Data Overview - Prices

Vintage	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
2019-03-31	6551.0	7100.0	7350.0	7300.0	4900.0	4950.0	4912.0	4730.0	5150.0	6024.0	4709.0
2019-04-30	6468.0	7100.0	7200.0	7270.0	4850.0	4850.0	4920.0	4800.0	5190.0	5964.0	4633.0
2019-05-31	6560.0	7200.0	7200.0	7200.0	4762.0	4850.0	4900.0	4800.0	5250.0	6050.0	4780.0
2019-06-30	6560.0	7200.0	7190.0	7300.0	4809.0	4950.0	4919.0	4824.0	5200.0	5760.0	4800.0
2019-07-31	6600.0	7200.0	7240.0	7450.0	4910.0	4950.0	4900.0	4850.0	5200.0	5950.0	4800.0

```
[(2007, Timestamp('2008-06-30 00:00:00')), (2008, Timestamp('2009-04-30 00:00:00')), (2009, Timestamp('2010-06-30 00:0
0:00')), (2010, Timestamp('2011-07-31 00:00:00')), (2011, Timestamp('2012-04-30 00:00:00')), (2012, Timestamp('2013-04
-30 00:00:00')), (2013, Timestamp('2014-04-30 00:00:00')), (2014, Timestamp('2015-05-31 00:00:00')), (2015, Timestamp(
'2016-06-30 00:00:00')), (2016, Timestamp('2017-06-30 00:00:00')), (2017, Timestamp('2018-06-30 00:00:00'))]
```

Shape 134 x 11

Data Overview - Scores

	Vintage	Lafite Rothschild	Carruades Lafite	Mouton Rothschild	Petit Mouton	Margaux	Pavillon Rouge (Margaux)	Haut Brion	Clarence Haut Brion	Latour	Forts Latour
0	2007	94	90	92	88	92	89	94	88	92	89
1	2008	98	91	95	89	94	88	96	91	95	91
2	2009	97	93	99	93	98	93	100	92	100	95
3	2010	98	94	99	93	99	94	100	93	100	97
4	2011	92	88	92	87	93	89	94	90	94	91
5	2012	91	89	96	90	96	90	96	90	-	92
6	2013	90	78	92	89	91	87	92	87	-	91
7	2014	95	88	95	92	95	90	96	91	-	-
8	2015	96	91	98	93	99	92	100	93	-	-
9	2016	99	92	100	93	99	93	100	93	-	-
10	2017	98	91	98	90	97	92	96	90	-	-

The Task

- Analysis of past key trends based on the statistical evidences drawn from the data
- Analysis of the price behaviour of these wines
- Conclusion about how to extrapolate future price evolution

Potential Target Audience

1. **Investors**

   Which wine will have appreciate in the future ?

   How to diversify investment portfolios?

Predictive Model	Decision Support	Pricing Strategies

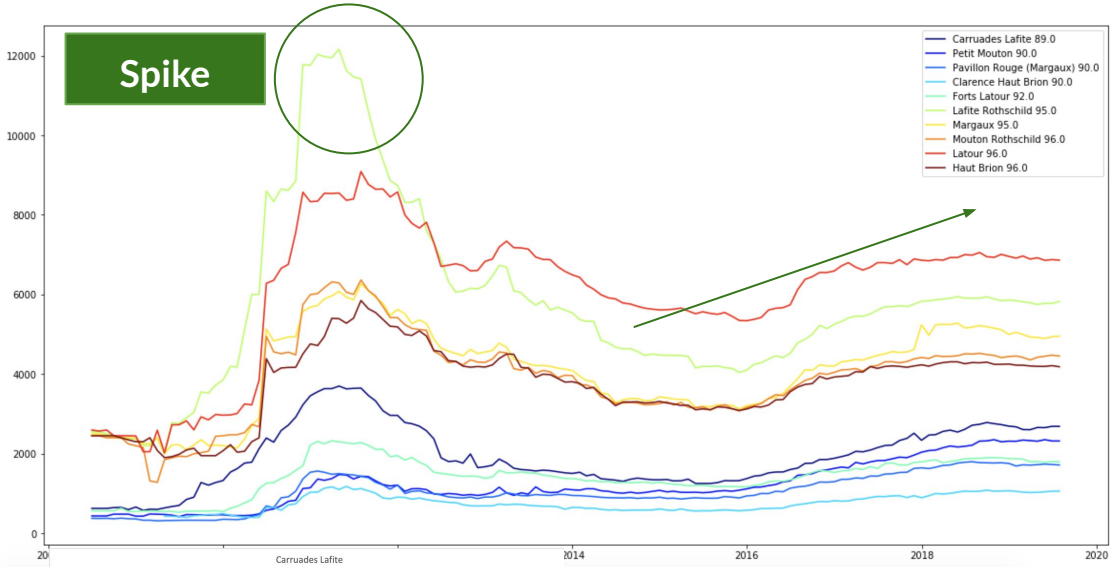
2. **Marketing Team**

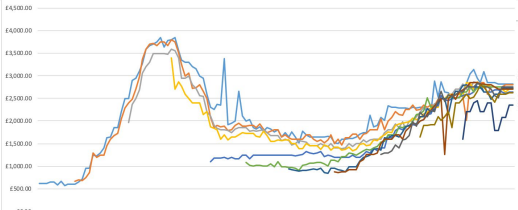
   Identify customers groups with targeted marketing campaigns

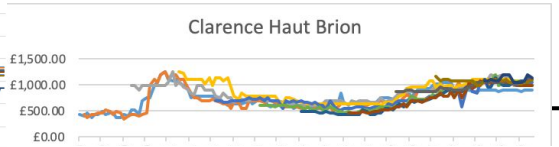
Planning and Forecasting Tools

Descriptive Analysis

Trend



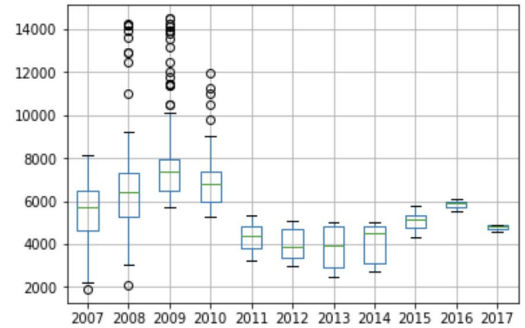




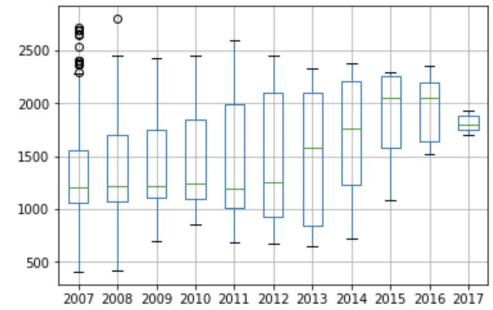
- First spike happened in 2011. The price went down rapidly especially for [Lafite Rothschild].
- For each wine, the volatility is different. The risk can be mitigated by having well-diversified portfolios.
- The spike can be treated as a market bubble and huge correction happened between 2012 and 2015.
- Prices are on the upward trend now.

Trend

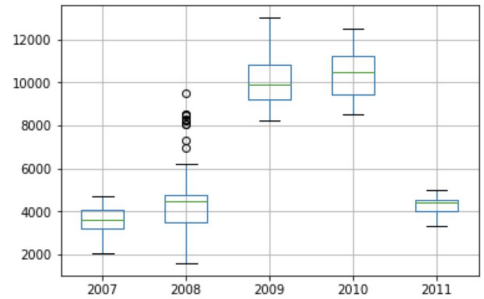
Lafite Rothschild



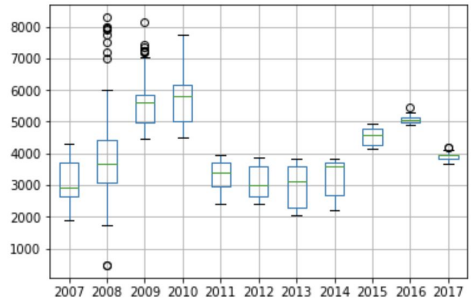
Petit Mouton



Latour

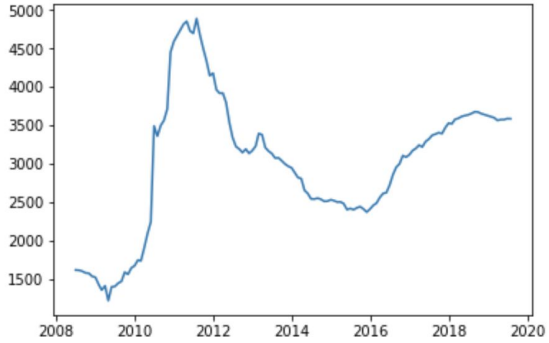


Mouton Rothschild



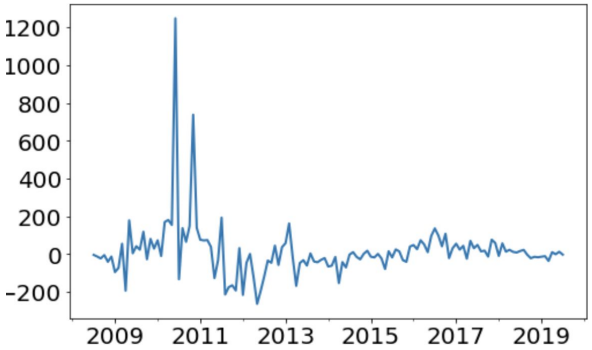
- It is easier to spot outliers before 2011.

Wine Market Trend



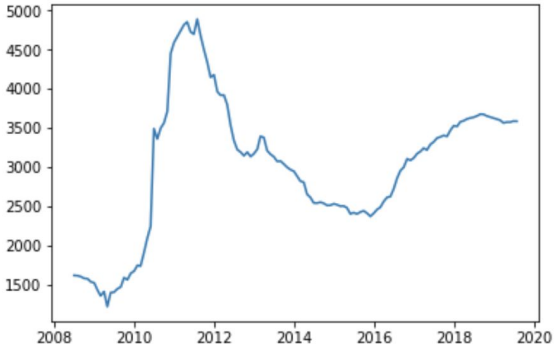
Seasonality

Using first-order differences: Removing the trend may reveal correlation in seasonality.
-> No strong sign of seasonality.



Cyclicity

Spike on 2011 and current highest point on 2018, both of them cannot be identified as a cycle yet.

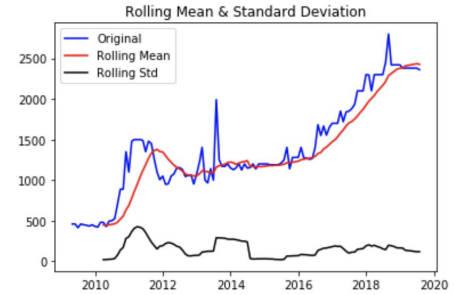


Stationary

Using Dickey-Fuller Test:

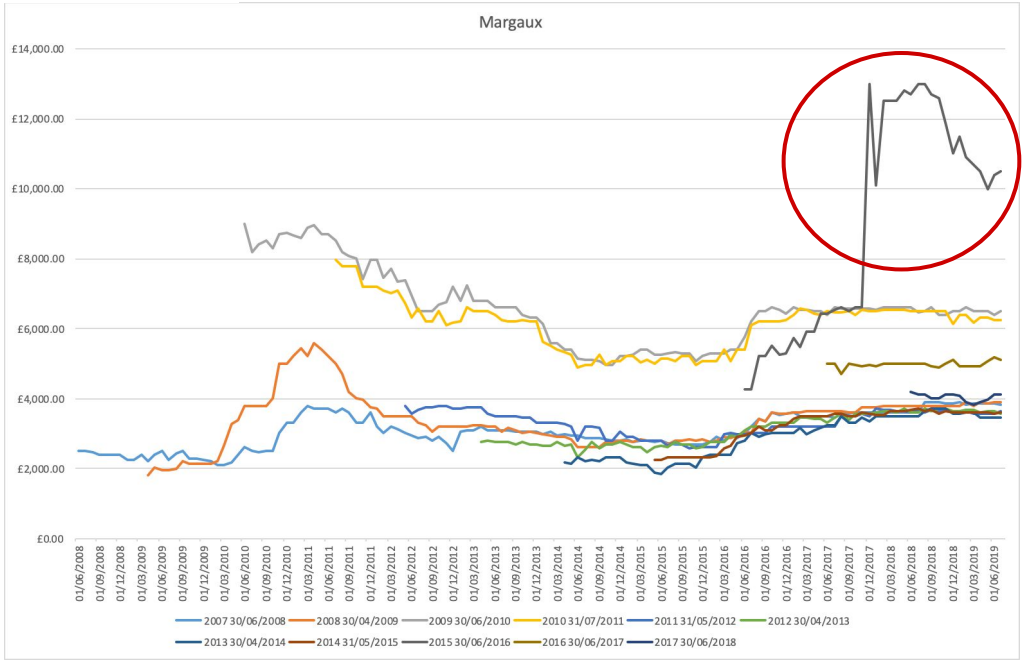
-> Generally they are not stationary

```
test_stationarity(wine_list[3][2008].dropna())
```




```
Results of Dickey-Fuller Test:
Test Statistic                  -0.930695
p-value                          0.777696
```

Irregularity - Marguax 2015 Score: 99



Question 1:

Is it correctly priced by its given score?

Question 2 :

Is it align with the trend ?

Conclusion:

Is it a bubble?

Irregularity Cont.

Ans1: Negative

Score 99 Maximum Price:

Margaux 2015 - £13000
Margaux 2010 - £7962
Margaux 2016 - £5168
Mouton Rothschild 2009- £ 8150 Mouton Rothschild 2010- £7749

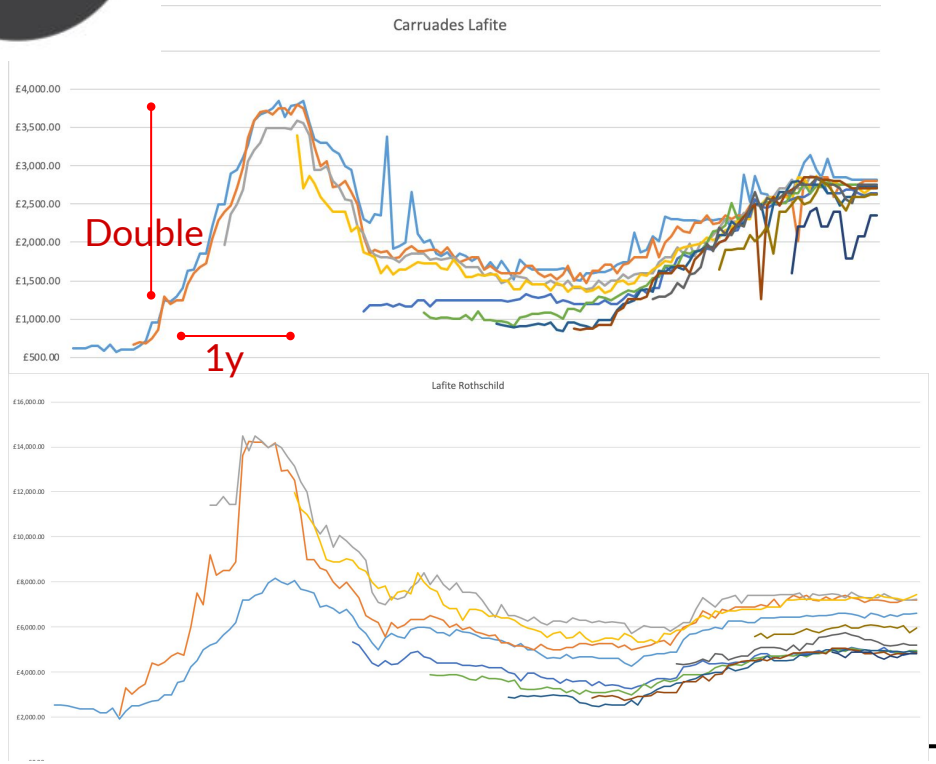
Ans2 : Negative

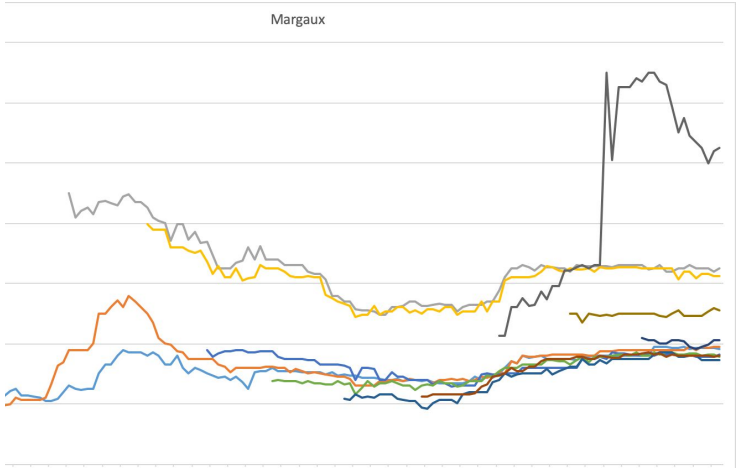
High Standard Deviation - 3170

Does not follow the trend.

Vintage	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
count	134.000000	124.000000	110.000000	97.000000	87.000000	76.000000	64.000000	51.000000	38.000000	26.000000	14.000000
mean	3056.820896	3370.129032	6597.118182	6109.556701	3300.264368	3093.723684	2881.953125	3212.549020	8907.157895	4977.230769	4033.571429
std	508.192431	762.815356	1076.754959	734.194059	367.409804	437.602334	604.199784	518.195419	3170.060836	90.068111	103.795445
min	2088.000000	1798.000000	4950.000000	4869.000000	2560.000000	2334.000000	1851.000000	2236.000000	4260.000000	4684.000000	3838.000000
25%	2700.000000	2820.000000	5626.250000	5400.000000	3000.000000	2691.000000	2293.000000	2925.000000	5899.250000	4927.750000	3974.750000
50%	3004.500000	3330.000000	6541.000000	6250.000000	3300.000000	3025.000000	3000.000000	3500.000000	10046.500000	5000.000000	4086.000000
75%	3560.000000	3800.000000	6780.000000	6500.000000	3650.000000	3571.000000	3477.500000	3600.000000	12350.000000	5000.000000	4100.000000
max	3900.000000	5580.000000	9000.000000	7962.000000	3800.000000	3700.000000	3700.000000	3700.000000	13000.000000	5168.000000	4200.000000

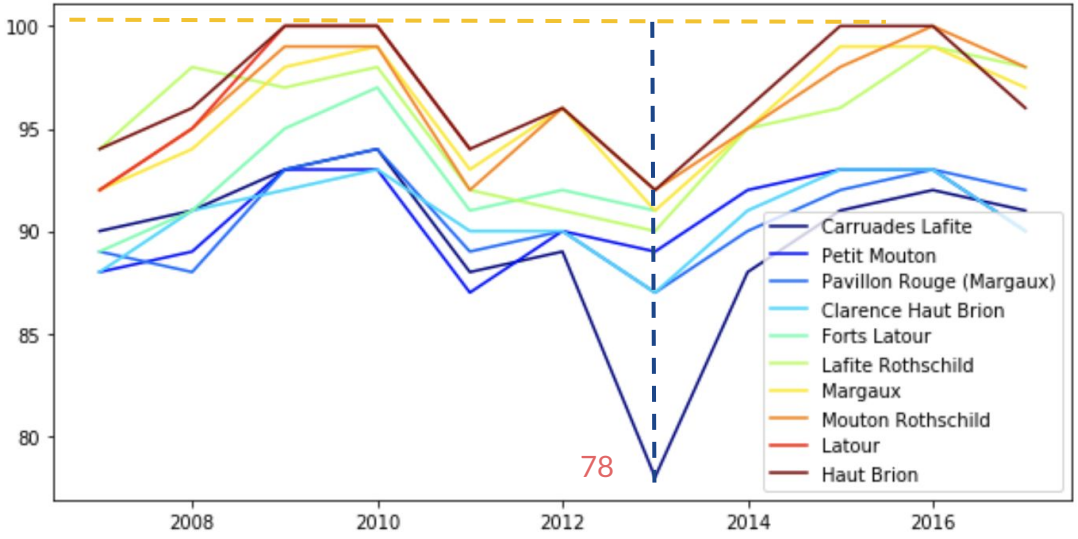
Bubble Detection





It is very suspicious….

Vintage Scores for Each Wine

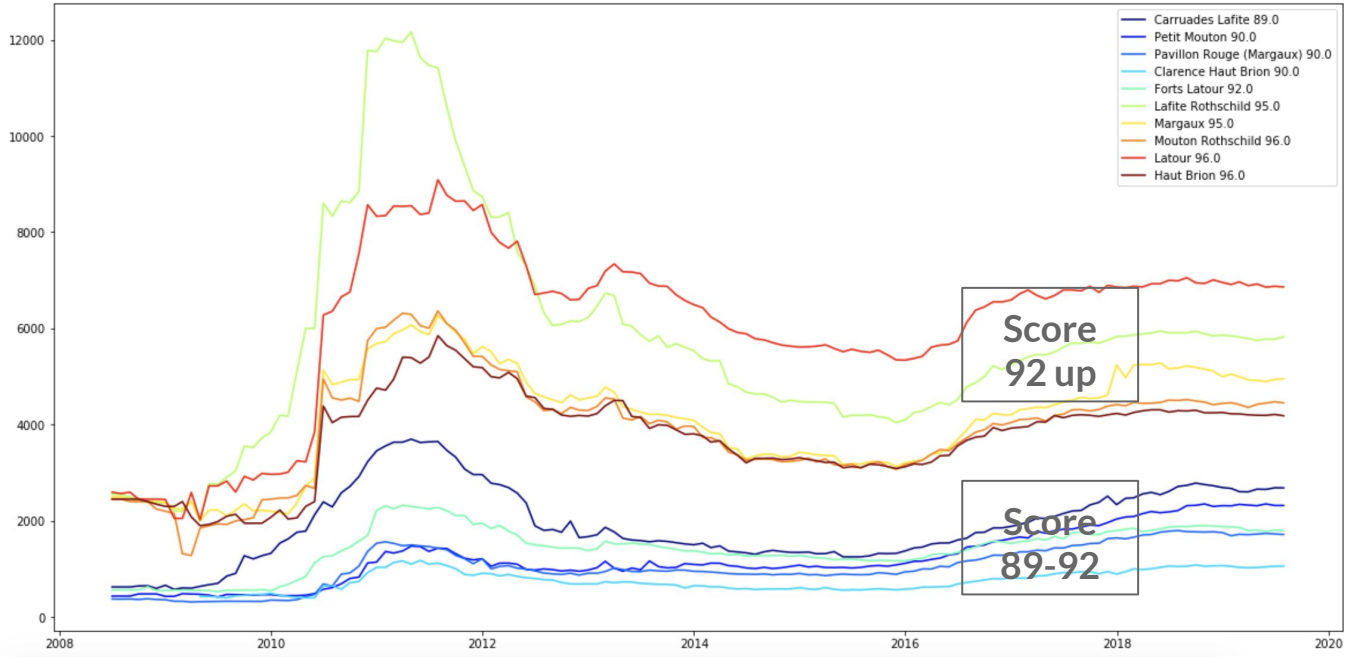


Haut Brion generally has very high scores. (100 - 4X)

Carruades Lafite has a very low score at 78

Carruades Lafite	89.5
Petit Mouton	90.6
Pavillon Rouge (Margaux)	90.6
Clarence Haut Brion	90.7
Forts Latour	92.3
Lafite Rothschild	95.3
Margaux	95.7
Mouton Rothschild	96.0
Latour	96.2
Haut Brion	96.7

Average Prices for Wines



Scores for wines indicate trends for different kind of wine.

Vintage 2013 from each wine

Name, score, starting price, wine_avg

```
[('Carruades Lafite', 78.0, 948.0, 2000.74),
 ('Clarence Haut Brion', 87.0, 499.0, 740.44),
 ('Forts Latour', 91.0, 1485.0, 1292.53),
 ('Haut Brion', 92.0, 2395.0, 3682.43),
 ('Lafite Rothschild', 90.0, 2896.0, 6513.12),
 ('Margaux', 91.0, 2157.0, 4036.97),
 ('Mouton Rothschild', 92.0, 2360.0, 3925.25),
 ('Pavillon Rouge (Margaux)', 87.0, 719.0, 799.22),
 ('Petit Mouton', 89.0, 738.0, 829.74)]
```

*wine-avg :
cat_ave[name].loc[:'2013-02-28']

Question 1:

Set starting price using scores?

Correlation: 0.57

Question 1:

Set starting price using wine_avg?

Correlation: 0.94

Regression Prediction

Features:

- Price average
- Scores

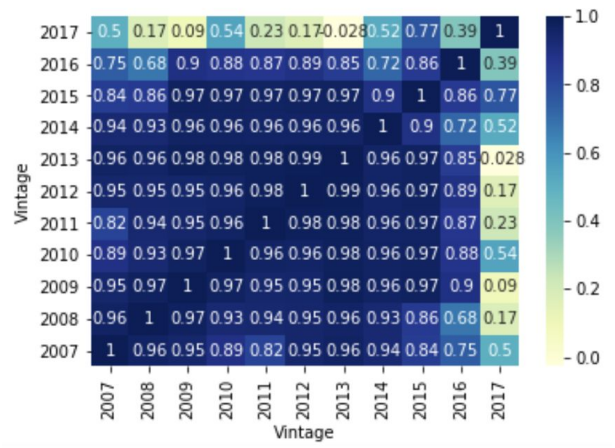
Label:

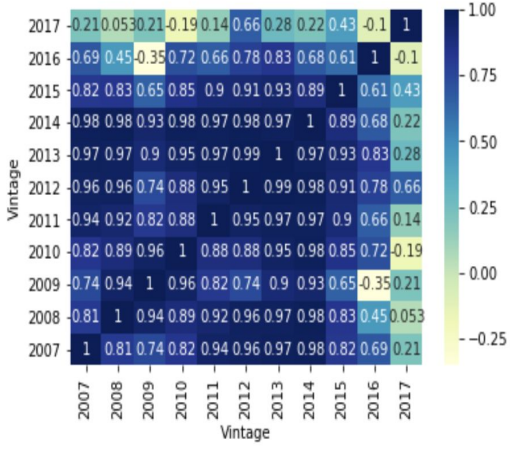
- Starting Price

Models:

- LinearRegression
- LogisticRegression
- DecisionTreeRegressor

Diversified the Portfolios





Avoid the wines with high correlation.

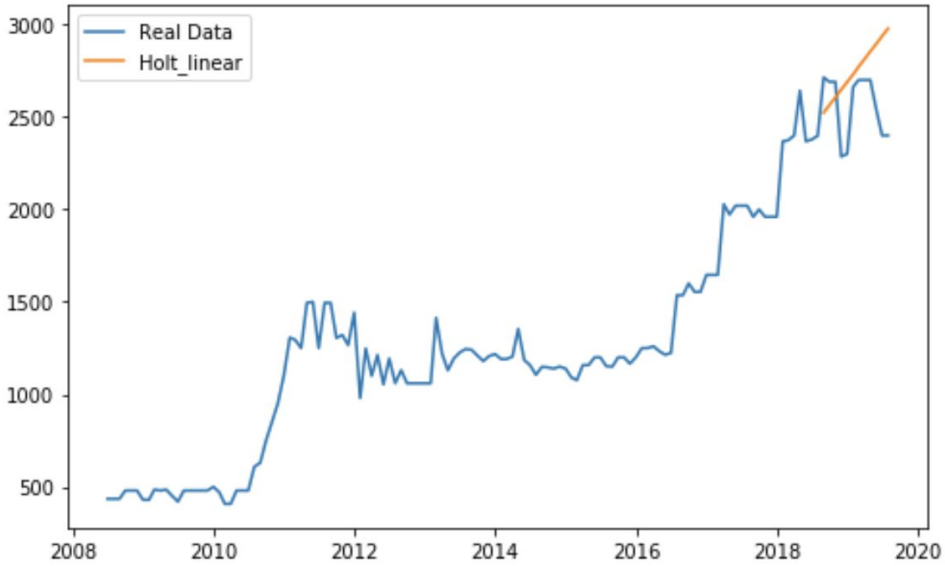
Diversified the portfolios by picking wines from different scores.

Predictive Modeling

Methodology

- Transposing data to the time series format. Reset data type to datetime
- Model checking with cross validation
- Hyperparameter Tuning with Grid Search
- Perform different models on time series data
    - Skip naive methods like [Single Exponential smoothing] or [MA]

Holt's Linear Trend



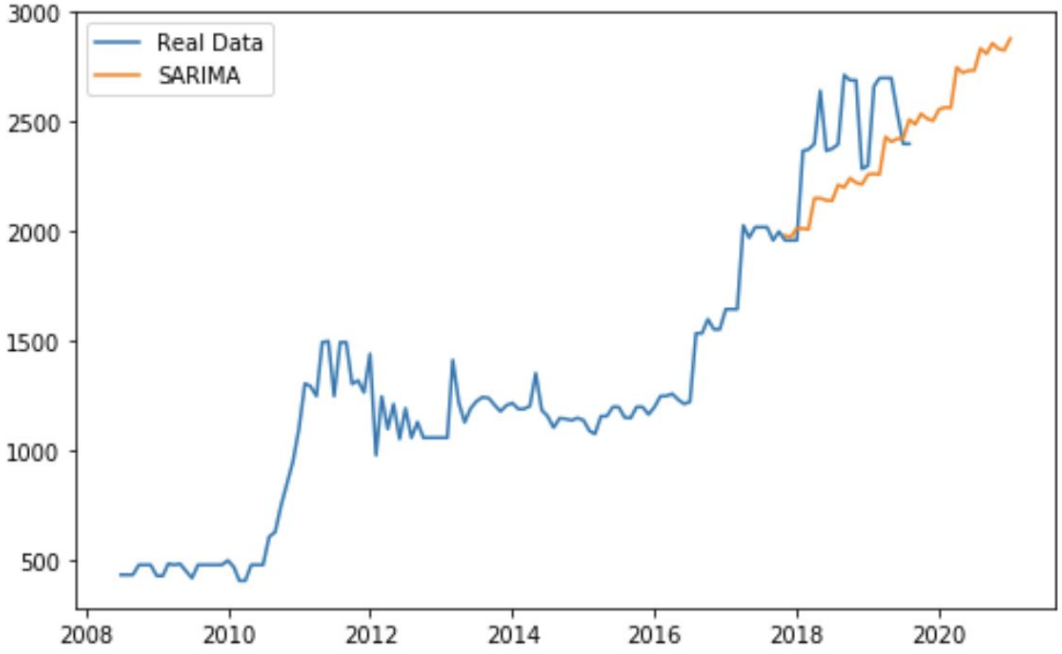
Forecast + Level + Trend

smoothing_level=0.3, smoothing_slope=0.15

- The data doesn't show strong seasonal trend. Holt-Winters Method will not be recommended here.

SARIMA

Seasonal Auto-Regressive Integrated Moving Average



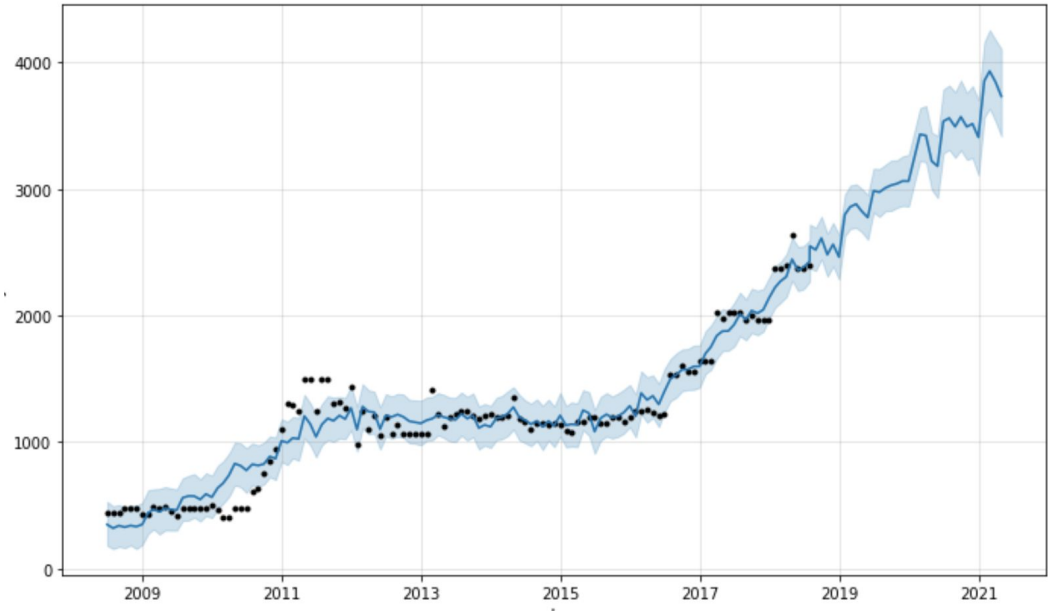
p is the number of autoregressive terms (AR part). It allows to incorporate the effect of past values into our model.

d is the number of nonseasonal differences needed for stationarity.

q is the number of lagged forecast errors in the prediction equation

- SARIMA is capturing the trend upward.

FBProphet



logistic + piecewise linear function

- It captures the upward trend and provide the prices for upcoming 3 months. The blue area indicates the confidence interval (80%).

Model Selection and Tuning

- Grid Search: SARIMA model takes 3 parameters (p,d,q)
  - Initiate q = d = range(0, 2) and p = range(0, 4)
  - Not Exhaustive. Require more search space to improve the model.
- Holt Linear: Try different Smoothing Level and slope
- FBprophet: `changepoint_prior_scale`

Model	RMSE
Holt	307.724
SARIMA	319.261
FBprophet	264.735

Coming Up Next

- Analysis of correlation with S&P. Nasdaq. Gold. USD. Crypto Currencies
- Portfolios Generators
- Online analysis

Improving Data Collection

- !! More Data !! Auction prices/Trading amount
- Performing NLP on the wine description
- Adding new features
- Adjusting models by experts