Datahacks S&P 500 Prediction Competition

Dataset from Vulcan Value Partners
Track done by Kevin Jay, Manan Surana

Your Task

Use the given datasets to attempt to predict S&P 500 prices.

1. Predict S&P 500 prices using only data from the past **week**.
2. Predict S&P 500 prices using only data from the past **month**.
3. Predict S&P 500 prices using only data from the past **year**.

Instructions

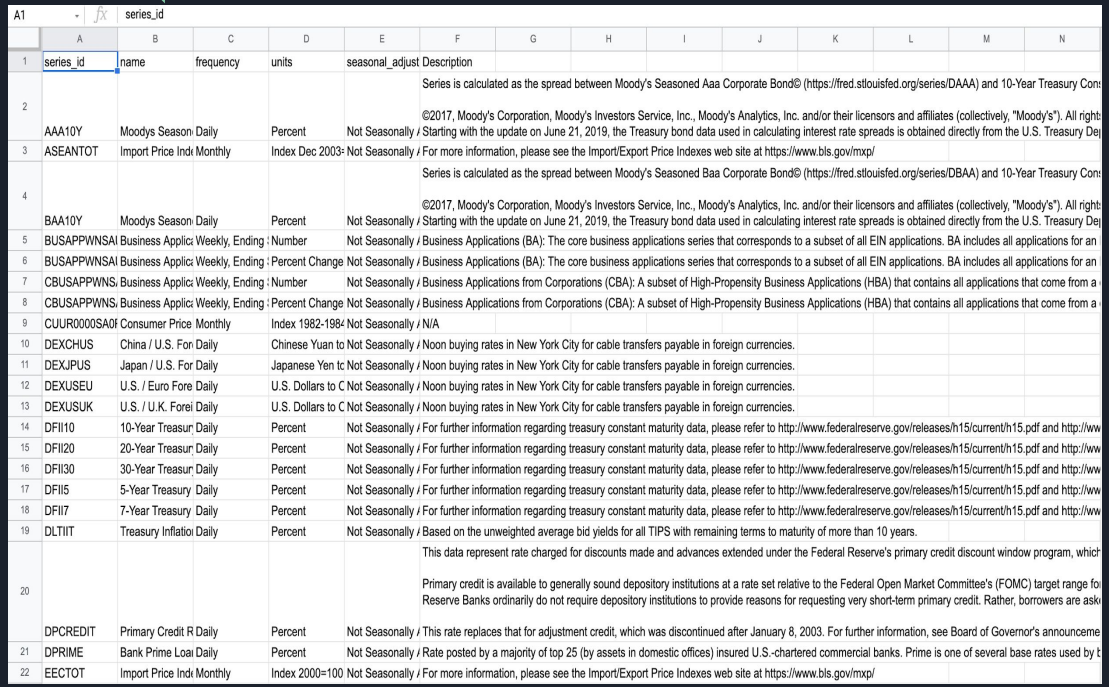
For each time frame, build a time-series model to attempt to predict pricing. You are free to use any model you deem appropriate, as long as you are not importing data from outside sources.

- Some features have a large number of missing values. Describe how you addressed this issue.
- Using your preferred visualization tool, provide several comparisons between actual prices and predicted prices.
- Which features did you find the most predictive? Which features did you find the least predictive? Provide a visualization of your most significant feature.
- Answer each question fully, as well as submit a legitimate time-series model

Instructions

- What kind of model did you choose to use? Why?
- Did you undergo hyperparameter tuning? What were the optimal values?
- What was your accuracy? Provide a visualization using your preferred tool comparing actual and predicted prices.
- What is the error on your prediction? Provide a visualization.
- Do your results differ among time frames? Give a brief explanation as to why or why not.
- What did you learn from this project?

series.csv



series_id	name	frequency	units	seasonal_adjust	Description
AAA10Y	Moodys Season	Daily	Percent	Not Seasonally A	Series is calculated as the spread between Moody's Seasoned Aaa Corporate Bond© (https://fred.stlouisfed.org/series/DAAA) and 10-Year Treasury Cons ©2017, Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "Moody's"). All right Starting with the update on June 21, 2019, the Treasury bond data used in calculating interest rate spreads is obtained directly from the U.S. Treasury Dep
ASEANTOT	Import Price Inde	Monthly	Index Dec 2003=	Not Seasonally A	For more information, please see the Import/Export Price Indexes web site at https://www.bls.gov/mxp/
BAA10Y	Moodys Season	Daily	Percent	Not Seasonally A	Series is calculated as the spread between Moody's Seasoned Baa Corporate Bond© (https://fred.stlouisfed.org/series/DBAA) and 10-Year Treasury Cons ©2017, Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "Moody's"). All right Starting with the update on June 21, 2019, the Treasury bond data used in calculating interest rate spreads is obtained directly from the U.S. Treasury Dep
BUSAPPWNSAI	Business Applic	Weekly, Ending	Number	Not Seasonally A	Business Applications (BA): The core business applications series that corresponds to a subset of all EIN applications. BA includes all applications for an
BUSAPPWNSAI	Business Applic	Weekly, Ending	Percent Change	Not Seasonally A	Business Applications (BA): The core business applications series that corresponds to a subset of all EIN applications. BA includes all applications for an
CBUSAPPWNS	Business Applic	Weekly, Ending	Number	Not Seasonally A	Business Applications from Corporations (CBA): A subset of High-Propensity Business Applications (HBA) that contains all applications that come from a
CBUSAPPWNS	Business Applic	Weekly, Ending	Percent Change	Not Seasonally A	Business Applications from Corporations (CBA): A subset of High-Propensity Business Applications (HBA) that contains all applications that come from a
CUUR0000SA0	Consumer Price	Monthly	Index 1982-1984	Not Seasonally A	N/A
DEXCHUS	China / U.S. For	Daily	Chinese Yuan to	Not Seasonally A	Noon buying rates in New York City for cable transfers payable in foreign currencies.
DEXJPUS	Japan / U.S. For	Daily	Japanese Yen to	Not Seasonally A	Noon buying rates in New York City for cable transfers payable in foreign currencies.
DEXUSEU	U.S. / Euro Fore	Daily	U.S. Dollars to C	Not Seasonally A	Noon buying rates in New York City for cable transfers payable in foreign currencies.
DEXUSUK	U.S. / U.K. Forei	Daily	U.S. Dollars to C	Not Seasonally A	Noon buying rates in New York City for cable transfers payable in foreign currencies.
DFII10	10-Year Treasur	Daily	Percent	Not Seasonally A	For further information regarding treasury constant maturity data, please refer to http://www.federalreserve.gov/releases/h15/current/h15.pdf and http://ww
DFII20	20-Year Treasur	Daily	Percent	Not Seasonally A	For further information regarding treasury constant maturity data, please refer to http://www.federalreserve.gov/releases/h15/current/h15.pdf and http://ww
DFII30	30-Year Treasur	Daily	Percent	Not Seasonally A	For further information regarding treasury constant maturity data, please refer to http://www.federalreserve.gov/releases/h15/current/h15.pdf and http://ww
DFII5	5-Year Treasury	Daily	Percent	Not Seasonally A	For further information regarding treasury constant maturity data, please refer to http://www.federalreserve.gov/releases/h15/current/h15.pdf and http://ww
DFII7	7-Year Treasury	Daily	Percent	Not Seasonally A	For further information regarding treasury constant maturity data, please refer to http://www.federalreserve.gov/releases/h15/current/h15.pdf and http://ww
DLTIIT	Treasury Inflatio	Daily	Percent	Not Seasonally A	Based on the unweighted average bid yields for all TIPS with remaining terms to maturity of more than 10 years.
DPCREDIT	Primary Credit R	Daily	Percent	Not Seasonally A	This data represent rate charged for discounts made and advances extended under the Federal Reserve's primary credit discount window program, which Primary credit is available to generally sound depository institutions at a rate set relative to the Federal Open Market Committee's (FOMC) target range fo Reserve Banks ordinarily do not require depository institutions to provide reasons for requesting very short-term primary credit. Rather, borrowers are aske
DPRIME	Bank Prime Loa	Daily	Percent	Not Seasonally A	Rate posted by a majority of top 25 (by assets in domestic offices) insured U.S.-chartered commercial banks. Prime is one of several base rates used by b
EECTOT	Import Price Inde	Monthly	Index 2000=100	Not Seasonally A	For more information, please see the Import/Export Price Indexes web site at https://www.bls.gov/mxp/

Each series has a corresponding series ID, which is important for the next dataset.

In addition, each series has:

-Name

-Frequency of collection

-Corresponding units

-Seasonal adjustment

-Brief description

observations.csv

A1 | fx | series_id

	A	B	C	D	E
1	series_id	date	value		
2	AAA10Y	2018-01-02 0:00	1.06		
3	AAA10Y	2018-01-03 0:00	1.06		
4	AAA10Y	2018-01-04 0:00	1.01		
5	AAA10Y	2018-01-05 0:00	1.03		
6	AAA10Y	2018-01-08 0:00	1.01		
7	AAA10Y	2018-01-09 0:00	1.02		
8	AAA10Y	2018-01-10 0:00	1.03		
9	AAA10Y	2018-01-11 0:00	0.99		
10	AAA10Y	2018-01-12 0:00	0.95		
11	AAA10Y	2018-01-15 0:00	NULL		
12	AAA10Y	2018-01-16 0:00	0.95		
13	AAA10Y	2018-01-17 0:00	0.93		
14	AAA10Y	2018-01-18 0:00	0.92		
15	AAA10Y	2018-01-19 0:00	0.94		
16	AAA10Y	2018-01-22 0:00	0.94		
17	AAA10Y	2018-01-23 0:00	0.94		
18	AAA10Y	2018-01-24 0:00	0.96		
19	AAA10Y	2018-01-25 0:00	0.92		
20	AAA10Y	2018-01-26 0:00	0.92		
21	AAA10Y	2018-01-29 0:00	0.89		
22	AAA10Y	2018-01-30 0:00	0.9		
23	AAA10Y	2018-01-31 0:00	0.87		
24	AAA10Y	2018-02-01 0:00	0.84		
25	AAA10Y	2018-02-02 0:00	0.89		
26	AAA10Y	2018-02-05 0:00	0.93		
27	AAA10Y	2018-02-06 0:00	0.91		
28	AAA10Y	2018-02-07 0:00	0.92		
29	AAA10Y	2018-02-08 0:00	0.94		
30	AAA10Y	2018-02-09 0:00	0.99		

Separated into 2 sets, training and testing set

Each row contains three values:

-Series ID

-Date of collection

-Corresponding value

Dataset has 24,000 rows and 68 unique features.

Keep in mind that this dataset is indexed by series_ID, so a given date will have a value for each series. It is likely you may need to group by date to do your analysis.

Questions?

Good luck!