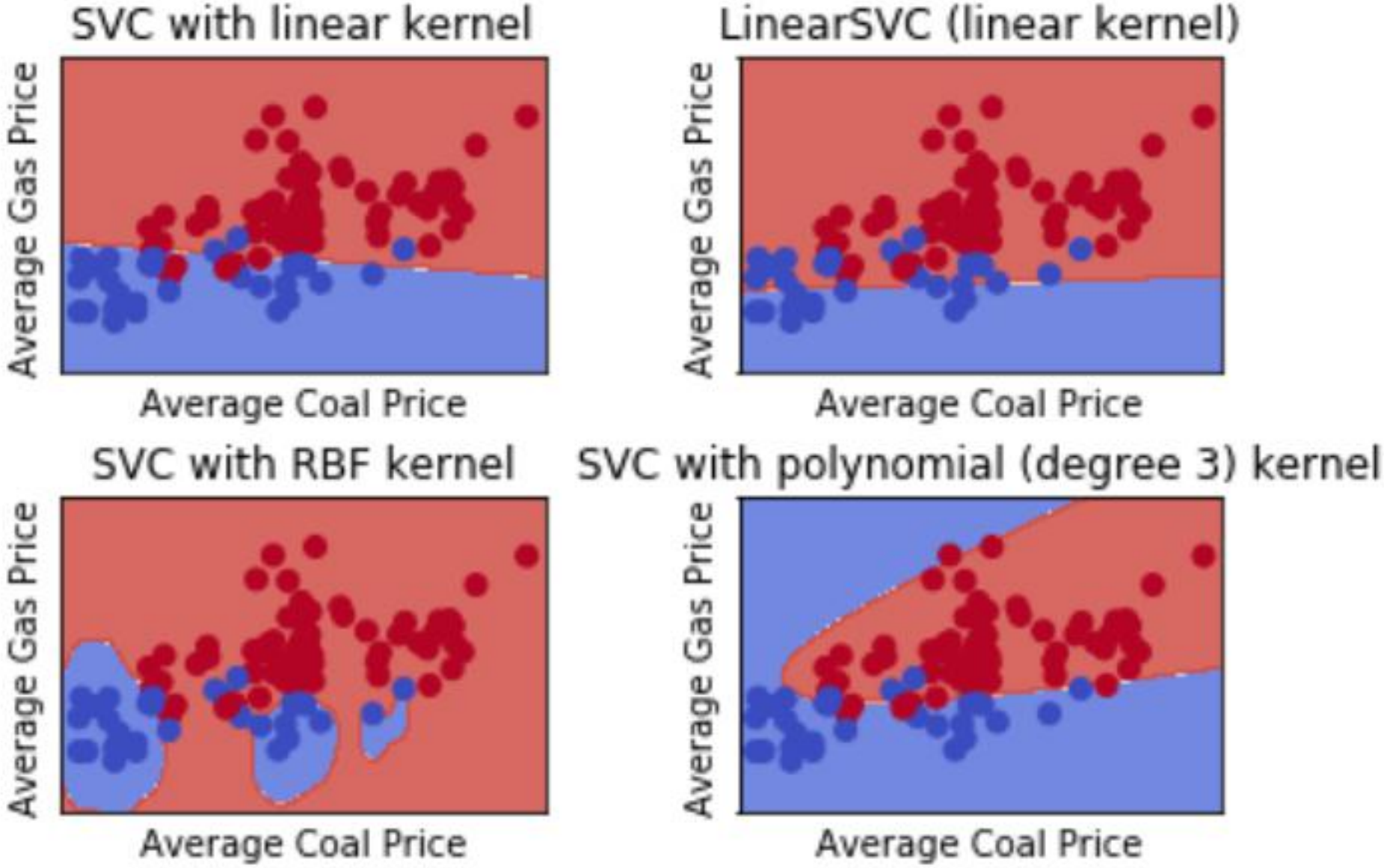
array([0.05515652, 0.00292577, -0.00167807, 0.00039166])



Multivariate Adaptive Spline

GasPrice vs AveCoalPrice

Supervised Vector Machine / Classification



Boosting accurracy of the machine learning model

Accuracy Score

Linear Regression	Decision Tree	Extremely Randomnized Tree	Random Forest	Extra Tree
0.47	0.68	0.69	0.78	0.73

Accuracy Score (Ada Boost = 0.70)

GradientBoostingClassifier (Score) - 92%

GradientBoostingRegression - MSE: 0.12

Summary and Recommendation

- Coal price remains the key indicator of increase in gas price. Therefore, it is a variable that investors should watch out for.
- Whereas Oil price was seen as the major determinant of natural gas prices, our study shows that its impact in the past decade is marginal.
- Gas demand is a key factor that determines US gas price. Gas demand is primarily driven by seasonal variation or occurence of natural disaster.

Recommendation

- Seasonal weather, including tropical storms and hurricanes, can have a great influence on natural gas production.
  - The seasonal effect of weather on gas price cannot be effectively analyse using logistic regressison and classification.
  - Time series Analysis is therefore recommended in order to explore the effect of seasonality.
- There are other factors that ontribute to gas price variation that have not been captured in this project.
  - Such variables as economic growth, natural gas storage and availability of gas pipelines have effect on gas prices.
  - It is recommended that future studies include these variables.

End