THE US INFLATION PHENOMENON |

It's Oil, silly

AUTHOR | Rand Sobczak Jr.

DATE | 22 April 2021

Table of contents

01 Problem Identification

02 Generated Deliverables

03 Data Pre-Processing

04 Model Description

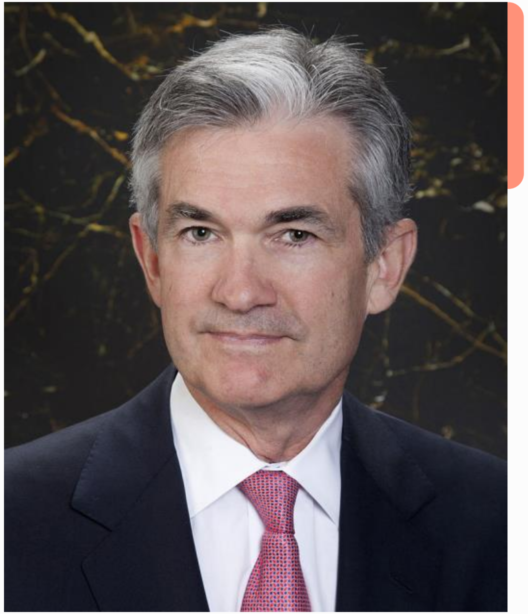
05 Model Findings

06 Next Steps

01

Problem Identification

Developing a model to explain & understand the phenomenon of Inflation



Inflation is important

as **stabilizing it is one of three objectives of the Federal Reserve who's decisions move the global financials markets**

The purpose & goal of this Data Science project is to

build a model to
explain & understand
the phenomenon of
Inflation

Problem Identification
(cont.)

I **shortlisted 19 variables** to determine their influence on Inflation

Items	Reported	API	API Source	Comments
Inflation	Monthly	Quandl	U.S. Bureau of Labor Statistics	The target variable
Wages CPI	Monthly	FRED	U.S. Bureau of Labor Statistics	A component of the target variable
WTI	Daily	Quandl	CME	West Texas Intermediate - One of many commodities
Heating Oil	Daily	Investpy	Investing.com	One of many commodities
Copper	Daily	Investpy	Investing.com	One of many commodities
Sugar	Daily	Investpy	Investing.com	One of many commodities
Natural Gas	Daily	Investpy	Investing.com	One of many commodities
Cattle	Daily	Investpy	Investing.com	One of many commodities
Lean Hogs	Daily	Investpy	Investing.com	One of many commodities
Soybeans	Daily	Investpy	Investing.com	One of many commodities
Lumber	Daily	Investpy	Investing.com	One of many commodities
Capacity Utilization	Monthly	FRED	Board of Governors of the Federal Reserve	The % of resources used by corporations
Corn	Daily	Investpy	Investing.com	One of many commodities
M2 Velocity	Quarterly	FRED	Federal Reserve Bank of St. Louis	Movement of money; state of the economy proxy
GDP	Quarterly	FRED	U.S. Bureau of Economic Analysis	A proxy for the state of the economy
Wheat	Daily	Investpy	Investing.com	One of many commodities
PMI	Monthly	Quandl	Institute of Supply Management	Manufacturing PMI - A proxy for the economy
USD Index	Daily	Quandl	Intercontinental Exchange Inc	(DXY) Proxy for potentially importing inflation
Unemployment Rate	Monthly	Quandl	U.S. Bureau of Labor Statistics	A proxy for the state of the economy
Initial Jobless Claims	Weekly	Quandl	U.S. Employment and Training Administration	A proxy for the state of the economy

02

Generated Deliverables

The power of API’s

Generated Deliverables

(cont.)

I **shortlisted 19 variables** to determine their influence on Inflation

Items	Reported	API	API Source	Comments
Inflation	Monthly	Quandl	U.S. Bureau of Labor Statistics	The target variable
Wages CPI	Monthly	FRED	U.S. Bureau of Labor Statistics	A component of the target variable
WTI	Daily	Quandl	CME	West Texas Intermediate - One of many commodities
Heating Oil	Daily	Investpy	Investing.com	One of many commodities
Copper	Daily	Investpy	Investing.com	One of many commodities
Sugar	Daily	Investpy	Investing.com	One of many commodities
Natural Gas	Daily	Investpy	Investing.com	One of many commodities
Cattle	Daily	Investpy	Investing.com	One of many commodities
Lean Hogs	Daily	Investpy	Investing.com	One of many commodities
Soybeans	Daily	Investpy	Investing.com	One of many commodities
Lumber	Daily	Investpy	Investing.com	One of many commodities
Capacity Utilization	Monthly	FRED	Board of Governors of the Federal Reserve	The % of resources used by corporations
Corn	Daily	Investpy	Investing.com	One of many commodities
M2 Velocity	Quarterly	FRED	Federal Reserve Bank of St. Louis	Movement of money; state of the economy proxy
GDP	Quarterly	FRED	U.S. Bureau of Economic Analysis	A proxy for the state of the economy
Wheat	Daily	Investpy	Investing.com	One of many commodities
PMI	Monthly	Quandl	Institute of Supply Management	Manufacturing PMI - A proxy for the economy
USD Index	Daily	Quandl	Intercontinental Exchange Inc	(DXY) Proxy for potentially importing inflation
Unemployment Rate	Monthly	Quandl	U.S. Bureau of Labor Statistics	A proxy for the state of the economy
Initial Jobless Claims	Weekly	Quandl	U.S. Employment and Training Administration	A proxy for the state of the economy

Generated Deliverables

Quandl

Quandl is a marketplace for financial, economic and alternative data

Investing.com

A financial platform & news website; one of the top 3 financial websites in the world

FRED

Federal Reserve Economic Data (FRED) a database maintained by the Research division of the Federal Reserve Bank of St. Louis

Generated Deliverables
(cont.)

Source Code

This can be found at my GitHub account referenced at the end

Research Report

Also can be found at my GitHub account referenced at the end

Presentation Report

This one...

03

Data Pre-Processing

Split it up...

Data Pre-Processing

Data Cleaning

Data Frames should talk to each other

- After pulling the data frame was **composed of variables with different lengths**

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 14166 entries, 1946-01-01 to 2021-04-20
Data columns (total 19 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Wages CPI              14157 non-null  float64
 1   WTI                    11952 non-null  float64
 2   Copper                 10304 non-null  float64
 3   Soybeans               9863 non-null   float64
 4   Natural Gas            9779 non-null   float64
 5   Heating Oil            12951 non-null  float64
 6   Corn                   12950 non-null  float64
 7   Wheat                  9865 non-null   float64
 8   Cattle                 12948 non-null  float64
 9   Lean Hogs              12953 non-null  float64
 10  Sugar                  12951 non-null  float64
 11  Lumber                 12953 non-null  float64
 12  Capacity Utilization   13897 non-null  float64
 13  GDP                    14159 non-null  float64
 14  M2 Velocity            14015 non-null  float64
 15  PMI                    14145 non-null  float64
 16  USD Index              11137 non-null  float64
 17  Initial Jobless Claims 13894 non-null  float64
 18  Unemployment Rate      14145 non-null  float64
dtypes: float64(19)
memory usage: 2.2 MB
```

Data Pre-Processing

Data Cleaning (cont.)

Data Frames should talk to each other (cont.)

- Different lengths
- **Cut the data to 18 April 1991**

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 9616 entries, 1991-04-18 to 2021-04-20
Data columns (total 19 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Wages CPI                9616 non-null   float64
 1   WTI                      9616 non-null   float64
 2   Copper                   9616 non-null   float64
 3   Soybeans                 9616 non-null   float64
 4   Natural Gas              9616 non-null   float64
 5   Heating Oil              9616 non-null   float64
 6   Corn                     9616 non-null   float64
 7   Wheat                    9616 non-null   float64
 8   Cattle                   9616 non-null   float64
 9   Lean Hogs                9616 non-null   float64
 10  Sugar                    9616 non-null   float64
 11  Lumber                   9616 non-null   float64
 12  Capacity Utilization     9616 non-null   float64
 13  GDP                      9616 non-null   float64
 14  M2 Velocity              9616 non-null   float64
 15  PMI                      9616 non-null   float64
 16  USD Index                9616 non-null   float64
 17  Initial Jobless Claims   9616 non-null   float64
 18  Unemployment Rate        9616 non-null   float64
dtypes: float64(19)
memory usage: 1.5 MB
```

Data Pre-Processing

Data Cleaning (cont.)

Data Frames should talk to each other (cont.)

- Different lengths
- Cut the Data
- **Concatenated with Inflation**
  - **Only 317 observations**

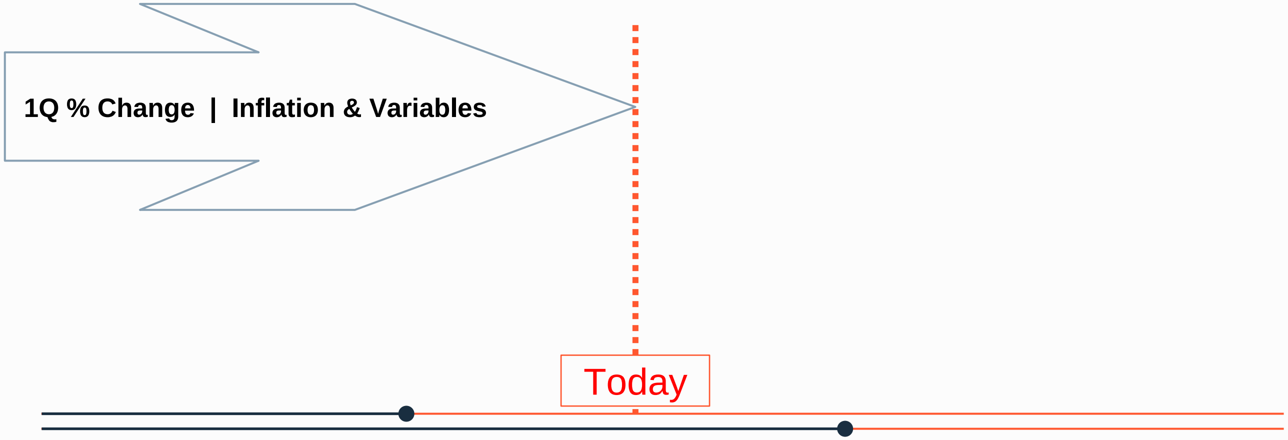
```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 317 entries, 1991-04-30 to 2021-03-31
Data columns (total 20 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Inflation                 317 non-null    float64
 1   Wages CPI                 317 non-null    float64
 2   WTI                       317 non-null    float64
 3   Copper                    317 non-null    float64
 4   Soybeans                  317 non-null    float64
 5   Natural Gas               317 non-null    float64
 6   Heating Oil               317 non-null    float64
 7   Corn                      317 non-null    float64
 8   Wheat                     317 non-null    float64
 9   Cattle                    317 non-null    float64
 10  Lean Hogs                 317 non-null    float64
 11  Sugar                     317 non-null    float64
 12  Lumber                    317 non-null    float64
 13  Capacity Utilization      317 non-null    float64
 14  GDP                       317 non-null    float64
 15  M2 Velocity               317 non-null    float64
 16  PMI                       317 non-null    float64
 17  USD Index                 317 non-null    float64
 18  Initial Jobless Claims    317 non-null    float64
 19  Unemployment Rate         317 non-null    float64
dtypes: float64(20)
memory usage: 52.0 KB
```

Data
Pre-Processing

Exploratory Data Analysis

Investigating the Time Relationships

- **Quarter on Quarter (for all)**
  - Compared a quarterly change on Variables against Inflation
- Month on Month (for all)
- Quarter on Quarter for Variables (past) & Inflation (forwards)
- Quarter on Quarter w/ Rolling Averages

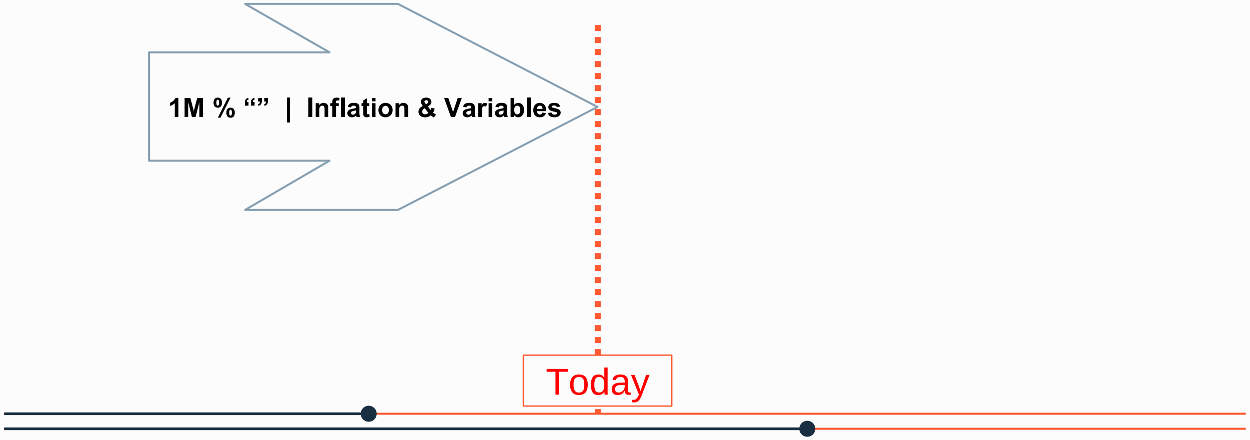


Data Pre-Processing

Exploratory Data Analysis (cont.)

Investigating the Time Relationships (cont.)

- Quarter on Quarter (for all)
- **Month on Month (for all)**
  - The same as the previous but looked at monthly change
- Quarter on Quarter for Variables (past) & Inflation (forwards)
- Quarter on Quarter w/ Rolling Averages

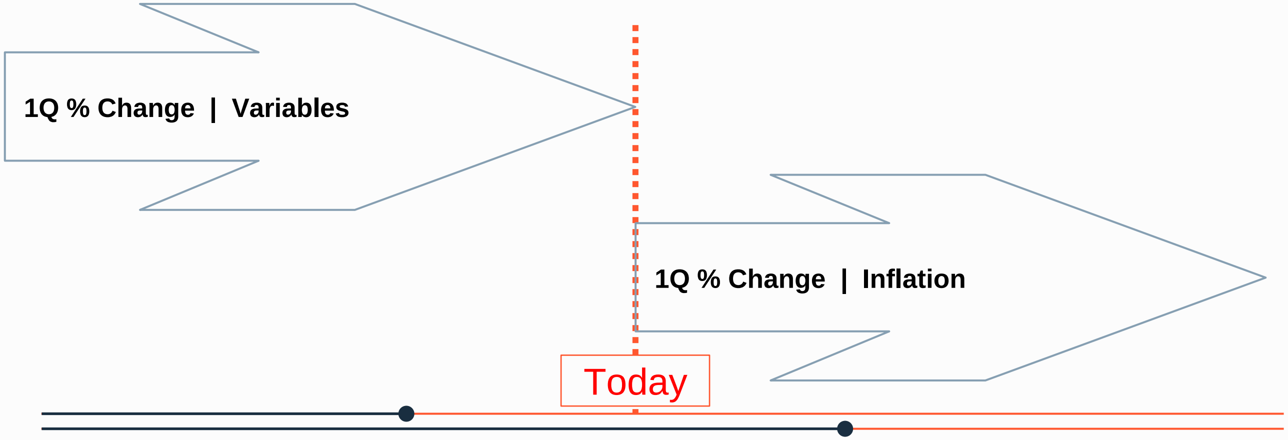


Data Pre-Processing

Exploratory Data Analysis (cont.)

Investigating the Time Relationships (cont.)

- Quarter on Quarter (for all)
- Month on Month (for all)
- **Q on Q for Variables (past) & Inflation (forwards)**
  - Looked at a previous 1 quarter change from variables to a 1 quarter future change in Inflation
- Quarter on Quarter w/ Rolling Averages

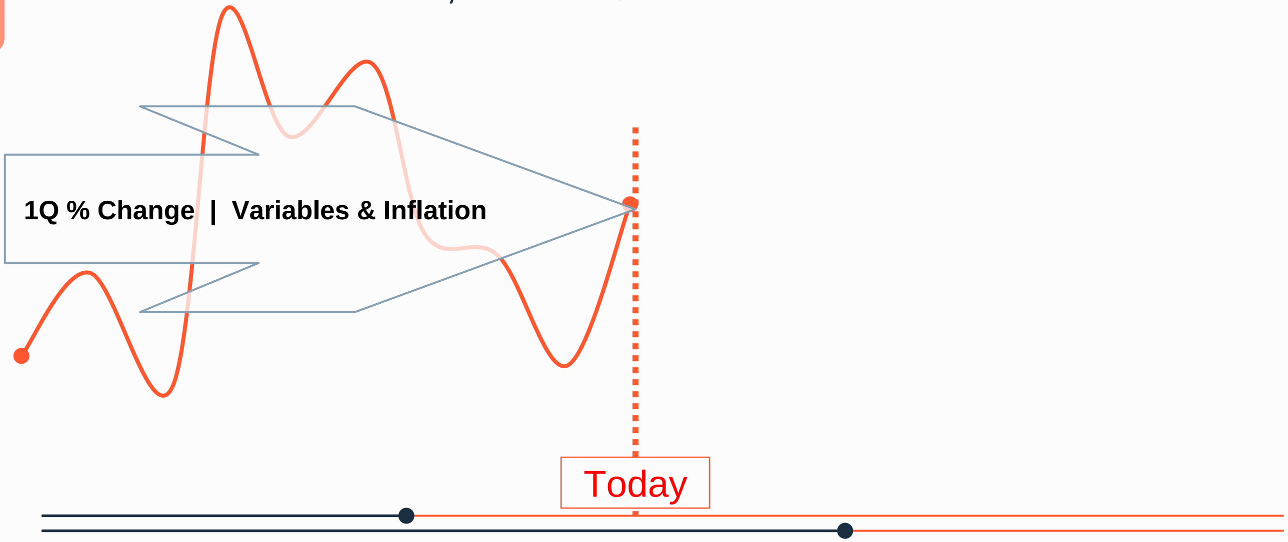


Data Pre-Processing

Exploratory Data Analysis (cont.)

Investigating the Time Relationships (cont.)

- Quarter on Quarter (for all)
- Month on Month (for all)
- Q on Q for Variables (past) & Inflation (forwards)
- **Quarter on Quarter w/ Rolling Averages**
    - Similar to # 1 albeit used a rolling average for those that were reported more often than once a Quarter as a Variable "may have had" a bad week or day when the Quarter ended

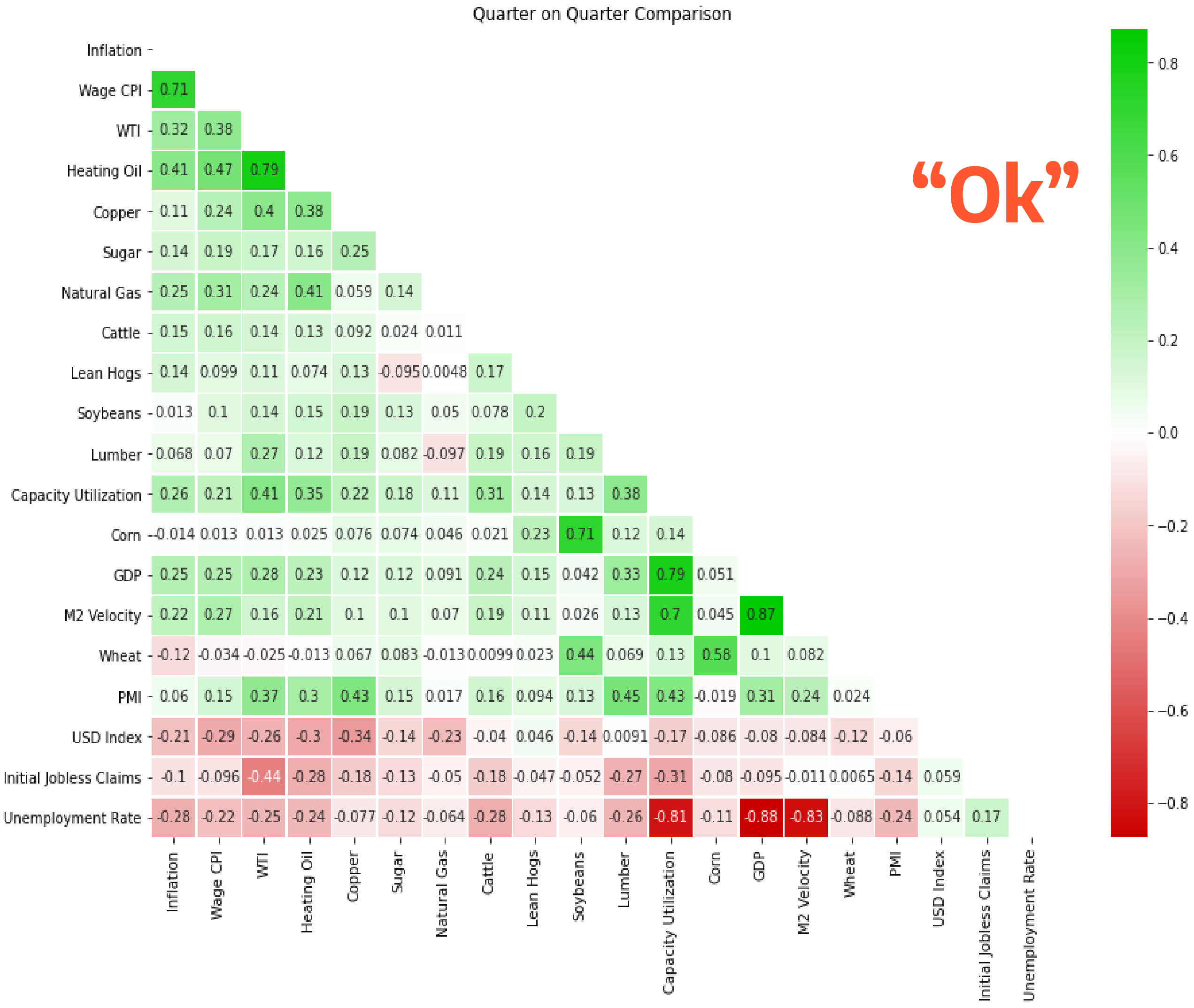


Data Pre-Processing

Exploratory Data Analysis

Quarter on Quarter (for all)

Feature Correlation Heat Maps with the Pearson correlation coefficients



Data Pre-Processing

Exploratory Data Analysis

Month on Month (for all)

Feature Correlation Heat Maps with the Pearson correlation coefficients (cont.)

Data Pre-Processing

Exploratory Data Analysis

Q on Q for Variables (past) & Inflation (forwards)

Feature Correlation Heat Maps with the Pearson correlation coefficients (cont.)

"Even Worse"

Data Pre-Processing

Exploratory Data Analysis

Quarter on Quarter w/ Rolling Averages

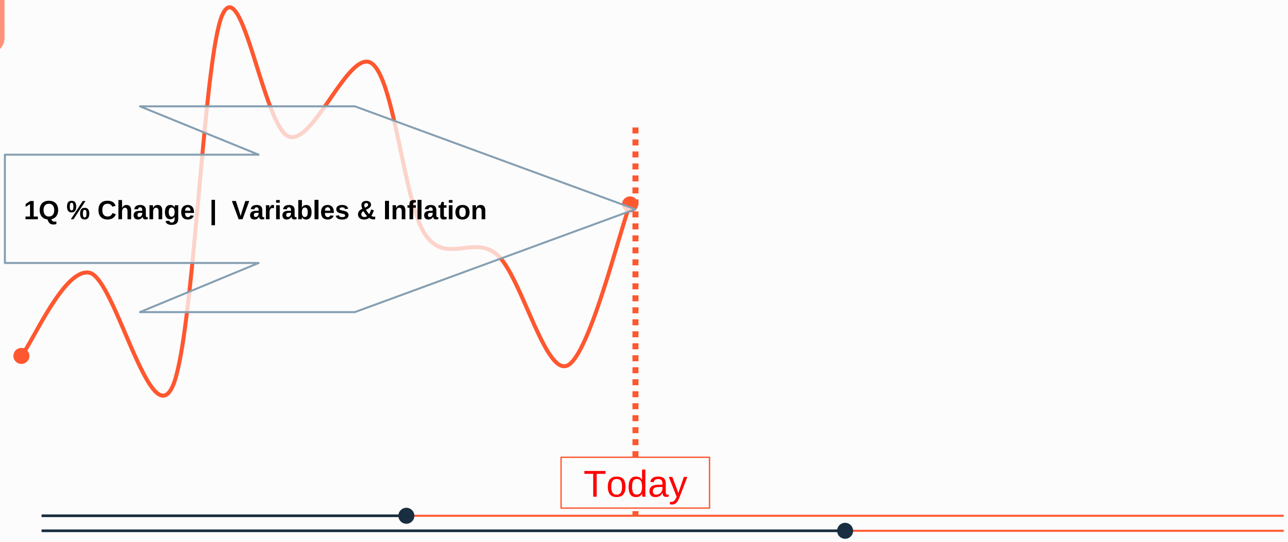
Feature Correlation Heat Maps with the Pearson correlation coefficients (cont.)

Data Pre-Processing

Pre-Processing

Splitting & Scaling

- **Chosen data frame**
  - The Quarter on Quarter w/ Rolling Averages was chosen
- Train, Test Split
- Scaling

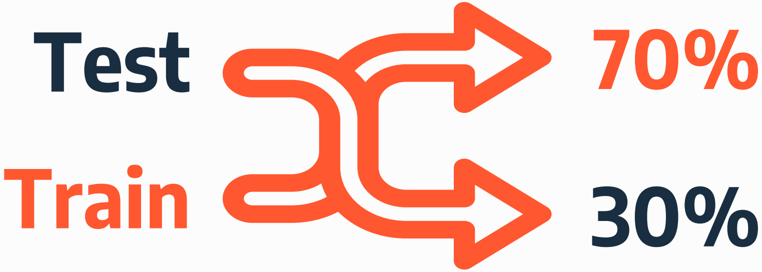


Data Pre-Processing

Pre-Processing (cont.)

Splitting & Scaling (cont.)

- Chosen data frame
- **Train, Test Split**
  - The data was then split for Training & Testing for the different Scaling Approaches
- Scaling



Data Pre-Processing

Pre-Processing (cont.)

Splitting & Scaling (cont.)

- Chosen data frame
- Train, Test Split
- **Scaling**
  - 3 scaling approaches were tried:
    - Standard Scaling (SS)
    - MinMax Scaling (MM)
    - Log Transformation (LG)

	Wages CPI_SS	WTI_SS
count	2.180000e+02	2.180000e+02
mean	-4.838128e-18	2.750094e-17
std	1.002301e+00	1.002301e+00
min	-6.977019e+00	-3.783391e+00
25%	-2.671202e-01	-5.665365e-01
50%	1.153214e-01	-3.959852e-02
75%	4.280369e-01	6.677299e-01
max	3.074376e+00	2.742497e+00

Wages CPI_MM	WTI_MM
218.000000	218.000000
0.694134	0.579751
0.099718	0.153589
0.000000	0.000000
0.667559	0.492937
0.705608	0.573683
0.736719	0.682071
1.000000	1.000000

Wages CPI_LG	WTI_LG
2.180000e+02	2.180000e+02
-2.340126e-16	-1.018553e-17
1.002301e+00	1.002301e+00
-4.203779e+00	-3.308051e+00
-3.922100e-01	-6.014282e-01
2.665979e-02	-8.488108e-02
3.947888e-01	6.501558e-01
4.675375e+00	3.071756e+00

Data Pre-Processing

Pre-Processing (cont.)

Splitting & Scaling (cont.)

- Chosen data frame
- Train, Test Split
- **Scaling**
  - 3 scaling approaches were tried:
    - Standard Scaling (SS)
    - MinMax Scaling (MM)
    - Log Transformation (LG)
  - MM posted poor results; thus removed

```
R² results for nothing scaled below
                    Test 0.254 ( nothing scaled )

R² results for X & y scaled below
SS Train | 0.3966   Test 0.2796
MM Train | 0.0424   Test -0.1085
LG Train | 0.4149   Test -23.8319

R² results for X only scaled below
SS Train | 0.4185   Test 0.254
MM Train | -0.2444  Test -0.0533
LG Train | 0.4142   Test -23.4693

R² results for the LG & SS combination below
SS Train | 0.4067   Test -22.811

R² averages of LG & SS X only scaled below
Av. Train | 0.4164   Test -11.6077
```

```
MAE results for nothing scaled below
                    Test 0.563 ( nothing scaled )

MAE results for X & y scaled below
SS Train | 0.5376   Test 0.684
MM Train | 0.0811   Test 0.0943
LG Train | 0.5478   Test 1.6306

MAE results for X only scaled below
SS Train | 0.4312   Test 0.563
MM Train | 0.6711   Test 0.6112
LG Train | 0.4381   Test 1.2897

MAE results for the LG & SS combination below
SS Train | 0.4377   Test 1.2751

MAE averages of LG & SS X only scaled below
Av. Train | 0.4346   Test 0.9263
```

```
MSE results for nothing scaled below
                    Test 0.7556 ( nothing scaled )

MSE results for X & y scaled below
SS Train | 0.6034   Test 0.8602
MM Train | 0.0105   Test 0.0146
LG Train | 0.5851   Test 30.4687

MSE results for X only scaled below
SS Train | 0.3727   Test 0.571
MM Train | 0.7976   Test 0.8061
LG Train | 0.3755   Test 18.7277

MSE results for the LG & SS combination below
SS Train | 0.3803   Test 18.2239

MSE averages of LG & SS X only scaled below
Av. Train | 0.3741   Test 9.6493
```

Data Pre-Processing

Pre-Processing (cont.)

Splitting & Scaling (cont.)

- Chosen data frame
- Train, Test Split
- **Scaling**
    - 3 scaling approaches were tried:
        - Standard Scaling (SS)
        - MinMax Scaling (MM)
        - Log Transformation (LG)
    - SS & LG posted the best results. These went forward & the data frame was divided into where SS & LG would be most appropriate

R^2 results for nothing scaled below
Test 0.254 (nothing scaled)

R^2 results for X & y scaled below
SS Train | 0.3966 Test 0.2796
~~MM Train | 0.0424 Test -0.1085~~
LG Train | 0.4149 Test -23.8319

R^2 results for X only scaled below
SS Train | 0.4185 Test 0.254
~~MM Train | -0.2444 Test -0.0533~~
LG Train | 0.4142 Test -23.4693

R^2 results for the LG & SS combination below
SS Train | 0.4067 Test -22.811

R^2 averages of LG & SS X only scaled below
Av. Train | 0.4164 Test -11.6077

MAE results for nothing scaled below
Test 0.563 (nothing scaled)

MAE results for X & y scaled below
SS Train | 0.5376 Test 0.684
~~MM Train | 0.0811 Test 0.0943~~
LG Train | 0.5478 Test 1.6306

MAE results for X only scaled below
SS Train | 0.4312 Test 0.563
~~MM Train | 0.6711 Test 0.6112~~
LG Train | 0.4381 Test 1.2897

MAE results for the LG & SS combination below
SS Train | 0.4377 Test 1.2751

MAE averages of LG & SS X only scaled below
Av. Train | 0.4346 Test 0.9263

MSE results for nothing scaled below
Test 0.7556 (nothing scaled)

MSE results for X & y scaled below
SS Train | 0.6034 Test 0.8602
~~MM Train | 0.0105 Test 0.0146~~
LG Train | 0.5851 Test 30.4687

MSE results for X only scaled below
SS Train | 0.3727 Test 0.571
~~MM Train | 0.7976 Test 0.8061~~
LG Train | 0.3755 Test 18.7277

MSE results for the LG & SS combination below
SS Train | 0.3803 Test 18.2239

MSE averages of LG & SS X only scaled below
Av. Train | 0.3741 Test 9.6493

Data Pre-Processing

Pre-Processing (cont.)

Splitting & Scaling (cont.)

- Chosen data frame
- Train, Test Split
- **Scaling**
  - 3 scaling approaches were tried:
    - Standard Scaling (SS)
    - MinMax Scaling (MM)
    - Log Transformation (LG)
  - SS & LG posted the best results & thus the data frame was divided into where SS & LG would be most appropriate
  - The resulting x5 Data frames went to a Random Forest Model

```
R² results for X & y scaled below
SS Train | 0.3966   Test 0.2796
LG Train | 0.4149   Test -23.8319

R² results for X only scaled below
SS Train | 0.4185   Test 0.254
LG Train | 0.4142   Test -23.4693

R² results for the LG & SS combination below
SS Train | 0.4067   Test -22.811
```

04

Model Description

The Random Forest

The Random Forest Model was then used

with the goal of determining what variables best explain & understand Inflation

05

Model Findings

What's moving Inflation¿

Model Findings

Where’s Inflation coming from?

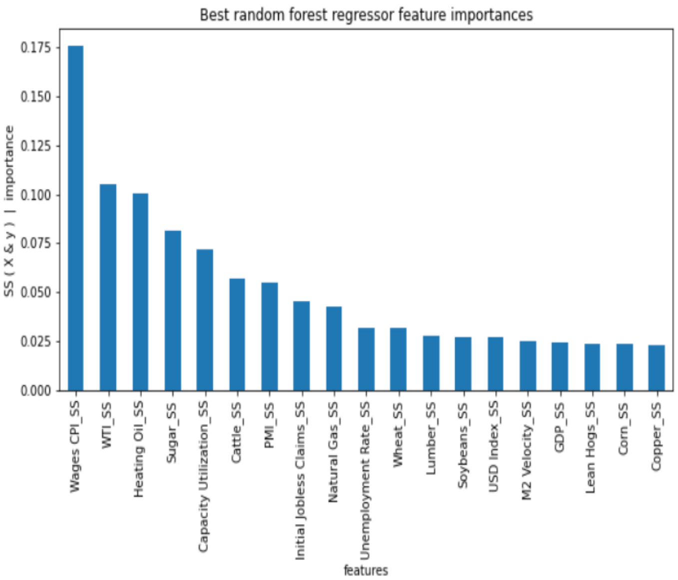
- **The standard process was taken on x5**
  - Grid Search
  - Random Forest
  - Hyperparameter search using Grid Search CV

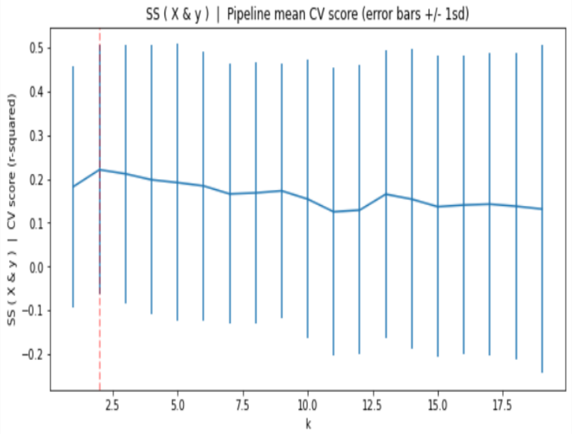
Model Findings

(cont.)

Where's Inflation coming from? (cont.)

- The standard process was taken on x5
- **The results**
  - Random Forest showed a ubiquitous 1st & 2nd place to Wages CPI & WTI respectively on all; other variables discounted the performance





Model Findings

(cont.)

Where's Inflation coming from? (cont.)

- The standard process was taken on x5
- **The results**
  - Random Forest showed a ubiquitous 1st & 2nd place to Wages CPI & WTI respectively on all; other variables discounted the performance
  - In the end, the SS approach on both X & y presented the best results with these two variables

R^2 results for X & y scaled below
SS Train | 0.2924   Test 0.424
LG Train | 0.2815   Test 0.3673

R^2 results for X only scaled below
SS Train | 0.2924   Test 0.3489
LG Train | 0.2778   Test 0.2979

R^2 results for the LG & SS combination
SS Train | 0.284   Test 0.3761

MAE results for X & y scaled below
SS Train | 0.5639   Test 0.5811
LG Train | 0.5727   Test 0.598

MAE results for X only scaled below
SS Train | 0.4515   Test 0.6127
LG Train | 0.4572   Test 0.6272

MAE results for the LG & SS combination
SS Train | 0.4572   Test 0.6272

MSE results for X & y scaled below
SS Train | 0.7076   Test 0.6877
LG Train | 0.7185   Test 0.7763

MSE results for X only scaled below
SS Train | 0.4536   Test 0.7774
LG Train | 0.4629   Test 0.8615

MSE results for the LG & SS combination
SS Train | 0.4629   Test 0.8615

Model Findings

(cont.)

Where's Inflation coming from? (cont.)

- The standard process was taken on x5
- **The results**
  - Random Forest showed a ubiquitous 1st & 2nd place to Wages CPI & WTI respectively on all; other variables discounted the performance
  - In the end, the SS approach on both X & y presented the best results with these two variables
  - & showed that the process presented notable improvement from where we started
    - After rolling averages on the unscaled 19 Variables

```
A 17.0 bps increase in R²; 66.94 % increase.

A 1.81 bps increase in MAE.

A -6.79 bps decrease in MSE.
```

Model Findings

(cont.)

Where's Inflation coming from? (cont.)

- The standard process was taken on x5
- **The results**
  - Random Forest showed a ubiquitous 1st & 2nd place to Wages CPI & WTI respectively on all; other variables discounted the performance
  - In the end, the SS approach on both X & y presented the best results
  - & showed that the process presented notable improvement from where we started
    - After rolling averages on the unscaled 19 Variables
    - After rolling averages on the SS X & y scaled 19 Variables

```
A 14.44 bps increase in R²; 51.64 % increase.

A -10.29 bps decrease in MAE.

A -17.24 bps decrease in MSE.
```

Model Findings

(cont.)

Where's Inflation coming from? (cont.)

- The standard process was taken on x5
- **The results**
  - Random Forest showed a ubiquitous 1st & 2nd place to Wages CPI & WTI respectively on all; other variables discounted the performance
  - In the end, the SS approach on both X & y presented the best results
  - & showed that the process presented notable improvement from where we started
  - So the verdict is that when you use these two variables alone you best position yourself to understand Inflation. While the Wages CPI is a component of Inflation itself, we will borrow the words to explain it on something that moves every day

The wise words of Bill Clintons' advisor to his 1992 political campaign

"It's the economy, stupid"

- James Carville

Play on words...

Our Conclusion

“It’s Oil, silly”

06

Next Steps

Keep going

Next Steps

Variables not included

- **Steel**
  - 2008 was the furthest I could pull
- Gasoline
- Growth in M2
- US Wages Hourly Earnings
- US Dollar Index: Broad, Goods & Services

Next Steps

Variables not included

- Steel
- **Gasoline**
    - 2005 was the furthest I could pull
- Growth in M2
- US Wages Hourly Earnings
- US Dollar Index: Broad, Goods & Services

Next Steps

Variables not included

- Steel
- Gasoline
- **Growth in M2**
  - Possible overlap with M2 Velocity
- US Wages Hourly Earnings
- US Dollar Index: Broad, Goods & Services

Next Steps

Variables not included

- Steel
- Gasoline
- Growth in M2
- **US Wages Hourly Earnings**
  - Limited Data as well
- US Dollar Index: Broad, Goods & Services

Next Steps

Variables not included

- Steel
- Gasoline
- Growth in M2
- US Wages Hourly Earnings
- **US Dollar Index: Broad, Goods & Services**
    - Only goes until 2006

Next Steps
(cont.)

More attention may be applicable to the below:

- **The SS & LG Divide**
  - Reassess the Variables which were chosen in the SS & LG divide; discussed in Pre-processing
- Scrape Variables
- Predict Wages CPI Itself
- Build a Better Imported / Exported USD

Next Steps
(cont.)

More attention may be applicable to the below:

- The SS & LG Divide
- **Scrape Variables**
    - Winsorizing way present better results
- Predict Wages CPI Itself
- Build a Better Imported / Exported USD

Next Steps
(cont.)

More attention may be applicable to the below:

- The SS & LG Divide
- Scrape Variables
- **Predict Wages CPI Itself**
    - Develop a model to remove ourselves from the US gov’t’s reporting
- Build a Better Imported / Exported USD

Next Steps
(cont.)

More attention may be applicable to the below:

- The SS & LG Divide
- Scrape Variables
- Predict Wages CPI Itself
- **Build a Better Imported / Exported USD**
    - **The DXY doesn't correctly address whether the US Imports or Exports Inflation** as it's weighting is a weighted geometric mean of the:
        - Eurozone (EUR),
        - Japan (JPY),
        - United Kingdom (GBP),
        - Canada (CAD),
        - Sweden (SEK) &
        - Switzerland (CHF)
    - **Doesn't take into account the US's largest trading partner, China**

Thanks

By **Rand Sobczak Jr.**
rand.sobczak@gmail.com
+1 313 447 8634

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**