International Journal of Management and Marketing Research
Vol. 15, No. 1, 2022, pp. 1-13
ISSN: 1931-0269 (print)
ISSN: 2157-0698 (online)

www.theIBFR.com

WHAT DRIVES HIGHER BEER RATINGS? EVIDENCE FROM BIG DATA

Andrew Manikas, University of Louisville
Michael Godfrey, University of Wisconsin Oshkosh
Jason Woldt, University of Wisconsin Oshkosh

ABSTRACT

The purpose of this paper is to present a data analytics model utilizing a large database of beer reviews. We are witnessing a big data revolution with companies' ability to capture and analyze large volumes, velocity, and variety of consumer spending patterns and purchase factors. Analysis of consumer reviews recently has shifted from focusing only on consumer ratings to also mining the content of those reviews. Using a sample of approximately 9 million reviews over 22 years from BeerAdvocate, we developed a multiple regression model to test the relationship of alcohol by volume (ABV) and review word length to the rating that a beer received. Our main finding was that the more (ABV) that a beer had, the higher was that beer's average rating. Although increased ABV was associated with higher consumer scores, we found diminishing returns. We also confirmed that negativity bias occurred in these reviews—lower-scoring beers tended to have longer reviews than higher scoring beers had. This suggests that breweries would be advised to have higher alcohol content in their products to meet consumer preferences (even if those preferences are subconscious). Given the supply chain disruptions plaguing the world today, big data analysis of consumer preferences for beer could enable both manufacturers with constrained capacity to better align product offerings with consumer tastes and preferences.

JEL: L31, L84, M11

KEYWORDS: Big Data, Supply Chain Analytics, Consumer Reviews

INTRODUCTION

According to the Brewers Association (2020), the beer market is valued at 94.1 billion dollars annually. Even though the overall beer market has been declining in recent years, the number of craft breweries still is increasing, resulting in more product variety and consumer choice (Brewers Association, 2020). This explosion of independent craft breweries has led to breweries pushing the limits of traditional brewing by experimenting with ingredients in unique ways to create differentiated product offerings. According to the National Brewers Wholesale Association (NBWA) (2019), while a typical beer distributor managed only 185 unique SKUs in 1999, the typical distributor managed 1,174 unique SKUs in 2018 (an increase of approximately 635 percent). As a result, selecting a beer that a consumer may enjoy becomes increasingly difficult. To help with this process, online reviews are a way that current and potential customers can obtain suggestions. Social proof theory states that a higher rated product is more likely to be selected by others. In their study, Shin and Darpy (2020) found that 42.28 percent of respondents indicated that they would choose a higher rated product even if it were more expensive. The purpose of this research is to present a data analytics model showing determinants of high consumer beer ratings. Using multiple regression analysis, we examined the data obtained from BeerAdvocate (posted on Kaggle, retrieved May 29, 2020). This data set contained over 9 million user reviews of beers from 1996 – 2018. Our research question was: What factors drive higher scores in online beer reviews? If we could find statistically significant independent variables that correlated with the final score, that information could be helpful for breweries in tailoring their products or targeted marketing toward customers.

The next section discusses the literature review. After that, we present the data and the methodology. The last two sections include results followed by discussion and conclusions.

LITERATURE REVIEW

Supply Chain Big Data Analytics

The term big data has been defined as data sets that are so large that they require complex methods of storage, management, analysis and technology for visualization (Chen et al., 2012). In addition, Erevelles et al. (2016) defined analytics as tools to help find hidden patterns in data. Using the combination of data and analytical tools, organizations are able to uncover valuable insights to make better business decisions (Power et al., 2018; Sharma et al., 2014). Big data and analytics have been particularly useful in manufacturing and supply chain management due to availability of data internally and externally in the organization (Sahay & Ranjan, 2008; Trkman et al., 2010). Big data analytics has been defined using the three Vs: volume, variety, and velocity (Gartner, 2012; Kwon & Sim, 2013; McAfee & Brynjolfsson, 2012). The volume of the data refers to the size of the data set that is available (Gandomi & Haider, 2015). Because data is easier to store and less expensive to collect, the sheer size of data sets available has increased exponentially. According to Holst (2021), the amount of data created, captured, copied, and consumed worldwide increased from 2 zettabytes in 2020 to 64.2 zettabytes in 2021, and is projected to reach 181 zettabytes in 2025. The velocity of data refers to the speed at which big data is generated (Gandomi & Haider, 2015). Because data is generated quickly, and in some cases, available in real time, organizations must quickly capture and analyze this data to extract maximum value (McAfee & Brynjolfsson, 2012). Finally, variety refers to the heterogeneity of the data (Gandomi & Haider, 2015). Today, data beyond numbers (structured) can be captured and analyzed in the form of sensor data, social media posts, images, audio, video, and even customer reviews (unstructured). Cukier (2010) found that only 5 percent of data is structured, while the other 95 percent exists in an unstructured form.

Bendle and Wang (2016) presented a taxonomy of big data divided into two types: structured data (e.g., ratings, questions with binary answers, or questions with a finite range of responses) and unstructured data (e.g., reviews written in free-form English). As the field of big data and analytics has evolved and has become more complex, explaining the field with only volume, variety, and velocity provided an incomplete description of the field. As such, the 3 Vs of big data analytics have since been expanded to 5Vs to also include veracity and value. Wamba et al. (2015) appear to have been the first to present the 5V big data elements from prior research in an integrated framework as: volume, variety, velocity, veracity, and value. Veracity refers to the correctness, precision, and reliability of the data (Gandomi & Haider, 2015; Richey et al., 2016), which researchers and practitioners have found to be a large barrier in implementing robust data analytics strategies (LaValle et al., 2011) and is critical to draw accurate conclusions about the data (Aman et al., 2014). Finally, value refers to the usefulness or worth of the data. In raw form, data holds little value (Gandomi & Haider, 2015), but as data transitions from unstructured data to structured data, and is later organized and analyzed, it holds much more value to be incorporated into models and strategies (Harris & Mehrotra, 2014; Veeramachaneni, 2016).

Sources of big data include point-of-sale data, in-store path data (e.g., customer browsing and purchase behaviors), and user-generated content (data created by unpaid contributors; e.g., internet searches as well as online reviews and ratings of products and services) (Boone et al., 2019). Recently the supply chain literature has experienced an increase in articles about big data analytics and literature reviews summarizing those articles (e.g., Ghalehkhondabi et al., 2020; Hazen et al., 2018; Mahya & Fereshteh, 2020; Talwar et al., 2021). Text mining at its core is a computer-automated process of word counting in free-form text

(Kulkarni et al., 2019). Ghalehkhondabi et al. (2020) divided the big data analytics research in supply chain management into the following categories: strategy development, operations improvement, sustainability, food supply chains, risk management, and marketing and sales. Given that our current study is an analysis of the effects of consumer reviews and product characteristics on demand, the next section focuses on research related to big data analytics for demand forecasting.

Big Data Analytics for Demand Forecasting

Sikora and Chauhan (2012) suggested that online reviews of products and reviewer-related data are considered by many as one of the most important knowledge-based systems associated with online commerce websites. These reviews could be mined for both structured and unstructured data, as defined by Bendle and Wang (2016). Several researchers have studied online reviews to improve forecasting effectiveness. Chong et al. (2017) compared the predictive ability of online promotional marketing and online reviews as predictors of consumer product demand. By extracting data from Amazon.com and using a neural network analysis, they were able to show that variables from promotional marketing strategies and online reviews were significant predictors of demand, and of the two, variables related to online reviews were better predictors of demand. For example, Schneider and Gupta (2016) studied reviews of tablets posted on Amazon.com. Using a bag-of-words method along with a random projections approach to reduce the amount of manual pre-processing of textual data (required to reduce the number of words or sequences of words to be mined), they found that their model performed better than a competitor's model for forecasting demand for existing and new products. Chern et al. (2015) analyzed online reviews to develop a new forecasting model using sales data from a cosmetic retail chain in Taiwan. They used variables such as the properties of reviews (polarity and text sentiment), characteristics of the reviewer, and responses from readers (number of readers who liked the review). Ting et al. (2017) used Python programming and data mining to study consumer reviews from Yelp.com and to assess how customer experiences posted online affect demand in the U.S. hospitality industry. They used text analytics to create a list of 36 words associated with online reviews by hotel customers related to rating, price, and rating/price. Furthermore, Salehan and Kim (2016) employed a sentiment mining approach of online consumer reviews of 20 different products posted on Amazon.com, and found that reviews with higher levels of positive sentiment in their titles exhibited higher readership. In addition, they found that reviews with neutral text polarity were perceived as being more helpful, and that the length of a review positively affected readership and perceived helpfulness of those reviews. Finally, Archak et al. (2011) used text mining of consumer reviews posted on Amazon.com for two different products (digital cameras and camcorders). They combined textual data with econometric and predictive modeling, finding that the content of the text in reviews was able to be used to predict consumer demand.

Beer Reviews

Some prior research has focused on beer reviews. For example, Colen and Swinnen (2016) analyzed the changes in beer consumption among countries over time. They found that per capita beer consumption fell in traditional beer-drinking countries, but increased strongly in emerging countries, particularly China. They suggested that beer consumption initially increases with rising incomes, but then decreases at higher levels of income. Donadini et al. (2014) conducted a study of 246 beer drinkers and 8 different beers in a tavern in Italy. They found that the highest percent (about 40 percent) preferred sweet and fruity samples, which were perceived as having a higher level of alcohol. Conversely, Thong et al. (2018), in a labeled choice experiment with beer drinkers in Vietnam, found that alcohol content had the least significant effect on consumer beer choice—beer brand, packaging format, and price were the most significant predictors of consumers' choice of beer. However, they noted that the alcohol percentage only ranged between 4.5% and 5.5%. Meyerding et al. (2019) surveyed German beer consumers to determine attributes of beer that consumers find critical to the purchasing process. They found that beer type had the highest influence on consumer beer choice, followed by price, the origin of beer, and alcohol content. In a study of beer consumption in the U.S., Schiff et al. (2021) analyzed manufacturer websites, consumer beer-related websites, and Nielsen Consumer Panel data. They found that higher alcohol content beer consumption increased steadily from 2004 – 2014, increasing from 9.6 percent of total consumption in 2004 to 21.6 percent in 2014. From 2011 on, however, total consumption of beer in the U.S. declined by 3.04 percent annually. Given the extant literature, we believe that a gap exists in examining if, and how, alcohol percentages affect the rating that consumers assign to beers.

DATA AND METHODOLOGY

In this section, the model is described first. Second, the variables are discussed. Third, the data are described. Fourth, the hypotheses are presented.

Model Description

Data on beer reviews were collected from Kaggle (https://www.kaggle.com/ehallmar/beers-breweries-and-beer-reviews) on May 29, 2020, at the commencement of this study. There were 9,069,130 reviews, of which 8,897,129 had complete data for our analysis (a text review and populated score data). Therefore, our final sample included 8,897,129 reviews. For each review, a user filled out a free-form text review, and rated five numeric components (Look, Smell, Taste, Feel, and Overall). Those components then were used to calculate Score.

Variables

Dependent Variable: Score. Score is a numeric rating overall for a beer based on five attributes (Look, Smell, Taste, Feel, and Overall) that users filled out based on a five-point Likert scale that ranged from 1 (least appealing) to 5 (most appealing). Per BeerAdvocate (BeerAdvocate, 2021), Score was calculated as 0.06 * Look + 0.24 * Smell + 0.40 * Taste + 0.10 * Feel + 0.20 * Overall.

Independent Variables: ABV (Alcohol by volume) is a measure of the percentage of alcohol, where a higher number indicates a stronger beer. *TotWords* is a count of how many words the user typed into their free-form review of the beer.

Controls: Year, Month, and Day are from the date that the review was completed. These control variables allow isolation of artifacts due to chronology (e.g., it may be that reviews during a recession year, or a certain month of the year, are systematically higher).

Brewery_Id: is a unique number to identify a particular brewery. It may be that beers from a certain brewery are rated consistently better or worse than other breweries, so this control variable was used to capture that effect.

Data Description

Figure 1 highlights the the dispersion of the original 9,069,130 beer reviews in the original database before we dropped incomplete records to get our final sample. The number of reviews by year in the database increased from 1996 through 2014, then tapered down through 2018.

Figure 1: Number of Reviews by Year

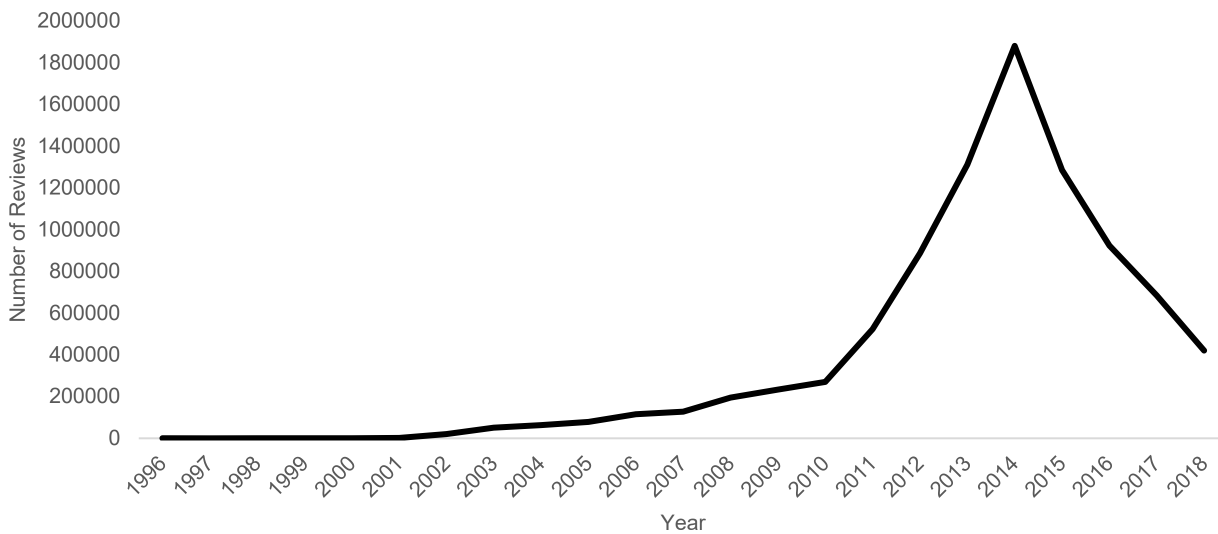


Figure 1 displays the number of reviews by year in the original dataset. Note that this number increased dramatically from 1996 to 2014, and then tapered down.

Table 1 presents the correlations between the variables included in the database.

Table 1: Correlations between Variables

	Score	ABV	TotWords	Year	Month	Day	Brewery_Id
Score	1						
ABV	0.3300*	1					
TotWords	-0.0277*	0.0100*	1				
Year	0.1038*	0.0588*	-0.4253*	1			
Month	-0.0050*	-0.0188*	-0.0177*	-0.0987*	1		
Day	0.0015*	0.0003	-0.0030*	-0.0057*	0.0189*	1	
Brewery_Id	0.1168*	0.0111*	-0.1202*	0.3831*	-0.0105*	0.0022*	1

*Table 1 shows the correlations between variables. Correlations that were significant at $p<0.05$ are denoted with *.*

As the alcohol content among beers differs widely, we wondered if that was a contributing factor to the variation in scores that reviewers gave. Therefore, we examined the possible influence of the alcohol percentage, alcohol by volume (*ABV*), on the reviewers' scoring of beers (*Score*). Further, we analyzed the length of the free-form review text (*TotWords*) and its relationship with that Score. We determined how our independent variables affected the Score that these beers received. Our two variables were *ABV* (alcohol by volume) and *TotWords* (the total number of words that the reviewer typed in their free-form review). Because drinking beer provides a vehicle for alcohol, we posited that higher ABV beers would be rated higher – after all, alcohol is likely a significant reason why people are drinking beer.

Hypothesis 1: Beers with a higher ABV will have higher Scores by reviewers.

Further, negativity bias would predict that people tend to say more about products that they dislike than those that they like. Telling someone that a beer is good conveys information to an audience. However, when a reviewer wants to convey that a beer is bad, they tend to explain in detail how it is "not good." Prior research has demonstrated a negativity bias, i.e., people tend to evaluate negative information more highly than positive information (Poncheri et al., 2008). Thus, we hypothesized that higher scoring beers would have significantly shorter free-form reviews (*TotWords*).

Hypothesis 2: Beers with longer (shorter) free-form reviews will correlate with lower (higher) **Scores** by reviewers.

The regression model for predicting ***Score*** is:

$$Score = \beta_0 + \beta_1 * ABV + \beta_2 * TotWords + \beta_3 * Brewary_{Id} + \beta_4 * Year + \beta_5 * Month + \beta_6 * Day \quad (1)$$

RESULTS

Using linear regression in Stata 17 software, we found the results shown in Table 2. Both of our hypotheses were supported; higher *ABV* correlated positively with higher Scores, and the longer the review (*TotWords*), the lower the Score (negative correlation). The non-significance of the year dummy variables indicates that the Scores did not change significantly from year to year, even though the number of reviews varied widely from year to year, as shown in Figure 1 above.

We found that higher *ABV* correlated significantly and positively with the Score that a beer received. Further, the total words (*TotWords*) used to describe the beer increased as the Score decreased (significant and negative), thereby confirming negativity bias. We conducted two post hoc analyses here to delve into the details of these relationships. For the first post hoc analysis, we analyzed the average number of words used for beers (*TotWords*) scoring below average (3.9 was the average score out of 5 for beers in our dataset) versus the average number of words for beers (*TotWords*) scoring above average. From the regression results shown in Table 2, we knew that there would be a difference, but we wanted to find out the average number of words in positive versus negative scoring beer reviews. The average total words for reviews (*TotWords*) with Scores below the mean was equal to 60.6 words, and the average number of words (*TotWords*) for reviews with Scores above the mean was equal to 30.3. We conducted a two-sample t-test with unequal variances on these two distributions, and rejected the null hypothesis that they were equal. Therefore, the alternate hypothesis that the difference was not equal to zero was assumed. Thus, there was a statistically significant difference ($p<0.01$) between the shorter reviews for above-average rated beers and the longer reviews for below-average rated beers. Further, by looking at the coefficients and significance of the months in the above regression, we can examine seasonality effects, as shown in Table 3. During July and August, Scores were systematically lower. Note that reviewers were from around the world, but the majority (82.9%) were from the United States. Thus, July and August in the northern hemisphere (U.S.) would be hotter months.

We found that higher *ABV* correlated significantly and positively with the Score that a beer received. Further, the total words (*TotWords*) used to describe the beer increased as the Score decreased (significant and negative), thereby confirming negativity bias. We conducted two post hoc analyses here to delve into the details of these relationships. For the first post hoc analysis, we analyzed the average number of words used for beers (*TotWords*) scoring below average (3.9 was the average score out of 5 for beers in our dataset) versus the average number of words for beers (*TotWords*) scoring above average. From the regression results shown in Table 2, we knew that there would be a difference, but we wanted to find out the average number of words in positive versus negative scoring beer reviews. The average total words for reviews (*TotWords*) with Scores below the mean was equal to 60.6 words, and the average number of words (*TotWords*) for reviews with Scores above the mean was equal to 30.3. We conducted a two-sample t-test with unequal variances on these two distributions, and rejected the null hypothesis that they were equal. Therefore, the alternate hypothesis that the difference was not equal to zero was assumed. Thus, there was

a statistically significant difference ($p<0.01$) between the shorter reviews for above-average rated beers and the longer reviews for below-average rated beers. Further, by looking at the coefficients and significance of the months in the above regression, we can examine seasonality effects, as shown in Table 3. During July and August, Scores were systematically lower. Note that reviewers were from around the world, but the majority (82.9%) were from the United States. Thus, July and August in the northern hemisphere (U.S.) would be hotter months.

Table 2: ABV and Review Total Words as Predictors of Score

Variables	Score
ABV	0.0810***
	(0.0000784)
TotWords	-0.000055***
	(0.00000331)
1998.year	-0.0657
	(0.576)
1999.year	-0.123
	(0.581)
2000.year	0.0873
	(0.575)
2001.year	-0.131
	(0.572)
2002.year	-0.200
	(0.572)
2003.year	-0.236
	(0.572)
2004.year	-0.211
	(0.572)
2005.year	-0.213
	(0.572)
2006.year	-0.233
	(0.572)
2007.year	-0.246
	(0.572)
2008.year	-0.217
	(0.572)
2009.year	-0.203
	(0.572)
2010.year	-0.211
	(0.572)
2011.year	-0.153
	(0.572)
2012.year	-0.235
	(0.572)
2013.year	-0.205
	(0.572)
2014.year	-0.188
	(0.572)
2015.year	-0.131
	(0.572)
2016.year	-0.132
	(0.572)
2017.year	-0.130
	(0.572)
2018.year	-0.138
	(0.572)
Constant	3.424***
	(0.572)
Observations	8,897,129
R-squared	0.126

*Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10*

Regression Equation: $\boldsymbol{Score} = \beta_0 + \beta_1 * ABV + \beta_2 * TotWords + \beta_3 * Brewary_{Id} + \beta_4 * Year + \beta_5 * Month + \beta_6 * Day$

*This table shows that certain breweries outperformed other breweries in a systemic manner (one brewery may make several highly rated beers). Per our hypotheses, higher ABV correlated with a higher Score by reviewers, and the total number of words (**TotWords**) correlated with a lower Score, i.e., fewer words were used to describe higher scoring beers. Note that we controlled for the year, month, and day of review with dummy variables and used a dummy variable for Brewery_Id (Brewery_Id, Month, & Day are not shown for ease of reading this table). Further note: The year dummy variables represent the coefficient and possible significanceof each year's average review Score to other years. As all year dummy variables were not significant, there were not significant differences in beer review Scores across the 1996 – 2018 time period. The year 1997 had no reviews.*

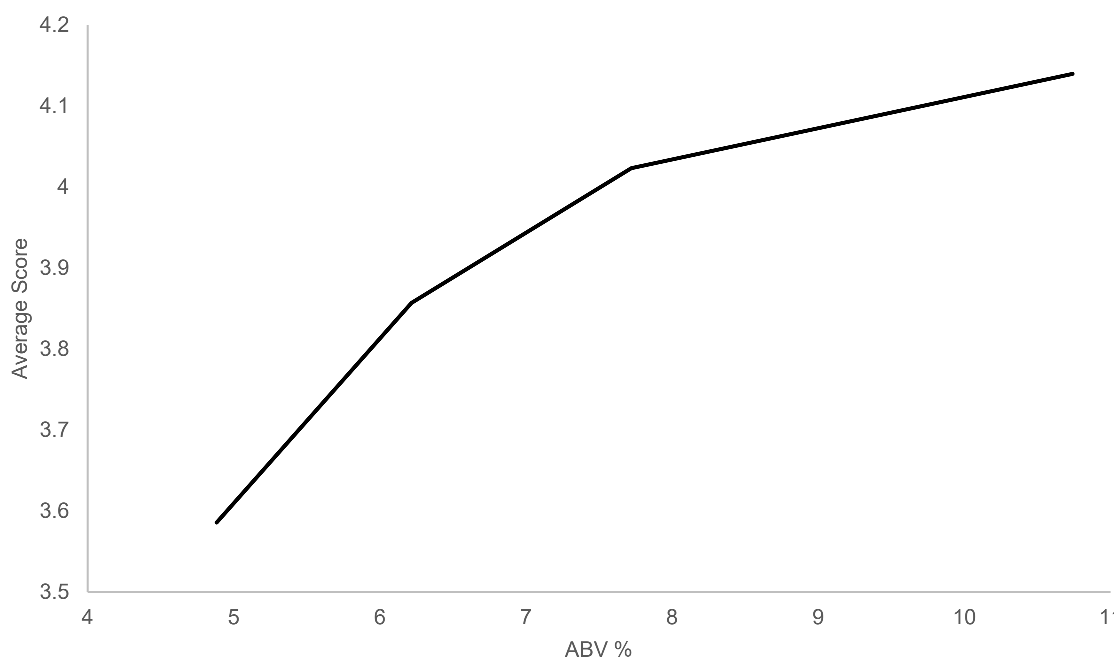
Table 3: Monthly Coefficients of Scores

January	Control Month
February	0.0078
March	0.0142
April	0.0133
May	0.0028
June	0.0048
July	-0.0037
August	-0.0026
September	0.0016
October	0.0089
November	0.0399
December	0.0092

All months significantly different from January at p<0.10

For the second post hoc analysis, we plotted the *ABV* against Score to see if there was a linear or non-linear relationship between the two variables. We included a squared *ABV* term to see the shape (convex versus concave). The multiple regression results showed that the square term had a significant and negative sign, indicating a concave shape to the curve. In effect, higher *ABV* correlated positively to higher Score, but with diminishing returns. It might be that beyond the alcohol limits of this data set, extremely strong beers (ABV) might score lower than the average Score for the highest *ABV* beers in our sample. Figure 2 below is illustrative of the curve (we picked four points of ABV corresponding to the midpoint of ABV quartiles and the average score associated with them).

Figure 2: ABV Related to Average Score



This figure lists ABV on the x-axis and beer Score on the y-axis. As shown above, the higher the ABV, the higher the average Score that those beers received. However, the increase in Score had diminishing returns, i.e., going from 5% to 6% alcohol had a greater positive impact on the Score than moving from 8% to 9%..

DISCUSSION AND CONCLUSIONS

The goal of this paper was to investigate what factors might drive higher ratings of beers by consumers. Using nearly 9 million reviews over 22 years from BeerAdvocate, we were able to find strong evidence supporting our hypotheses. First, Hypothesis 1 was supported, indicating that beers with higher *ABV* tended to have higher scores from the reviewers. It may be that all the marketing related to small-batch, craft breweries, import versus domestic, etc., is not what is important to consumers' wants and needs – whether or not they realize it. Our data give evidence that consumers like strong beers better. This one factor, *ABV*, drives the ratings, and we would advise any breweries to focus on stronger beverages if they seek to get higher reviews than those of competitor products. Further, in case consumers are aware that they prefer higher ABV, breweries could list the ABV more prominently on the label. However, consumers may not be aware of why they happen to prefer a certain beer, and conversely, be dissuaded from trying a high alcohol beer, as it may seem less "social."

Although negativity bias is a known phenomenon, with Hypothesis 2, we confirmed that negativity bias plays a role in online beer reviews. Consumers find negative aspects of a product more salient than positive ones, so they will tend to write longer reviews for products that they dislike compared to the reviews for products that they prefer. One limitation of our approach is that we used data for beer only. We cannot be sure that *ABV* would have the same effect for wine or bourbon drinks, as those may have other prestige factors such as the age of the wine, the location where grown, the barrel for the bourbon, how many years aged, etc. We would expect the negativity bias to show in wine/bourbon reviews, as that phenomenon has been documented previously in a wide variety of consumer review contexts. Beers typically have a limited range of ABV compared to wines and hard liquor. A study of a wider variety of alcoholic drinks with a large range of ABV would be a good future study. We also cannot be sure of the mechanism of ABV that is causing higher ratings – is it the sweetness, the increased effect on the consumer, or something else? Further, different regions of the world (country, climate, etc.) may have different preferences for ABV, which could be explored in future research.

Increasing the ABV of a beer comes with additional expense. This requires additional ingredients such as fruit, honey, maple syrup, and other simple sugars (Lewis, 2020). In addition, specialty beers that require aging in bourbon or oak barrels require the additional expenses of the equipment and the time and space to age the beer. These additional expenses must be weighed against a range of factors, including higher ABV content and higher consumer reviews.

Many of the reviews are rich with additional information that is relevant to consumer tastes and preferences. A more careful analysis of the free-form text could uncover trends as well as categories that are important to consumers beyond the categories that were used in the reviews (Look, Smell, Taste, Feel, and Overall). Because product reviews are influential in purchasing decisions, there are growing concerns about the veracity of product reviews. While it is clear that BeerAdvocate has processes in place to minimize the number of fraudulent reviews (requiring user accounts, only one review per beer, "report" button (for other users to report other reviews), disclaimer pertaining to review rules and consequences, etc.), fraudulent reviews still exist. Perhaps reviewers should have to take a picture of their UPC to ensure that the timing of their review aligns with the timing of consumption, and that the person writing the review actually consumed the beer (veracity). Logistic regression models also could be built to identify inaccurate or fraudulent reviews from those that are accurate. For instance, a large number of misspelled words, reviews where the text matches the exact text of another review (duplicated reviews), or reviews that fall outside of realistic tolerance ranges may be variables that could be incorporated into a model to identify reviews that should be flagged and potentially removed from the sample. For example, a review from a user who rated look, smell, taste, and feel as 5, but then provided an overall rating of 1 could be flagged as an inaccurate review.

REFERENCES

Aman, S., C. Chelmis and V. Prasanna, V. (2014) "Addressing Data Veracity in Big Data Applications," *2014 IEEE International Conference on Big Data*, p. 1-3. Retrieved November 30, 2021 from the U.S. Department of Energy Office of Scientific and Technical Information web site: https://www.osti.gov/servlets/purl/1333162

Archak, N., A. Ghose and P.G. Ipeirotis (2011) "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science*, vol. 57(8), p. 1485-1509.

BeerAdvocate (2021). *How to Review a Beer*. Retrieved November 22, 2021 from the BeerAdvocate web site: https://www.beeradvocate.com/community/threads/how-to-review-a-beer.241156/

Bendle, N.T. and X. Wang (2016) "Uncovering the Message from the Mess of Big Data," *Business Horizons*, vol. 59(1), p. 115-124.

Boone, T., R. Ganeshan, A. Jain and N.R. Sanders (2019) "Forecasting Sales in the Supply Chain: Consumer Analytics in the Big Data Era," *International Journal of Forecasting*, vol. 35(1), p. 170-180.

Brewers Association (2020). *National Beer Sales and Production Data*. Retrieved November 28, 2021 from the Brewers Association web site: https://www.brewersassociation.org/statistics-and-data/national-beer-stats/

Chen, H., R.H.L. Chiang and V.C. Storey (2012) "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly*, vol. 36(4), p. 1165-1188.

Chern, C., C. Wei, F. Shen and Y. Fan (2015) "A Sales Forecasting Model for Consumer Products Based on the Influence of Online Word-of-Mouth," *Information Systems and eBusiness Management*, vol. 13(3), p. 445-473.

Chong, A.Y.L., E. Ch'ng, M.J. Liu and B. Li (2017) "Predicting Consumer Product Demands via Big Data: The Roles of Online Promotional Marketing and Online Reviews," *International Journal of Production Research*, vol. 55(17), p. 5142-5156.

Colen, L. and J. Swinnen (2016) "Economic Growth, Globalisation and Beer Consumption," *Journal of Agricultural Economics*, vol. 67(1), p. 186-207.

Cukier, K. (2010) "Data, Data Everywhere," *The Economist*, February 27, Accessed November 30, 2021 at: https://www.economist.com/special-report/2010/02/27/data-data-everywhere

Donadini, G., M.D. Fumi and I.R. Newby-Clark (2014) "Consumers' Preference and Sensory Profile of Bottom Fermented Red Beers of the Italian Market," *Food Research International*, vol. 58, p. 69-80.

Erevelles, S., N. Fukuwa and L. Swayne (2016) "Big Data Consumer Analytics and the Transformation of Marketing," *Journal of Business Research*, vol. 69(2), p. 897-904.

Gandomi, A. and M. Haider (2015) "Beyond the Hype: Big Data Concepts, Methods, and Analytics," *International Journal of Information Management*, vol. 35(2), p. 137-144.

Gartner (2012). *Big Data*. Retrieved November 28, 2021 from the Gartner web site: https://www.gartner.com/en/information-technology/glossary/big-data

Ghalehkhondabi, I., E. Ahmadi and R. Maihami (2020) “An Overview of Big Data Analytics Application in Supply Chain Management Published in 2010-2019,” *Production*, vol. 30. Retrieved November 3, 2021 from: http://dx.doi.org/10.1590/0103-6513.20190140

Harris, J. G. and V. Mehrotra (2014) “Getting Value from Your Data Scientists,” *MIT Sloan Management Review*, vol. 56(1), p. 15-18.

Hazen, B.T., J.B. Skipper, C.A. Boone and R.R. Hill (2018) “Back in Business: Operations Research in Support of Big Data Analytics for Operations and Supply Chain Management,” *Annals of Operations Research,* vol. 270(1-2), p. 201-211.

Holst, A. (2021). *Amount of Data Created, Consumed, and Stored 2010-2025*. Retrieved November 28, 2021 from the statista web site: https://www.statista.com/statistics/871513/worldwide-data-created/

Kaggle (2020). *Beers, Breweries, and Beer Reviews*. Retrieved May 29, 2020 from the Kaggle web site: https://www.kaggle.com/ehallmar/beers-breweries-and-beer-reviews

Kulkarni, S., P. Verma and R. Mukundan (2019) “Assessing Manufacturing Strategy Definitions Utilising Text-Mining,” *International Journal of Production Research*, vol. 57(14), p. 4519-4546.

Kwon, O. and J.M. Sim (2013) “Effects of Data Set Features on the Performances of Classification Algorithms,” *Expert Systems with Applications*, vol. 40(5), p. 1847-1857.

LaValle, S., E. Lesser, R. Shockley, M.S. Hopkins and N. Kruschwitz (2011) “Big Data, Analytics and the Path from Insights to Value,” *MIT Sloan Management Review*, vol. 52(2), p. 21-32.

Lewis, R. (2020). *Tips for Increasing the ABV in your Homebrew*. Retrieved November 28, 2021 from the MRBEER web site: https://www.mrbeer.com/blog/post/tips-when-increasing-the-alcohol-level-in-your-homebrew

Mahya, S. and M. Fereshteh (2020) “Predictive Big Data Analytics for Supply Chain Demand Forecasting: Methods, Applications, and Research Opportunities,” *Journal of Big Data,* vol. 7(1), p. 1-22. Retrieved November 2, 2021 from: http://dx.doi.org/10.1186/s40537-020-00329-2

McAfee, A., and E. Brynjolfsson (2012) “Big Data: The Management Revolution,” *Harvard Business Review*, vol. 90(10), p. 60-68.

Meyerding, S.G.H., A. Bauchrowitz and M. Lehberger (2019) “Consumer Preferences for Beer Attributes in Germany: A Conjoint and Latent Class Approach,” *Journal of Retailing and Consumer Services*, vol. 47, p. 229-240.

NBWA (2019). *Industry Fact Facts Packaging and SKUs*. Retrieved November 28, 2021 from the NBWA web site: https://www.nbwa.org/resources/industry-fast-facts

Poncheri, R.M., J.T. Lindberg, L.F. Thompson and E.A. Surface (2008) “A Comment on Employee Surveys: Negativity Bias in Open-Ended Responses,” *Organizational Research Methods*, vol. 11(3), p. 614-630.

Power, D. J., C. Heavin, J. McDermott and M. Daly (2018) “Defining Business Analytics: An Empirical Approach,” *Journal of Business Analytics*, vol. 1(1), p. 40-53.

Richey, R. G., T.R. Morgan, K. Lindsey-Hall and F.G. Adams (2016) "A Global Exploration of Big Data in the Supply Chain," *International Journal of Physical Distribution & Logistics Management*, vol. 46(8), p. 710-739.

Sahay, B.S. and J. Ranjan (2008) "Real Time Business Intelligence in Supply Chain Analytics," *Information Management & Computer Security*, vol. 16(1), p. 28-48.

Salehan, M. and D.J. Kim (2016) "Predicting the Performance of Online Consumer Reviews: A Sentiment Mining Approach to Big Data Analytics," *Decision Support Systems,* vol. 81, p. 30-40.

Schiff, M.D., D.D. Mendez, T.L. Gary-Webb, J.J. Inman and A. Fabio (2021) "A Decade of Drinking: Temporal Trends in Apparent Household Beer Intake and Standard Drink Consumption in the United States," *Substance Use & Misuse*, vol. 56(9), p. 1363-1373. Retrieved November 3, 2021 from: https://doi.org/10.1080/10826084.2021.1928208

Schneider, M.J. and S. Gupta (2016) "Forecasting Sales of New and Existing Products Using Consumer Reviews: A Random Projections Approach," *International Journal of Forecasting*, vol. 32(2), p. 243-256.

Sharma, R., S. Mithas and A. Kankanhalli (2014) "Transforming Decision-Making Processes: A Research Agenda for Understanding the Impact of Business Analytics on Organisations," *European Journal of Information Systems*, vol. 23(4), p. 433-441.

Shin, D. and D. Darpy (2020) "Rating, Review and Reputation: How to Unlock the Hidden Value of Luxury Consumers from Digital Commerce?" *The Journal of Business & Industrial Marketing*, vol. 35(10), 1553-1561.

Sikora, R.T. and K. Chauhan (2012) "Estimating Sequential Bias in Online Reviews: A Kalman Filtering Approach," *Knowledge-Based Systems*, vol. 27, p. 314-321.

Talwar, S., P. Kaur, S.F. Wamba and A. Dhir (2021) "Big Data in Operations and Supply Chain Management: A Systematic Literature Review and Future Research Agenda," *International Journal of Production Research*, vol. 59(11), p. 3509-3534.

Thong, N.T., B.Q. Thanh, H.S. Solgaard and Y. Yang (2018) "The Role of Packaging Format, Alcohol Level and Brand in Consumer's Choice of Beer: A Best-Worst Scaling Multi-Profile Approach," *Food Quality and Preference*, vol. 65, p. 92-100.

Ting, P.L., S. Chen, H. Chen and W. Fang (2017) "Using Big Data and Text Analytics to Understand How Customer Experiences Posted on Yelp.com Impact the Hospitality Industry," *Contemporary Management Research*, vol. 13(2), p. 107-130.

Trkman, P., K. McCormack, M.P.V. de Oliveira and M.B. Ladeira (2010) "The Impact of Business Analytics on Supply Chain Performance," *Decision Support Systems*, vol. 49(3), p. 318-327.

Veeramachaneni, K. (2016) "Why You're Not Getting Value from Your Data Science," *Harvard Business Review*, December 7, Accessed December 1, 2021 at: https://hbr.org/2016/12/why-youre-not-getting-value-from-your-data-science

Wamba, S.F., S. Akter, A. Edwards, G. Chopin and D. Gnanzou (2015) "How 'Big Data' Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study," *International Journal of Production Economics*, vol. 165, p. 234-246.

BIOGRAPHY

Dr. Manikas earned his B.S. in Computer Science and his MBA in Materials and Logistics Management from Michigan State University, and his Ph.D. from The Georgia Institute of Technology. Before that, he was an instructor for supply chain optimization courses for i2 Technologies and worked as a management consultant for KPMG Peat Marwick, CSC, and Deloitte Consulting. Dr. Manikas is an Associate Professor in the Information Systems, Analytics, and Operations Department at the University of Louisville.

Dr. Godfrey earned his B.S. in Operations Management and M.S. in Management Information Systems from Northern Illinois University, and his Ph.D. in Production & Operations Management from the University of Nebraska - Lincoln. Dr. Godfrey is the Department Chair of the Supply Chain Management Department at the University of Wisconsin Oshkosh.

Dr. Woldt earned his BBA in Operations Management from the University of Wisconsin Whitewater, his MBA from Marquette University, and his DBA from the University of Wisconsin Whitewater. Prior to academia, Dr. Woldt worked in various supply chain management positions for Kohler Company. He is currently an Associate Professor in the Supply Chain Management Department at the University of Wisconsin Oshkosh.