Northwestern

# MSiA – [Client] Practicum

Sophie Du
James Fan
Michael Fedell
Jonathan Lewyckyj
Nora Xu

An Optimized Pricing Tool for Commercial Contracts

# Project Introduction

- **Context**:
  [Client] operates in a saturated and competitive environment with tight margins.
- **Objective**:
  Build more advanced analytical capabilities into the existing Bid Pricing workflow.
- **Resources**:
  Vast amount of time-series data on index prices, rack posters, and bid outcomes.

# Overview

## In Summary

Our team worked with the Planning and Optimization team to understand the current business process and to build out robust analytical tools in [Client]'s cloud environment  that will scale into the future.

1. Survey of Market
2. Modeling Capabilities
3. Revamped Workflow

# Survey of Market

1. Branded oil prices comparisons & similar competitor selections for each material type
2. Unbranded oil product recommendations referring to similar competitors' prices

# Branded vs Unbranded

**Branded**

Displaying 8 rows

| COMPETITOR_ID | count |
|---|---|
| [redacted] | 3759 |
| | 3759 |
| | 3759 |
| | 3759 |
| | 3759 |
| | 3754 |
| | 3750 |
| | 3270 |

**Unbranded**

Displaying 16 rows

| COMPETITOR_ID | count |
|---|---|
| [redacted] | 3759 |
| | 3748 |
| | 3720 |
| | 3510 |
| | 2601 |
| | 2405 |
| | 2065 |
| | 1952 |
| | 1360 |
| | 1253 |
| | 1029 |
| | 643 |
| | 536 |
| | 531 |
| | 247 |
| | 89 |

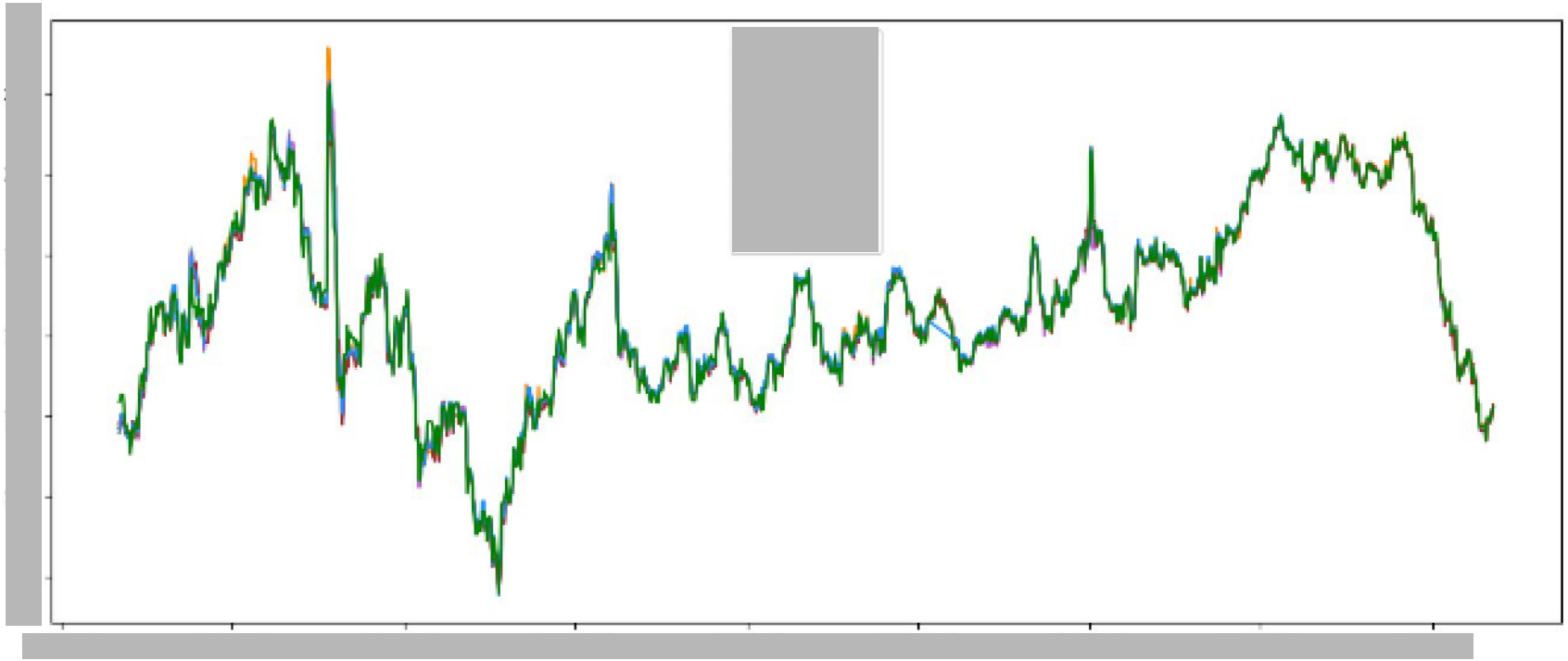
***Who reported both branded AND unbranded?***

- [redacted]

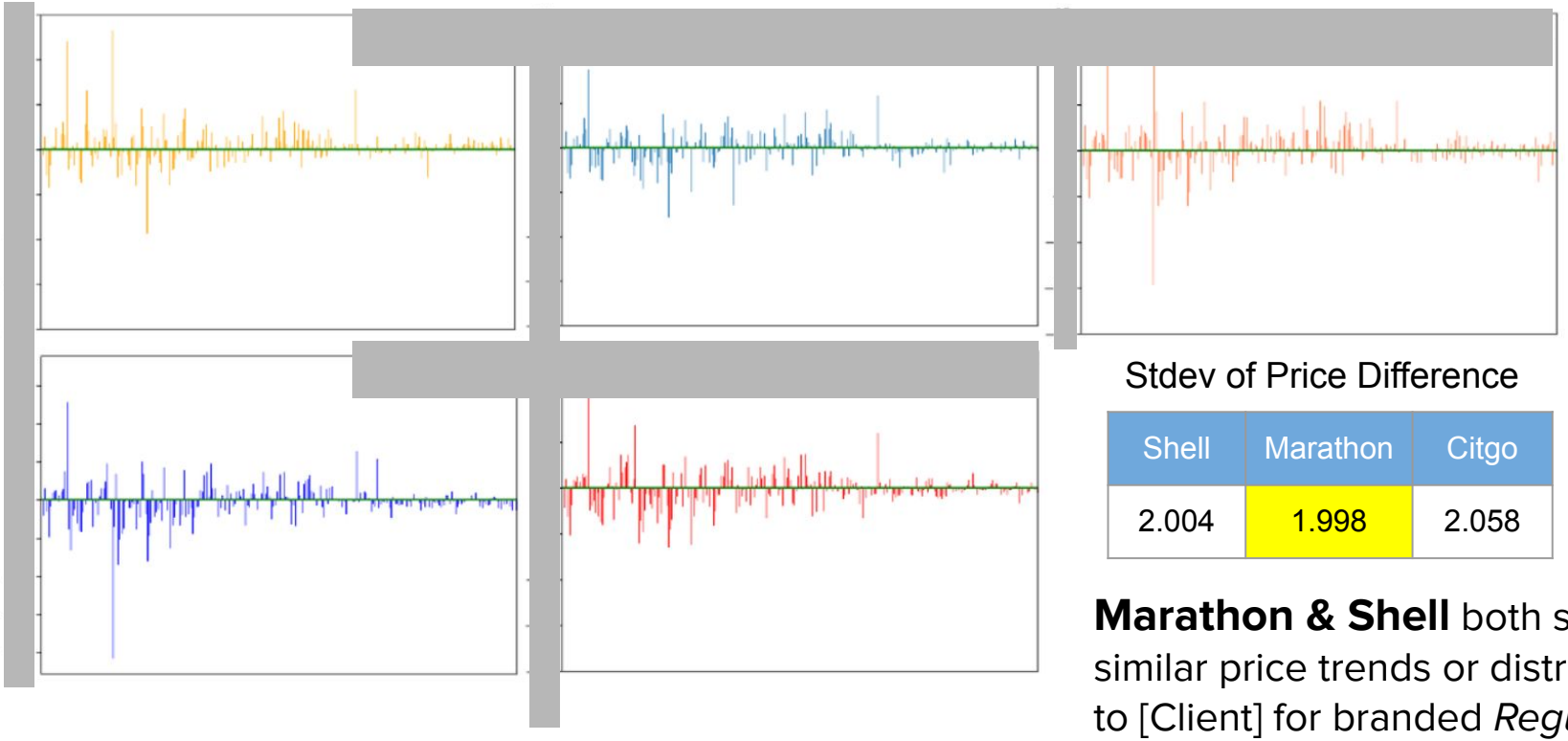
# Branded vs Unbranded

- Mainly focus on competitors who have ***both*** branded and unbranded oil prices reported
- Find competitors who have *similar branded* oil price trends or distributions to [Client] for each of the 3 materials
- Provide [Client] with recommendations for unbranded oil prices based on the *competitors' unbranded* oil prices for each of the 3 materials

[redacted] (87 R9+ Regular E10)



[redacted] (87 R9 + Regular E10)

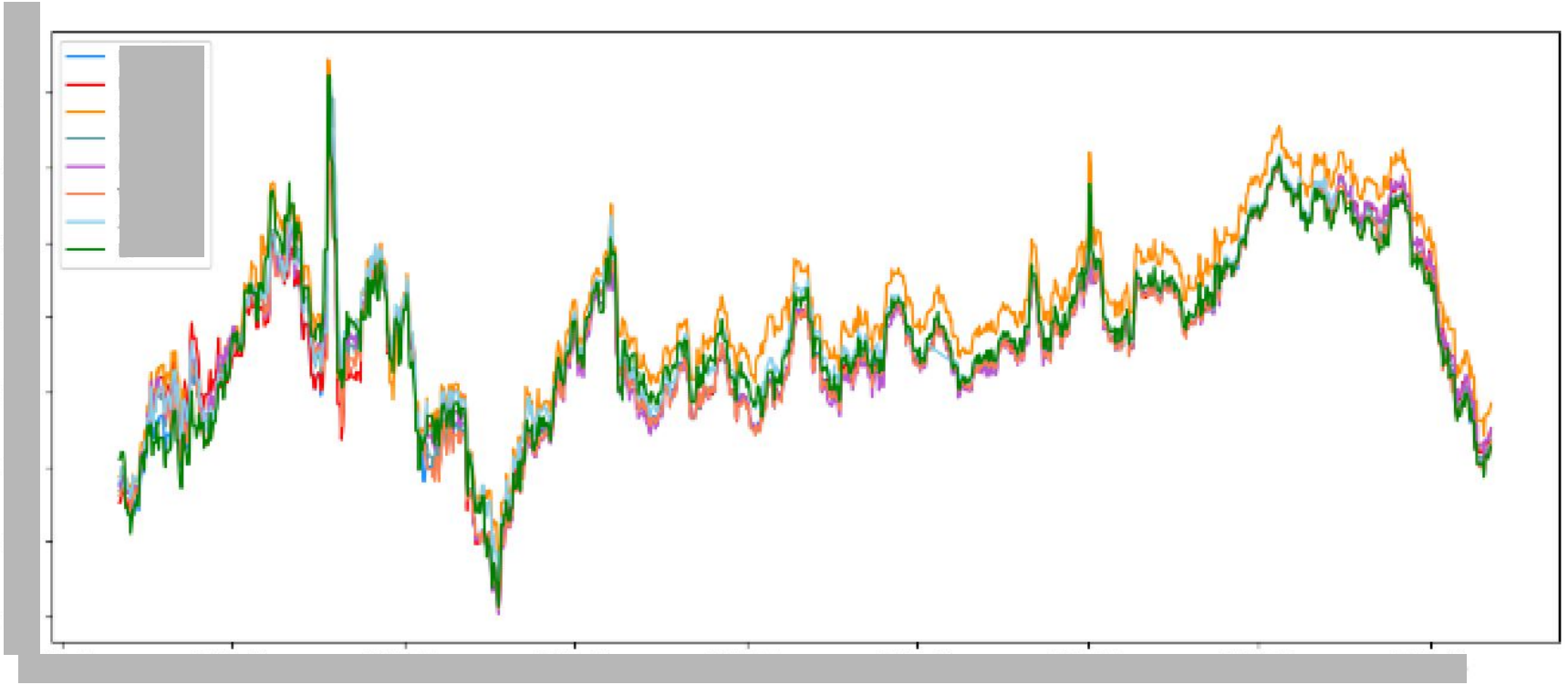


Stdev of Price Difference

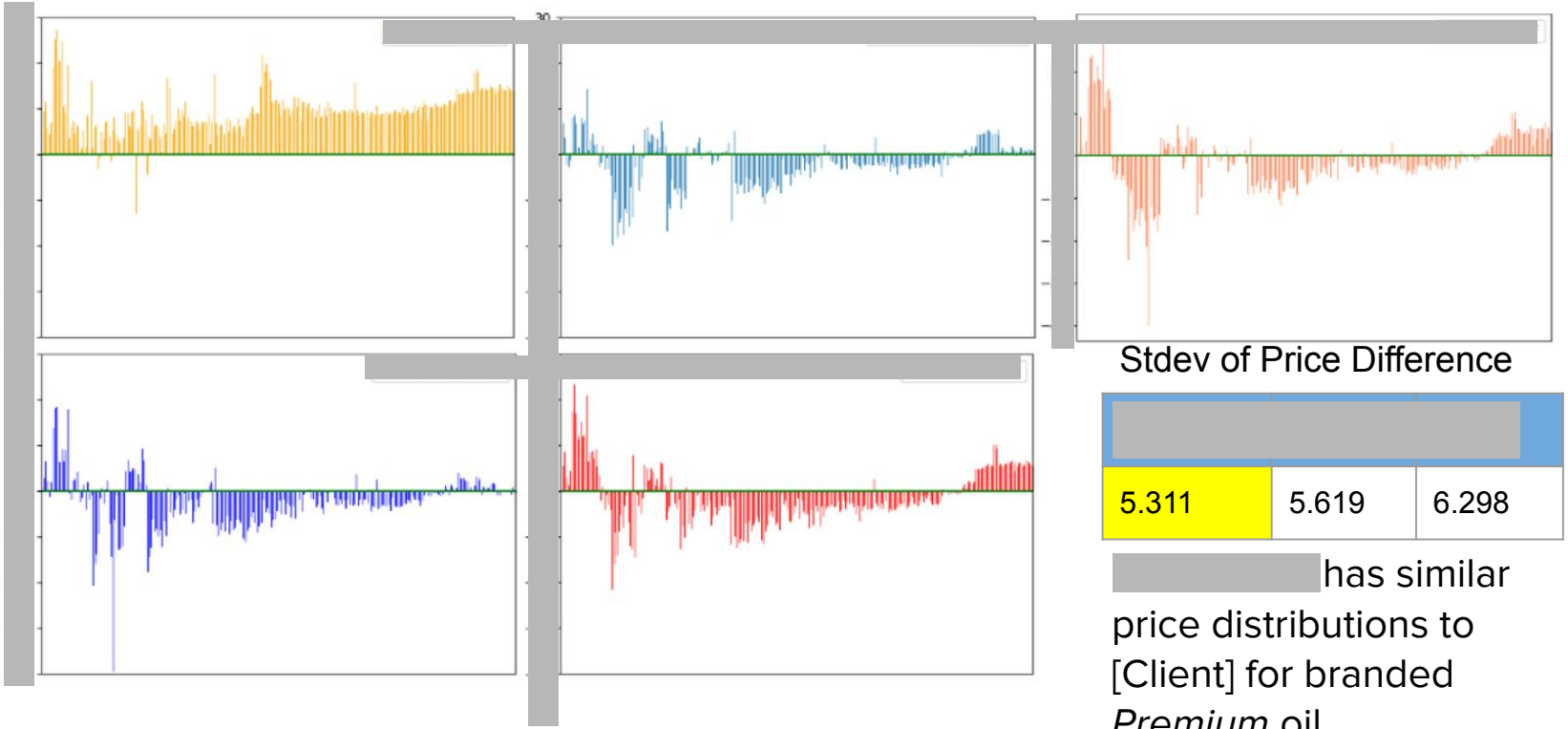
| Shell | Marathon | Citgo |
|---|---|---|
| 2.004 | 1.998 | 2.058 |

**Marathon & Shell** both show similar price trends or distributions to [Client] for branded *Regular* oil

(93 R9+ Premium E10)



[redacted] (93 R9 + Premium E10)

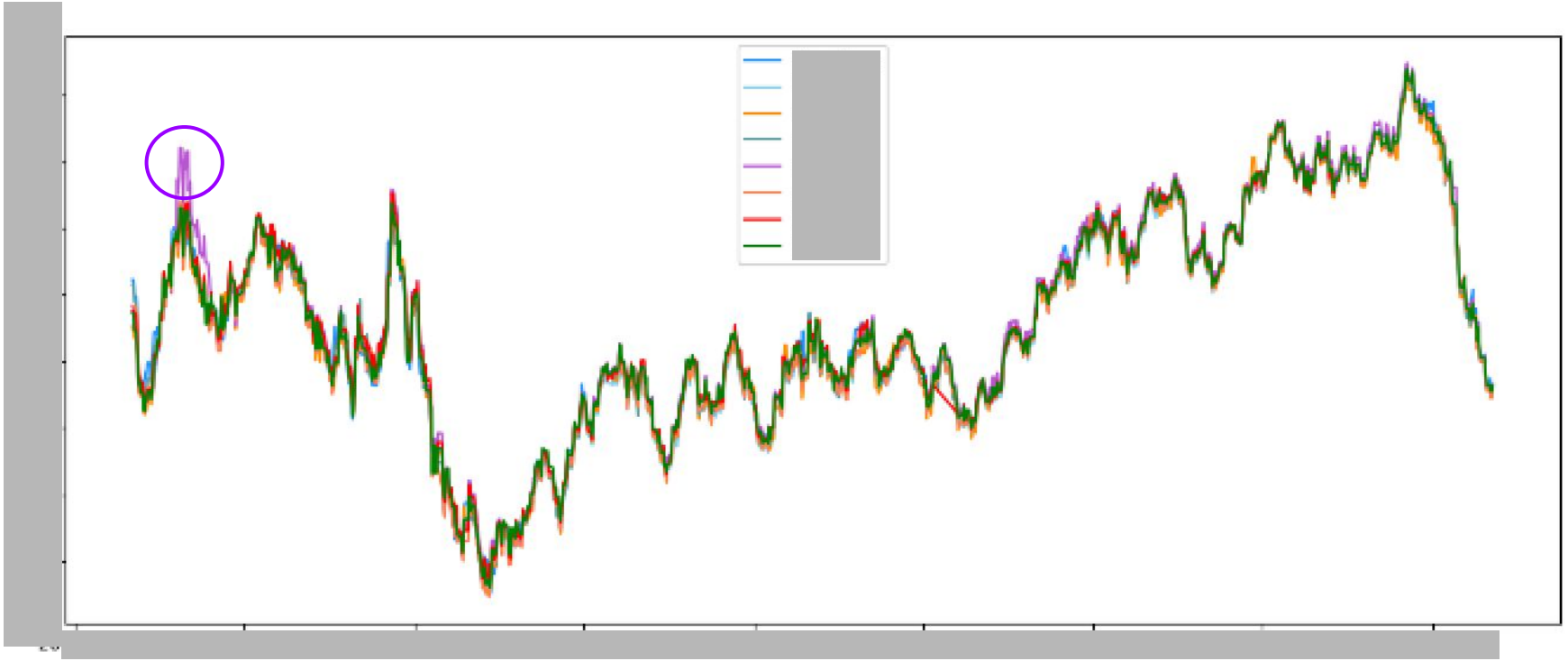


Stdev of Price Difference

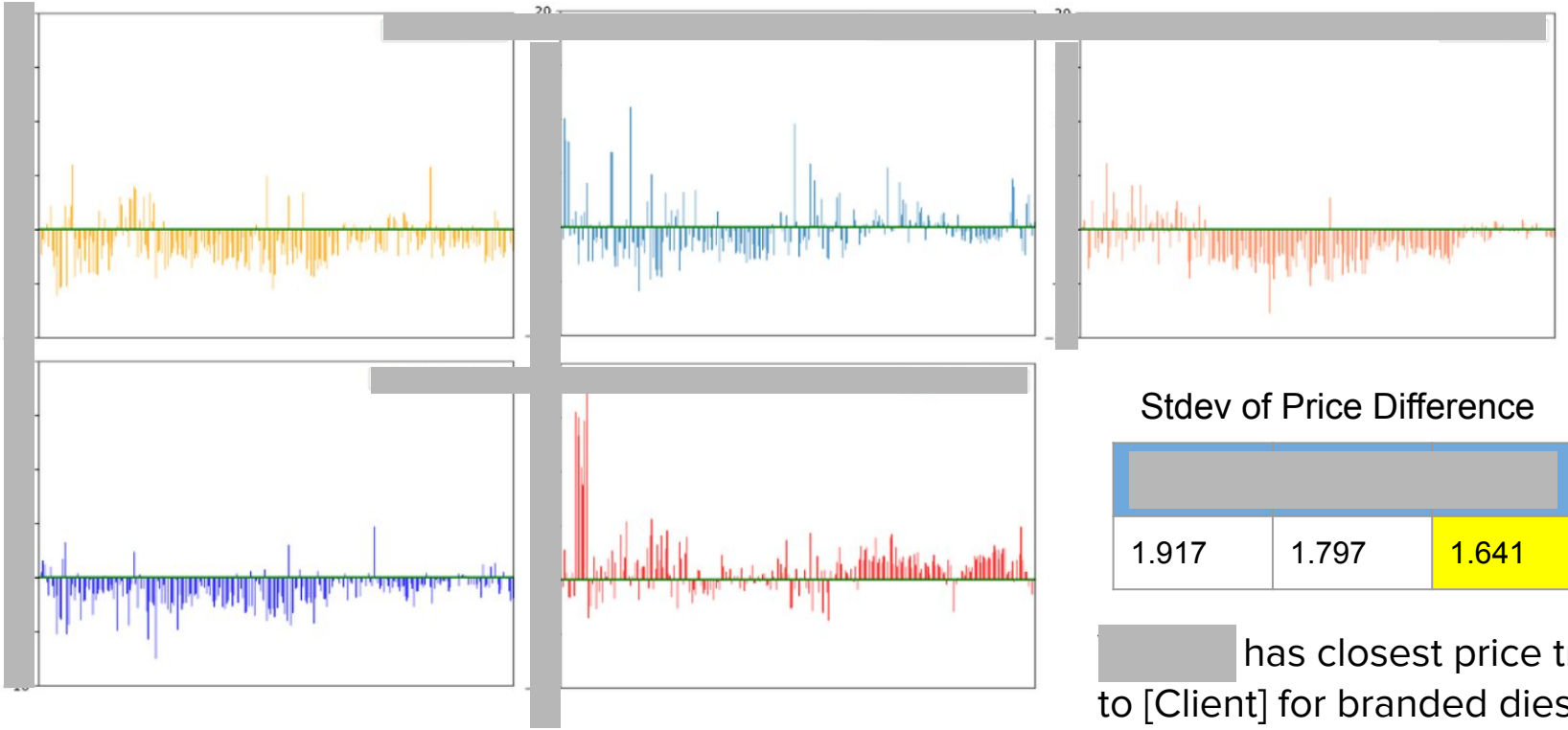
| [redacted] | | |
|---|---|---|
| 5.311 | 5.619 | 6.298 |

[redacted] has similar price distributions to [Client] for branded *Premium* oil

[redacted] (ULSD #2 Diesel)



[redacted] (ULSD #2 Diesel)

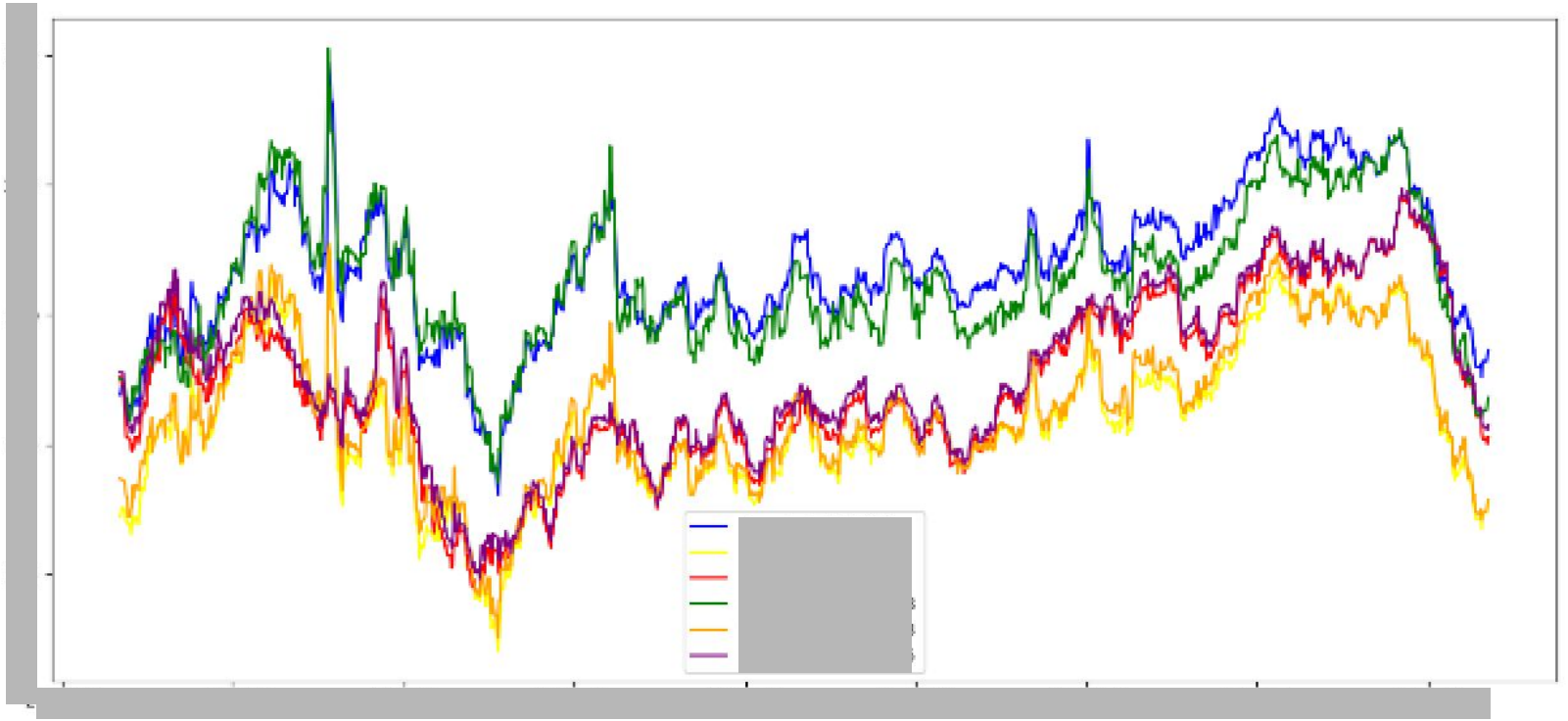


Stdev of Price Difference

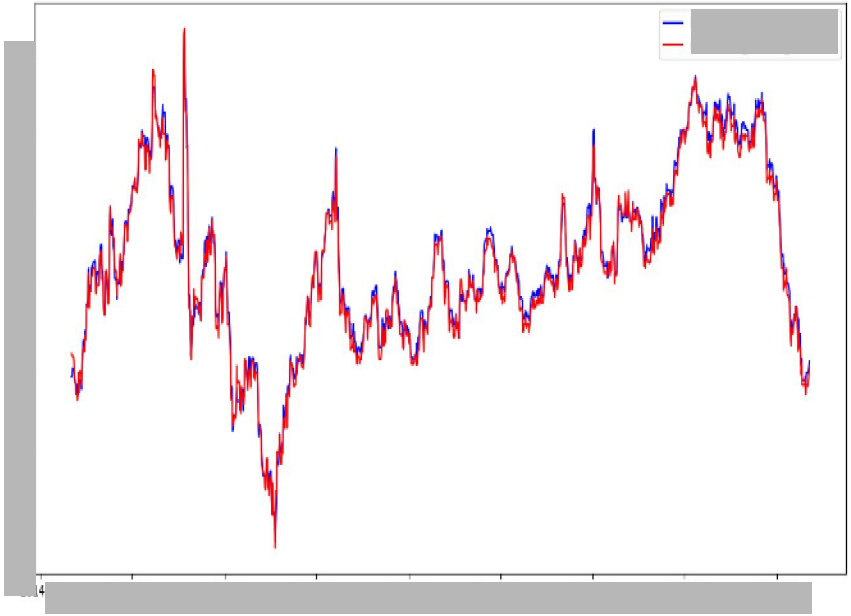
| [redacted] | | |
|---|---|---|
| 1.917 | 1.797 | 1.641 |

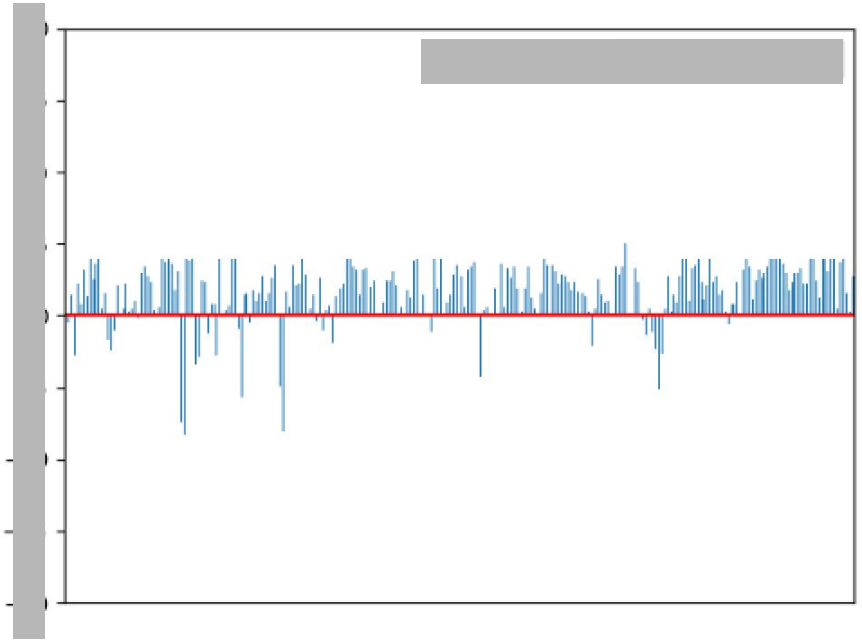
[redacted] has closest price trends to [Client] for branded diesel oil

# Industry - Branded VS Unbranded

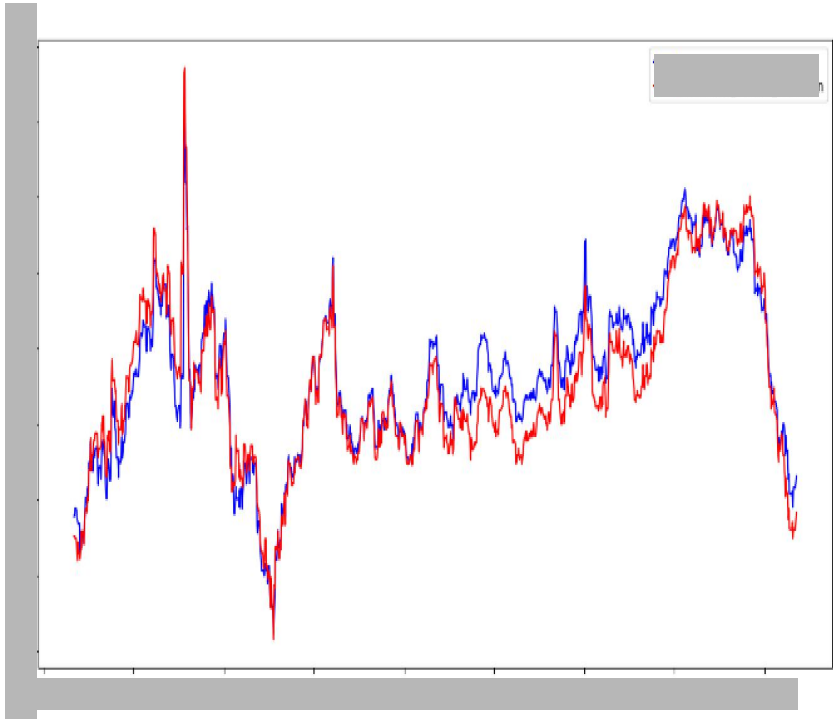


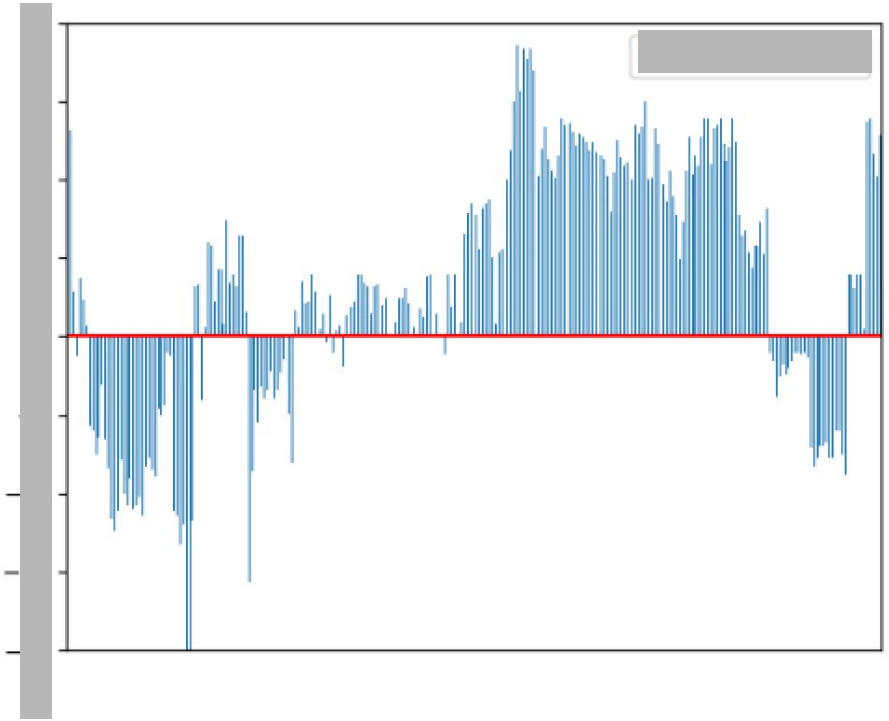
[redacted] (87 R9 + Regular E10)



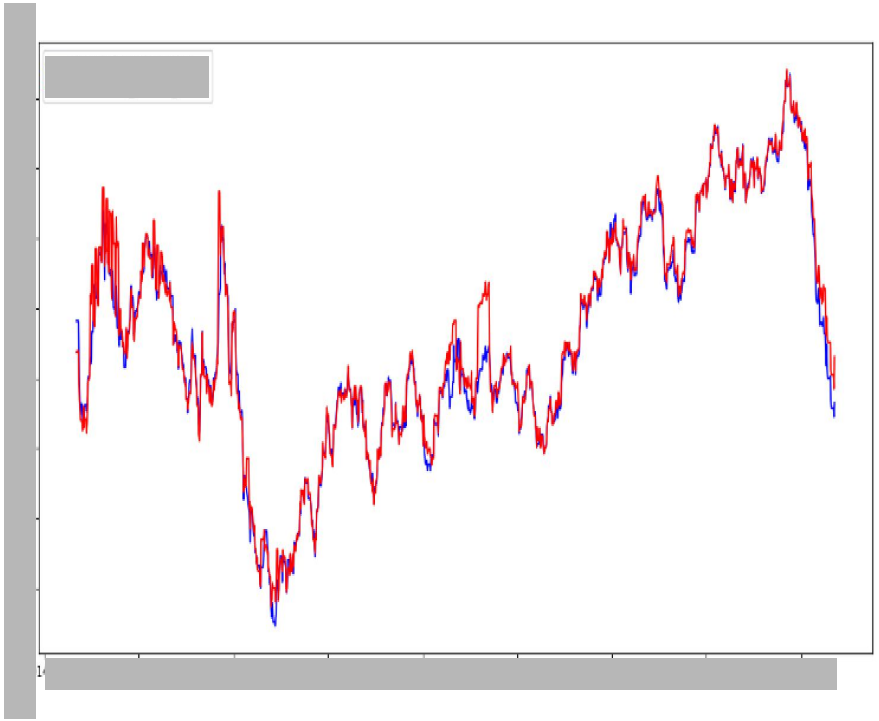


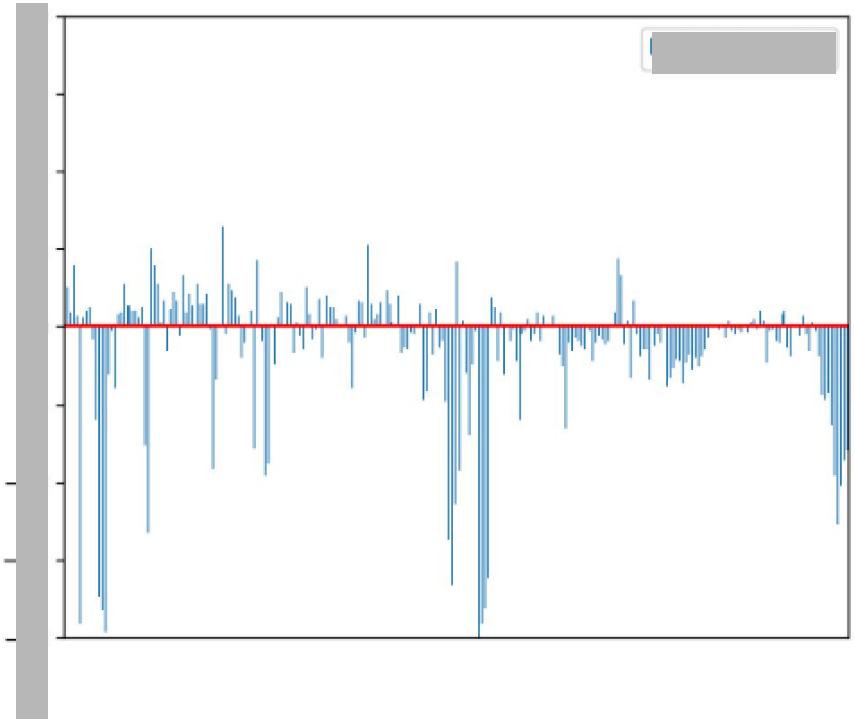
(93 R9 + Premium E10)





(ULSD #2 Diesel)





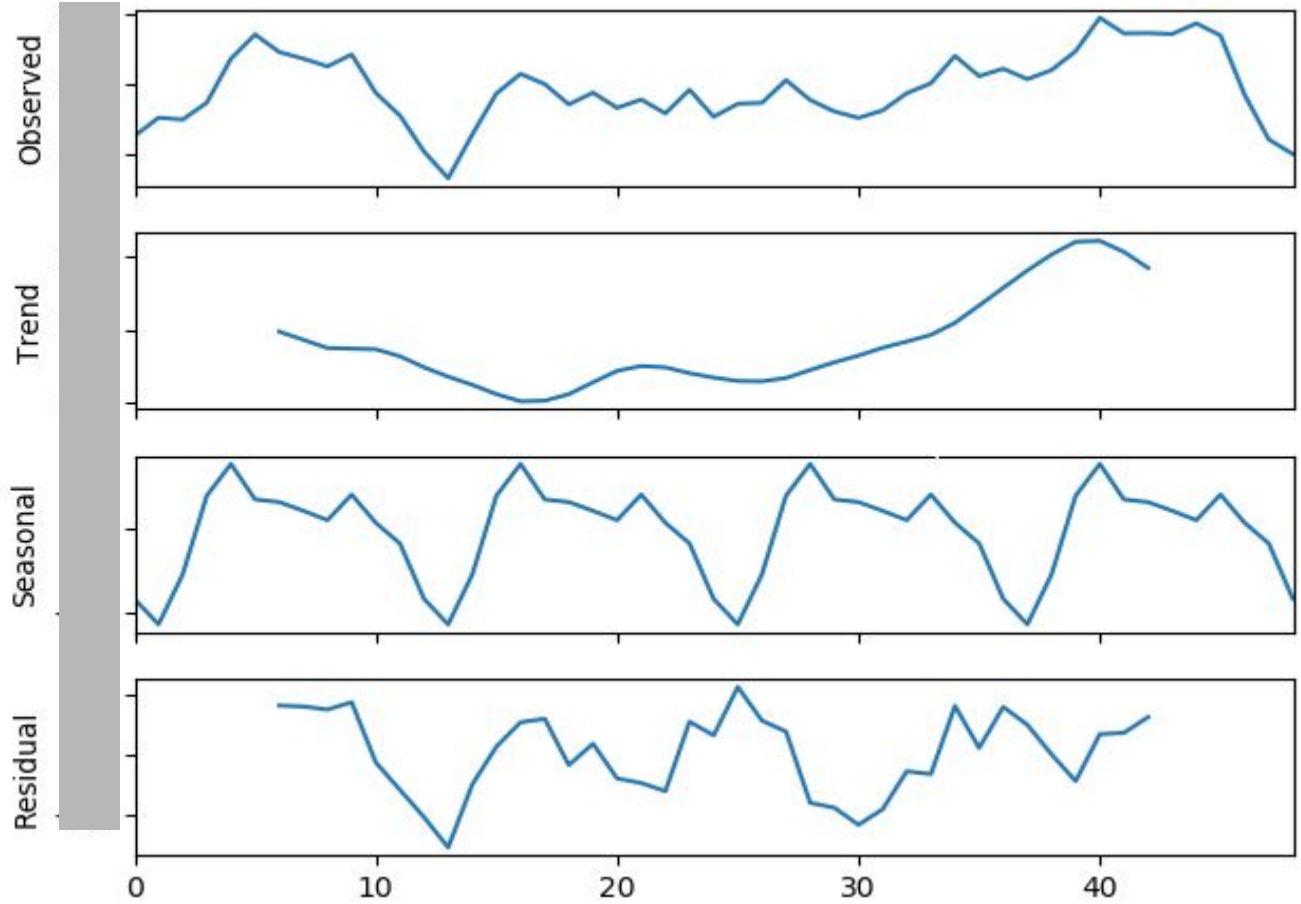
# Recommendations

- [Client] should evaluate how to sell more unbranded premium and diesel oil during the period with high oil price based on l
- Merchandise unbranded regular and diesel oil for more pricing benefits with regards to close prices for both rack types and additional packages and advertising costs on branded gas
- Consider clients' willingness to spend more on branded gasoline as an advantage of negotiating contracts

# Predictive Modeling

1. Time Series Decomposition and Forecasting
2. Bid Outcome Predictions

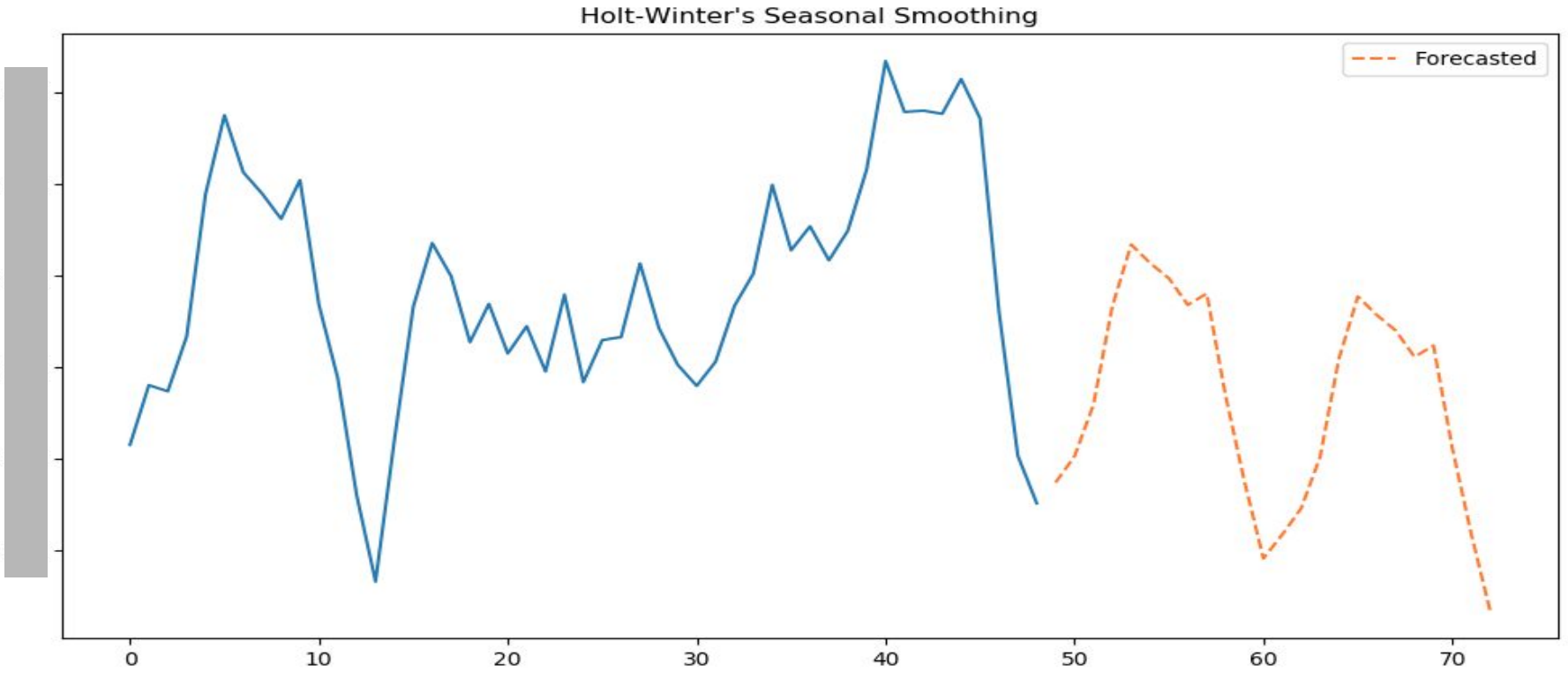
# Model: Seasonal Decomposition Model - Monthly Average Prices, OPIS Low




Seasonality Index:

| | |
|---|---|
| January | -16.75 |
| February | -22.87 |
| March | -10.85 |
| April | 8.06 |
| May | 15.69 |
| June | 7.15 |
| July | 6.51 |
| August | 4.42 |
| September | 2.17 |
| October | 8.34 |
| November | 1.53 |
| December | -3.40 |

# Model: Holt-Winters Forecasting Model - Monthly Average Prices, OPIS Low

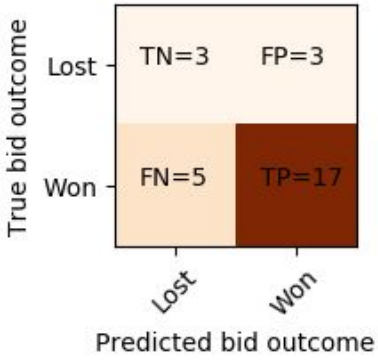


# Bid Outcomes Prediction

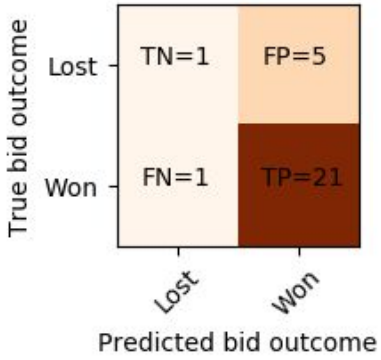
- Models
  - Logistic regression
  - Decision tree classifier
  - Random forest classifier
  - Gradient boosting classifier
- Variable used
  - Won_or_Lost
  - Customer type
  - SAP material
  - Existing Gallon Per Month
  - Existing KBD
  - Index summary
  - Gross/Net
  - Adder merged
  - Floor expected margin

# Model Comparison

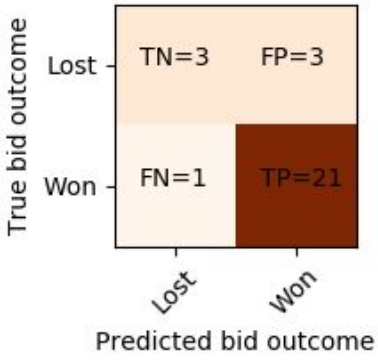
Logistic Classification: 0.714285714286



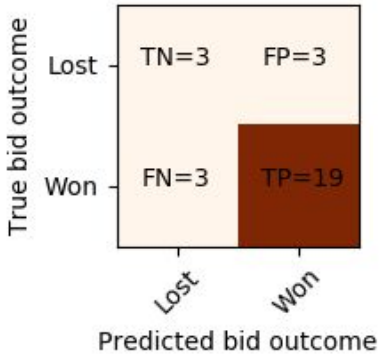

Decision Tree: 0.785714285714




Random Forest: 0.857142857143




Gradient Boosting: 0.785714285714



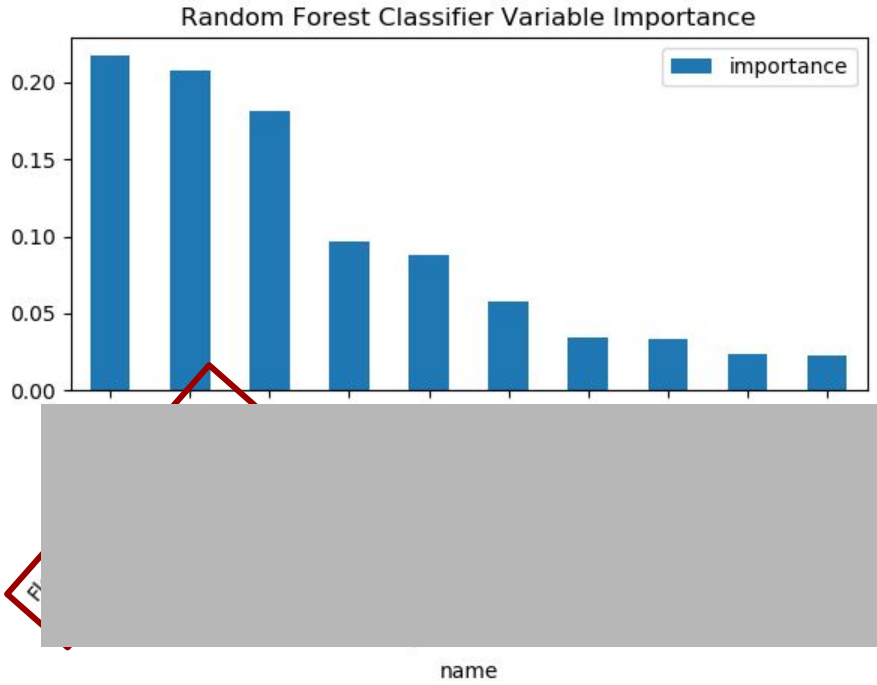

TN: true negative
TP: true positive
FN: false negative
FP: true positive

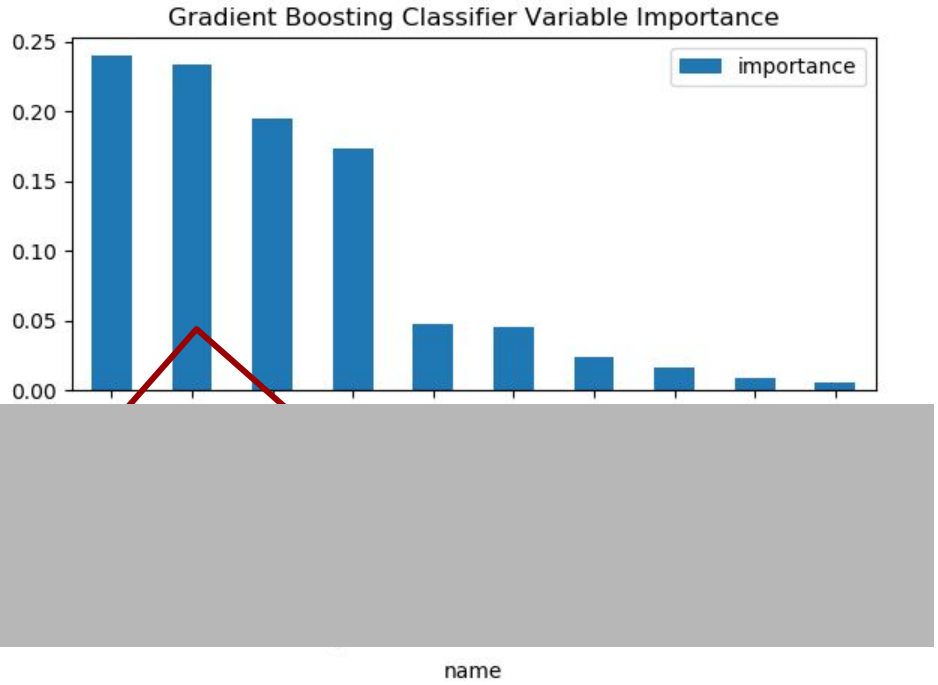
# Model Comparison

| | Model | Classification | F1Score | Precision | Recall |
|---|---|---|---|---|---|
| 1 | RandomForest | 0.8571428571428571 | 0.9130434782608695 | 0.9545454545454546 | 0.875 |
| 2 | GradientBoosting | 0.7857142857142857 | 0.8636363636363636 | 0.8636363636363636 | 0.8636363636363636 |
| 3 | Logistic | 0.7142857142857143 | 0.8095238095238095 | 0.7727272727272727 | 0.85 |
| 4 | DecisionTree | 0.7857142857142857 | 0.875 | 0.9545454545454546 | 0.8076923076923077 |

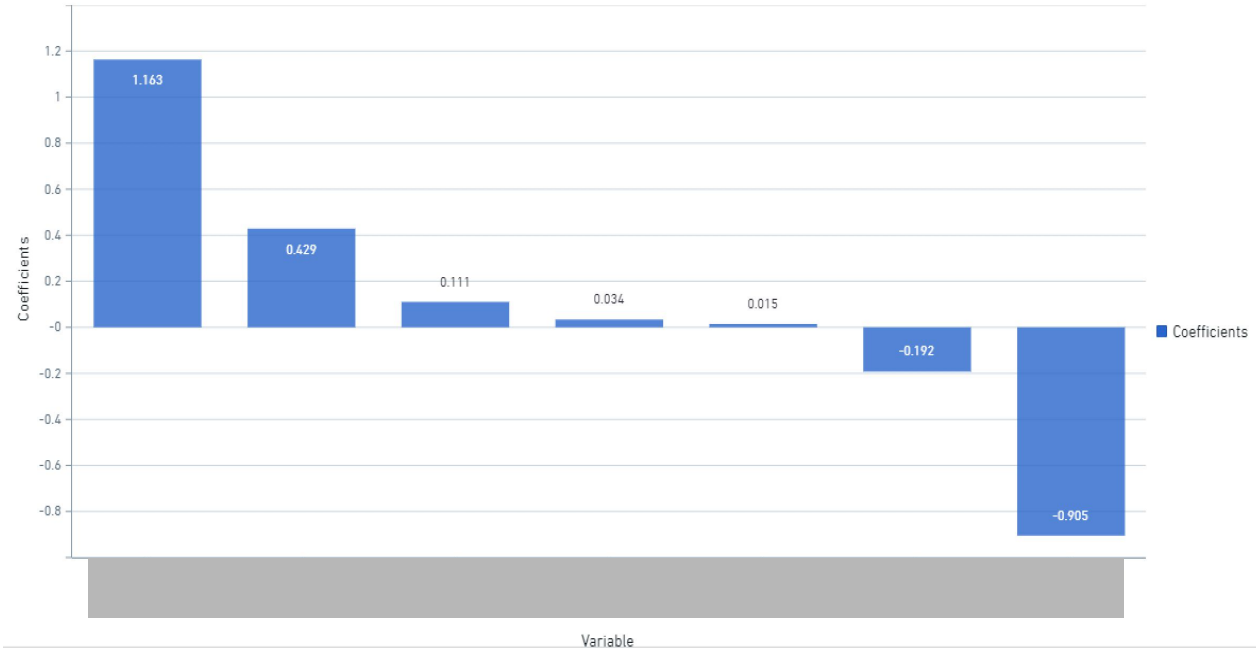
- **Classification**:
  - Percentage of all correct decisions
- **Precision**:
  - Percentage of predicted winning results actually won
  - weigh on minimizing the number of classifying to win while actually lose
- **Recall**:
  - Percentage of total winning results correctly classified
  - weigh on minimizing the number of classifying to lose while actually win
- **F1 Score**:
  - Weighted average of Precision and Recall

# Variable Importances in Random Forest & Gradient Boosting



# Variable Coefficients in Logistic Regression




Benchmark of categorical variable:

- Customer Type
  - Reseller
- Material
  - Regular
- Index
  - OPIS
- Gross/Net
  - Net

# Summary

- Random forest classifier performs the best in terms of different metrics among all 4 classifiers
- Expected margin and adder are the top 2 important variables
  - However, counterintuitively, with expectation to have higher floor margin, probability of winning the bid contract will be higher under logistic regression
  - Increasing the amount of adder is more likely to result in winning
- [Client] is more likely to win the bid contract when bidding on Gross and dealing with reseller customer

# Bringing It All Together

A new workflow for bid pricing

# Current Process

## Summary

Excel workbook stores snapshots of index data

Bid structure settings change calculations/lookups in complicated ways

Historical margin profile analyzed to judge bid viability

## Limitations

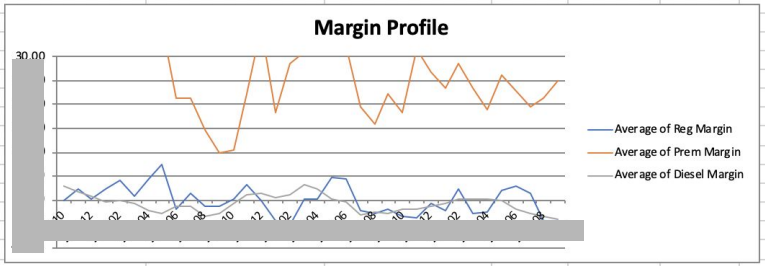
- Slow
- Excel formulas opaque
- Data limitations
- Versioning
- External dependencies

| INPUTS FROM MJ: | Indianapolis |
|---|---|
| Pricing Index | OPIS |
| OPIS Low G/N | Gross |
| Gross/Net Gallons | Net |
| Same/Prior Day | Prior |
| Seasonality | None* |
| | TOP TIER |
| * | |

| INPUTS FROM COST STACK: | | | |
|---|---|---|---|
| | 9# | Prem | ULSD |
| ise Product | | | |
| ETOH Diff | | | |
| Primary Trans | | | |
| S&H | | | |
| Additive Cost/Lubricity | | | |
| Add Sys/Eth/Bio Fee | | | |
| G/N Temp | | | |
| Variable Cost | | | |
| Integrated Value | | | |
| Total | | | |

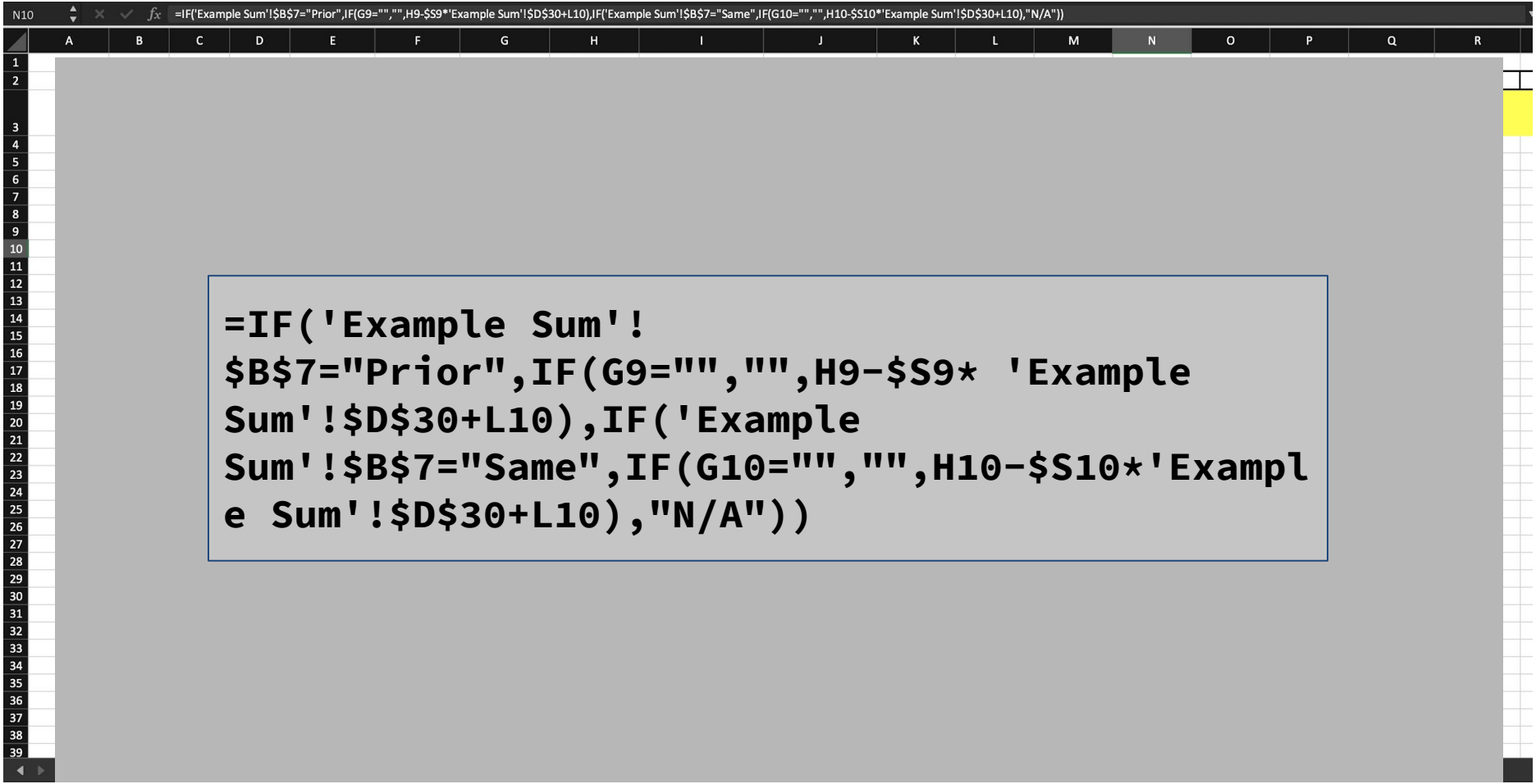
| | OPIS Rack Price | Product Cost | Price | Expected Margin | BP Branded Margin |
|---|---|---|---|---|---|
| 9# Regular | | | | | |
| 9# Premium Adder | | | | | |
| ULSD Adder | | | | | |

| | OPIS Adder | Platts | RIN Share | Current Adder/Index |
|---|---|---|---|---|
| 9# Regular | | | | |
| 9# Premium Adder | | | | |
| ULSD Adder | | | | |




| Row Labels | Average of Reg Margin | Average of Prem Margin | Average of Diesel Margin |
|---|---|---|---|
| 201510 | | | |
| 201511 | | | |
| 201512 | | | |
| 201601 | | | |
| 201602 | | | |
| 201603 | | | |
| 201604 | | | |
| 201605 | | | |
| 201606 | | | |
| 201607 | | | |
| 201608 | | | |
| 201609 | | | |
| 201610 | | | |
| 201611 | | | |
| 201612 | | | |
| 201701 | | | |
| 201702 | | | |
| 201703 | | | |
| 201704 | | | |
| 201705 | | | |
| 201706 | | | |
| 201707 | | | |
| 201708 | | | |
| 201709 | | | |
| 201710 | | | |
| 201711 | | | |
| 201712 | | | |
| 201801 | | | |
| 201802 | | | |
| 201803 | | | |
| 201804 | | | |
| 201805 | | | |
| 201806 | | | |
| 201807 | | | |
| 201808 | | | |
| 201809 | | | |
| Grand Total | | | |

| | Volume | Average Margin | Integrated Margin |
|---|---|---|---|
| 9# Regular | | | |
| Midgrade | | | |
| 9# Premium | | | |
| ULSD | | | |
| TOTAL | | | |


N10
=IF('Example Sum'!$B$7="Prior",IF(G9="","",H9-$S9*'Example Sum'!$D$30+L10),IF('Example Sum'!$B$7="Same",IF(G10="","",H10-$S10*'Example Sum'!$D$30+L10),"N/A"))
A B C D E F G H I J K L M N O P Q R
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
=IF('Example Sum'!
$B$7="Prior",IF(G9="","",H9-$S9* 'Example
Sum'!$D$30+L10),IF('Example
Sum'!$B$7="Same",IF(G10="","",H10-$S10*'Exampl
e Sum'!$D$30+L10),"N/A"))

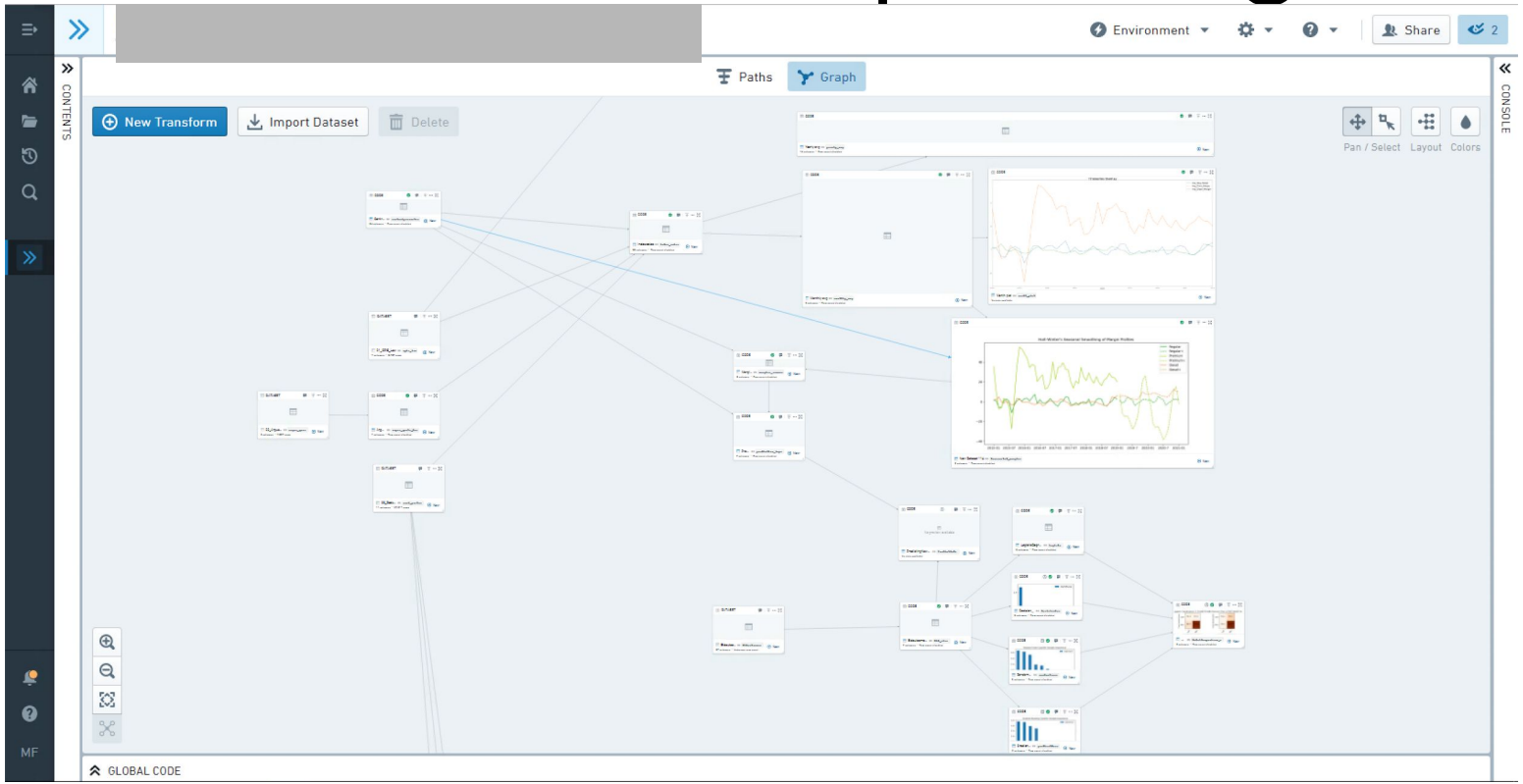
# PySpark Advantage

- Built on existing infrastructure in the cloud
- Scales with large amounts of data
- Allows for more complex modeling
- More maintainable code

# Workbook Demonstration

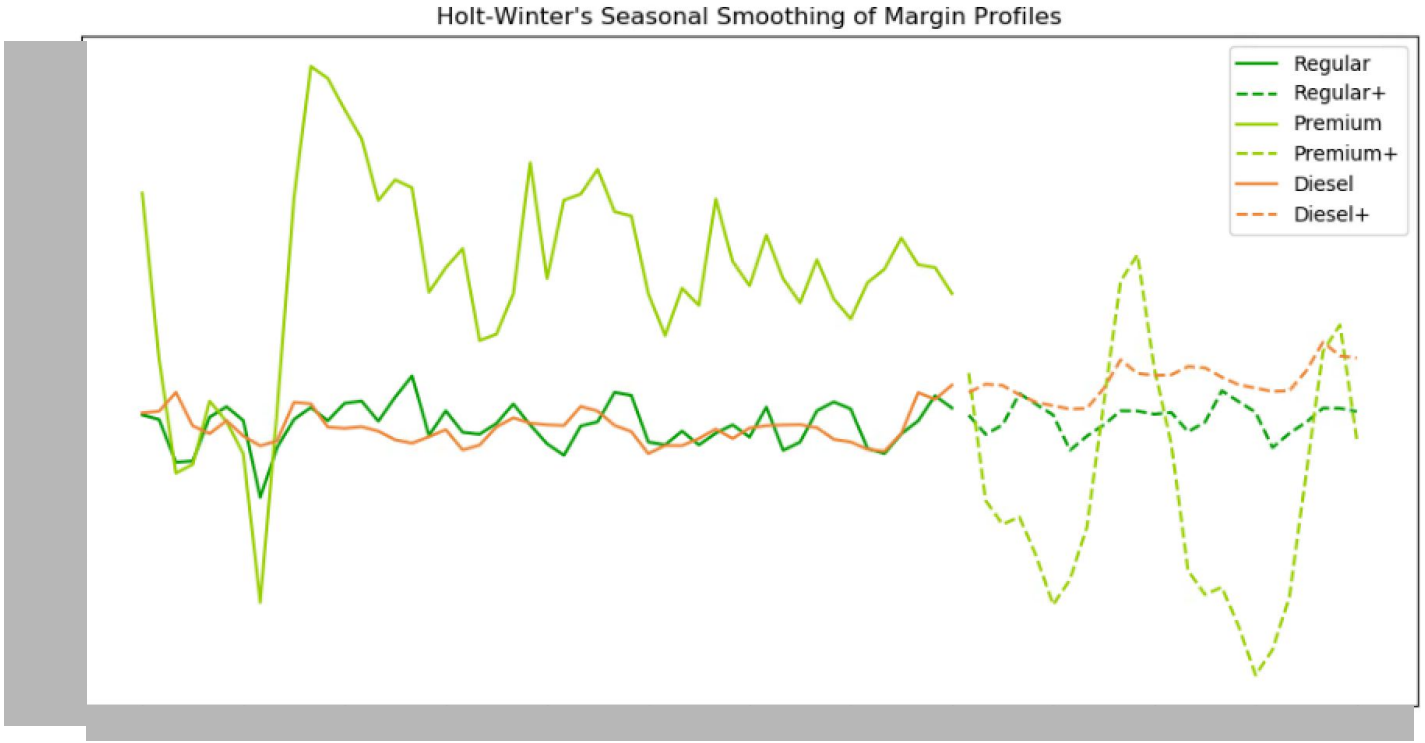
[Live demonstration in Palantir]

# Workbook Sample Images

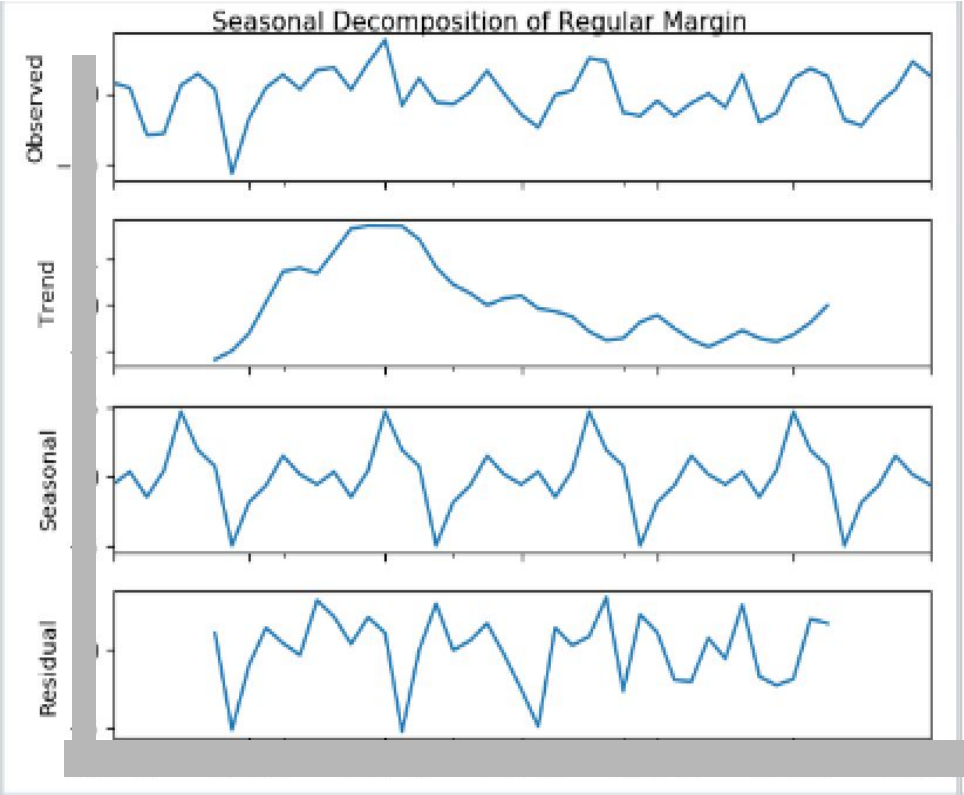


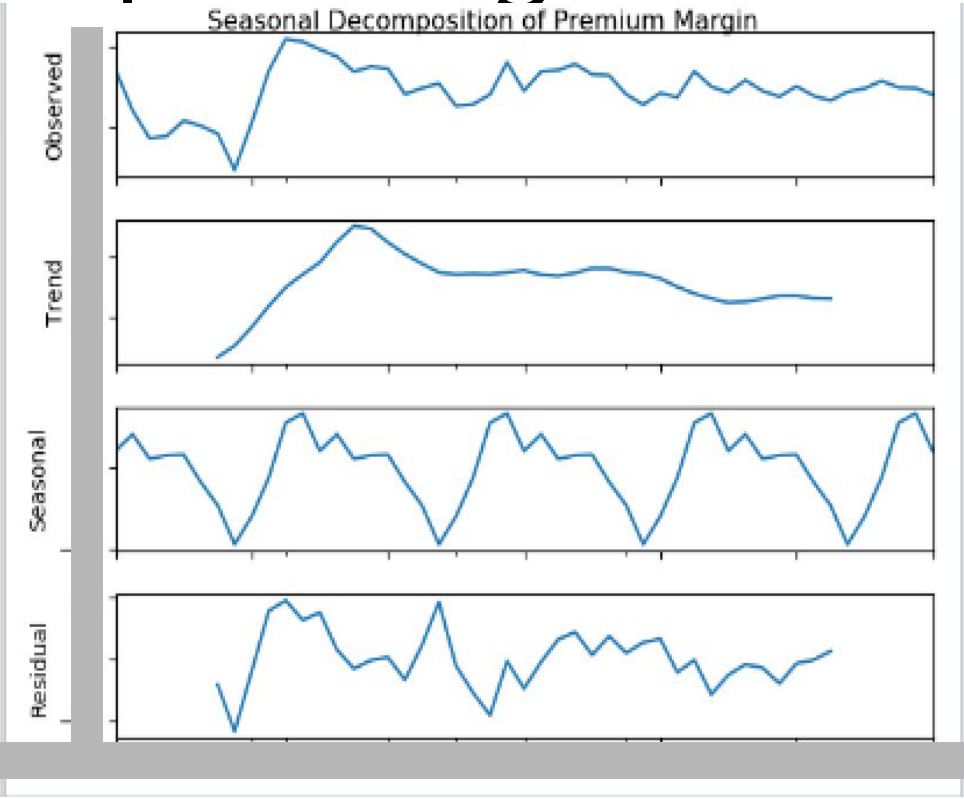
# Workbook Sample Images

# Workbook Sample Images



# Workbook Sample Images



# Revamped Process

- Data flows through Palantir as available
- Notebook provides graphical view of flow
- Infused with Analytics and Machine Learning
- Ultimately leads to more positive outcomes

# What We Have Gained

*“This project helped me to understand the difference between branded and unbranded gas, and to build visualizations in PySpark” -* **James**

*“I’ve learned to build tables and graphs with hierarchy in PySpark and have better sense of the data both from a general sense and through comparisons” -* **Sophie**

*“I’ve enjoyed being able to apply concepts related to time-series data to real world applications” -* **Jonathan**

*“I’ve learned a lot about Palantir functionalities with implementation of PySpark” -* **Nora**

*“This project helped me to understand how enterprise cloud products can fit into existing business workflows” -* **Michael**

# Thank You!