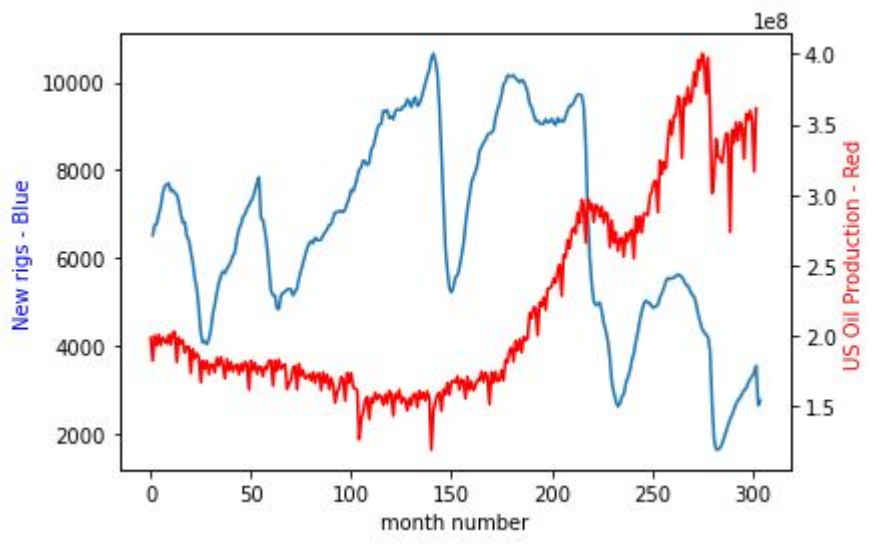
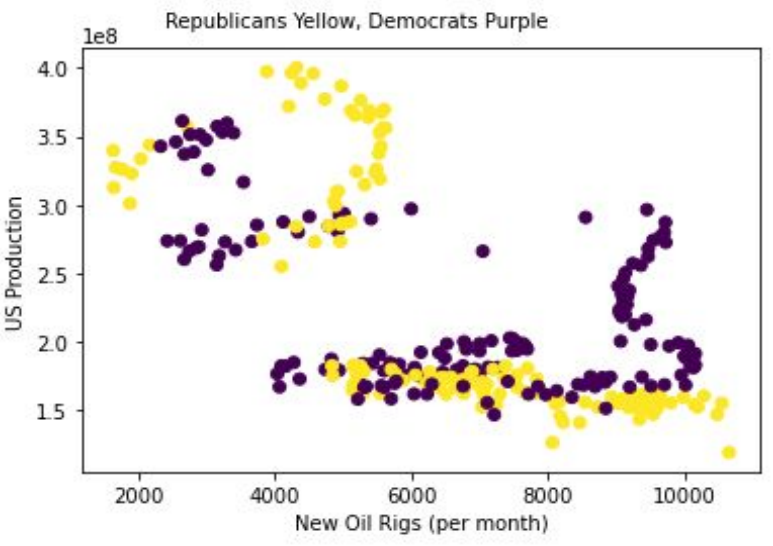
Oil and Gas Analysis

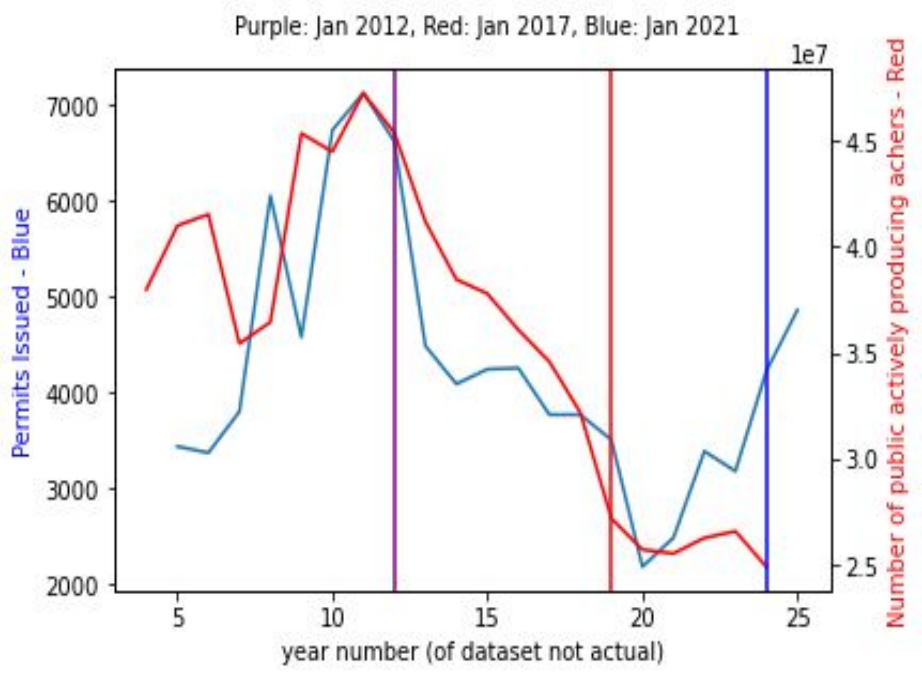
US Oil Health: Construction of New Oil Rigs

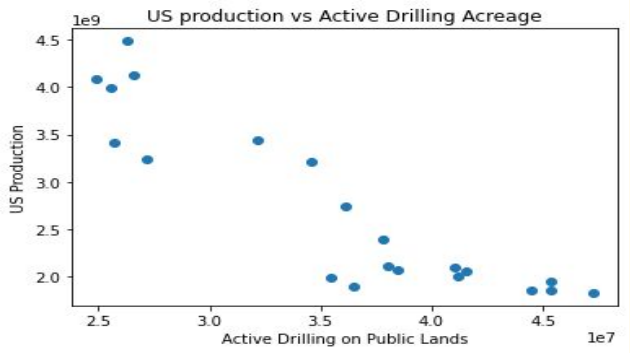


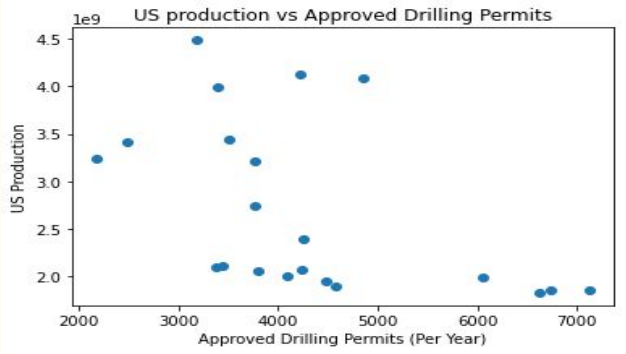


- No correlation between new oil rigs and US oil production
- No correlation between new oil rigs and presidential party (1997-2021)

US Oil Health: Gov't Drilling Permits and Active Acreage



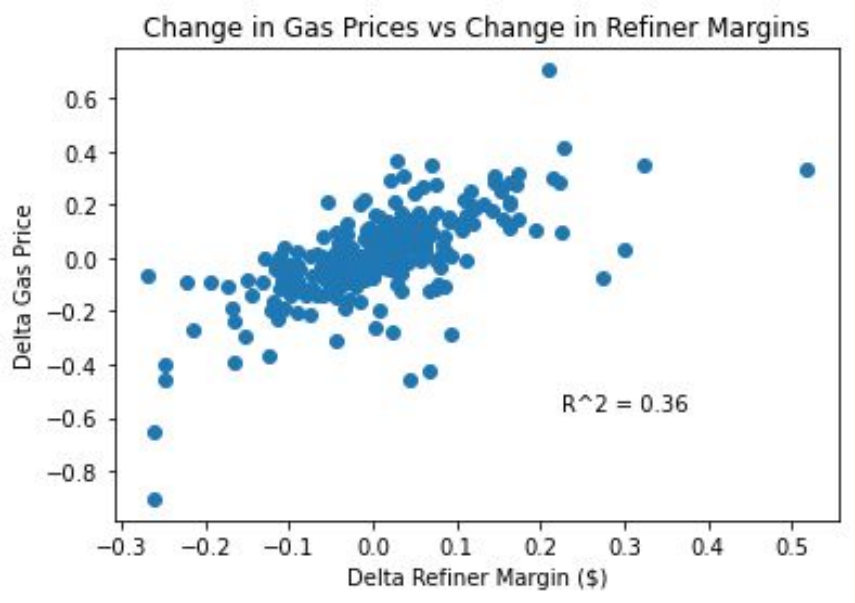


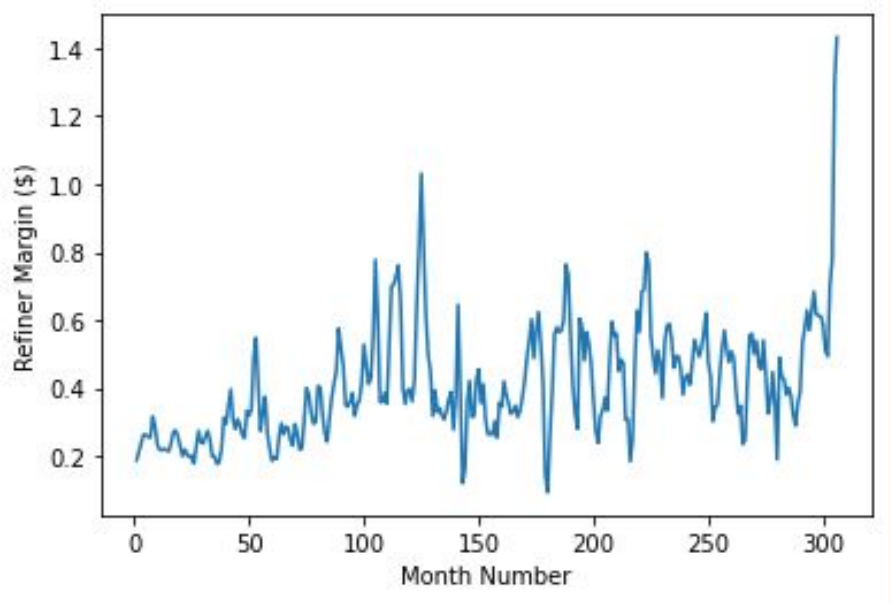


- Permits correlate to active drilling
- Active Drilling is negatively correlated with production (not all wells produce & data for public drilling only)

- No correlation between approved permits and US production
- Biden approved more Drilling permits than Trump YoY

Refiner Profit Margin Analysis





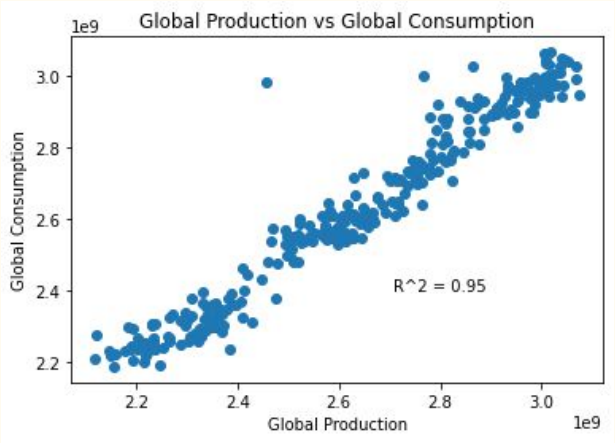
- Change in refinery profit margins correlate to change in gas prices (1 cent in margins = 3 cents in gas price on average)
- Highest refinery margins since at least 1997 have occurred in every month of 2022 (Data ends in May)

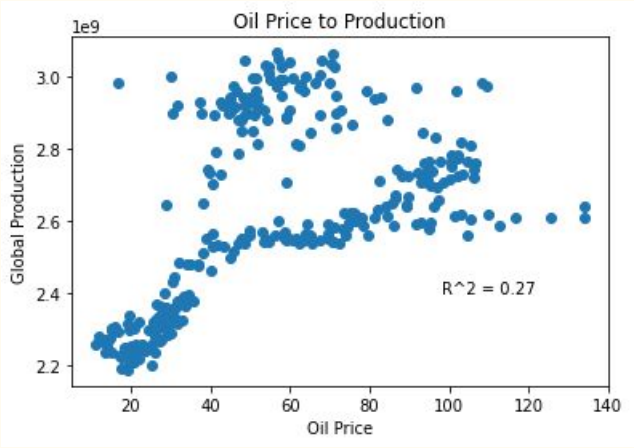
Oil Production Correlations

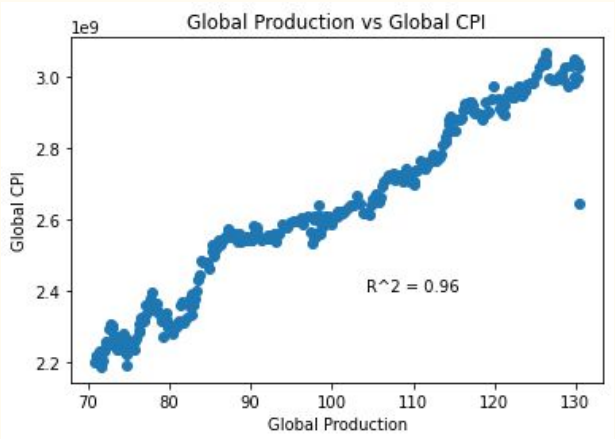
Index	Global Production	US Production	Global Production Ind	US Production Ind
glob_prod_Mbbld	1	0.77072	True	True
glob_prod	1	0.77072	True	True
indust_prod	0.98425	0.761299	True	True
core_cpi	0.982567	0.640456	True	True
world_cpi	0.979158	0.781785	True	True
us_cpi	0.978998	0.757001	True	True
other_prod	0.975828	0.730539	True	True
nonus_prod	0.973788	0.605587	True	True
us_indust_prod	0.97304	0.702612	True	True
glob_con	0.972591	0.722943	True	True
glob_con_Mbbld	0.972591	0.722943	True	True
month_no	0.955457	0.793276	True	True
us_retail_sales	0.955146	0.801938	True	True
year	0.954066	0.792457	True	True
gas_margin	0.852076	0.763995	True	True
us_core_cpi	0.819912	0.588247	True	True
opec_prod_Mbbld	0.79091	0.293632	True	False
opec_prod	0.79091	0.293632	True	False
us_prod_Mbbld	0.770733	0.999999	True	True
us_prod	0.77072	1	True	True

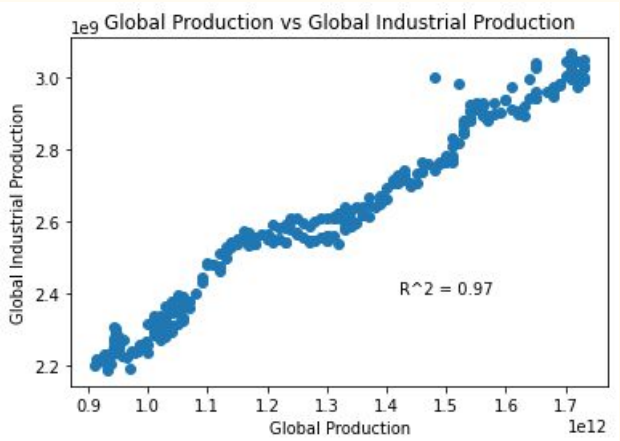
Index	Global Production	US Production	Global Production Ind	US Production Ind
retail_sales	0.723486	0.212515	True	False
gas_raw	0.65915	0.227268	True	False
refiner_sale	0.589584	0.135331	True	False
refiner_cost	0.544375	0.0944077	True	False
oil_price	0.523936	0.0551597	True	False
refiner_margin	0.500384	0.255492	True	False
us_con_Mbbld	0.124893	0.0172744	False	False
us_con	0.124893	0.0172744	False	False
republican	0.0281541	-0.00205397	False	False
glob_deficit_Mbbld	0.0192602	0.12507	False	False
glob_deficit	0.0192602	0.12507	False	False
margin_dif	0.00559738	-0.00520498	False	False
gas_adj	-0.00478892	-0.107354	False	False
gas_dif	-0.0180716	-0.0116632	False	False
oced_gdp	-0.0539358	-0.177773	False	False
new_rigs	-0.212452	-0.583333	False	True
unem_rate	-0.496772	-0.507057	False	True
running_def_Mbbl	-0.914923	-0.684189	True	True
running_def	-0.914923	-0.684189	True	True

Oil Production Correlations: Global Production

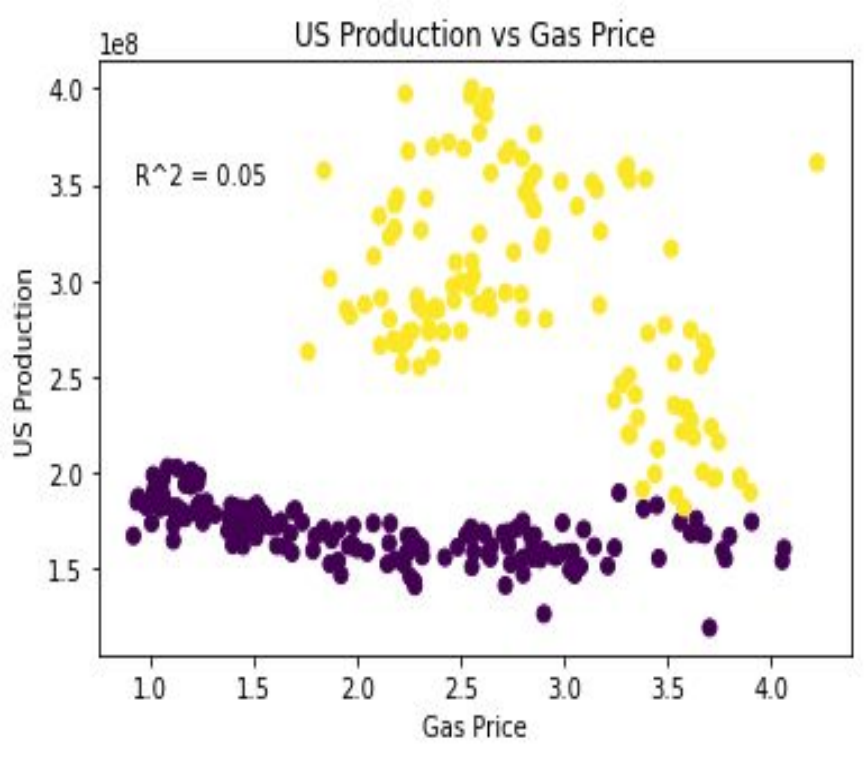


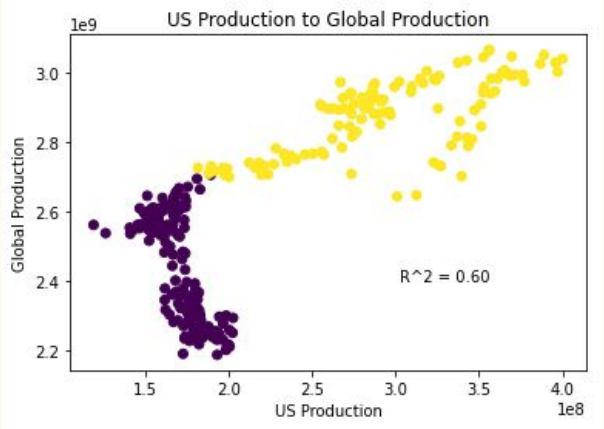


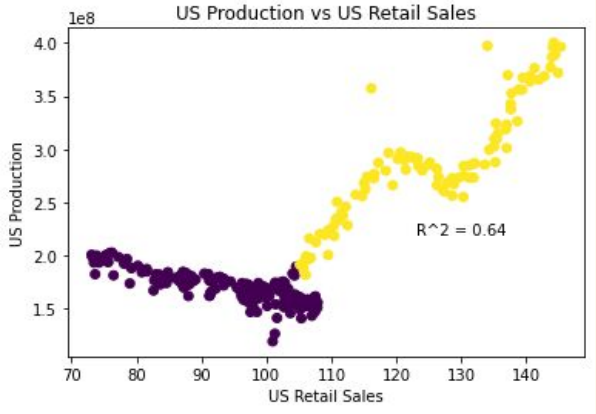




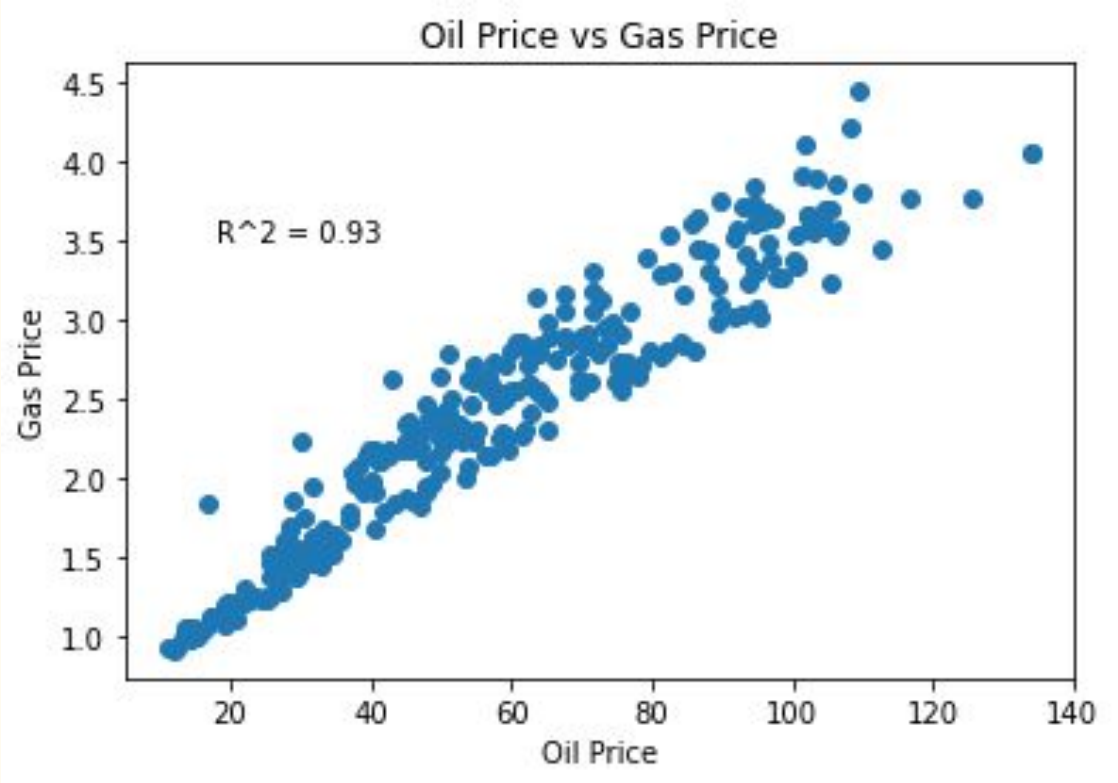
Oil Production Correlations: US Production Purple is pre 2012







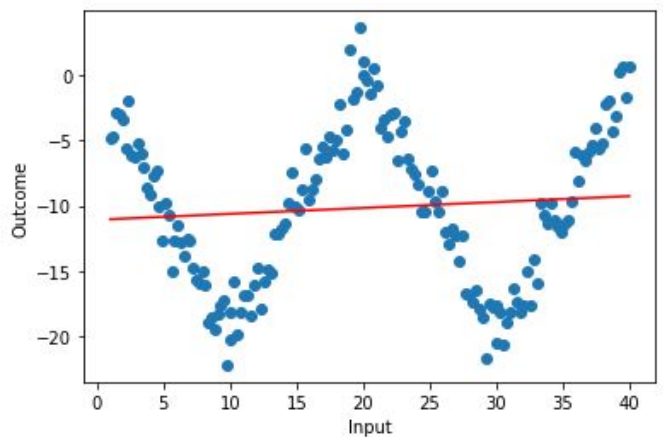
What Drives Gas Price? Oil Price!



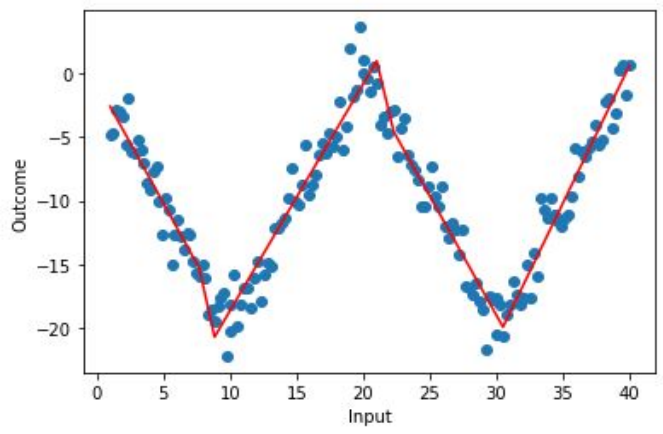
MARS Introduction

- MARS = Multivariate Adaptive Regression Splines
  - Machine Learning method to rapidly identify both linear and non-linear trends
  - “Splines” are knots in otherwise linear data (vertices of the W)
  - MARS can find:
    - Linear Trends
    - Non-linear combinations (think variable * variable)
    - Linear Splines

Standard Linear Regression



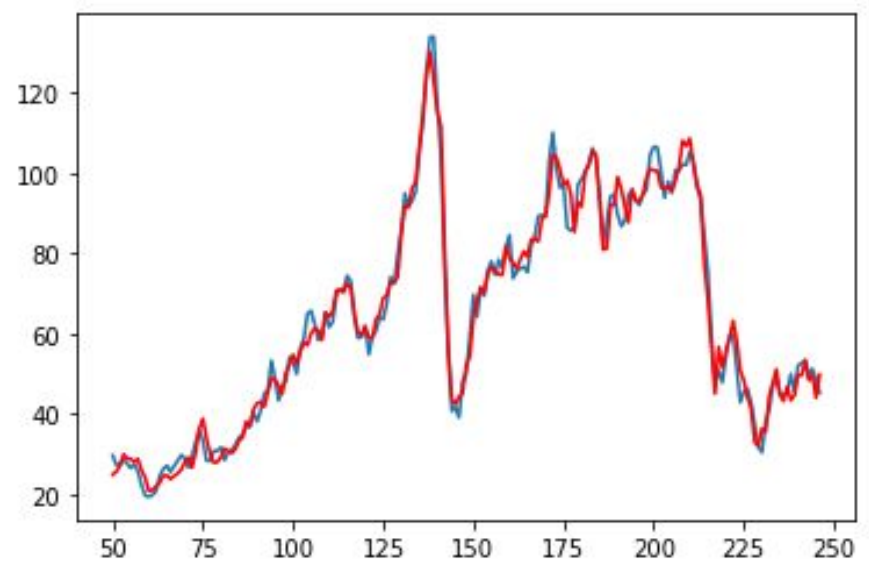
MARS Solution



MARS Data Mining

- Training over whole dataset MARS fits nearly perfectly
- MARS alone struggled to deal with **out of sample** large drops (overfit to the data - next slide)
- Can MARS be used as a data mining application?
    - Run 200 80-20 Train/Discard splits and assess frequency of each feature incorporated into the model

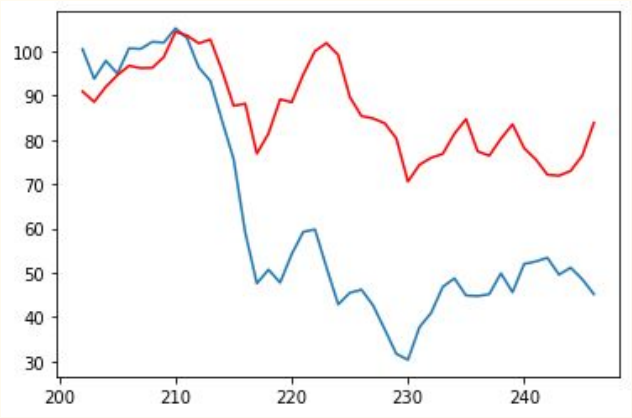
Full Model Fit with MARS (Blue is actual oil price red is MARS prediction)



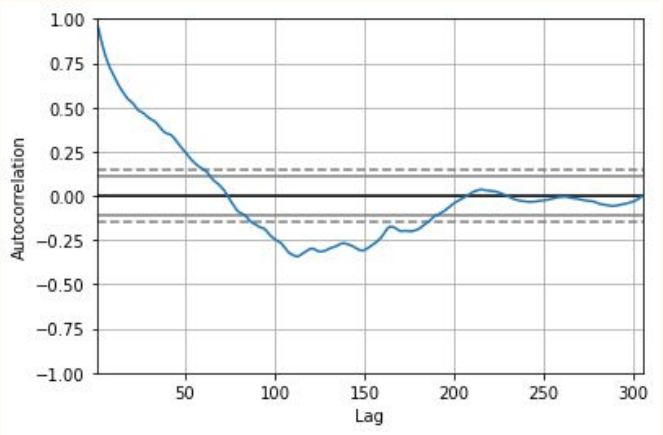
Multiple Regression Optimization

- Used MARS feature frequency to narrow down feature options
  - 4,000 possible combinations instead of 4.1 Million
  - Added autoregressive term to 1000 with lowest Mean Squared Error (MSE)
    - Autoregressive term set arbitrarily at R = 0.75 which correspond to regression over last 7 months
- Final Model chosen with lowest net MSE
  - Net MSE = MSE on oil price prediction + MSE when piped into regression to predict gas price (next slide)
  - Autoregression ultimately stabilized predictions rather than dictated them

MARS Out of Sample

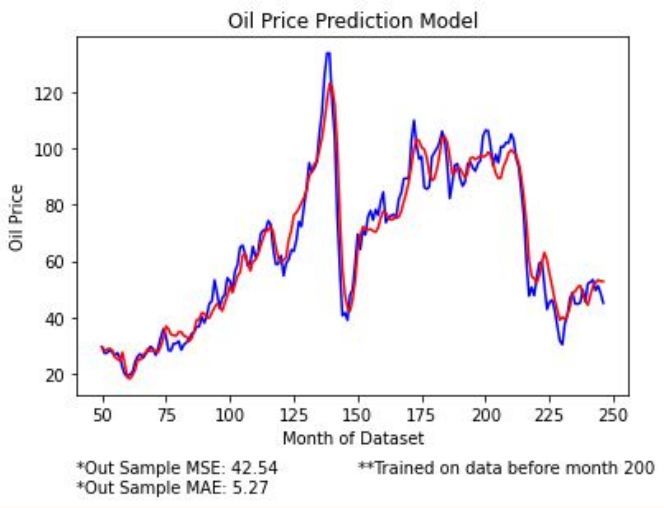


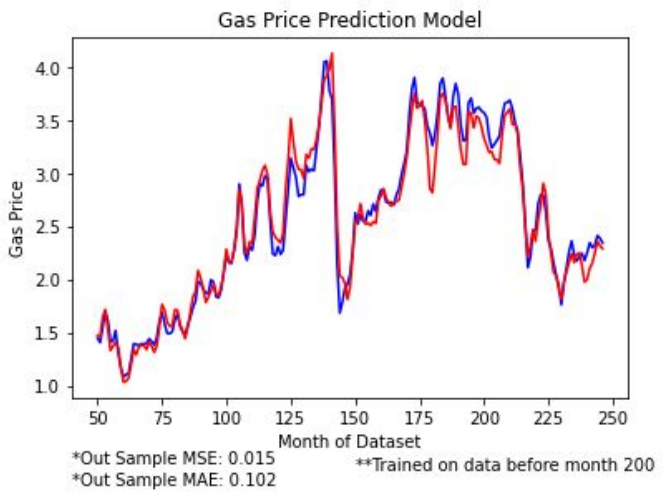
Oil Price Autocorrelation



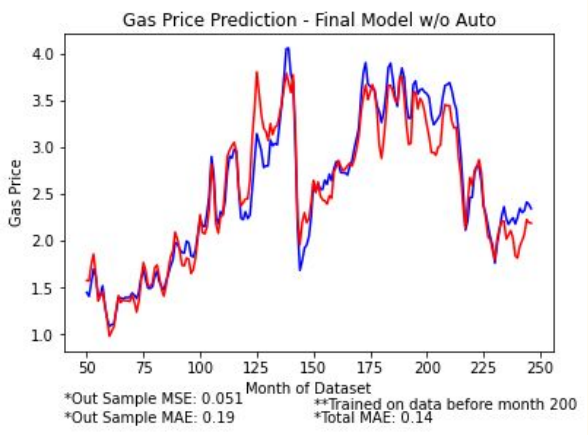
Final Model

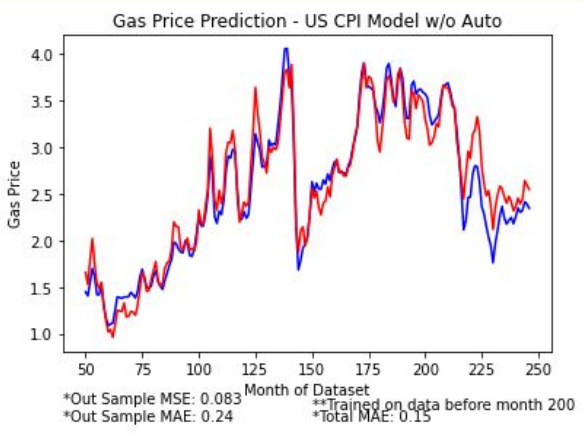
- Most Predictive Features:
  - Running Deficit (Oil production - Oil consumption from 1997 to 2022)
    - Crude estimation of supply/demand
  - US and Global Industrial Production
  - US Retail Sales
  - Global Overall CPI
  - Global Core CPI
- Final Model for Oil Price results piped into gas regression
  - Gas Price = Oil Price + Refinery Profit Margins
  - R^2 = 0.933 → 0.972 (Just oil → oil + refiner margins)
- Whole Model Average Absolute Errors:
  - Oil: $4.63
  - Gas: $0.11
- Total Absolute Errors:
  - Oil: $912.27 (7.2%)
  - Gas: $21.12 (4.2%)

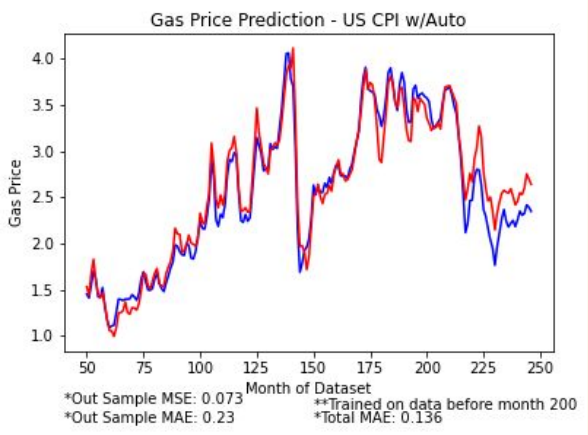




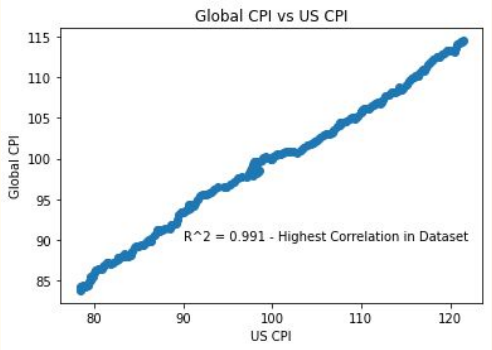
Other Models - US CPI preforms mildly worse, but US CPI nearly perfectly correlated to global CPI





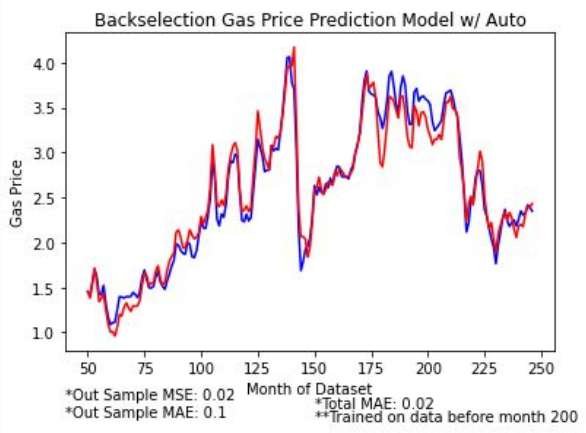


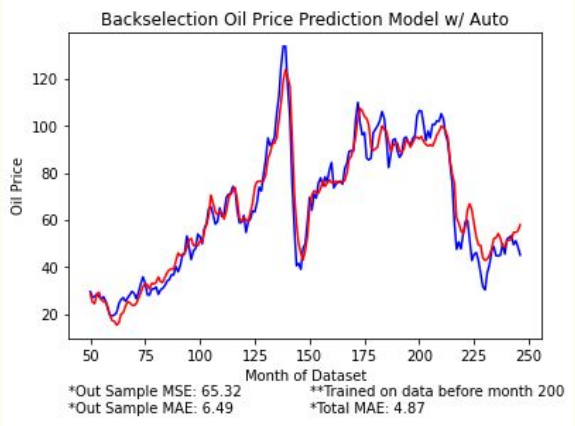
** Auto = autoregressive term

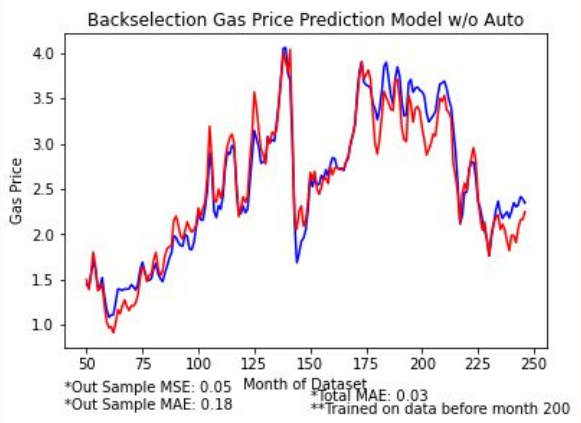


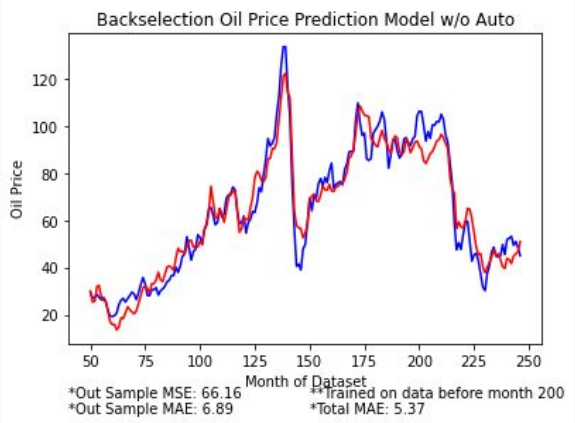
** Auto = autoregressive term

Other Models - Standard backwards model selection performs worse than MARS data mining solution









Correlation is not Causation

- The final model is the most accurate of all the candidates for both oil and gas predictions
- However many models perform similarly
- Does Oil price cause global inflation?
- Do **any** of the variables cause oil price changes?
  - If so, how many are correlated to each other?
- Is there any other factor tying them all together?
- Causation does require correlation