Master's Thesis

Prediction Methodologies for Undervalued Assets: Finding the Next Big Drink

Juliet Rose Shi Black and Ma. Theresa Pareno

Prediction Methodologies for Undervalued Assets: Finding the Next Big Drink

Master's Thesis

at Frankfurt School of Finance & Management

Supervised by

Prof. Gianluigi Giustiziero

Prof. Levente Szabados

Submitted by

Juliet Rose Shi Black and Ma. Theresa Pareno

Master of Applied Data Science, Program Code: 1821

Student ID's: 8400906 and 8406220

Wiesenstrasse 25 and Hauser Gasse 12

+12088302348 and +491783146223

julietblack@u.boisestate.edu and rezangpareno@gmail.com

Frankfurt am Main, June 2020

Table of Contents

Abstract

This paper explores the impact of climate change upon existing viticultural regions and wine quality. It includes a study of the literature about the wine scoring process and climate change. More importantly, this paper provides a methodology to predict when undervalued assets might appear, while also offering a solution to the uncertainty around the effects climate change will have on wine regions. The undervalued assets being viticulture regions and grape varieties. This research will benefit wine investors and vineyard owners. Our model was able to predict the temperatures in 2021 for regions in France, Italy, and Spain with an overall RMSE of 1.4 degrees Celsius.

Keywords: climate change, viticultural regions, wine quality, wine investments, wine appreciation, prediction models, machine learning, deep learning, temperature prediction

Introduction

Motivation: Climate Change

Understanding climate change and its environmental impacts has become significantly important as uncertain levels of greenhouse gases bring about complex shifts that drastically affect our planet. Global warming, or the rising of Earth's average temperature, is the major impact associated with climate change (Richards & Melford, 2019). Research over the past fifty years has revealed that the average global temperature has increased at the fastest rate in recorded history (MacMillan, 2020). This has led to catastrophic consequences such as severe weather developments, shifting wildlife populations and habitats, rising sea levels, melting glaciers, and various other impacts (Richards & Melford, 2019). Needless to say, climate change has clearly altered human-based systems as well.

These observed temperature increases influence agricultural production viability due to changes in winter hardening potential, frost occurrence, and growing season periods and lengths (Jones, Storchmann, & White, 2005). Jones, Storchmann, and White state in their

paper, “The importance of understanding climate change impacts on agriculture is especially evident with viticulture (the science of the cultivation of grapevines)” (2005). Historical data shows that distinct viticulture regions in Mediterranean climates have produced the finest wines (Jones, Storchmann, & White, 2005).

Climate and weather in these regions tremendously influence the production of quality grapes and thus high-quality wine. In general, there exists a baseline climate for various grapes and overall wine style, but vintage-to-vintage quality differences are determined from climate variability (Jones & Hellman, 2003). While there are a multitude of individual weather and climate elements that can affect the quality of wine and grape growth (e.g., solar radiation, precipitation, and wind), the duration of the growing season and temperatures are the most critical aspects (Jones G. V., 2003). This is because of their considerable effect on the ability to ripen grapes to optimum levels of acid, flavor, and sugar (Jones, Storchmann, & White, 2005). Each grape variety will have a “sweet spot” that produces grapes with these optimum levels. Part of our approach will be using these sweet spots to find undervalued grape varieties and regions.

Climate change is attracting attention in portfolio management and asset pricing research as well (Diaz-Rainey, Robertson, & Wilson, 2017). Agriculture may be especially affected, however. Even though climate change found its way into financial economics, the topic remains underrepresented and seemingly underappreciated, given its vast policy and business implications (Diaz-Rainey, Robertson, & Wilson, 2017). Some negative side effects include more underwriting fees and long term municipal bonds (Painter, 2020). The threat of climate change also adds a new type of risk for investors in the stock market. There is even a newly released climate index, Actuaries Climate Index (ACI) (Jiang, 2020). The ACI is used as proxies for climate change risk (Jiang, 2020).

Motivation: Wine Investments

High-quality wine can be bought, stored, aged, and then resold in hope for price appreciation. Consequently, this has led to people viewing wine as financial investments. According to Lee W. Sanning and Sherrill Shaffer (2008), “Wine possesses characteristics that allow it to be considered and analyzed as investment vehicles, with some wine even selling for 15 million US dollars” . With appreciation spikes being in the millions in some exceptional cases, it is important to understand how climate change will influence overall wine quality and how this will benefit or adversely affect individual wine regions.

Climate change effects on viticulture have been and are likely to continue to be highly variable geographically (Jones, Storchmann, & White, 2005). Maracchi, Sirotenko, and Bindi’s state that, “Early spatial modeling research has indicated possible geographical shifts and/or expansion of viticultural regions with parts of southern Europe becoming too hot to produce high-quality wines and northern regions becoming viable once again” (2018). This thesis will provide an approach to predict when these viticultural regions will occur by using machine learning to create prediction models based on global temperature and wine quality data. In other words, we would like to be able to predict “The next big drink”. Our research will serve as an exploratory study that will benefit wine investors and owners of wineries, by providing them with the methodology to study the effects climate change will have on their land and grapes.

Background

Wine Quality and Scoring

How it Works

One factor that plays into creating an expensive wine is determined from the score it receives from an expert or wine publications (such as *Wine & Spirits Magazine, Wine Spectator, Wine Advocate, Decanter Magazine, Vinous, The Wine Cellar*, and *The Wine Cellar Insider*), and therefore a subjective quality score. Along with the wine score, there will also be reviews or tasting notes. Experts, like Robert Parker, Allen Meadows, and John

Gilman, became professional “wine tasters” by gathering extensive knowledge about the various wines in existence, wine production, and how to properly consume wine. Furthermore, Robert Parker is famous for creating the 100-point wine rating system.

According to this system, wine can be scored on a scale from 50 to 100, with 50 being the lowest possible evaluation and 100 being the best (see Appendix A). It is based on the American high-school grading system, which is why the scale does not start at 0 (Godden & Gawel, 2008). This infers a rough correlation for receiving A-F wine grades. However, even though the 100-point system is by far the most commonly used, there exist other evaluations like the 20-point scale and the 5-point scale (see Appendix B & C). We are using the 100-point wine rating system in this thesis.

Regardless of which point system one chooses to use, all experts follow a standardized procedure when evaluating a wine. The methodology is as follows: blind tasting, scorecard metrics, and score calculation. The blind tasting requires the critic to blindly sample the wine. This is done through the removal of bottle labels and the retail price. The scorecard contains a series of criteria and traits to analyze while tasting that will later be assigned a numeric score. Generally speaking, the traits to be judged can be appearance, consistency, nose (smell), taste, and complexity. A maximum value of 5 points can be given for the appearance, 15 points for the smell (nose), 20 points for the taste, and another 10 points for the overall impression (complexity and consistency) (Auvimar, 2017).

Appearance includes, but is not limited to, core and rim, color, viscosity, and opacity. Consistency, or the wine’s “mouthfeel”, is a wide-ranging term that identifies the wine’s body and density. The nose refers to the bouquet and aroma of a wine. Taste is one of the most defining features when rating a wine (Godden & Gawel, 2008). A few definitive tasting metrics found on the majority of scoring cards are acidity, dryness, sweetness, primary flavors, intensity, balance, depth, and aftertaste. If a wine tasting experience is described as dynamic, then the complexity is being measured (Godden & Gawel, 2008). Complexity is

one of the final scoring metrics used to define the overall impression the wine leaves the taster with. However, it is important to note that this is not a granular list of traits, but instead a high-level overview of traits scored.

The last step, the score calculation, has two different methods for calculating the final score: single ratings and average ratings. For example, a store may choose between exclusively advertising a single review from a publication or professional, or the store can calculate an average score from multiple publications or professionals and advertise the weighted total.

Research

Our modelling relies heavily on the assumption that wine quality correlates with wine prices. Due to this being a key assumption, we have provided thorough research on the subject.

Study 1

In a study done by Jean-Marie Cardebat and Jean-Marc Figuet, they discovered a positive relationship between wine pricing and wine scoring (2004). They anonymously bought 254 wines from twenty six Bordeaux appellations (Cardebat & Figuet, 2004). Six juries of five to six experts (brokers, wine waiters, and oenologists) then blind-tasted these wines (Cardebat & Figuet, 2004). Comments were solicited in regards to the bouquet and taste for each wine (Cardebat & Figuet, 2004). The experts were then asked to give scores for the quality of the wine (Cardebat & Figuet, 2004).

In order to find the correlation between price and quality, Cardebat and Figuet had to measure the contribution of the characteristics of the wine to its price (Cardebat & Figuet, 2004). This was done by regressing the price of the wines on their attributes. The study determined not only the influence on wine prices of sensory factors, but also the impact of the objective variables detectable before purchase. In other words, their research helped discover the value that consumers attach to the innate quality of a wine. To further prove

their findings, they noticed the upward flexibility of prices for non-ranked wines to be severely limited. The opposite was observed for ranked wines.

Study 2

Marco Costanigro, Jill J. McCluskey, and Ron C. Mittelhammertook (2007) took a different approach to confirm the relationship between price and quality. Their method required the wine market to be segmented by price and then have a hedonic regression applied to the segmented data. The goal of this research was to prove the implicit assumption that there lies a positive relationship between wine attributes and price. The dataset was composed of California and Washington red wines. This included 13,024 rating scores that were collected over 10 years (1991–2000) from the *Wine Spectator* magazine (Costanigro, McCluskey , & Mittelhammer, 2007).

The resulting segments were: commercial wines (price less than $13), semi-premium (between $13 and $21), premium (between $21 and $40) and ultra-premium ($40) (Costanigro, McCluskey, & Mittelhammer, 2007). Empirical results showed for all estimated models that price increased with increasing rating scores over the range of the data. Costanigro, McCluskey, and Mittelhammer confirmed numerous previously published results. Another key result of their research was that they found that segmentation models have greater capability explaining the variability of the data, and produced more tenable and descriptive estimates of the hedonic link between prices and wine quality. This finding is particularly valuable given our similar method of clustering.

Study 3

The last wine score study we will present was done more recently (2018) by The Global Wine Score (GWS). Their research compared various factors that are already known to affect wine prices. The key variables were Age, Vintage, Color, AVA, Grape Variety (or blend type if a blend), and Global Wine Score (The Global Wine Score, 2018). The dataset contained 851 Global Wine Scores from Napa Valley from the vintages 2006–2015 (The

Global Wine Score, 2018). They developed an equation for the price of wine using a multiple linear regression model.

The GWS's results identified the wine score as the most impactful and significant variable when determining a wine's price. Their model also showed the effects were higher if the score was higher when the wines were separated into two groups based on points. It is important to note that the Global Wine Score combines the scores of many critics. Therefore, the results from the study shows how the consensus of expert opinions impacts wine prices.

Problems

Argument 1

There are countless studies that support our methodology and research. However, there exists arguments as well. The biggest conflict with our research is the validity of the wine scoring process. Robert T. Hodgson went out to explore this very problem with his study, "An Examination of Judge Reliability at a major U.S. Wine Competition".

From 2005 to 2008, Hodgson analyzed wine judges at a premier wine competition in the United States (Hodgson, 2008). Flights of thirty wines embedded with samples poured from the same bottle were distributed to panels of four expert judges, resulting in sixty five to seventy judges being tested each year (Hodgson, 2008). The goal of Hodgeson's (2008) research was to answer these four questions:

1) Why does a particular wine win a gold medal at one competition and fails to win any award at another?
2) Is this caused by bottle-to-bottle variability of the wine?
3) To what extent is the variability caused by differing opinions within a panel of judges?
4) Could the variability be caused by inability of individual judges to reproduce their scores?

Even though the experiment was designed to maximize the likelihood in favor of the judges' ability to reproduce their scores, the end results showed that only about 10% of the judges were able to replicate their scores (Hodgson, 2008). Please note that the results were divided into two categories: group (panel) performance and individual judge performance. Another intriguing finding revealed that judges tend to be more consistent in what they do not like than what they do. Overall, the study indicated that less than half of the panels presented awards based exclusively on wine quality (Hodgson, 2008).

Argument 2

Along with the issue of expert reliability, there also lies a problem known as the "reputation effect". A significant amount of research infers that there is a stronger positive correlation between a wine's price and a wine's reputation, rather than price and quality (wine score). Bith-Hong Ling & Larry Lockshin conducted research in Australia comparing the contribution that brand reputation, wine quality, region, vintage, and winery size have on wine prices (2003).

A hedonic price analysis was used to measure the contribution to the price for various wines. The four major wine varieties were shiraz, cabernet, chardonnay, and riesling (Ling & Lockshin, 2003). There were a total of 1880 observations of bottled wines (Ling & Lockshin, 2003). Their conclusion was that brand reputation has a dominant effect on wine prices, whereas vintage has the most minor effect. However, like many other studies, wine quality was close behind brand reputation.

Our thesis highly depends on the premises that wine quality is the most significant independent variable when predicting price. The research also relies on the validity of these wine scores. These arguments that we presented are the most common and bring some concern, but do not completely make our research invalid. While many have found reputation to be the biggest price driver, wine quality is almost always the second biggest factor. We are confident that solely using wine quality and age will be sufficient for our

models. Furthermore, despite wine experts' reviews being put into question, this is the only system existing to measure wine quality. At the moment, there still lies a positive correlation between a wine's price and it's score regardless of how accurate the score is or trustworthy the expert is. Until a more reliable system is created, this research will stand.

Climate Change

Research

> Eric Asimov, writer for the New York Times, says,
>
> "Climate change will inevitably transform the way the world produces goods. Farmers who produce wine grapes, an especially sensitive crop, are already feeling those effects." (2019)

According to Asimov, wine is among the most sensitive agricultural commodities when it comes to climate (2019). The environmental impact has transformed local economies in places like England. Producing exceptional wine has also become easier in well-established regions like Germany and Burgundy. That is to say, the wine map is expanding. So how fast can we expect to see established areas further grow (or plummet) and new viticulture regions develop?

As mentioned previously, good viticulture regions have sweet spot temperatures for specific grapes. A study done by Aysun Sener, Ahmet Canbas, and M. Ümit Ünal proved temperature to be the driving factor for optimal wine fermentation (2006). This was because of how complex the biochemical process is for turning grape juice into wine. The process begins when yeasts utilize sugars and other components of grape juice as substrates for their growth. The yeast will then convert those constituents into ethanol, carbon dioxide, and other metabolic end products. This entire process contributes to the chemical composition and sensory quality of the wine. Since yeast is such a crucial element in the fermentation process, it is no surprise why temperature plays such a vital role given yeast's sensitivity to temperature.

Therefore, one can assume that having the ideal, or sweet spot, temperature leads to higher quality and more expensive wines. The study also determined ranges for the desired sweet spot temperatures depending on the grape type. Between 18 and 25°C was recommended for farmers growing white wine grapes, and between 29 and 32°C was the ideal temperature for red wine grapes (Sener, Canbas, & Ünal, 2006).

Methodology

Data Collection

Our research consists of mainly 2 datasets from different sources. The first dataset is the Wine Review dataset that was retrieved from Kaggle.com. Kaggle is a data platform that manages data competitions and hackathons that has been acquired by Google (Lardinois, Mannes, & Lynley, 2017). The Wine Review dataset was scraped by Zack Thoutt from WineEnthusiast last November 22, 2017.

The second dataset used is the Climate Change dataset which we scraped from NASA POWER. The POWER meteorology data that we gathered was collected by NASA from NASA's GMAO MERRA-2 assimilation model and GEOS 5.12.4 FP-IT (NASA, 2019). The POWER solar data is from NASA's GEWEX/SRB archive and numerous versions of NASA's CERES FLASHFlux project (NASA, 2019). Simply put, for both meteorology and solar data, NASA, through its Earth Science research program, utilized satellite systems to study and collect global climate and climate processes information.

Tools

A variety of open source tools were used to assist in the data collection, analysis, and modelling processes. All coding was performed with Python within Jupyter Notebooks. The plotting and visualizations used the Seaborn and Matplotlib libraries, DataFrame construction required the NumPy and Pandas libraries, statistical modelling was done with Statsmodels API, latitude and longitude coordinate data was converted with GeoPy and the

NASA POWER API, and modeling required Sklearn. PowerBI Desktop was also utilized for more in depth analysis and visualization.

While the majority of the tools mentioned above are well known, there are, however, two that need further explanation: Statsmodels API and GeoPy. In 2009, at the Google Summer Code, statsmodel was invented and released (Perktold, Seabold, & Skipper, 2010). However, the package was originally created as a part of scipy. Statsmodels API is a Python module that provides functions and classes for the assessment of quite a few different statistical models. The module can also be used for conducting statistical tests and statistical data exploration. Statsmodels creates its models by utilizing R-style formulas and Pandas dataframes. The bulk of our statistical modeling used statsmodels' OLS regressions.

GeoPy is a Python 2 and 3 service for a considerable amount of popular geocoding web providers. It includes geocoder classes for Google Maps, ArcGIS, AzureMaps, Bing, and more. Geocoding is the computational development of altering a physical address depiction (home address, city, region, country, etc.) to a location on the Earth's surface (Abdishakur, 2019). In other words, geographical representations are transformed into numerical coordinates. Our regional data for the wine needed to match up with our Climate Change data. The Climate Change data contained coordinates instead of regional location. Therefore, the Wine Review dataset was put through Geopy so both datasets would contain longitude and latitude data.

DataFrames

Three DataFrames were created from the Wine Review and Climate Change datasets. There is one cleaned[1] dataframe for each dataset and an additional DataFrame that contains "Region", "Country", "Grape_Variety", "Coefficient_Points", "Coefficient_Years",

[1] Cleaned refers to removing null values, removing countries and regions with little data, fixing spelling errors, and filtering out specific data points (e.g. rows that contain Classification "depreciation").

“P-value_Points”, “P-value_Years”, “R”, “R-Squared”, “Wine_Score”, and “Classification”[2] (see Appendix D,E, and F).

The cleaned Wine Review dataframe (wine_review_2) contains thirteen columns and 38,459 rows of wine data. Wine_Review_2 has information about the country a wine was produced, wine reviews, a wine’s designation, the points a wine received, a wine’s provincial location, the region a wine was from, taster (judges) name’s, the Twitter handle for each of the tasters, a wine’s title, the variety description of a wine, the winery where a wine was made, and the production year. This particular DataFrame was used for exploratory data analysis (EDA) and modelling.

The cleaned Climate Change dataframe consists of six columns: “lat”, “lon”, “parameters”, “year”, “month”, and “annual_temperature”. The temperature is in Celsius and is the average temperature per month. There are 544 rows to be joined with the All_Countries_Classification_Filter_10_Region_Count dataframe.

The self-created dataframe (All_Countries_Classification_Filter_10_Region_Count) was necessary for EDA. There are nine columns and 544 rows of statistical data within the All_Countries_Classification_Filter_10_Region_Count dataframe. This dataframe was made using a threshold of regional count. If there were ten regions or more within a country, all data for that country was added to the dataframe. Anything less was deemed as an insufficient amount of regional data to produce significant results. The All_Countries_Classification_Filter_10_Region_Count dataframe was used as a validation step during EDA.

Analysis

EDA

The EDA for this research had eight purposes that are also outlined in the Jupyter Notebook which you will find near the bottom of this paper:

[2] “Classification” column is used to classify the wines as appreciating or depreciating in price.

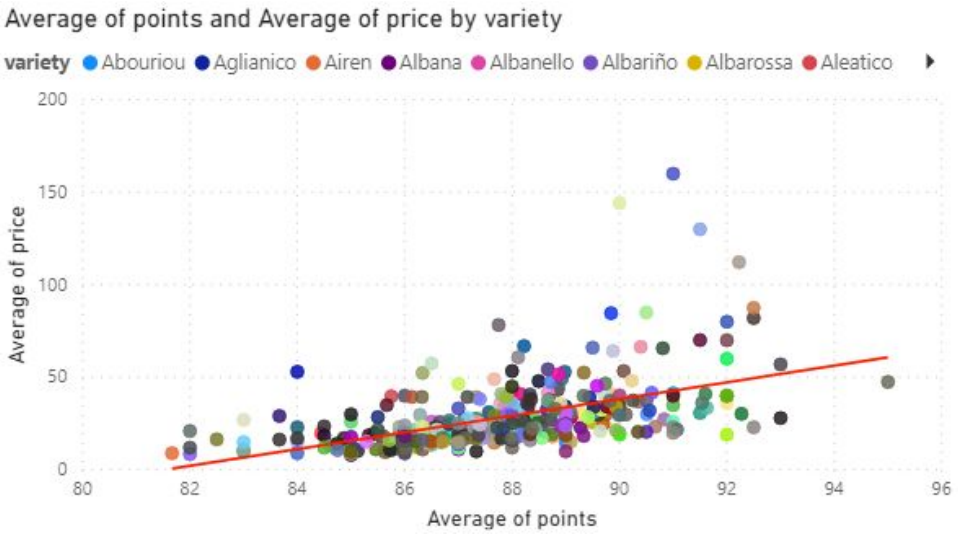
1. Find patterns on each region based on price of the wine and year.
2. Find patterns on each region based on the points of the wine and year.
3. Find patterns on each region based on the price of the wine and year per variety.
4. Find patterns on each region based on the points of the wine and year per variety.
5. Know the distribution of regions in each country.
6. Know the distributions of regions that have price filled up.
7. Know the distributions of regions that have points filled up.
8. Support our hypothesis that there lies a correlation between temperature and wine prices.

Overall, the end goal was to observe these patterns and distributions to make inferences that helped guide us towards creating methodologies for predicting undervalued assets (regions and grape varieties), and to validate previous research from other intellectuals.

Notable findings worth mentioning were the price and points relationship, points by country distribution, wineries distribution by country, and wine taster distribution. Additionally, it is important to note that this is nowhere near all of the results from the EDA. However, within this document is a link to the PowerBI file used for the visualizations.[3] All one has to do is download the free version of PowerBI Desktop and then open the PowerBI file provided. You can then experiment on your own by clicking on various aspects of the dashboards within the file.

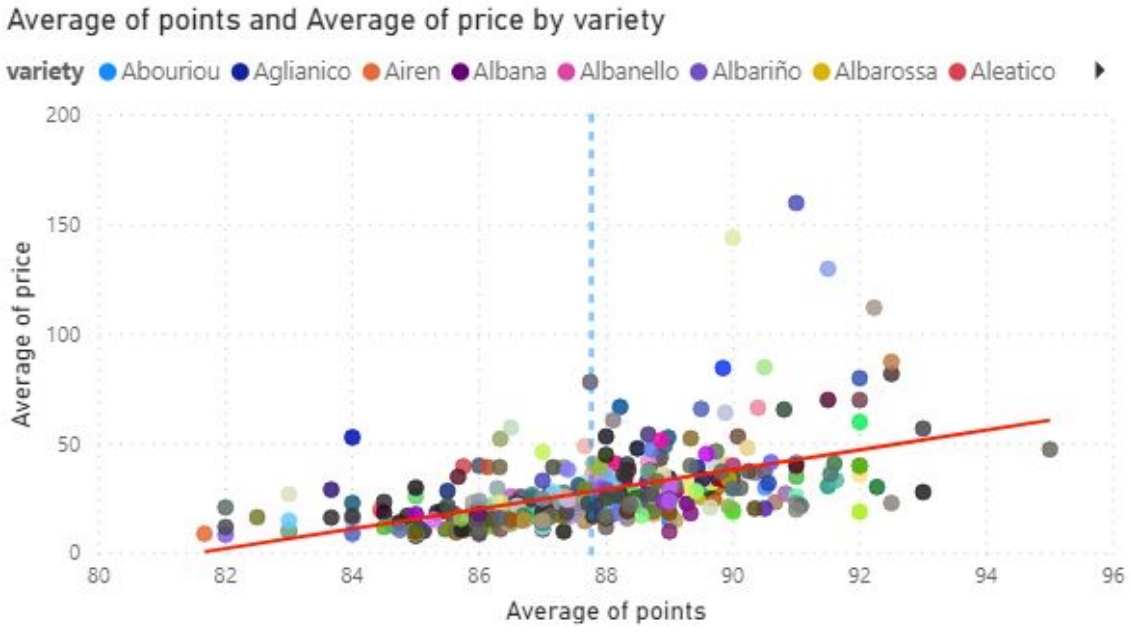
Observing the relationship between price and points was necessary to confirm previous statements that higher points lead to higher wine prices. If you refer to Plot 1, then you will notice our dataset followed this assumption, with average points being the independent variable on the X-axis, and average price being the dependent variable located on the Y-axis. Nonetheless, there were two intriguing discoveries.

[3] https://drive.google.com/file/d/1_2JYzlpgv8X8cAJlN3Jsr5qVitLDR3oG/view?usp=sharing



Plot 1

First of all, you will notice that our dataset has low variance. Since so much of our data is around the fitted line, the data displays homoscedasticity, meaning our data is homogeneous. This makes sense, though, given that the data comes from the same source. Second, the chart seems to imply a decrease in returns with how the shape of the data almost flattens out near the end. Refer to Plot 2 to see optimal points.

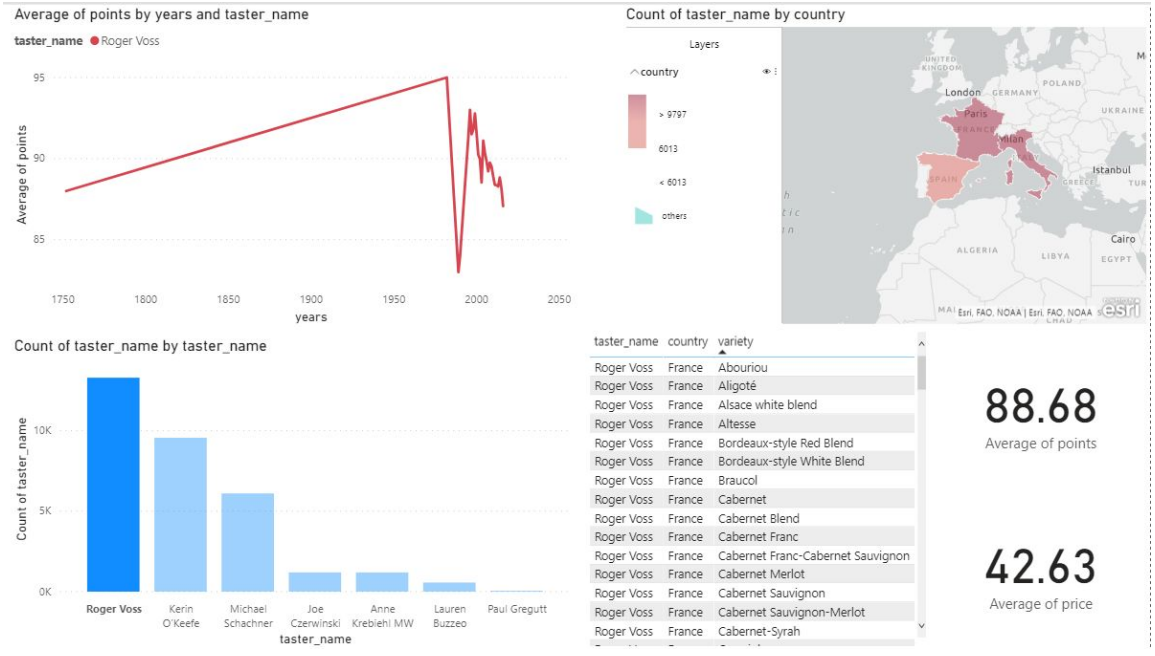


Plot 2

If you look at the chart, you will notice a blue dashed line and a red solid line. The red solid line is the trend and the blue dashed line represents the average points given to all wines. The optimum price and point combination lies where the red line and blue dashed line

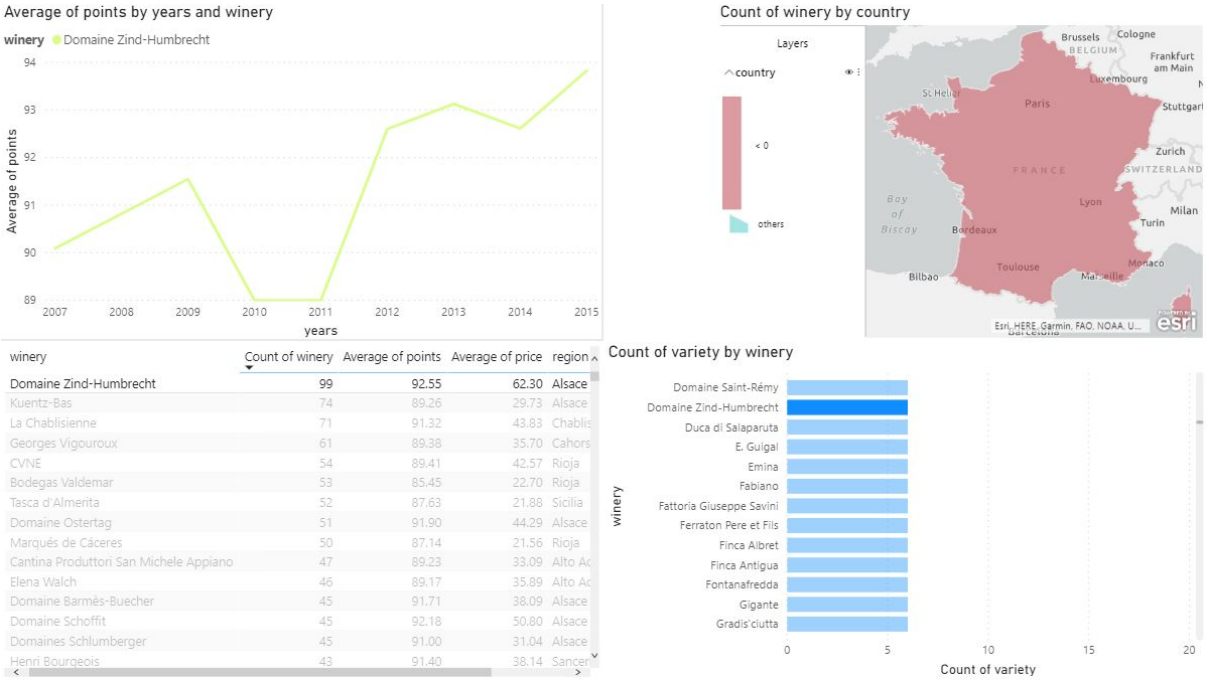
intersect. So anything to the right of the blue dashed line and below the red line are receiving diminishing returns. These wines may have higher than average points, but they are selling below the average price and for nearly the same price as wines with scores below the average points. This is obviously not an ideal situation for wineries to be in. Based on this information, the optimal wine score would be to receive 88 points in order to maximize wine prices.

With that being said, the goal of this paper is not to research optimal wine scores. This sighting is interesting, though, and it would be equally interesting to find the reason why the wine data is suggesting there is a point where the score possibly no longer matters. We did notice an unequal distribution of wine tasters, and this could be one of the reasons. In Plot 3 you will see the dashboard created for the wine tasters. The dashboard contains a distribution chart where you will see the name Roger Voss. This wine taster made up the vast majority and had an average wine score of 88.68. It may be coincidence that Voss's average score is about the optimal score, but it could very well be one of the reasons why the optimal point exists, given how many wines Voss scored. Nevertheless we digress as the focus of this paper is finding undervalued assets and not optimization.



Plot 3

The next analysis worth mentioning is the distribution of wineries. You can see the visualizations in Plot 4. This dashboard contains the information about the scores over the year, the average score per winery, the count of wines from each winery, how the winery is distributed throughout the three countries, and the count of unique varieties for each winery. The image in Plot 4 specifically displays Domaine Zind-Humbrecht winery data. Domaine Zind-Humbrecht had the largest count of wines in the dataset, so gathering insights from this winery was valuable given the amount of data it had compared to the others. Domaine Zind-Humbrecht also behaved unusually in terms of the relationship between high quality wines and time.

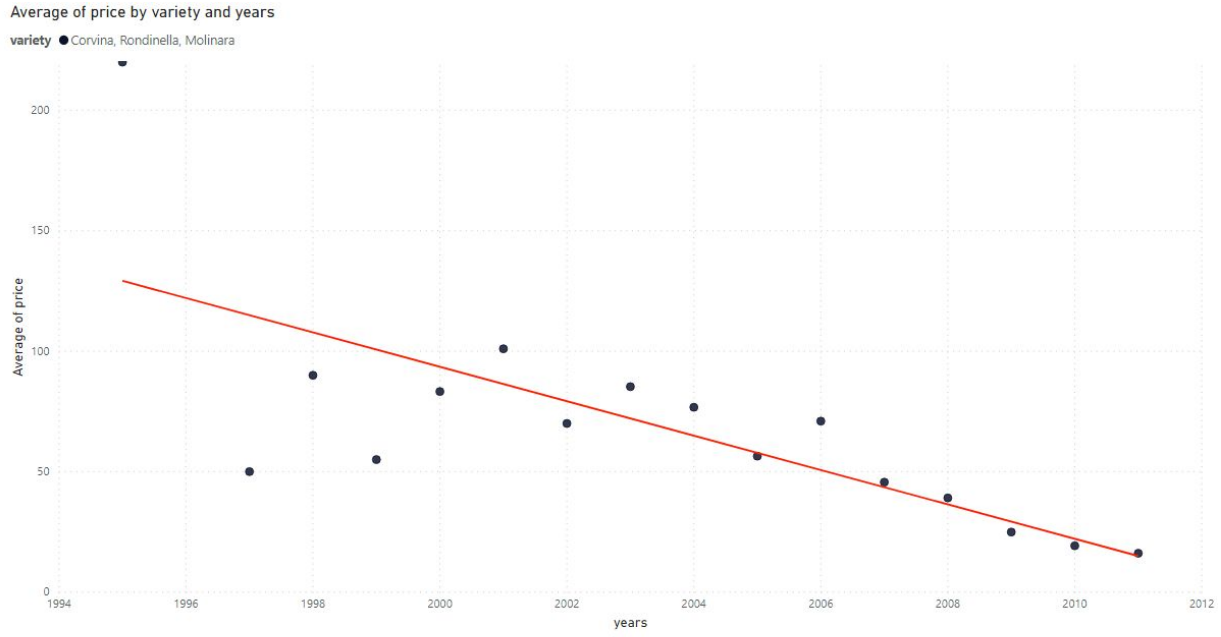


winery	Count of winery	Average of points	Average of price	region
Domaine Zind-Humbrecht	99	92.55	62.30	Alsace
Kuentz-Bas	74	89.26	29.73	Alsace
La Chablisienne	71	91.32	43.83	Chablis
Georges Vigouroux	61	89.38	35.70	Cahors
CVNE	54	89.41	42.57	Rioja
Bodegas Valdemar	53	85.45	22.70	Rioja
Tasca d'Almerita	52	87.63	21.88	Sicilia
Domaine Ostertag	51	91.90	44.29	Alsace
Marqués de Cáceres	50	87.14	21.56	Rioja
Cantina Produttori San Michele Appiano	47	89.23	33.09	Alto Ad
Elena Walch	46	89.17	35.89	Alto Ad
Domaine Barmès-Buecher	45	91.71	38.09	Alsace
Domaine Schoffit	45	92.18	50.80	Alsace
Domaines Schlumberger	45	91.00	31.04	Alsace
Henri Bourgeois	43	91.40	38.14	Sancer

Plot 4

Normally, younger wines receive lower wine scores compared to older wines. However, with this winery the opposite was observed. There were only six different varieties of grapes being grown so we suspected that the Alsace region had not hit its “sweet spot” with regards to temperature until later. The prospect of other factors like production changes, new farming techniques, or more brand recognition playing a part was also possible, but research suggested climate change would be the most important variable.

In order to be able to add in the climate data, we needed to make certain that wine (grape) variety was not an influencing factor for appreciation. To accomplish this, we had to first classify appreciating and depreciating wines. Refer to Plot 5 for an example of an appreciating wine. With this variety (Corvina, Rondinella, Molinara) you will notice a downward slope. For this dataframe, a negative slope actually suggests appreciation. This is due to the fact that that year is actually the production year. Recall that the data was scraped from 2017. The prices on the Y-axis were the average sell prices from 2017. The first point tells us that the average price was $220.00 and that the wine was produced in 1995. This also happens to be the most expensive wine for this variety. We know from previous research that generally older wines have higher selling prices, so this chart makes sense.



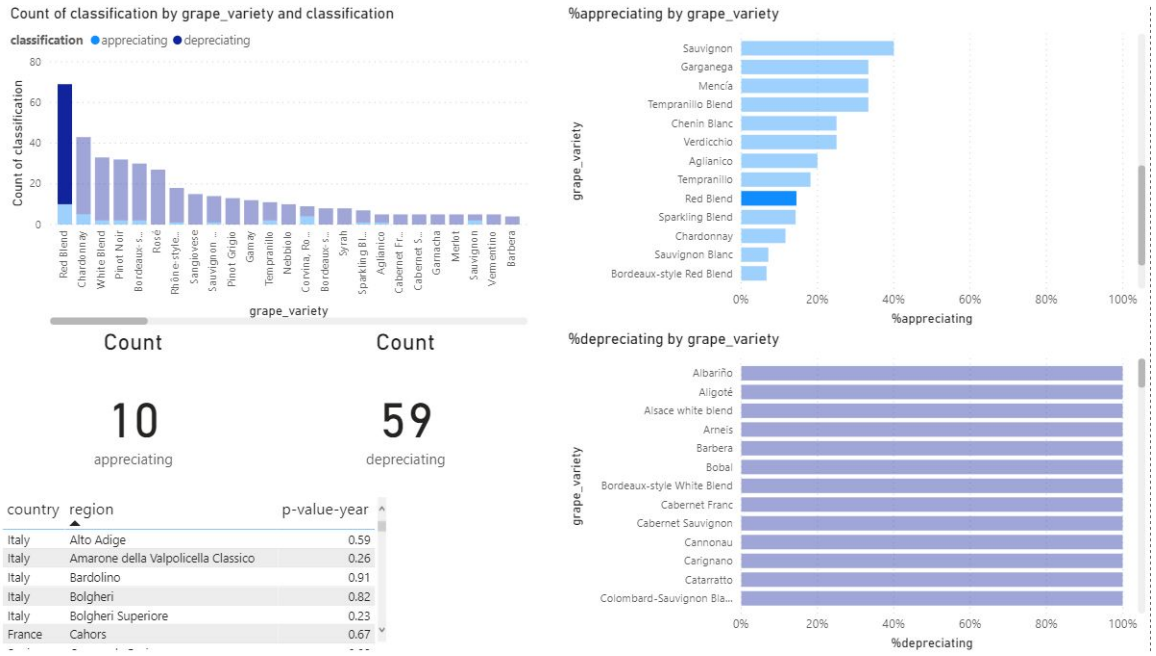
Plot 5

Furthermore, we can also assume that due to its high price, the wine received a high score and is therefore a high quality wine. For those that are curious, the average wine score for 1995 was 94, and 85.11 in 2011 for data contained in this chart. So if you go backwards in time (look at the chart from right to left vs. left to right), then you will see the price appreciation trend. In other words, we were able to come to the same conclusion as

previous research that as time goes on, price will appreciate and wine will be considered higher quality.

After confirming this theory, we wrote a code that created a classification of “appreciating” or “depreciating” based on the slope for every average price and year combination grouped by variety. Once we had this classification we developed visualizations to check for dominating grape varieties. Next we created two different measures: appreciation count and %appreciation. Appreciation count was the count of appreciated wines for each variety, and %appreciation told us what percent of that variety was appreciating. A grape variety was considered dominating if it had an average total count, high %appreciation, and one of the highest appreciation counts. Having a dominating variety would insinuate that the highest indicator for appreciation would be the variety.

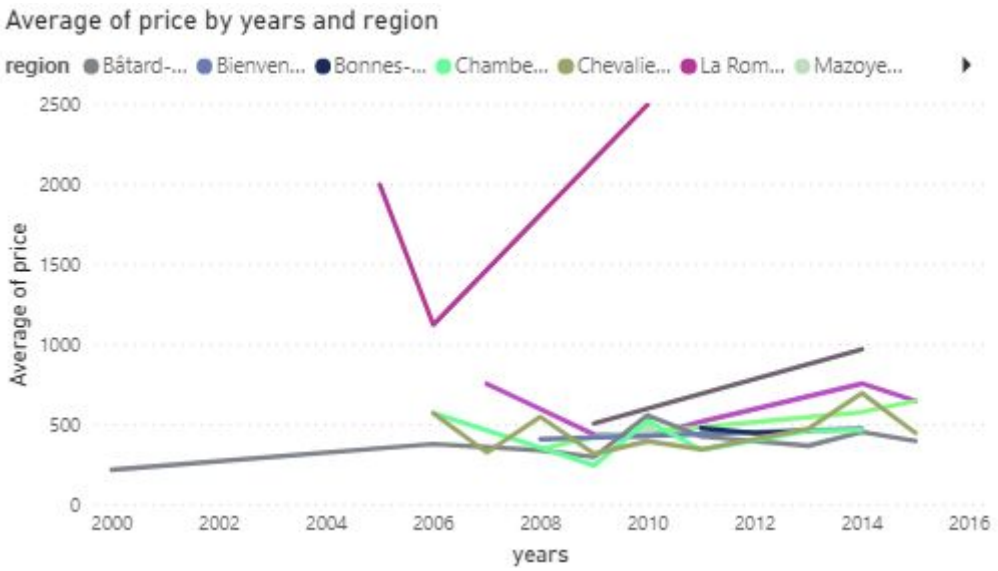
For example, if you refer to Plot 6, you will notice that at first glance Red Blend has the highest appreciation count of 10. Without looking into what percent of the Red Blends are appreciating, and without knowing how the total count compares to the other total counts of the rest of the varieties, you could assume that Red Blends tend to appreciate regardless of factors like climate change. However, you can see in the dashboard that Red Blends have the highest total count of all the varieties, so it is not surprising that it would also contain the highest count of appreciated wines. You will also notice that the %appreciation is only %14.49 and near the very bottom in terms of ranking %appreciation values. Luckily, after thoroughly examining this dashboard, we were able to validate that there were not any dominating grape varieties. With that being said, it is crucial to note that this was not our methodology for predicting appreciating wines. This specific method was only used to validate this dataframe.



Plot 6

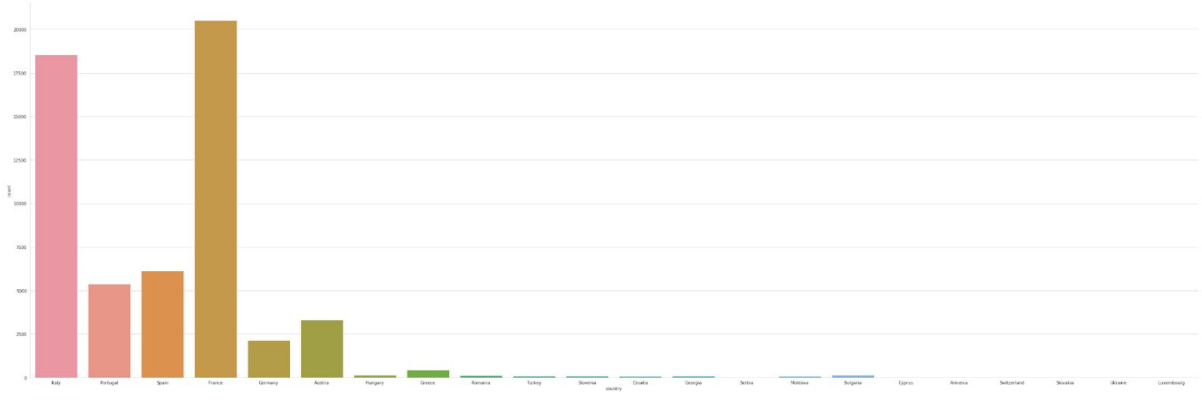
In addition to knowing the classification, we also calculated the p-score to determine the significance. By filtering with the p-value we were able to determine which wine varieties were appreciating and had a significant relationship between year and price. The year is a critical variable since our climate data can only be concatenated based on year. We found an indicative amount of significant p-values for the year in each variety. This pattern helped to prove our theory that certain years may have had events (e.g. ideal climates) that lead to better or worse prices.

The other analysis we performed to reinforce our theory about years was to find any patterns with years and price with all wines. Unfortunately, not a lot could be interpreted or understood from this chart. At first, having all the regions on one chart was incredibly messy and unreadable. We then added a filter that only showed the top ten regions by average price (see Plot 7). Even with only ten regions, there were not any distinct patterns that were detectable.



Plot 7

With regards to the descriptive analytics, the country distribution showed that the top three countries were France (20,511 rows), Italy (18,550 rows), and Spain (6,142) (see Plot 8). According to Italian Wine Central, in 2017, Italy, France, and Spain were the world's top producing countries for wine (2020). Italy produced 17%, France produced 15%, and Spain produced 13% of the world's wine (Italian Wine Central, 2020). Altogether that's 45% of the world's wine production in those three countries, which matches our dataset (before it was cleaned). Therefore, we were not concerned with the skewed distribution.

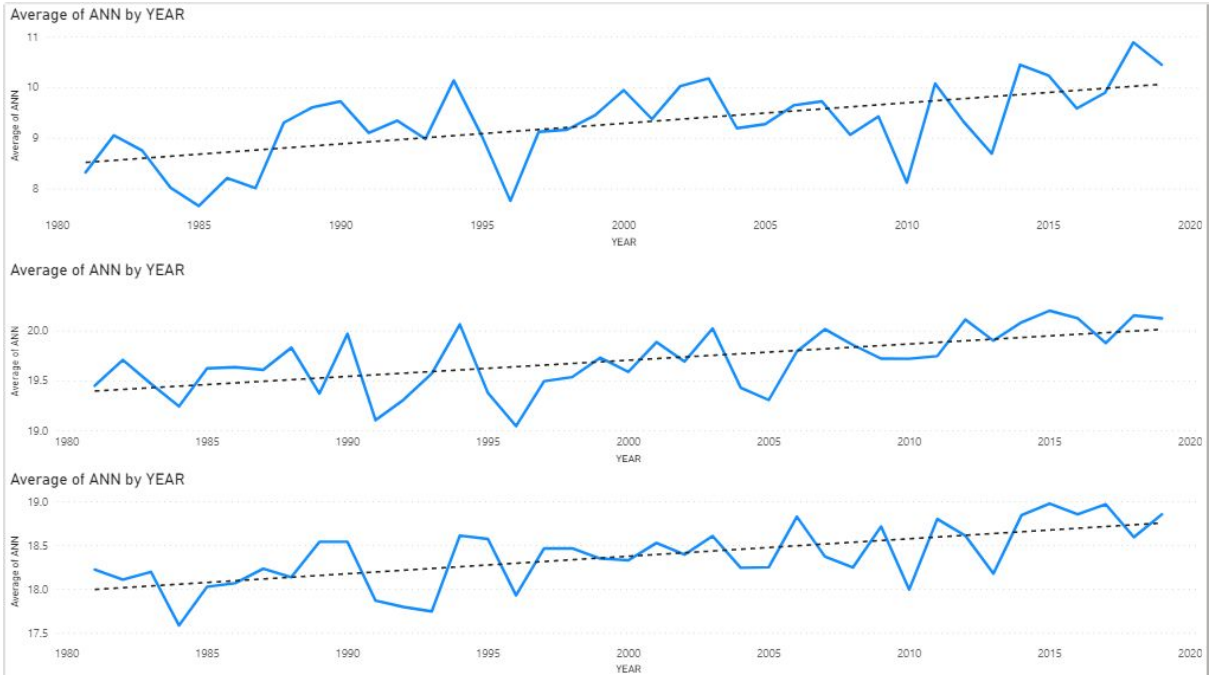


Plot 8

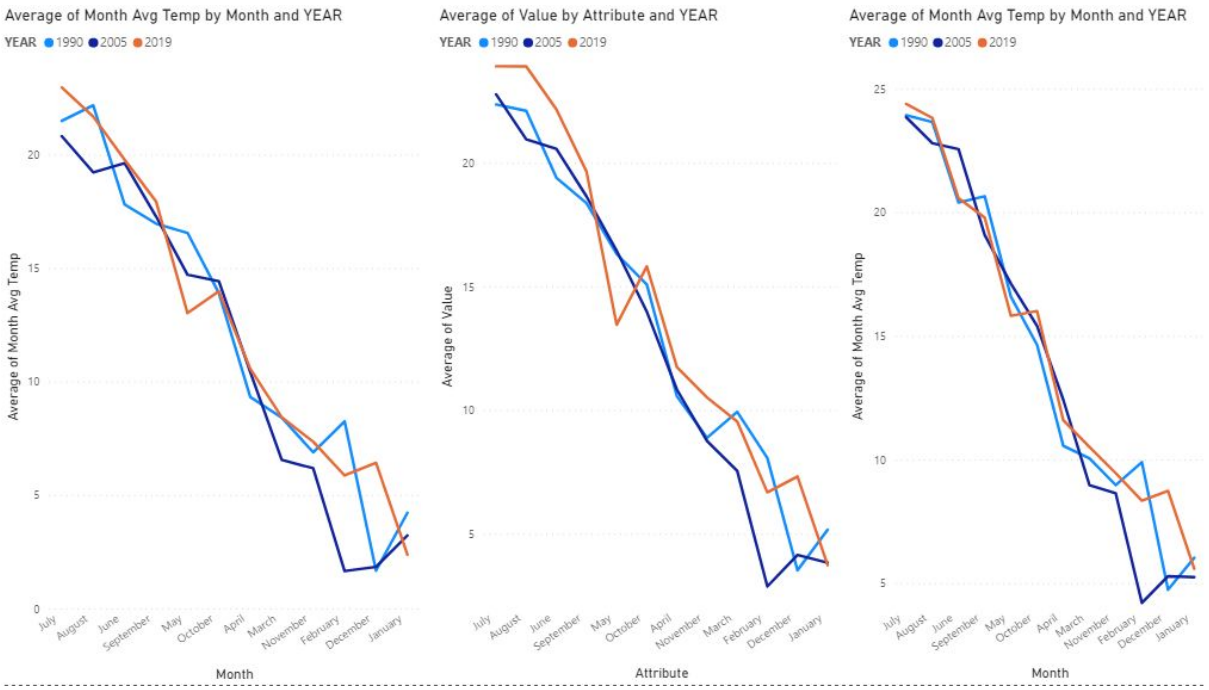
In the end, we were able to validate our dataset and confirm that it contained no major issues. The general patterns we found, regardless of how the data was grouped, proved that previous assumptions held true for our datasets. This was shown in the many

plots with time and price, and points and price as well, where price and points were the independent variables. Sadly, little evidence was found to support our theory that some years may perform better in terms of price because of weather. The lack of patterns by year meant that it was time to add the climate data to truly test this hypothesis.

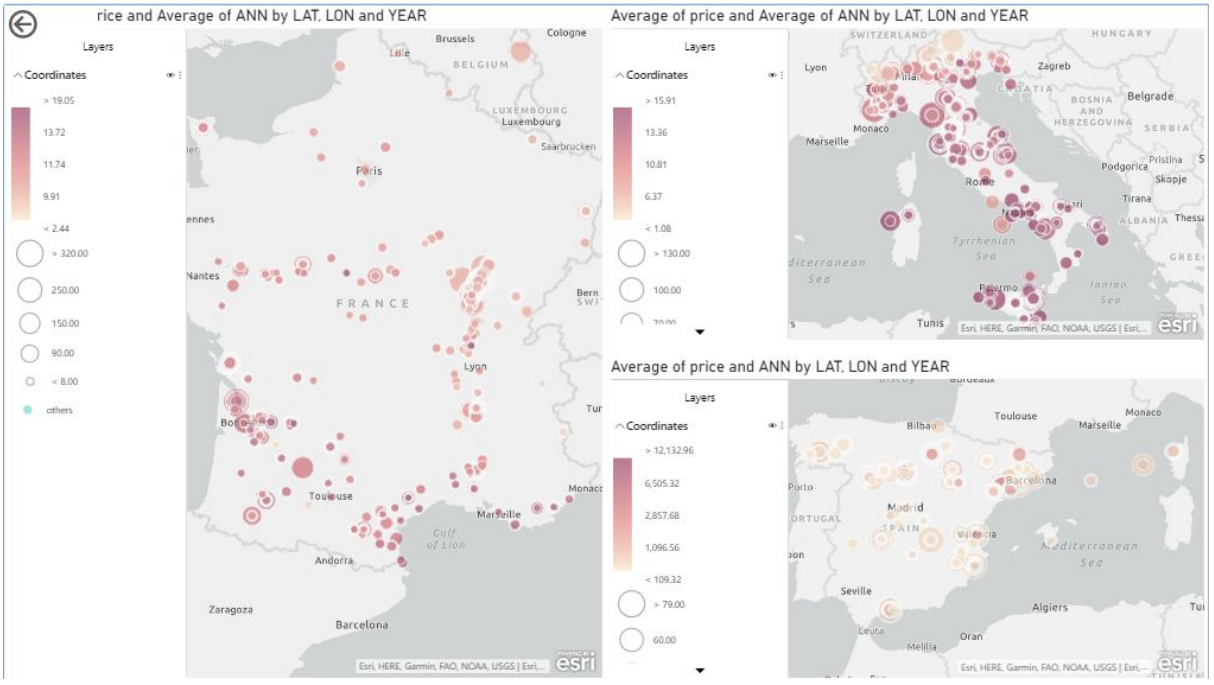
Based on the data, there is no question that the Earth's surface temperature is increasing (see Plot 9). In all three countries you can observe an upward trend from the years 1981 to 2019. In Plot 10, you will also notice that all three countries display seasonality as well. July seems to be the hottest month while January produces the coldest temperatures. Plot 11 contains basic correlation inferences. Our prediction model depends on temperature being a driving force for wine prices. This hypothesis was critical to confirm. In this image the larger the bubble the more expensive the wine, and the darker the red the hotter the temperature. In all three countries the data is broken down by region. There seemed to be slight patterns between price and temperature. There is also a line graph for each country that shows the possible relationship between price and temperature over time (see Plot 12). Only a true correlation test would prove our hypothesis, however.



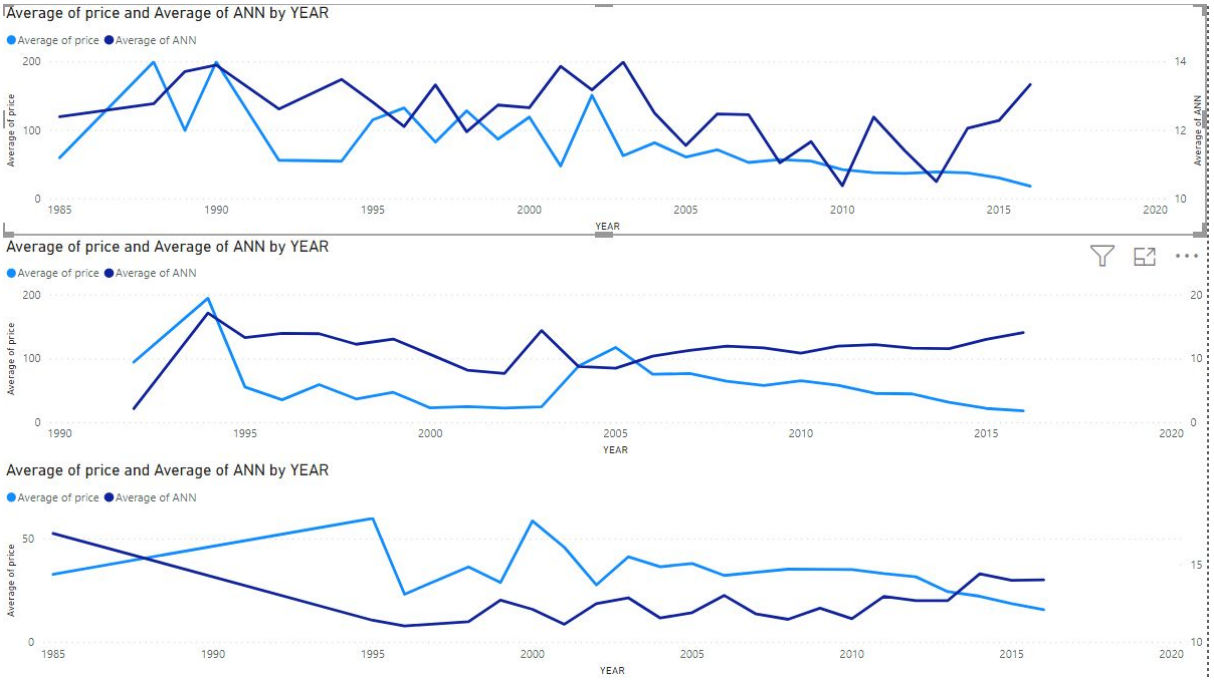
Plot 9



Plot 10

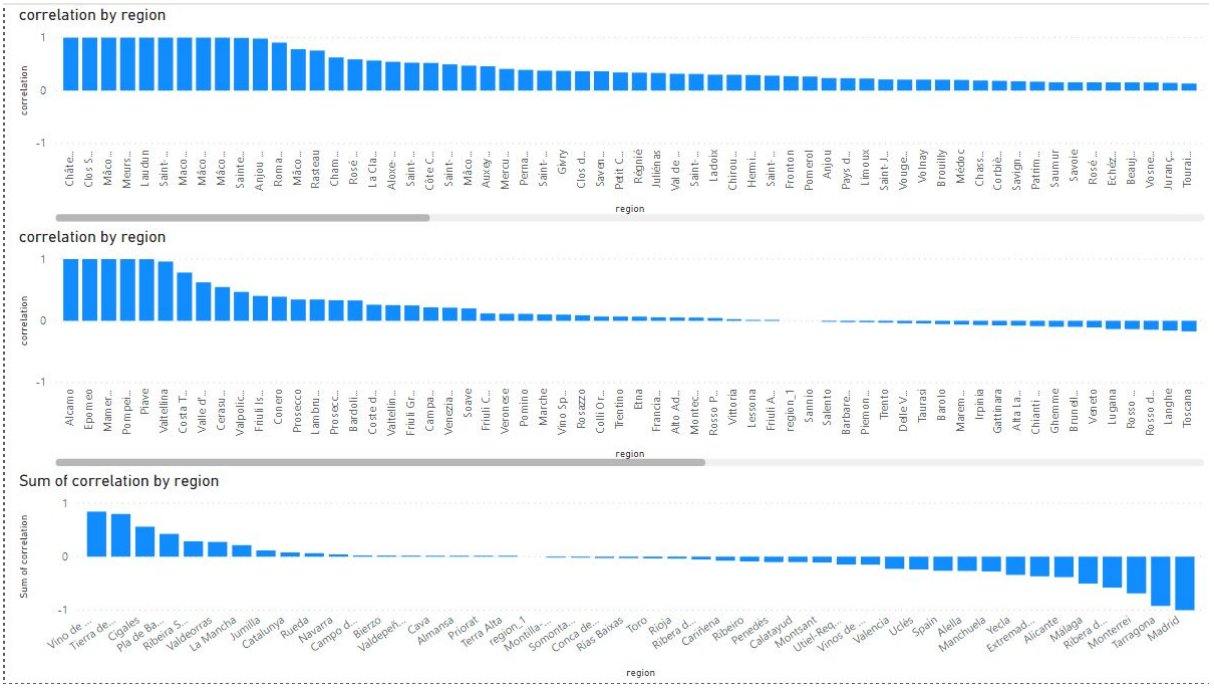


Plot 11



Plot 12

With NumPy, we were able to calculate the correlation coefficients for regional prices and annual average temperatures for each country (see Plot 13). The results were shocking. France contained seventeen regions with perfect (nine positive and eight negative) correlations, Italy had nineteen regions with perfect (five positive and fourteen negative) correlations, and Spain had one region that was perfectly negatively correlated. We expected there would be regions with strong correlation, but we did not predict perfect correlations. The next step was to break it down by grape variety.



Plot 13

We again witnessed surprising results. France had eleven grape varieties with perfect correlations, Italy had ten grape varieties with perfect correlations, and Spain had six grape varieties with perfect correlations (see Plot 14). The most valuable finding in this analysis was the fact that in all three countries there was a cabernet with a perfect correlation. If you look at the charts, you'll see that France has a Merlot-Cabernet Sauvignon with a -1.00 correlation coefficient, Italy has a Cabernet with a -1.00 correlation coefficient, and Spain has a Cabernet Blend with a 1.00 correlation coefficient. Even though they are three different cabernets from different vineyards, it was still fascinating to see that the cabernet was the only grape variety with a perfect correlation in all three countries. This finding strongly supported our theory of grape varieties having a sweet spot.

France

Variety	Sum of Correlation
Carignan	1.00
Colombard-Ugni Blanc	1.00
Merlot-Malbec	1.00
Pinot Meunier	1.00
Mourvèdre	1.00
Chardonnay-Viognier	1.00
Pinot Noir-Gamay	1.00
Grenache Blanc	-1.00
Sémillon	-1.00
Merlot-Cabernet Sauvignon	-1.00
Sciaccerellu	-1.00

Italy

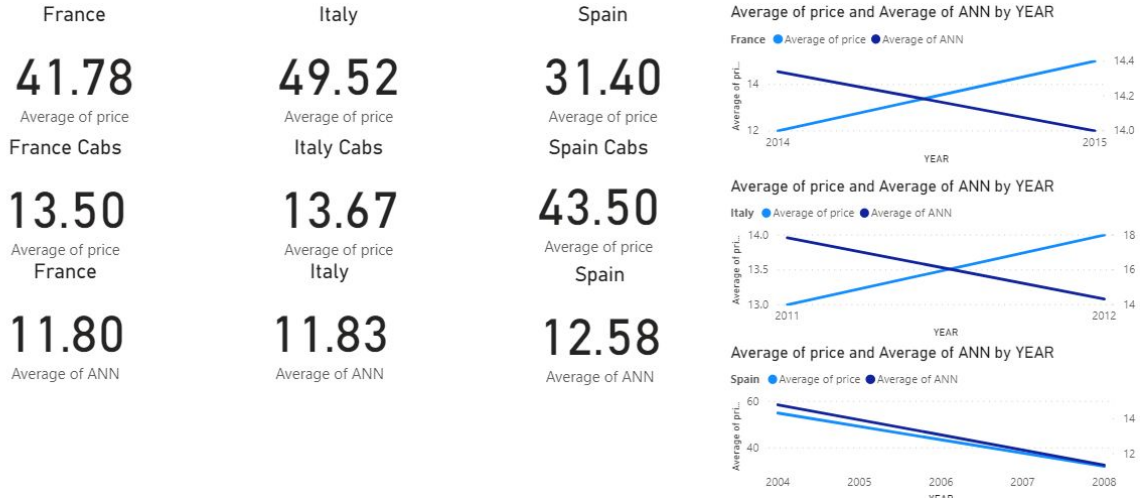
Variety	Correlation
Moscato Rosa	1.00
Nosiola	1.00
Petit Manseng	1.00
Refosco	1.00
Traminer	1.00
Bordeaux-style Red Blend	1.00
Barbera	1.00
Cabernet	-1.00
Gaglioppo	-1.00
Petit Verdot	-1.00

Spain

Variety	Sum of Correlation
Cabernet Blend	1.00
Shiraz	1.00
Tinta del Pais	1.00
Tinta del Toro	1.00
Malvasia-Viura	-1.00
Rhône-style Red Blend	-1.00

Plot 14

A further investigation of the cabernet discovery can be found in Plot 15. We decided to compare the price of the cabernets with perfect correlations to the average price of wines in the specific country. You will also find the average annual temperatures of each country. On the right of the dashboard, you will see the line charts to visually show the relationships between the cabernet's prices and annual temperatures. For France and Italy, it looks like the sweet spot temperature is around 14 degrees Celsius. Spain, however, behaved the opposite compared to the other countries. We unfortunately lacked the data to further investigate this.



Plot 15

Algorithms

There exists a variety of methods for predicting climate change. For this paper, we tested two different algorithms that are well suited for regression and time series analysis.

The first model was machine learning based and used as the baseline prediction, and the second was deep learning based used to compare with the baseline prediction. However, we will only discuss in detail the algorithm with the best performance based on the root mean square error (RMSE). RMSE is a statistical metric for evaluating the error of the model. The RMSE also determined which model was used for the final temperature prediction. The two algorithms that were compared were support vector regression (SVR) and the LSTM (Long-Short Term Memory), which is a variant of recurrent neural networks (RNN). Based on the initial evaluation on one temperature dataset, SVR produced an RMSE of 2.61 and LSTM had an RMSE of 1.68. Therefore, we decided to use the LSTM multivariate algorithm for the temperature prediction of specific regions in three countries: France, Italy, and Spain.

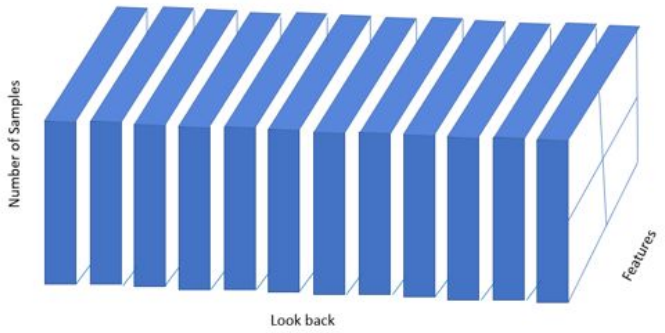
Data Pre-processing and Feature Engineering

The data used for temperature prediction is the climate change dataset which was collected from NASA POWER. The data contains monthly average temperatures from January 1981 to December 2019. Since we were working with monthly temperature data, the data is naturally cyclical. Therefore, we performed feature engineering for the monthly cycles.

Since the data consist of months that are written in words, we converted the words to numbers in order for the model to understand the data. January became 1, February became 2, March became 3, and so on until December as 12. As for the cyclicity, January represented as 1 should not be far from December which is represented as 12. Due to this, we calculated the sin and cosine component of the months that serves as an X, Y coordinates in a circle.

To predict the next temperature, we transformed the data, which was still a sequence, into a supervised dataset as required for the LSTM model to learn. A supervised dataset has a set of features and a target variable {*feature,target}* format. Here we will introduce the term look back. A look back is the number of samples we will use to predict the

next output. In our case, it is the number of previous months to predict the next month. For example, a look back of twelve uses twelve months (ie., January – December) to predict the next year's January temperature. In our case, swe used a multivariate LSTM with four variables as inputs. Those inputs being year, average temperature, month_sin, and month_cos. The data was prepared with a similar structure, but was represented as windows. The dimensionality of the data (number of samples, lookback, number of features) is represented in the picture below. The format follows the requirements set by Keras.



Data representation for multivariate LSTM

The prepared data was divided into a training set, validation set, and test set. The shape of the training set is (374,4) and were collected from 1981-01-01 to 2012-02-01, and (94,4) is the shape of the validation set collected from 2012-03-01 to 2019-12-01.The training set was used for training and the validation set was used for validating how well the model performed on the training set. The test set is the dataset used to predict. For our purpose, this meant predicting the years 2020 and 2021. The model was not trained on the validation set, however, the model is tuned based on the validation set performance. Consequently, in some way it implicitly affects the model's performance. We set a threshold for the performance of the validation set and it should be below the RMSE of two. Lastly, if the performance in the validation set passed the threshold, we then used the model to predict the test set and record the predictions for the years 2020 and 2021. While testing the years of 2020 - 2021, we used twelve samples from the year 2019 to predict January of

2020. We then used a loop where we move 1-time step at a time to predict February of 2020 and so on. The prepared dataset was then fed to the LSTM model

RNN

In order to properly understand an LSTM, one must also understand the RNN model as well. Imagine a sequence of data like a time series. Time series, especially climate data, is incredibly dependent on the previous temperature. For instance, temperature in February is dependent on the temperature of January and temperature in January is dependent on the temperature of December. Feeding this kind of data to a traditional neural network is technically possible, however, the model cannot determine the sequential dependencies of the data.

Recurrent neural networks (RNNs) are designed to process sequential data more effectively by taking into attention the sequential nature of the data (Elsworth & Güttel, 2020). A recurrent neural network can process one sequence of elements one at a time (Elsworth & Güttel, 2020). The simplest RNN is composed of one neuron that consists of an input (X_t) and produces an output (h_t), thus sending it back to itself (Geron, 2019). The Figure A (right) shows an unrolled version of the single recurrent neuron at each time step. It receives the inputs (X_t) as well as its own output from the previous time step (h_{t-1}).This design can make the information persist.

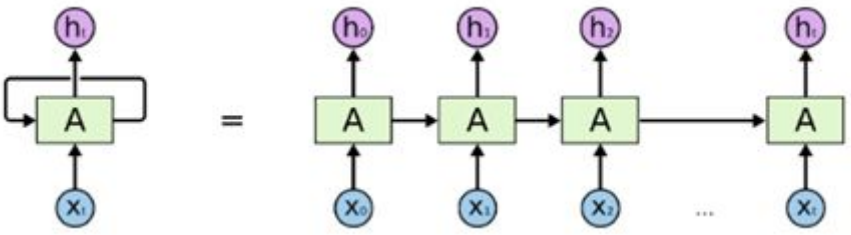


Figure A. A recurrent neuron (left) unrolled through time (right).[4]

[4] Image taken from: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

One limitation of RNN, however, and all other neural networks is the vanishing gradient problem (Hochreiter, 1998). Recurrent neural networks have a problem with long sequences because they have a difficult time carrying the information from the early time steps to the current one (Phi, 2018). This led to the development of LSTM (Long short-term Memory) cells (Hochreiter & Schmidhuber,1997).

LSTM

Long short-term memory is a kind of RNN that is capable of long-term dependencies (Olah, 2015). The main concept of the LSTM is the memory cell or cell state which allows the transfer of important information.

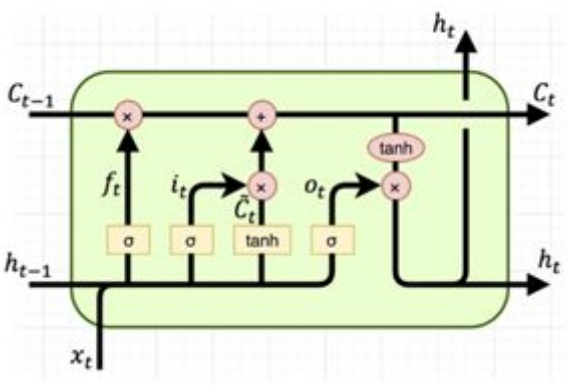


Figure B. Basic Unit of LSTM cell. [5]

Figure B shows the basic structure of the LSTM cell. Similar to the simple RNN cell, LSTM also has a repeating module, but with a more complex structure. There are four neural network layers inside an LSTM cell that are interacting in an incredibly special way to retain important information from the earlier time steps up to the current time step. As mentioned earlier, what makes the LSTM special is the cell state denoted as C. This serves as a carrier of information that runs through the entire chain with some interactions. These interactions include the careful adding or removal of information that are regulated by the gates. LSTM cells have three important gates being the forget gate, update gate and output gate.

[5] Image taken from https://towardsdatascience.com/grus-and-lstm-s-741709a9b9b1

The first step in a LSTM is to decide what information is to be kept and what is to be thrown away from the cell state (Olah, 2015).

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

The forget gate layer.

The "forget gate" layer denoted by (f_t) is the output of the concatenation of the previous output (h_{t-1}) and the current input (X_t). This decision is made by the sigmoid layer which is denoted by (σ). This gate outputs a number in the range of 0-1. A value near to 1 means the information is going to be kept, and a value near 0 means the information will be thrown away.

The second step is composed of a sigmoid layer and a tanh layer. This step decides what kind of new information is going to be kept in the cell state. The sigmoid layer is called the "input gate" which decides the values to update. The other layer is called the tanh layer that creates candidate values that could be added in the state. The proper formula is written below.

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

The input gate layer and the tanh layer with candidate values.

To update the old cell state (C_{t-1}) into a new cell state (C_t), we multiply (C_{t-1}) by (f_t). The multiplication of the old cell state and forget gate will allow the network to forget the information that it wants to forget. We will then add the product of the input gate layer and the new candidate values. The new candidate values are the result of the update of the cell state. The formula is written below.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

The new candidate values.

The last step is for the “output gate” denoted by (O_t). This step outputs the information that we want to output. First, we run the sigmoid layer that determines which part of the cell state we are going to output.

$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$

The output gate layer.

The result of the output gate layer is pushed to the tanh layer in order to make the values between -1 and 1. The values of the tanh layer are then multiplied to the output gate.

$$h_t = o_t * \tanh\left(C_t\right)$$

The tanh layer.

Finding the Sweet Spot

Before running regional data into the models for prediction, we filtered the data in the essence of time. As gleaned from EDA, we know there exists a relationship between price and temperature. However, certain regions contain stronger correlations than others. This leads us to assume there also exists sweet spots for optimal grape growth. Due to time constraints, it was not realistic to run every region through the prediction models. Instead, we filtered for regions with sweet spots. To determine the sweet spot temperatures for the regions, we chose the most recent year with sufficient region and grape variety counts. Counts were then filtered with a threshold of 10 data points or more. Next we grouped the filtered data by region and grape type and sorted by median of the price. After sorting we

then took the top 50% of the price per region getting the grape varieties that had the highest price on that given year. We then mapped the list of these top priced regions to the temperature datasets from NASA POWER to get the average annual (ANN) temperature of the given year as a sweet spot. This methodology was repeated for all of the countries (France, Italy, Spain) to get the sweet spot temperature. All the regions that had the highest prices in that given year are then fed to the LSTM model for temperature prediction. The regions with perfect correlations were also included.

Results

LSTM

Two approaches were taken for the LSTM predictions: univariate and multivariate. The main difference between the two is the amount of variables being analyzed. Univariate models examine a single variable, while multivariate models allow for the examination of multiple variables. Below, a link to the GitHub repository can be found, where all modelling results are stored for both univariate and multivariate LSTM's.[6]

Univariate

In Images A,B, and C you will see snapshots of the univariate results for France, Italy, and Spain respectively. Overall the performance was acceptable. The average RMSE (valid) for France was 1.84, Italy was 1.76, and Spain was 1.90. Recall temperature was recorded in Celsius so the RMSE results are also in Celsius. See Image D for an example of 2021 predictions for the region Salento in the country of Italy.

[6] https://github.com/MADS-THESIS/Thesis

region_found	_temperatn_humidi		RMSE_train	RMSE_valid	model evaluation
Alsace	[10.45]	[81.68]	2.22	2.05	0.0002934987423941493
Anjou	[10.99]	[79.24]	2.15	2.07	0.0027467464096844196
Bandol	[15.56]	[74.19]	1.61	1.46	0.013371596112847328
Beaujolais	[20.04]	[77.74]	1.62	1.92	0.0023579788394272327
Beaujolais Blanc	[11.5]	[80.59]	2.17	2.01	0.004762586206197739
Beaujolais-Villages	[11.5]	[80.59]	2.20	2.00	0.003740040585398674
Beaune	[10.67]	[84.29]	2.18	2.07	0.0006676570628769696
Bordeaux	[14.34]	[77.93]	1.81	1.62	0.0035557840019464493
Bordeaux Blanc	[19.81]	[64.19]	2.09	1.91	1.27E+08
Bordeaux Rosé	[13.96]	[77.67]	1.88	1.99	0.005178520921617746
Bordeaux Supérieur	[11.06]	[85.06]	1.86	1.74	0.0020468260627239943
Bourgogne	[10.58]	[84.54]	2.38	1.96	5.82E+10
Bourgueil	[12.34]	[81.81]	2.35	1.93	0.00010986885172314942
Brouilly	[11.5]	[80.59]	2.17	1.19	0.0022584092803299427
Cahors	[13.51]	[76.56]	1.98	1.82	0.0030729700811207294
Chablis	[11.23]	[83.17]	2.13	1.78	0.003603180404752493
Chambolle-Musigny	[10.67]	[84.29]	2.10	1.79	0.0036137248389422894
Chassagne-Montrachet	[11.19]	[82.54]	2.01	1.98	0.0035716535057872534

Image A

region_found	_temperatn_humidi		RMSE train	MAE train	RMSE valid	MAE valid	Model Evaluation	uni/multi
Abruzzo	[11.86]	[69.34]	1.69		1.8		9.12E-07	univariate
Alto Adige	[2.28]	[81.85]	2.27		2.28		2.12E-06	univariate
Barbaresco	[13.06]	[69.09]	2.03		1.75		1.13E-06	univariate
Barolo	[11.44]	[66.9]	1.99		1.93		1.02E-05	univariate
Basilicata	[15.87]	[63.15]	1.54		1.69		5.53E-07	univariate
Bolgheri	[15.61]	[70.43]	1.56		1.58		1.26E-05	univariate
Brunello di Montalcino	[14.24]	[68.93]	1.84		1.64		7.82E-08	univariate
Campania	[16.15]	[70.02]	1.39		1.61		7.32E-07	univariate
Chianti	[14.12]	[62.57]	1.67		2.06		0.001407124	univariate
Chianti Classico	[9.05]	[85.13]	2.47		2.05		4.72E-07	univariate
Collio	[10.03]	[70.96]	1.94		2.04		4.05E-05	univariate
Etna	[16.72]	[67.92]	1.39		1.24		2.39E-06	univariate
Gavi	[13.57]	[71.06]	1.87		1.8		1.75E-06	univariate
Irpinia	[15.28]	[64.69]	1.59		1.88		7.62E-07	univariate
Langhe	[13.47]	[66.42]	1.75		1.6		1.41E-07	univariate
Lugana	[11.79]	[69.43]	1.93		1.87		5.63E-07	univariate
Marche	[14.32]	[68.26]	1.58		1.66		1.82E-07	univariate
Maremma Toscana	[12.93]	[70.54]	1.87		1.79		8.54E-07	univariate

Image B

region_found	_temperatn_humidi	RMSE train	MAE train	RMSE valid	MAE valid	Model Evaluation	uni/multi
Bierzo	[10.79] [72.54]	1.83		1.78		0.0008539	univariate
Calatayud	[12.41] [62.05]	2.33		2.31		2.85E-06	univariate
Cava	[10.05] [66.49]	2.53		2.5		4.12E-05	univariate
Jumilla	[16.74] [60.21]	1.93		1.69		7.10E-05	univariate
La Mancha	[11.96] [84.]	1.61		1.55		3.84E-06	univariate
Montsant	[16.15] [70.08]	1.9		1.85		3.78E-05	univariate
Navarra	[12.46] [76.05]	1.86		1.66		0.000495297	univariate
Priorat	[15.92] [66.62]	1.8		1.8		0.000724594	univariate
Rías Baixas	[22.17] [76.8]	0.96		1.08		1.08E-06	univariate
Ribera del Duero	[11.55] [67.19]	2.13		2.03		1.29E-06	univariate
Rioja	[11.82] [69.46]	1.96		1.78		5.94E-04	univariate
Rueda	[8.73] [78.68]	2.65		2.83		0.000318662	univariate
Toro	[12.64] [65.52]	1.92		1.62		3.85E-05	univariate
Valdeorras	[11.14] [70.34]	2.17		2.01		0.000429088	univariate
Valencia	[17.71] [67.59]	1.85		1.94		0.000154931	univariate

Image C

	YEAR	month	month_sin	month_cos	predicted_temp
12	2021	1	0	1	17.10285301
13	2021	2	0.5	0.8660254038	17.10889893
14	2021	3	0.8660254038	0.5	17.11461863
15	2021	4	1	0	17.12004708
16	2021	5	0.8660254038	-0.5	17.12522324
17	2021	6	0.5	-0.8660254038	17.13018527
18	2021	7	0	-1	17.13494905
19	2021	8	-0.5	-0.8660254038	17.13951718
20	2021	9	-0.8660254038	-0.5	17.14388144
21	2021	10	-1	0	17.14804823
22	2021	11	-0.8660254038	0.5	17.15202883
23	2021	12	-0.5	0.8660254038	17.15580916

Image D

Multivariate

In Images E, F, and G you will see snapshots of the multivariate results for France, Italy, and Spain respectively. In almost every instance multivariate outperformed univariate. The average RMSE (valid) for France was 1.52, Italy was 1.37, and Spain was 1.31. At first glance these results may not seem to significantly improve from the univariate LSTM, but given how sensitive grapes are to temperature, these differences in temperature prediction errors could make all the difference between a bad wine and good wine. See Plot 16 and Image H for a prediction vs. actual plot and the predictions for 2021 in the Salento region of Italy. Images I, J, and K contain snapshots of the average annual temperature predictions for 2021 in all sweet spot regions for every country.

region_found	_temperatn_humidi		MAE_train	RMSE_train	MAE_valid	RMSE_valid
Alsace	[10.45]	[81.68]	1.41	1.80	1.36	1.69
Anjou	[10.99]	[79.24]	1.34	1.70	1.35	1.67
Bandol	[15.56]	[74.19]	0.955227006	1.19	0.992003044	1.24
Beaujolais	[20.04]	[77.74]	1.10	1.42	1.31	1.65
Beaujolais Blanc	[11.5]	[80.59]	1.36	1.75	1.41	1.72
Beaujolais-Villages	[11.5]	[80.59]	1.36	1.75	1.41	1.72
Beaune	[10.67]	[84.29]	1.34	1.71	1.26	0.53
Bordeaux	[14.34]	[77.93]	1.21	1.53	1.16	1.38
Bordeaux Blanc	[19.81]	[64.19]	1.34	1.69	1.42	1.74
Bordeaux Rosé	[13.96]	[77.67]	1.29	1.63	1.26	1.49
Bordeaux Supérieur	[11.06]	[85.06]	1.21	1.60	1.16	1.45
Bourgogne	[10.58]	[84.54]	1.33	1.71	1.26	1.53
Bourgueil	[12.34]	[81.81]	1.31	1.68	1.30	1.55
Brouilly	[11.5]	[80.59]	1.36	1.75	1.41	1.72
Cahors	[13.51]	[76.56]	1.33	1.69	1.31	1.59
Chablis	[11.23]	[83.17]	1.35	1.73	1.27	1.58
Chambolle-Musigny	[10.67]	[84.29]	1.34	1.71	1.26	1.53
Chassagne-Montrachet	[11.19]	[82.54]	1.35	1.73	1.34	1.65

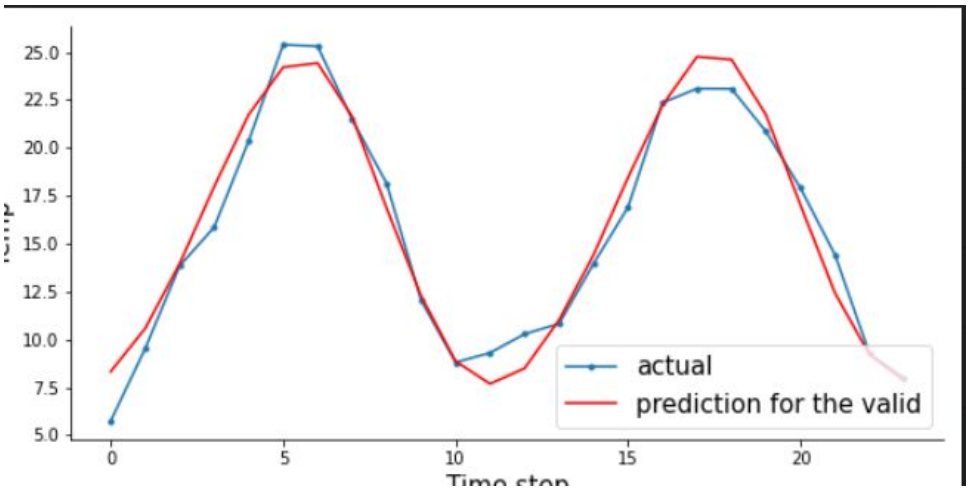
Image E

region_found	_temperatn_humidi		RMSE train	MAE train	RMSE valid	MAE valid	Model Evaluation	uni/multi
Abruzzo			1.42179543	1.1269958	1.4896958	1.1699046		multivariate
Alto Adige			1.84976736	1.4783664	1.8	1.38		multivariate
Barbaresco			1.17	1.5	1.5	1.17		multivariate
Barolo			1.53	1.19	1.5	1.15		multivariate
Basilicata			1.39	1.11	1.43	1.14		multivariate
Bolgheri			1.22	0.97	1.3	1.02		multivariate
Brunello di Montalcino			1.39	1.1	1.45	1.14		multivariate
Campania			1.15	0.9	1.17	0.92		multivariate
Chianti			1.39	1.1	1.23	0.99		multivariate
Chianti Classico			1.76	1.32	1.64	1.27		multivariate
Collio			1.54	1.23	1.53	1.21		multivariate
Etna			1.13	0.91	1.15	0.9		multivariate
Gavi			1.45	1.13	1.45	1.13		multivariate
Irpinia			1.33	1.05	1.37	1.07		multivariate
Langhe			1.44	1.13	1.43	1.1		multivariate
Lugana			1.47	1.18	1.5	1.18		multivariate
Marche			1.3	1.04	1.32	1.05		multivariate
Maremma Toscana			1.46	1.17	1.54	1.21		multivariate

Image F

region_found	_temperatn_humidi		RMSE train	MAE train	RMSE valid	MAE valid	Model Evaluation	uni/multi
Bierzo			1.4306189	1.1301028	1.3741371	1.1253533		multivariate
Calatayud			1.5322872	1.2228402	1.4868748	1.211347		multivariate
Cava			1.5600486	1.2514565	1.6377949	1.2945724		multivariate
Jumilla			1.2145786	0.9789441	1.0662213	0.8286919		multivariate
La Mancha			1.2505966	0.9670859	1.1495622	0.8720931		multivariate
Montsant			1.1374176	0.9063848	1.1160597	0.9097195		multivariate
Navarra			1.4383639	1.1524328	1.3429348	1.0986349		multivariate
Priorat			1.260806	1.0105442	1.2079484	0.9933915		multivariate
Rías Baixas			0.8998228	0.7186248	0.9110695	0.736864		multivariate
Ribera del Duero			1.4902925	1.1927308	1.4293558	1.2100601		multivariate
Rioja			1.4633038	1.1786238	1.4283522	1.1744184		multivariate
Rueda			2.1705007	1.6534561	1.8002381	1.355395		multivariate
Toro			1.4668734	1.1734747	1.3391205	1.1064971		multivariate
Valdeorras			1.4548095	1.1485483	1.3802013	1.1496549		multivariate
Valencia			1.015661	0.8171146	0.9169447	0.7340275		multivariate

Image G



Plot 16

	year	predicted_temp	month_sin	month_cos
12	2021	9.696842421	0	1
13	2021	8.439691437	0.5	0.8660254038
14	2021	9.198606409	0.8660254038	0.5
15	2021	11.65241421	1	0
16	2021	15.08344037	0.8660254038	-0.5
17	2021	19.07531535	0.5	-0.8660254038
18	2021	22.87414615	0	-1
19	2021	25.20226958	-0.5	-0.8660254038
20	2021	25.20381889	-0.8660254038	-0.5
21	2021	22.40573328	-1	0
22	2021	17.73969806	-0.8660254038	0.5
23	2021	13.12003588	-0.5	0.8660254038

Image H

region	YEAR	Average of predicted_temp
Alsace.xlsx	2021	12.74747469
Anjou.xlsx	2021	9.932715972
Bandol.xlsx	2021	15.24960926
Beaujolais.xlsx	2021	21.1907248
Beaujolais-Villages.xlsx	2021	11.11288547
Beaune.xlsx	2021	8.933677493
Bordeaux Blanc.xlsx	2021	19.85279369
Bordeaux Rosé.xlsx	2021	12.36602829
Bordeaux Supérieur.xlsx	2021	9.658063383
Bordeaux.xlsx	2021	13.22989691
Bourgogne.xlsx	2021	9.277891003
Bourgueil.xlsx	2021	11.19715796
Brouilly.xlsx	2021	10.79503783
Cahors.xlsx	2021	12.83431747
Chablis.xlsx	2021	10.76273748
Chambolle-Musigny.xlsx	2021	10.11833698
Chassagne-Montrachet.xlsx	2021	10.11581182
Châteauneuf-du-Pape.xlsx	2021	12.63389758

Image I

region_found	_temperahn_humidi	RMSE train	MAE train	RMSE valid	MAE valid	Model Evaluation	uni/multi	Predicted Annual Temperature (2021)
Abruzzo		1.42179543	1.1269958	1.4896958	1.1699046		multivariate	12.42434637
Alto Adige		1.84976736	1.4783664	1.8	1.38		multivariate	4.00531305
Barbaresco		1.17	1.5	1.5	1.17		multivariate	14.18338473
Barolo		1.53	1.19	1.5	1.15		multivariate	12.5039122
Basilicata		1.39	1.11	1.43	1.14		multivariate	16.3141733
Bolgheri		1.22	0.97	1.3	1.02		multivariate	16.641001
Brunello di Montalcino		1.39	1.1	1.45	1.14		multivariate	14.99083516
Campania		1.15	0.9	1.17	0.92		multivariate	16.47641032
Chianti		1.39	1.1	1.23	0.99		multivariate	11.17503112
Chianti Classico		1.76	1.32	1.64	1.27		multivariate	15.23387948
Collio		1.54	1.23	1.53	1.21		multivariate	11.12410614
Etna		1.13	0.91	1.15	0.9		multivariate	17.30873623
Gavi		1.45	1.13	1.45	1.13		multivariate	14.66035802
Irpinia		1.33	1.05	1.37	1.07		multivariate	15.57896591
Langhe		1.44	1.13	1.43	1.1		multivariate	15.01142046
Lugana		1.47	1.18	1.5	1.18		multivariate	12.84997131
Marche		1.3	1.04	1.32	1.05		multivariate	14.81924674
Maremma Toscana		1.46	1.17	1.54	1.21		multivariate	13.7580865

Image J

region_found	_temperahn_humidi	RMSE train	MAE train	RMSE valid	MAE valid	Model Evaluation	uni/multi	Predicted Annual Temperature (2021)
Bierzo		1.4306189	1.1301028	1.3741371	1.1253533		multivariate	11.20945613
Calatayud		1.5322872	1.2228402	1.4868748	1.211347		multivariate	12.8070648
Cava		1.5600486	1.2514565	1.6377949	1.2945724		multivariate	10.13990022
Jumilla		1.2145786	0.9789441	1.0662213	0.8286919		multivariate	17.33737478
La Mancha		1.2505966	0.9670859	1.1495622	0.8720931		multivariate	12.38976731
Montsant		1.1374176	0.9063848	1.1160597	0.9097195		multivariate	16.38948563
Navarra		1.4383639	1.1524328	1.3429348	1.0986349		multivariate	12.66550517
Priorat		1.260806	1.0105442	1.2079484	0.9933915		multivariate	16.12850352
Rías Baixas		0.8998228	0.7186248	0.9110695	0.736864		multivariate	23.17334656
Ribera del Duero		1.4902925	1.1927308	1.4293558	1.2100601		multivariate	11.78655977
Rioja		1.4633038	1.1786238	1.4283522	1.1744184		multivariate	12.00324192
Rueda		2.1705007	1.6534561	1.8002381	1.355395		multivariate	10.65296159
Toro		1.4668734	1.1734747	1.3391205	1.1064971		multivariate	12.88653846
Valdeorras		1.4548095	1.1485483	1.3802013	1.1496549		multivariate	11.65399542
Valencia		1.015661	0.8171146	0.9169447	0.7340275		multivariate	18.21171647

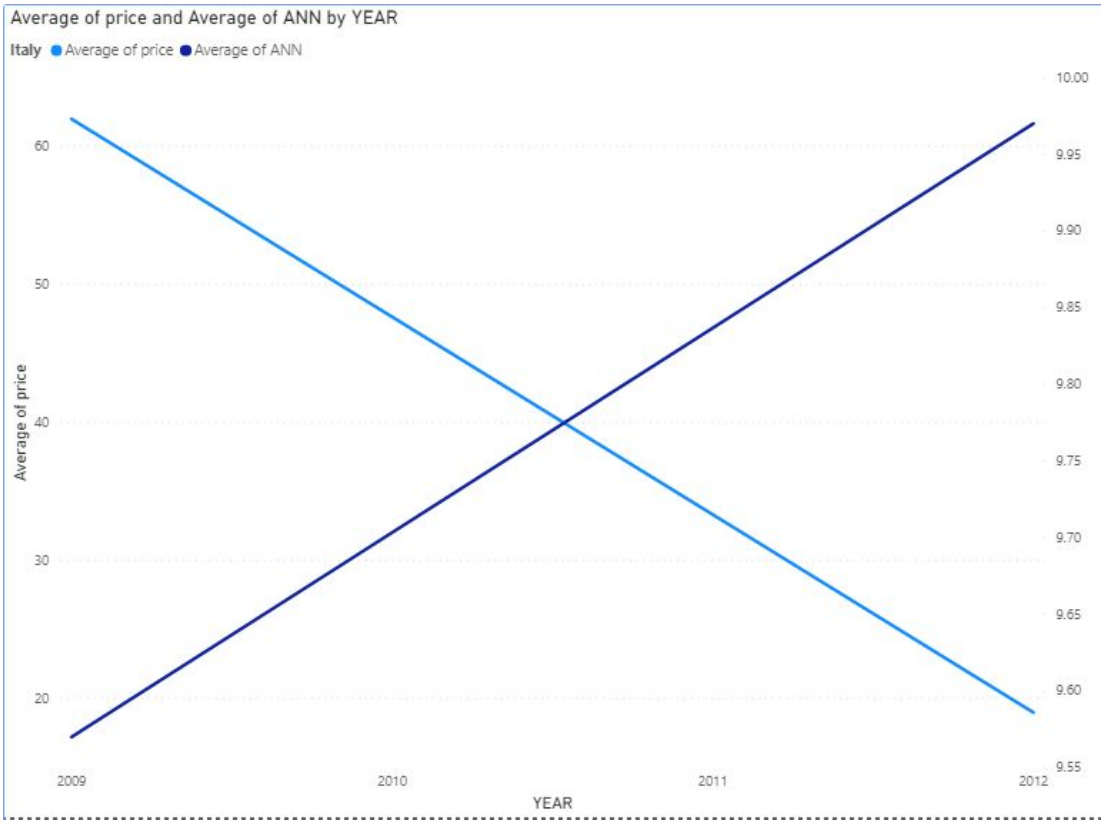
Image K

Discussion

Evaluating our methodology for finding the “Next Big Drink” requires many assumptions. In addition to these assumptions, it is also important to note that our initial goal changed throughout the course of our research. Instead of finding the “Next Big Drink”, we are actually finding the “Next Big Region”. Initially we were attempting to be able to predict the next wine to invest in for the best financial return. Unfortunately, after performing EDA it became clear we did not have sufficient correlation data to confidently predict specific grape variety appreciations. We did, however, have enough regional correlation data to predict which regions to buy wines from if you assume certain conditions.

Let us look at the Rosso del Veronese region in Italy. From EDA we know that this region contains wine varieties that are depreciating in value and that there is a perfect

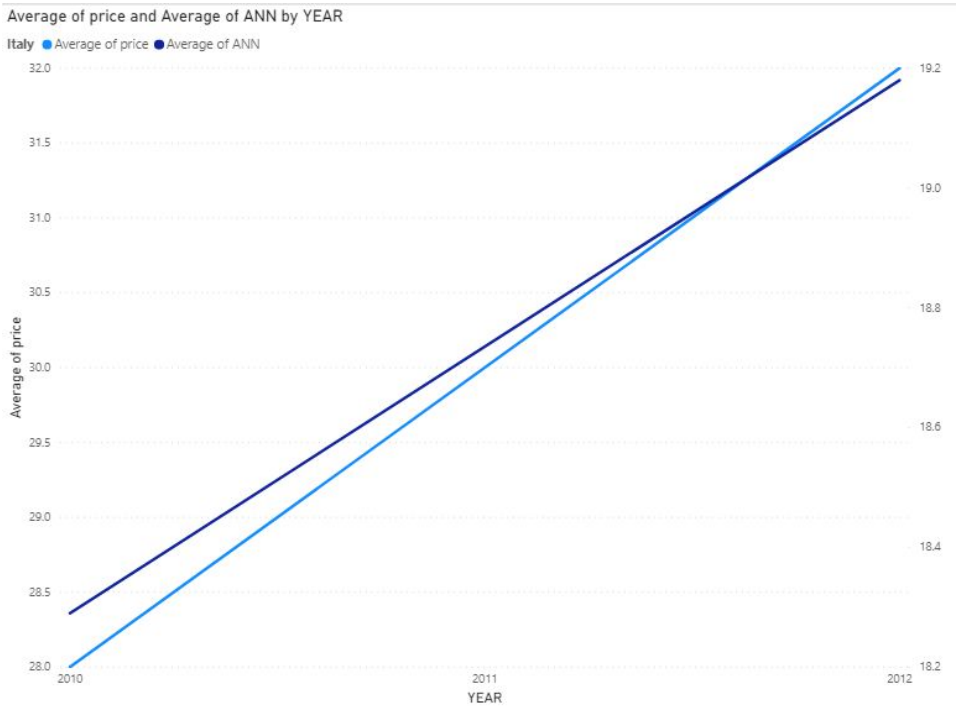
negative correlation (see Plots 17 & 18). You will also notice that annual temperature has been increasing in this region over time. Based on the multivariate LSTM prediction model, Ross del Veronese is expected to increase in temperature to about 11.08 degrees Celsius by the year 2021. We would therefore not recommend investing in wine from this region at the rate the temperature is predicted to rise, especially Red Blends.



Plot 17

Plot 18

Now let us observe the Pompeiano region in Italy. This region has a perfect positive correlation between price and temperature (see Plot 19). The multivariate LSTM model predicted that temperature will increase to 19.38 degrees Celsius by the year 2021. One must now assume that there lies a threshold for temperature. Eventually temperature will increase too much making the region not viable for quality grape growing.



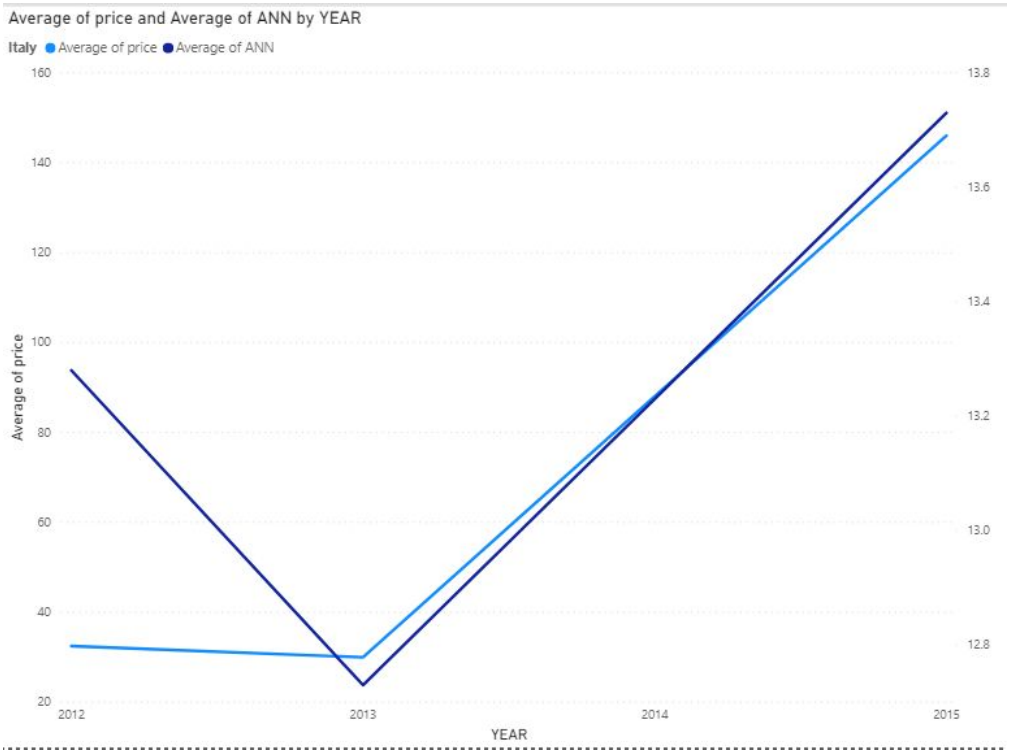
Plot 19

To determine the upper bound temperature threshold of the regions, we also want to see at which temperature the price goes down. After looking at the data, we noticed that there are some regions that exhibit strong negative correlations between temperature and price. This means that the increase in temperature made the price go down. Furthermore, we noticed that there is a strong negative correlation at temperatures near 20 degrees Celsius, but our data is not sufficient to conclude all varieties follow this pattern. Research done by ThirtyFifty (2020), however, shows that the ideal temperature for wines lies between 13 and 20 degrees Celsius. Assuming this holds true, then we would recommend looking into investing in wines from the Pompeiano region since the temperature seems to still lie within the sweet spot.

ThirtyFifty Figure 1

However, for a more confident return on investment for vineyard owners, it is recommended that one does more research about specific grape varieties and their ideal temperatures to verify that the predicted temperature will not surpass the upper threshold. In terms of rate of appreciation, one method would be to calculate the slope of the temperature line from the plot found in Plot 19. Recall that Pompeiano has a perfect positive relationship, so the rate of temperature increase will be similar to the rate of price increase. In this case, by using the first point on the plot and the most recent prediction in 2021, the rate of increase is about 0.10. The owner must decide if this is an ideal rate for increase in prices.

As a next step, we can now proceed to showing how our methodology could have benefitted winemakers eight years ago. During 2012 in the wine region of Costa Toscana, Italy, wine was selling for an average price of €32.50. The next year, 2013, average prices dropped to €30.00. Simultaneously, temperature also experienced a similar downward decline (see Plot 20). With these prices and unpredictable temperatures, nothing suggested investing more in the winemaking production or to even increase production. However, if our methodology would have been applied, then vineyard owners could have had a lot more insight into the future of their wines.

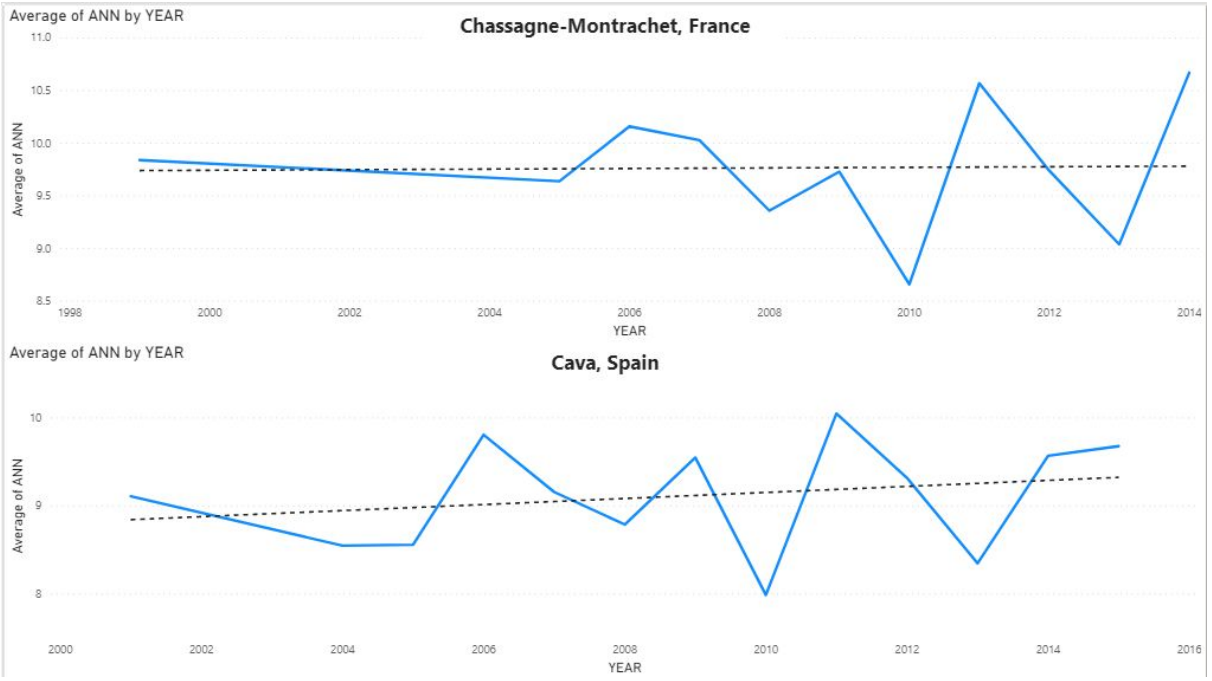


Plot 20

First off, the regional correlation between price and temperature could have been tested. Our research suggests a significantly high correlation of 0.78. With our prediction model, they then would have been able to see that in 2014 and 2015 there were going to be huge temperature spikes, therefore suggesting a surge in prices of wines in the Costa Toscana region. From 2013 to 2015, the average price of wine skyrocketed to €146.00. That is almost a 387% increase in prices. With this foresight, winemakers could have planned accordingly and received a much higher ROI from their grapes than they did without this knowledge.

On a final note, our research is also capable of providing convincing evidence for regions that lie in the unclear temperature category. In Plot 21 you will see line graphs of average annual temperatures for the regions Chassagne-Montrachet, France and Cava, Spain. Chassagne-Montrachet temperatures appear to be unpredictable and contain a less than ideal trendline. With ThirtyFifty's recommended lower bound temperature for wine production being 10 degrees Celsius, winemakers in this region most likely found it

unsettling to have temperatures switch between 9 degrees Celsius and 10 degrees Celsius every year. Over in Spain the Cava region is experiencing similar issues with consistently staying above 10 degrees. Planning for the next few years' wine production would prove difficult without some sort of temperature estimation.



Plot 21

Our prediction model is able to take on this uncertainty and confirm if that year the region will be viable or not. In 2021, we predicted that Chassagne-Montrachet would be 10.12 degrees Celsius on average and Cava would experience an average temperature of 10.14 degrees Celsius. Knowing the region's wine growth viability could potentially be the difference between a profitable year and unprofitable year for wine producers.

Conclusion

Summary

To summarize, we have created a methodology that provides wine investors and vineyard owners the ability to combat the uncertainty surrounding climate change. Our methodology allows users to predict future temperatures of different wine regions in France, Italy and Spain. These predicted temperatures allow people to anticipate regions with high

quality wines, and therefore good investments. The overall accuracy of our best performing model (multivariate LSTM) is an RMSE of 1.4 degrees Celsius. Our methodology requires the following:

1. Gather wine and temperature data similarly structured to our dataframes
2. Determine the correlation between price and temperature broken down by varieties and regions
3. Predict temperatures by using our multivariate LSTM notebook
4. Compare predicted annual temperatures to current temperature data of regions and varieties with strong correlations.
5. Use previous research to determine if grapes will stay within the sweet spot temperatures
6. Calculate the slope of temperature using predictions and previous temperature data to find rate of possible appreciation
7. Choose wine to invest in based on results

By following these steps, one can take much of the guesswork out of choosing which wines to invest in with significantly higher accuracy. This will ensure a higher probability of your wine or wines appreciating, giving you higher returns on investment. Vineyard owners will also receive a better understanding of what their crops are expected to yield in terms of grape quality and possible sales.

Our research also helps to confirm regions that will be viable viticulture regions in the future. Even if regions were previously thought to either not be viable wine growing land, or if they seemed to be showing signs of exiting the sweet spot temperatures based on ThirtyFifty's research, our prediction model can prove otherwise. Our model is also able to carry out the opposite and provide evidence that the land may not be as fertile as it once was due to climate change.

Limitations

This study includes limitations in regards to wine production, brand awareness data, other climate data, and time considerations. Our methodology neglects wine production factors. We do not consider variables such as farming techniques, fertilizer brands, machines, tools, and production methods. We also do not take into consideration the effects that brand awareness has on wine quality and price. We acknowledge that certain brands just tend to be more “high end” and therefore expensive, but collecting this data is incredibly difficult. The approach we took only utilizes temperature data in regards to climate. There are multiple other aspects like precipitation, soil, solar radiation, air pressure, wind speed, wind direction, humidity, and frost.

Additionally, our dataset only contains information about wines sold and judged in 2017. Lastly, due to time constraints, we were unable to provide sufficient research in regards to specific sweet spot temperatures for specific grape varieties. For example, with our dataset, we cannot confidently give an upper threshold where temperatures are too hot to produce certain grape varieties. Our methodology instead relies on many assumptions based on our dataset.

Recommendations

If you would like to further build upon our prediction methodology, then we recommend considering the above mentioned limitations. Taking into consideration these variables could increase model accuracy and performance. Another interesting approach could be to factor in altitude as there is research that finds once inhospitable altitudes to now be great areas for grape vines. One could also more seriously drill down into the data to look at specific bottle data. If more data was gathered than just the 2017 data, then you would be able to look at exactly how much each bottle is appreciating and build more predictive models based off of that information. You could even use clustering models and cluster by year depending on what the weather is doing in those years. We also recommend gathering data from the source to make concatenation easier as well.

In addition to those suggestions for improvement, finding more concrete research about grape varieties and their ideal growing temperatures would greatly contribute to the probability of higher ROI's. This would in turn create more confidence about the exact rate of appreciation. Knowing the sweet spots and upper threshold temperatures for each grape variety is crucial data for the most accurate forecasting of wine appreciation given how sensitive grapes are to temperature.

Another recommendation would be to research some of our interesting findings. Earlier we discussed that there may be a possible point optimization for wine. This would make for an exciting topic and benefit many vineyard owners and wine producers. One could also further study the correlation between temperature and price. The perfect correlations were fascinating results and understanding the why would be extremely beneficial.

References

Abdishakur. (2019, September 15). *Geocode with Python*. Retrieved from Towards Data Science: https://towardsdatascience.com/geocode-with-python-161ec1e62b89

Asimov, E. (2019, October 14). *The Pour: How Climate Change Impacts Wine*. Retrieved from The New York Times: https://www.nytimes.com/interactive/2019/10/14/dining/drinks/climate-change-wine.html

Auvimar. (2017, March 28). *Parker Points: what they are and their scoring system*. Retrieved from Auvimar M. Garcia S.L.: https://www.auvimar.com/en/blog/23_robert-parker.html

Berkeley Earth. (2017, May 1). *Climate Change: Earth Surface Temperature Data*, Version 2. Retrieved May 30, 2020, from Kaggle.com: https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data

Cardebat , J.-M., & Figuet, J.-M. (2004). What explains Bordeaux wine prices? *Applied Economics Letters*, 293-296.

Costanigro, M., McCluskey , J. J., & Mittelhammer, R. C. (2007). Segmenting the Wine Market Based on Price: Hedonic Regression when Different Prices mean Different Products. *Journal of Agricultural Economics*, 454-466.

Diaz-Rainey, I., Robertson, B., & Wilson, C. (2017). Stranded research? Leading finance journals are silent on climate change. *Climactic Change*, 143, 243–260.

Elsworth, S., & Güttel, S. (2020). Time Series Forecasting Using LSTM Networks: A Symbolic Approach. *The University of Manchester, UK*, 3.

Geron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras & Tensorflow: Concepts, Tools and Techniques to Build Intelligent Systems* (2nd edition ed.). California: O'Reilly Media Inc.

Godden, P. W., & Gawel, R. (2008, March 24). *Evaluation of the consistency of wine quality assessments from expert wine tasters.* Retrieved from Wiley Online Library: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1755-0238.2008.00001.x

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *Neural computation.*, 9. 1735-80.

Hodgson, R. T. (2008). An Examination of Judge Reliability at a major. *Journal of Wine Economics*, 105-113.

Italian Wine Central. (2020, April). *Top Fifteen Wine-Producing Countries*. Retrieved from Italian Wine Central: https://italianwinecentral.com/top-fifteen-wine-producing-countries/

Jiang, R. (2020, January 20). *Climate Change Risk in Stock Markets.* Retrieved from UWSPACE: https://uwspace.uwaterloo.ca/handle/10012/15504

Jones, G. V. (2003). Climate and Terroir: Impacts of Climate Variability and Change on Wine. *Geoscience Canada*, 1-2.

Jones, G. V., Storchmann, K., & White, A. M. (2005, December). *Climate Change and Global Wine Quality.* Retrieved from Research Gate: https://www.researchgate.net/profile/Gregory_Jones3/publication/226578343_Climate_Change_and_Global_Wine_Quality/links/02e7e51a95fc606ceb000000/Climate-Change-and-Global-Wine-Quality.pdf

Jones, G., & Hellman, E. (2003). *Oregon Viticulture.* Corvallis: Oregon State University Press.

Lardinois Frederic, L. M. (2017, March 8). *Google is acquiring data science community Kaggle.* Retrieved from https://techcrunch.com/: https://techcrunch.com/2017/03/07/google-is-acquiring-data-science-community-kaggle/

Lardinois, F., Mannes, J., & Lynley, M. (2017, march 8). *Google is acquiring data science community kaggle*. Retrieved may 29, 2020, from https://techcrunch.com/: https://techcrunch.com/2017/03/07/google-is-acquiring-data-science-community-kaggle/

Ling, B.-H., & Lockshin, L. (2003). *Components of Wine Prices for Australian Wine: How Winery Reputation,Wine Quality, Region, Vintage, and Winery Size Contribute to the Price of Varietal Wines.* Retrieved from Academia.edu: https://s3.amazonaws.com/academia.edu.documents/47738042/Components_of_Wine_Prices_for_Australian20160802-5940-ro2xpt.pdf?response-content-disposition=inline%3B%20filename%3DComponents_of_Wine_Prices_for_Australian.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-A

MacMillan, A. (2020, April 3). *The Natural Resources Defense Council*. Retrieved from Global Warming 101: https://www.nrdc.org/stories/global-warming-101#warming

Maracchi, G., Sirotenko, O., & Bindi, M. (2018). Impacts of Present and Future Climate Variability on Agriculture and Forestry in the Temperate Regions: Europe. *Springer*, 2.

NASA. (2019, December 19). *Data Questions*. Retrieved from NASA POWER: https://power.larc.nasa.gov/docs/faqs/data/

Olah, C. (2015, August 27). *Understanding LSTM Networks. Retrieved from https://colah.github.io/: https://colah.github.io/posts/2015-08-Understanding-LSTMs/*

Painter, M. (2020). An inconvenient cost: The effects of climate change on municipal bonds. *Journal of Financial Economics*, 468-482.

Perktold, J., Seabold, & Skipper. (2010). statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference.*

Phi, M. (2018, September 24). *Illustrated Guide to LSTM's and GRU's: A step by step explanation. Retrieved from https://towardsdatascience.com/: https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21*

Richards, C., & Melford, M. (2019, February 25). *What is Global Warming?* Retrieved from National Geographic: https://www.nationalgeographic.com/environment/global-warming/global-warming-overview/

Sanning, L. W., & Shaffer, S. (2008, March). *Bordeaux Wine as a Financial Investment.* Retrieved from Reasearch Gate: https://www.researchgate.net/profile/Sherrill_Shaffer/publication/228574900_Bordeaux_Wine_as_a_Financial_Investment/links/02e7e51d34cb73d6b0000000/Bordeaux-Wine-as-a-Financial-Investment.pdf

Sener, A., Canbas, A., & Ünal, M. Ü. (2006). *The Effect of Fermentation Temperature on the Growth.* Balcal›, Adana, Turkey: University of Çukurova.

The Global Wine Score. (2018, September 26). *How do critic scores affect wine prices? A study of Napa Valley wines.* Retrieved from Medium: https://medium.com/the-global-wine-score/how-do-critic-scores-affect-wine-prices-a-study-of-napa-valley-wines-226d8d08adb7

ThirtyFifty. (2020). *Climate Change and Wine Overview*. Retrieved from ThirtyFifty: https://www.thirtyfifty.co.uk/spotlight-climate-change.asp

WineEnthusiast. (2017, November 22). *Wine Enthusiast*. Retrieved May 29, 2020, from winemag.com: https://www.winemag.com/

Appendix

A

[7]

Score	Explanation
95–100	Classic: a great wine
90–94	Outstanding: a wine of superior character and style
85–89	Very good: a wine with special qualities
80–84	Good: a solid, well-made wine
75–79	Mediocre: a drinkable wine that may have minor flaws
50–74	Not recommended

B

[8]

Score	Explanation
20	Truly exceptional
19	A humdinger
18	A cut above superior
17	Superior
16	Distinguished
15	Average
14	Deadly dull
13	Borderline faulty or unbalanced
12	Faulty or unbalanced

C

[9]

Score	Explanation
5 Stars	Superlative. A Cape Classic
4 Stars	Excellent
3 Stars	Good Everyday Drinking
2 Stars	Casual Quaffing
1 Star	Very Ordinary

D[10]

country	description	designation	points	price	province	region_1	taster_name	taster_twitter_handle	title	variety	winery	years
Spain	Blackberry a	Ars In Vitro	87	15	Northern Spain	Navarra	Michael Schachner	@wineschach	Tandem 2(	Tempranillo-Merlot	Tandem	2011
Italy	Here's a brig	Belsito	87	16	Sicily & Sardinia	Vittoria	Kerin O'Keefe	@kerinokeefe	Terre di Gi	Frappato	Terre di Giurfo	2013
France	This dry and	restrained w	87	24	Alsace	Alsace	Roger Voss	@vossroger	Trimbach 2	Gewürztraminer	Trimbach	2012
France	This has gre	Les Natures	87	27	Alsace	Alsace	Roger Voss	@vossroger	Jean-Bapti	Pinot Gris	Jean-Baptiste Adam	2012
France	This is a dry	wine, very sp	87	30	Alsace	Alsace	Roger Voss	@vossroger	Leon Beye	Gewürztraminer	Leon Beyer	2012
Spain	Desiccated t	Vendimia Se	87	28	Northern Spain	Ribera del Duero	Michael Schachner	@wineschach	Pradorey 2	Tempranillo Blend	Pradorey	2010
Italy	Delicate aro	Feiligno	87	19	Sicily & Sardinia	Sicilia	Kerin O'Keefe	@kerinokeefe	Baglio di P	White Blend	Baglio di Pianetto	2007
Italy	Aromas of pi	Aynat	87	35	Sicily & Sardinia	Sicilia	Kerin O'Keefe	@kerinokeefe	Canicatti 2	Nero d'Avola	Canicatti	2009
Italy	Pretty aroma	Dalila	87	13	Sicily & Sardinia	Terre Siciliane	Kerin O'Keefe	@kerinokeefe	Stemmari	White Blend	Stemmari	2013
Italy	Aromas recall	ripe dark b	87	10	Sicily & Sardinia	Terre Siciliane	Kerin O'Keefe	@kerinokeefe	Stemmari	Nero d'Avola	Stemmari	2013
Italy	Aromas sugg	Mascaria Bar	87	17	Sicily & Sardinia	Cerasuolo di Vittoria	Kerin O'Keefe	@kerinokeefe	Terre di Gi	Red Blend	Terre di Giurfo	2011
Italy	This concent	Missoni	86	21	Sicily & Sardinia	Sicilia			Feudi del I	Cabernet Sauvignon	Feudi del Pisciotto	2010
Italy	Inky in color	I Tratturi	86	11	Southern Italy	Puglia			Feudi di S	Primitivo	Feudi di San Marzano	2011
Italy	Part of the n	Purato Made	86	12	Sicily & Sardinia	Sicilia			Feudo di S	Nero d'Avola	Feudo di Santa Tresa	2011
Italy	Catarratto is	one of Sicily'	86	17	Sicily & Sardinia	Sicilia			Feudo Mor	Catarratto	Feudo Montoni	2011
France	This is a fest	Nouveau	86	9	Beaujolais	Beaujolais	Roger Voss	@vossroger	Henry Fess	Gamay	Henry Fessy	2012
Italy	Spicy, fresh	Sallier de la	86	13	Sicily & Sardinia	Sicilia			Tasca d'Alr	Inzolia	Tasca d'Almerita	2011
France	Soft and frui	Eté Indien	86	14	Beaujolais	Brouilly	Roger Voss	@vossroger	Vignerons	Gamay	Vignerons de Bel Air	2011
Italy	The Monica	Dolia	85	14	Sicily & Sardinia	Monica di Sardegna			Cantine di	Monica	Cantine di Dolianova	2010
France	Fruity and liq	La Fleur d'Ar	85	15	Bordeaux	Bordeaux Blanc	Roger Voss	@vossroger	Château di	Bordeaux-style White Blend	Château de Sours	2011
Italy	There's a tot	Sallier de la	85	13	Sicily & Sardinia	Sicilia			Tasca d'Alr	Grillo	Tasca d'Almerita	2011
Italy	This densely	Prugneto	86	17	Central Italy	Romagna	Kerin O'Keefe	@kerinokeefe	Podere da	Sangiovese	Podere dal Nespoli	2015
France	From the warm	2015 vinta	86	24	Burgundy	Chablis	Roger Voss	@vossroger	Simonnet-	Chardonnay	Simonnet-Febvre	2015
France	This soft, rou	nded wine i	86	15	Burgundy	Mâcon-Milly Lamartine	Roger Voss	@vossroger	Vignerons	Chardonnay	Vignerons des Terres Secrètes	2015
Italy	Aromas of bl	Daginestra	86	32	Southern Italy	Aglianico del Vulture	Kerin O'Keefe	@kerinokeefe	Grifalco 20	Aglianico	Grifalco	2013
Spain	Bland, mature	aromas of	86	16	Galicia	Rias Baixas	Michael Schachner	@wineschach	Spyro 2014	Albariño	Spyro	2014
France	This fruity, s	La Réserve	86	11	France Other	Vin de France	Roger Voss	@vossroger	Lionel Osm	Petit Manseng	Lionel Osmin & Cie	2016
Italy	Subdued arc	Le Mandorle	86	29	Tuscany	Vernaccia di San Gimignano	Kerin O'Keefe	@kerinokeefe	Poggio Alle	Vernaccia	Poggio Alloro	2014
Italy	Made primarily	from Sang	88	19	Tuscany	Toscana	Kerin O'Keefe	@kerinokeefe	Fattoria Sa	Rosato	Fattoria Sardi	2015
France	This is a dense	wine, pack	88	20	Beaujolais	Juliénas	Roger Voss	@vossroger	Henry Fess	Gamay	Henry Fessy	2015

E[11]

R-squared	classification	coefficient_points	coefficient_years	country	grape_variety	p-value-points	p-value-year	region
0.06034112	depreciating	0.537949697	0.342611252	France	Gamay	0.775560622	0.603018294	Moulin-à-Vent
0.257533398	depreciating	2.677433064	1.227865137	France	Pinot Noir	0.494187541	0.481827648	Aloxe-Corton
0.507892319	depreciating	1.976676236	0.070587547	France	Chardonnay	0.043413981	0.823231915	Mâcon-Villages
0.454937558	depreciating	2.857927005	-0.064684727	France	Sparkling Blend	0.171346845	0.366602601	Crémant d'Alsace
0.3049519	depreciating	14.59181096	1.289485692	France	Bordeaux-style Red Blend	0.067908125	0.515859354	Saint-Julien
1	depreciating	2.076923077	-0.346153846	France	Rosé			Corse
0.336879342	depreciating	1.74672827	-4.279854541	France	Rhône-style White Blend	0.819048521	0.198170037	Hermitage
0.546942972	depreciating	-6.2761672	2.897076342	France	Syrah	0.435454397	0.101620067	Hermitage
0.836819684	depreciating	4.074680074	-0.600918651	France	Chardonnay	0.029855255	0.162299322	Beaujolais Blanc
0.190050179	depreciating	1.531958785	-0.195273835	France	Bordeaux-style Red Blend	0.181572277	0.579044389	Graves
0.097358226	depreciating	1.106188087	0.186714279	France	Bordeaux-style White Blend	0.361148791	0.689187305	Graves
1	depreciating	-54.76190476	-79.23809524	France	Provence red blend			Bandol
0.850151063	depreciating	3.421222996	-0.470750532	France	Rhône-style Red Blend	0.07803432	0.691993174	Bandol
0.48918731	depreciating	2.04264409	-0.500945586	France	Rosé	0.084381235	0.456048923	Bandol
0.212406028	depreciating	4.268355392	-2.137093233	France	Rhône-style Red Blend	0.402786857	0.575440292	Gigondas
0.615791007	appreciating	7.866601805	-2.093687612	France	Sauvignon Blanc	0.001031404	0.016997897	Pouilly-Fumé
0.49761636	depreciating	2.055788028	0.53410063	France	Gamay	0.082816447	0.224725908	Chiroubles
0.255902615	depreciating	0.284275767	0.325872246	France	Gamay	0.777096462	0.344028649	Juliénas
0.256875643	depreciating	0.391629927	0.06426474	France	Bordeaux-style White Blend	0.307483326	0.587260232	Entre-Deux-Mers
0.624851955	appreciating	0.34703546	-1.7060857	France	Red Blend	0.72869482	0.023025488	Madiran

F[12]

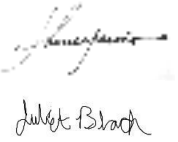
title	variety	winery	YEAR	LAT	LON	PARAMETER	ANN	Month	Month Avg Temp
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	January	*4.62*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	February	*4.92*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	March	*7.34*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	April	*10.73*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	May	*12.78*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	June	*17.6*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	July	*19.17*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	August	*16.65*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	September	*16.37*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	October	*13.03*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	November	*8.23*
La Chablisienne 2014 La Sereine (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	December	*3.02*
La Chablisienne 2014 Fourchaume Premier Cru (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	January	*4.62*
La Chablisienne 2014 Fourchaume Premier Cru (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	February	*4.92*
La Chablisienne 2014 Fourchaume Premier Cru (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	March	*7.34*
La Chablisienne 2014 Fourchaume Premier Cru (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	April	*10.73*
La Chablisienne 2014 Fourchaume Premier Cru (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	May	*12.78*
La Chablisienne 2014 Fourchaume Premier Cru (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	June	*17.6*
La Chablisienne 2014 Fourchaume Premier Cru (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	July	*19.17*
La Chablisienne 2014 Fourchaume Premier Cru (Chablis)	Chardonnay	La Chablisienne	*2014*	*47.81522*	*3.80018*	T2M	*11.23*	August	*16.65*

[11] Preview of All_Countries_Classification_Filter_10_Region_Count Dataframe
[12] Preview All_Country_Data (all previous data with temperature data)

Statement of Certification Joint Master's Thesis

We hereby confirm that the Group Work presented has been prepared independently according to the agreed work plan, using no other sources, resources and other aids than those mentioned. All parts – literally or by their meaning – taken from published or non-published sources are credited as such. The Group Work in its current or similar form has never been submitted as a graded assignment. Changes to the topic or the work plan have been agreed upon with the first assessor.

Frankfurt am Main, 12-07-2020
Boise, 12-07-2020



City, Date | Authors' Signatures