Tasting Notes and Top Dollar: the Correlation between Flavor Notes, Prices, and Ratings of Teas.

Samuel Wang

December 11, 2022

Introduction

Almost every year, the global consumption and market value of tea increases (Statista, 2022a). Apart from 2020, the global market value of tea grew every year from $142.2 billion in 2014 to $207.1 billion in 2021, a 45.6% increase. The US alone imported more than $530 million of tea in 2021, with the largest fraction of tea imported being black tea (Statista, 2022b). In the US, the specialty tea market was more than $3 billion in 2020, which is defined as teas sold in specialty tea shops or from tea vendors that aren't traditional markets like supermarkets or drink shops.

Although other beverages made from brewed plant matter are also known as tea (eg. chamomile tea), this study will only look at teas made solely from the Camellia sinensis plant as they make up the largest share of the global market. While all tea is made from the Camellia sinensis plant, the way in which the fresh tea is processed determines the type of tea that is ultimately produced (Liang et al., 21). This study will look at three categories of specialty tea: green tea, black tea, and pu'er. Teas that have been flavored with other compounds (eg. earl gray, jasmine) have been excluded to avoid comparisons between flavored and unflavored teas, as the focus of this paper is on unflavored, specialty teas. Each category of tea has a broad range of possible flavors, external, and internal cues about the quality of the tea. While previous studies have looked at external cues such as geographic labeling and marketing, no studies have looked at how internal cues like flavor notes affect perceived price and quality. Using a crowd-sourced review website, this study will determine correlations between the presence of various flavor notes and the prices and the average ratings of teas for the three aforementioned categories.

Specialty teas

Originating from southwest China, tea has a long history in China and around the world, with the earliest usage of tea as a medicine dating back to at least 141 B.C. (Zhang et al., 2019; Cherfas, 2016). As it has spread across the world, it has become one of the most popular caffeinated beverages. Today, specialty tea meant to be brewed by the end consumer is sold in three types of packaging: bagged, loose, or compacted. Compacted teas are usually made of fermented teas pressed into "cakes" meant to be broken up before being brewed (Figure 1). All three types of tea packaging were included in this study together in each of the six tea categories.

Figure 1 Cake of pu'erh tea (Gorazd, 2012)

There are also different cultivars of Camellia sinensis, the most well-known being var. Assamica (Lee et al, 2013). Since the specific cultivar of tea is often not specified to the consumer or even known by the producer, no distinction was made in the statistical analysis.

Processing and growing region

The single biggest factor affecting the finished tea is the processing of the freshly picked tea. Previous studies have found differing concentrations of caffeine and amino acids in various types of tea depending on processing method, which greatly affect the drinking experience of the consumer (Zhang et al., 2019). Depending on the level of processing and desired properties, teas can have a wide range of flavors from savory and fresh to sweet and buttery. In order of

oxidation, green tea is the least oxidized, followed by white tea, yellow tea, oolong tea, black tea, and then fermented teas like pu'er and heicha. Similar to beer production, many tea factories have a long history of tea production with cakes from the most famous factories often fetching hundreds of US dollars for a single 357g "cake" of tea (Zhang and Sivaramakrishnan, 2013).

Along with the production process, the source of the fresh tea leaf and the terroir (growing conditions) of the farming region also play a large role in the price of a tea. Similarly to how different wine-producing regions are marketed depending on the terroir of the vineyard, teas are also marketed depending on their unique elevation and growing climate (Atkin and Johnson, 2010). Some of the most prized tea growing regions in the world include the Alishan mountains in Taiwan, the Wuyishan mountains in China, and Darjeeling in India (Wang et al., 2021). Teas from these regions can often fetch far higher prices. As a result, there is much more tea that is claimed to be from these regions than is actually produced, leading to some government imposed regulations on teas from certain growing areas.

Tea is also processed differently in different countries. In China, green teas are most commonly pan fried, whereas in Japan the most common way of "killing" the green is by steaming. These variations have large effects on the brewed tea even if the base material is the same.

Flavor notes in teas

As processing methods and growing regions vary from country to country, there are large differences in the flavors associated with different countries even within the same tea category. A study on green teas from different parts of the world found that Japanese green teas were much more likely to exhibit notes of seaweed or grass, Chinese green teas were more likely to have

notes of tobacco or apricot (Lee et al., 2013). The consistent preparation methods and tasting group used by the study allowed large amounts of information about each tea. Using data clusters, the study found flavors that were much more prevalent in certain countries and characteristics that were prevalent in processing methods that were shared between countries. This sheds insight on the attributes produced by certain methods of production and allows for analysis of the specific attributes and their effects on perceived value and quality.

Methodology

The goal of this paper is to analyze the correlation between various flavor notes and prices, and to a lesser extent their perceived quality or rating. This is important because, as stated earlier, internal cues about a tea's geographic origin and quality can influence a consumer's likelihood to purchase similar teas or more teas from the same vendor (Wang et al., 2021). To gather a sufficient amount of data from the tea-drinking public, this study will use the website Steepster (steepster.com), an online forum that publishes crowd-sourced reviews about teas. Similar to restaurant review websites like Yelp, Steepster allows anyone that has an account to write a review for a tea, along with a rating and flavor notes. Steepster then aggregates the reviews to find an average rating and the most popular notes for each tea.

The reason that these three specific types of tea were chosen (green, black, pu'er) is that they represent a broad spectrum of the possible processing methods of tea. As mentioned earlier, green tea is the least oxidized, while black tea is fully oxidized but not fermented, while pu'er has undergone some level of fermentation (Xu, 2021). Any specialty tea will fall somewhere in between these three tea types, and the trends in data can be broadly applied to other categories of

tea depending on where they fall on the spectrum (e.g. an oolong that's between a green and a black may have flavor characteristics of both).

Data gathering

The page for each tea contains the name of the tea, the vendor that sells it, the category of the tea, the ingredients of the tea, and the most common flavor notes given by reviewers (Figure 2). When writing a review for a tea, each reviewer has the option of leaving a rating, description, brewing parameters, and a list of preset flavor notes from a dropdown menu (Figure 3). The reason that this website was chosen is because it offers a listing of over 80,000 teas and more than 200,000 ratings (steepster.com). It is the only way to gather a large enough sample size of reviews on each tea to come to a statistically significant result. While the website does have affiliate links to many of the teas listed,

Figure 2 Steepster tea review (steepster.com)

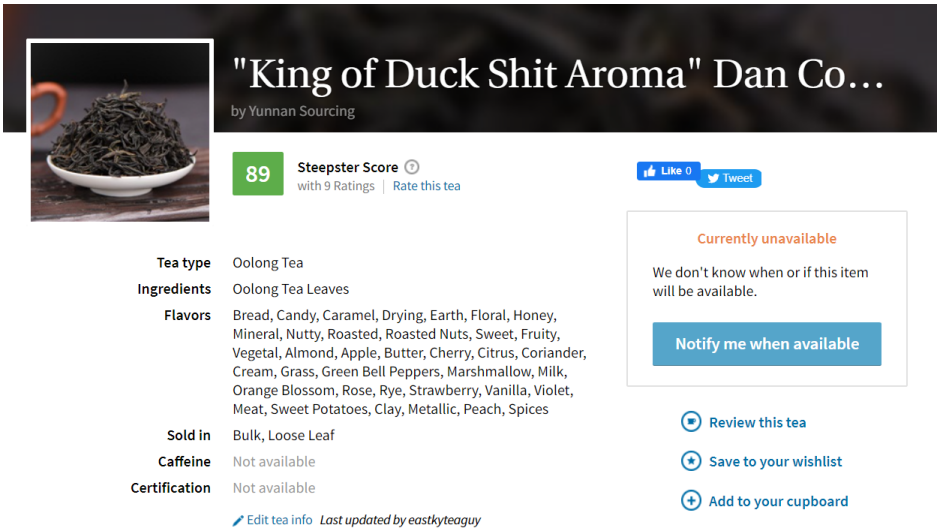
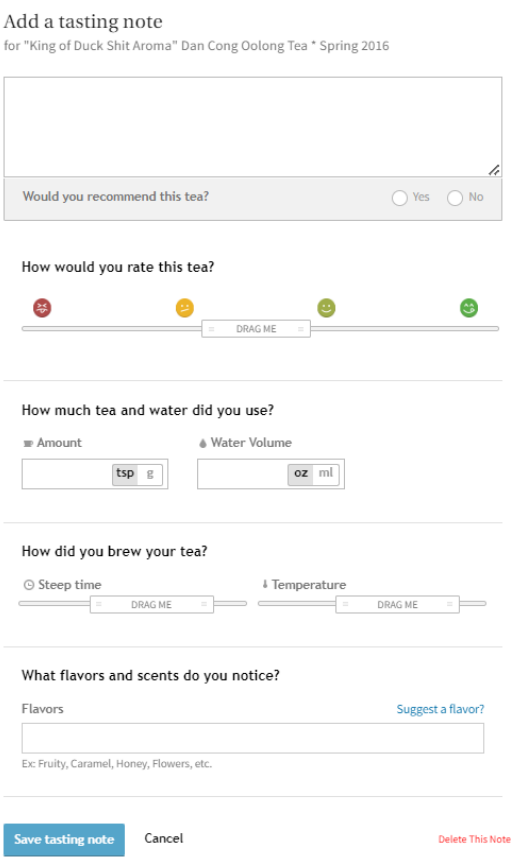


Figure 3 Review options (steepster.com)



Steepster itself doesn't prioritize tea with affiliate links in searches or when sorting by popularity.

To gather data points for analysis, the names, flavor notes, and ratings for the 30 most popular, unflavored teas that are still being sold for the three categories mentioned earlier were entered into a spreadsheet. For example, a black tea would only be included if its only ingredients were black tea leaves. The reason that flavored teas and blended teas were excluded from the study was that including flavored teas would result in a sample population that compared the prices of flavored teas to unflavored teas and dilute the key tasting notes of a tea category that most influence price. Only teas that are still currently being sold are included because it's difficult to find prices for teas that aren't being sold by the original vendor. Price was determined by finding the unit price in dollars per gram for the smallest amount possible for purchase that was 20 grams or greater. The reason 20 grams was chosen as a minimum is that many tea vendors sell samples of teas for less than 20 grams at a premium, so 20 grams was chosen as a baseline to make comparing true prices fairer. In cases where the tea could be purchased directly from the vendor, the price was calculated using the price listed by the vendor. In all other cases, the price was calculated using the lowest price listing on Amazon (amazon.com).

In some cases, there were multiple iterations of a tea that were on sale from the same vendor from different years. This was especially prevalent for pu'er teas, as factories and vendors often repeat the same recipe and market under the same name to build name recognition. Some recipes, like the Dayi 7542 are very well-known among pe'er enthusiasts and can be sold for much more than cakes of similar quality (Zhang and Sivaramakrishnan, 2013). In order to keep consistency between teas, the most recent batch available for purchase was chosen for price calculations.

Data analysis

Although all flavor notes were copied from the Steepster tea page onto a spreadsheet, not all of the flavors listed were used in the analysis (Figure 4). Many individual teas had more than fifteen flavor characteristics listed, so a custom Python function was written to create a list of the significant ones (Van Rossum et al., 2009). The order that the flavor notes are listed is first in order of number of reviews that included a certain tasting note, then in alphabetical order. A method that only takes the first ten notes may miss some notes that were mentioned just as frequently because of alphabetical ordering. To solve this issue, the function takes the first fifteen notes and continues adding notes until the alphabetical order resets to an earlier first letter. The reason that not all flavor notes were considered is because considering all tasting notes would allow flavor notes that were only mentioned once or twice to affect the analysis, and by only taking the significant flavor notes the main attributes of each tea are better represented.

Figure 4 Screenshot of part of spreadsheet

Name	Type	Vendor	Price $/g	Ratings	Flavor notes
English Bre	Black	Twinings	0.13	75	Biting, Bitter, Tannic, Malt, Tannin, Autumn Leaf Pile, Bread, Cinn
Laoshan B	Black	Verdant Te	0.22	88	Malt, Mineral, Bitter, Cacao, Chocolate, Floral, Roasted Barley, R
Bailin Gon	Black	Teavivre	0.21	88	Caramel, Malt, Cocoa, Dark Chocolate, Honey, Bread, Burnt Suga
Awake Eng	Black	tazo	0.25	59	Tannic, Tea, Astringent, Autumn Leaf Pile, Bitter, Leather, Raisins,
Irish Break	Black	twinings	0.13	89	Bread, Creamy, Honey, Malt, Molasses, Raisins, Smooth, Sweet P
PG Tips	Black	PG Tips	0.22	74	Dark Wood, Heavy, Malt, Tea, Green Beans, Metallic, Sawdust, C
Black Drag	Black	adagio	0.28	82	Cocoa, Hay, Honey, Leather, Chocolate, Earth, Apricot, Bread, M
Lipton Bla	Black	lipton	0.02	85	Wood, Lemon, Astringent, Bitter, Herbs, Medicinal, Tea, Honey, S
Fengqing D	Black	teavivre	0.18	85	Jasmine, Chocolate, Leather, Toffee, Dark Chocolate, Wood, Ber
Golden Fle	Black	Verdant Te	0.47	88	Butter, Grapes, Hay, Medicinal, Sour, Almond, Graham Cracker, G
Irish Break	Black	adagio	0.11	78	Cocoa, Drying, Malt, Metallic, Sweet Potatoes, Caramel, Cardboa
Award Wir	Black	Teavivre	0.27	89	Bread, Creamy, Honey, Malt, Molasses, Raisins, Smooth, Sweet P
Zhu Rong \	Black	Verdant Te	0.25	89	Apricot, Caramel, Dark Chocolate, Wet Moss, Yams, Nuts, Nutty,
Premium C	Black	Teavivre	0.2	87	Malt, Caramel, Dark Chocolate, Bitter, Sweet Potatoes, Bread, Co
nepal blac	Black	davids tea	0.3	83	Ash, Honey, Malt, Smoke, Tannin, Wet Earth, Wood, Wet Wood,
Golden Mo	Black	teavana	0.81	71	Caramel, Cocoa, Honey, Smoke, Sweet, Astringent, Bitter, Hay, M
Yu Lu Yan	Black	Verdant Te	0.25	86	Chocolate, Cinnamon, Honey, Stonefruit, Cocoa, Honeysuckle, M
Queen Cat	Black	harney and	0.08	83	Chocolate, Cream, Smoke, Honey, Smooth, Bergamot, Malt, Swe
Golden Mo	Black	adagio	0.25	81	Apple, Chestnut, Malt, Smoked, Chocolate, Honey, Dark Chocolat

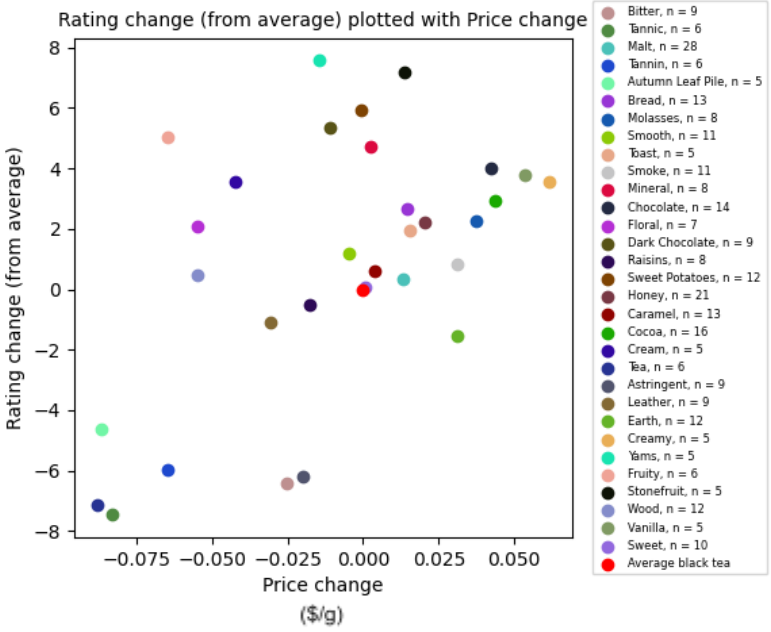
After all the teas were entered in dictionaries along with all their relevant information, the price and rating combination for each significant taste characteristic was also organized into

dictionaries, from which the averages for each flavor note could then be found. Only data from characteristics with at least a sample size of five teas were used in the statistical analysis, as anything less would lose all explanatory power. The tool used to create the graphs was Matplotlib, which is a library for Python (Hunter, 2007). Due to the nature of crowd-sourced tea reviews, most reviewers only listed a couple flavor notes for each tea. This combined with the sheer number of possible flavor notes resulted in only a handful of tasting notes that were listed on at least five teas in the fifteen most frequent positions. Thus, no test for statistical significance was performed due to a small sample size. Still, the results can be useful as a preliminary indicator of possible flavor notes for future investigation, as well as being useful for analyzing possible trends in attributes that were perceived as more premium or higher quality.

Results

Black tea results

Figure 5 Black tea graph



The presence of key flavor characteristics of black tea and their effects on price and rating can be seen in the graph in comparison to the price and rating of the average black tea of all 30 teas in the sample (red dot) (Figure 5). The most common flavor note was malt (n = 28). Nearly all thirty of

the black teas had malt listed as a flavor note, which is consistent with the complete self-oxidation of black tea from enzymes already existing within the cell walls of the leaf. This shares some similarity to malted grain, which also produces “malty” flavors from enzymes contained within the seed itself. Both the oxidized tea leaf and grain are then roasted to highlight the malty flavor. This suggests malted grain could be used as a flavor enhancer or substitute for black teas in beverages such as bubble tea or in other black tea flavored foods.

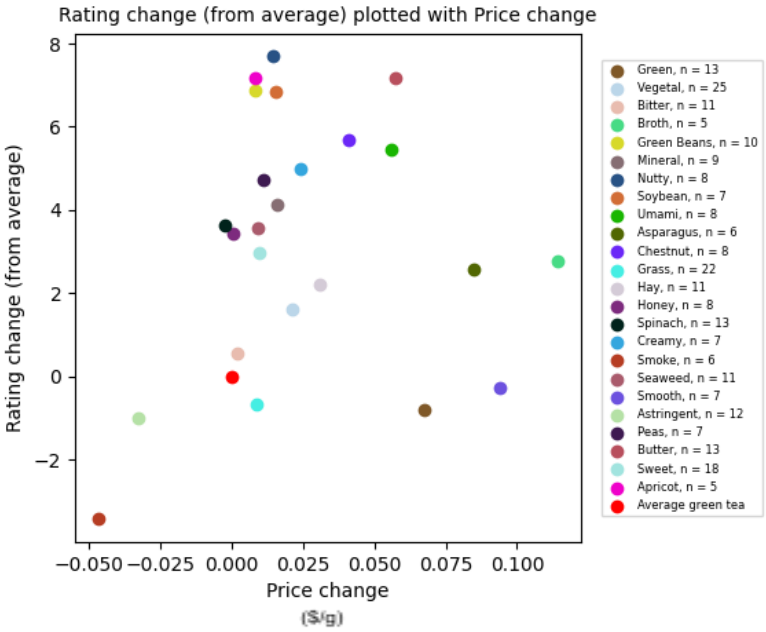
The most negative rating characteristic, tannic ($r = -7.5$), and the most negative price characteristic, “tea” ($p = -0.088$ \$/g), are both common in cheap, bagged teas from brands like Lipton or Twining. This provides a possible explanation for why many associate black teas so strongly with harshly astringent brews that need to be tamed with milk. Other extremely negative descriptors like tannins, bitter and astringent also trigger similar taste responses. The note “Autumn leaf pile” has the second lowest price correlation ($p = -0.086$ \$/g) and a strong negative rating correlation. Interestingly, the notes autumn leaf pile and astringency have a neutral or positive rating correlation for the other two tea types, which suggests they are not inherently negative, rather they are perceived negatively in the context of black teas because they may be perceived as off-notes or unexpected.

On the other hand, the note connected with the most expensive teas is creaminess ($p = 0.062$ \$/g). This is different from cream flavor in that one is a milky taste (cream), while the other is mainly used to describe mouthfeel (creamy). This is reflected in traditional Chinese tea culture, where some tasters place heavy emphasis on the mouthfeel and viscosity of a tea as an indicator of high quality base material (Li, 1993). Roasted notes such as yams, bread, and chocolate were correlated very strongly with better ratings. This argues that most specialty tea consumers prefer black teas with darker fruit notes such as sweet potatoes or fruits with a deeper

sweetness like stone fruits. Notes like wood, smoke, and earthiness did not correlate strongly with ratings either way, although they did vary with prices.

Green tea results

Figure 6 Green tea graph



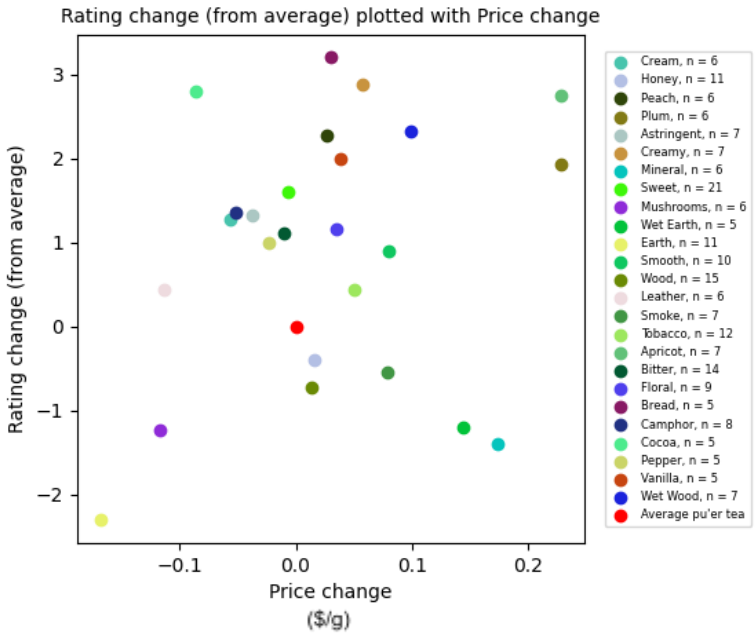
Similarly to black tea, the descriptor that had the biggest correlation with higher prices was a mouthfeel descriptor, namely brothiness (p = 0.11 $/g) (Figure 6). Creaminess and brothiness both describe the feeling of the brewed tea in the mouth. Another attribute with a strong positive correlation with price is smoothness (p = 0.094 $/g).

While astringency was still associated with a negative price difference and slightly negative rating difference, the gap is much smaller than with black teas. Unlike with black teas, smokiness is connected with both the strongest negative rating change (r = -3.4) and the strongest negative price change (p = -0.047 $/g). This is in line with traditional green tea processing methods, as they are never smoked traditionally, while they may be naturally astringent due to the lack of oxidation.

Some of the flavors most strongly associated with positive rating changes were nuttiness (r = 7.7), apricot (r = 7.2), and chestnut (r = 5.7). These flavors are commonly associated with Longjing tea, a flat green tea popular in China for its walnut skin aftertaste (Gascoyne et al., 2011). These data points suggest that consumers prefer similar types of teas with the same unique fruity tannic aftertaste. Other attributes associated with savoriness, like umami, seaweed, and soybeans were associated strongly with positive ratings and slight price increases. These notes, most common in Japanese green teas, are perhaps indicative of the space that Japanese green teas have carved out in the market as premium alternatives to Chinese green teas that offer a savory bite. While creaminess was still correlated with an increase in price (p = 0.024 $/g), the change is smaller than with black teas. This again indicates the idea that many descriptors are not always inherently positive or negative, rather they interact with other flavors and textures that paint a larger picture of the quality of the tea.

Pu'er tea results

Figure 7 Pu'er tea graph



Compared to the graphs for black tea and green tea, the range of rating differences for pu'er is much smaller, while the range of price differences is much larger. There are three

possible explanations for this phenomenon.

The first possible explanation is that because pu'er is a much smaller market, at least in the West, compared to the markets for green and black tea, there aren't any large brands that can dominate the market with efficient supply chains, so vendors that sell pu'er specialize in pu'er, so the differences between the worst and best pu'er teas are smaller while the already high prices exacerbate differences in prices.

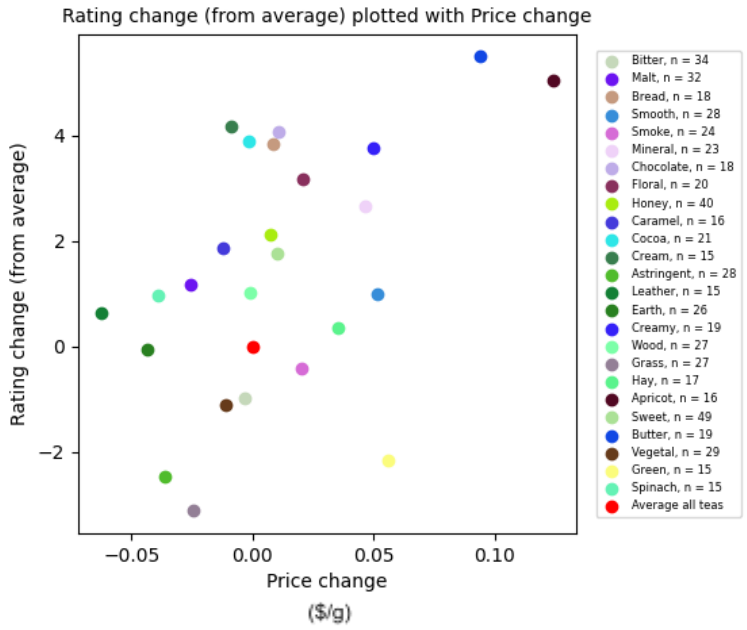
A second explanation could be that because much less people drink pu'er, proportionally more of the reviews are written by the same people. This averages out the ratings while the prices, which the reviewers can't control, diverge.

A third explanation could be that consumers are willing to pay much higher unit prices for marginal gains in drinking experience because that's the crowd the pu'er attracts. In the past, pu'er prices have shot up and down, and owning an expensive cake of tea could be a symbol of wealth and power (Zhang and Sivaramakrishnan, 2013). Perhaps there is a portion of pu'er drinkers that don't notice much difference between the teas and buy teas primarily for collecting rather than drinking.

Interestingly, earth and wood are both negatively associated with rating but wet wood and wet earth are associated positively. The mineral note tends to be expensive but also appears to be undesirable in context. The most expensive characteristics are apricot (p = $0.23) and plums (p = 0.35 $/g). These are both stone fruits, similar to the stone fruit descriptor in figure 5. The note of bread also correlated with higher ratings similarly to black teas. Even the most negative note, earthiness (r = -2.3), is correlated with less than half of the negative change as with tannic and the average black tea.

All teas

Figure 8 All teas graph



Combining all the teas together into one dictionary, then considering tasting notes that appeared in at least fifteen teas out of the ninety total teas produces the graph on the left (Figure 8). The most positive rating correlation was butter (r = 5.5), the most positive price correlation was apricot (p = 0.12 $/g), the least positive ratinge correlation was grass (r = 3.1), and the least positive price correlation was leather (p = -0.063 $/g). While the results are interesting to look at, they have limited usefulness because we've already determined that most attributes by themselves aren't necessarily positive or negative; they depend on the other flavors and textures and the type of tea that the consumer expects.

Discussion

In general, the attributes with the biggest implications (creamy, brothy, buttery) for price have to do with mouthfeel and the viscosity of the tea. As mentioned previously, this falls in line with the traditional belief the mouthfeel is directly related to the quality and picking standard of

the fresh tea leaf. In contrast, the flavor notes that have the greatest implications on the rating of the tea tend to be taste and smell oriented (yams, apricots, bread).

Implications

The most obvious application of the results is in marketing of teas. Based on the analysis, it appears that vendors should focus on creamy textures and sweet, darker notes like sweet potatoes and chocolate when marketing or describing their black teas. Vendors should market their green teas as brothy and nutty or brothy and savory, as those two clusters of tea appear to have the highest ratings and price differences. While no single flavor not was as prevalent for pu'er tea, the smaller range of rating changes and large price differences suggest that vendors should focus more on building their brand and reputation for certain recipes through marketing strategies like unique geographic locations and processing methods with a focus on stone fruit notes like apricot and plum.

One interesting possible application of this data is artificially boosting or recreating the flavors of teas in ready to drink (RTD) beverages or other snacks. The similarities between black tea and malted grain could open the door for more use of roasted, malted grains to be used in conjunction with black tea in beverages such as milk teas or boba drinks.

The most common flavor note for green teas was vegetal (n = 25), which shares no direct comparison as it's caused by a mixture of tannins and amino acids (Mukai, 1992). The prevalence of savory flavors such as seaweed and soybeans could be used to add other sources of umami to green tea flavored foods to amplify the flavor of the tea itself.

Limitations

A major limitation of all the data in this study is a small sample size, or a lack of statistical significance. Only a certain number of teas have enough reviews before a lack of reviews creates too much individual variance within each aggregate tea review as a whole.

The nature of the reviews presents a problem in that the tastings are not done blind. The reviewers are exposed to marketing that primes them to pick certain flavor notes. Reviewers are also more likely to rate a tea higher if they paid more for it, creating a self-fulfilling prophecy.

Another major limitation is the source of the sample population, steepster.com. The users on the website are not evenly distributed among all teas of a category. If a user buys one expensive tea, they are also much more likely to buy other expensive teas, and the same is true for teas from a certain country. That means that ratings are likely to be amplified for tea that are similar to each other.

Areas for future research

Future research could investigate the combinations of flavor notes that make a flavor desirable in one circumstance, and not desirable in another. For example, astringency was very undesirable in the sample population of black teas, but was almost neutral in the sample of green teas. An experiment could be conducted on the circumstances that make an off note in one tea a positive flavor in another.

There's also much more room for research between the three types of teas picked for this study. Oolong teas alone have dozens of processing methods, geographic locations, and picking standards. Doing research on the other types of teas would bring much needed context into other factors and attributes that contribute to a tea's perceived cost and quality.

Finally, there's a gap in combining the findings of a study on external cues and a study on internal cues. For example, if a green tea is described as grassy, how does that priming affect the tasting experience of the taster? Similar experiments could also be done for geographic labeling, awards given to the tea, etc.

Reference

Atkin, T., & Johnson, R. (2010). Appellation as an indicator of quality. International Journal of Wine Business Research, 22(1), 42–61. https://doi.org/10.1108/17511061011035198

Cherfas. (2016, January 26). World's Oldest Tea Discovered In An Ancient Chinese Emperor's Tomb. NPR.org. https://www.npr.org/sections/thesalt/2016/01/26/464437173/worlds-oldest-tea-discovered-in-an-ancient-chinese-emperors-tomb

Gascoyne, K., Marchand, F., Desharnais, J., & Americi, H. (2018). Tea: History, Terroirs, Varieties (Third). Firefly Books.

Gorazd. (2012, July 4). *Puerh cha*. Flickr. Retrieved December 11, 2022, from https://www.flickr.com/photos/ignatgorazd/7500656780

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(03), 90-95.

Lee, J., Chambers, D. H., & Chambers, E. (2013). A comparison of the flavor of green teas from around the world. Journal of the Science of Food and Agriculture, 94(7), 1315–1324. https://doi.org/10.1002/jsfa.6413

Li. (1993). Chinese Tea Culture. The Journal of Popular Culture, 27(2), 75–90. https://doi.org/10.1111/j.0022-3840.1993.00075.x

Liang, S., Granato, D., Zou, C., Gao, Y., Zhu, Y., Zhang, L., Yin, J. F., Zhou, W., & Xu, Y. Q. (2021). Processing technologies for manufacturing tea beverages: From traditional to advanced hybrid processes. Trends in Food Science &Amp; Technology, 118, 431–446. https://doi.org/10.1016/j.tifs.2021.10.016

Mukai, Horie, & Goto. (1992). Differences in Free Amino Acids and Total Nitrogen Contents among Various Prices of Green Tea. Chagyo Kenkyu Hokoku (Tea Research Journal), 1992(76), 45–50. https://doi.org/10.5979/cha.1992.76_45

Statista. (2022a, October 10). Global tea market size 2012-2025. https://www.statista.com/statistics/326384/global-tea-beverage-market-size/

Statista. (2022b, October 10). Leading tea importers worldwide 2021. https://www.statista.com/statistics/258620/main-import-countries-for-tea-worldwide/

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Wang, T. S., Liang, A. R. D., Ko, C. C., & Lin, J. H. (2021). The importance of region of origin and geographical labeling for tea consumers: the moderating effect of traditional tea processing methods and tea prices. Asia Pacific Journal of Marketing and Logistics, 34(6), 1158–1177. https://doi.org/10.1108/apjml-02-2021-0121

Zhang, J., & Sivaramakrishnan, K. (2013). Puer Tea: Ancient Caravans and Urban Chic (Culture, Place, and Nature). University of Washington Press.

Zhang, L., Ho, C., Zhou, J., Santos, J. S., Armstrong, L., & Granato, D. (2019). Chemistry and Biological Activities of Processed Camellia sinensis Teas: A Comprehensive Review. Comprehensive Reviews in Food Science and Food Safety, 18(5), 1474–1495. https://doi.org/10.1111/1541-4337.12479