Uncovering Consumer Preferences Through Beer Review Analytics

Ramon Benavides, Rory Covarrubia, Cesar Fabrega Tapia, Angie Camila Perez, Alejandra Sevilla
Team #143

I. INTRODUCTION

The craft beet market is experiencing significant growth as consumer behavior shifts towards more diverse styles and flavors [3] [4]. Research by Aquilani et al. [14] highlights the importance of factors such as aroma, quality, and draft beer availability in influencing craft beer consumption. To thrive in this competitive landscape, breweries must understand consumer preferences [2].

Our project has conducted an analysis of consumer preferences in the craft beer market using an extensive dataset of beer reviews. The specific objectives of our project were:

1. Analyze a dataset of 1.5 million beer reviews to identify key factors influencing consumer preferences, such as beer style, aroma, taste, and alcohol by volume (ABV).
2. Apply machine learning techniques, particularly clustering algorithms, to segment consumers based on their preferences [8] and identify distinct consumer groups.
3. Investigate geographical influences on consumer preferences and brewery popularity by integrating brewery location data from the Google Places API.
4. Develop interactive visualizations using Tableau to show our findings on top-rated beers, popular styles, leading breweries, and the regional distribution of consumer preferences.
5. Provide recommendations [3] to breweries, marketers, and beer drinkers based on the identified consumer preference segments and market trends. Understanding these trends is critical for breweries during the craft beer boom [2].

All team members contributed a similar amount of effort to this project, resulting in a collaborative and effective workflow.

The incorporation of Tableau facilitated an engaging exploration of our findings, allowing users to interact with dynamic visualizations and gain understanding of the craft beer landscape.

Our project builds upon prior research, such as the study by Manikas et al. [13], which identified a positive correlation between ABV and beer ratings. We have extended this analysis by considering additional factors shaping beer preferences and using a more extensive dataset. Ultimately, our findings provide knowledge for stakeholders in the craft beer industry to help them adapt to evolving consumer tastes and strengthen their market positions.

II. PROBLEM DEFINITION

The primary challenge of this project involved analyzing the subjective and diverse preferences within the craft beer market [1]. We addressed this by acquiring a Kaggle dataset (https://www.kaggle.com/datasets/rdoume/beerreviews/data) with over 1.5 million beer reviews, providing a rich source of consumer feedback.

Data preprocessing was essential to maintain data integrity and manage the dataset size. This included handling missing values (e.g., in brewery names, ABV), addressing outliers, and optimizing code for computational efficiency.

To further enrich our analysis, we integrated latitude and longitude data for breweries, obtained from the Google Places API, enabling us to investigate geographical trends in consumer preferences and brewery popularity. This integration facilitated the process of geocoding brewery names to obtain spatial coordinates, enabling the creation of an interactive and informative map visualization in Tableau. This visualization showed geographical distribution of breweries and consumer preferences within the craft beer market.

A key project objective was the implementation of clustering algorithms to segment consumer preferences. By identifying distinct clusters within beer reviews, we intended to reveal underlying patterns in consumer tastes.

Ultimately, this project intended to provide breweries with knowledge derived from our findings. Our goal was to help breweries align their product offerings with identified consumer preference segments, which could improve customer satisfaction and strengthening their market position [11].

III. LITERATURE SURVEY

Studies have explored the relationship between alcohol content and beer ratings. Manikas, Godfrey, and Woldt [13] found a positive correlation between alcohol by volume (ABV) and ratings, with diminishing returns, and a negativity bias in lower-rated beers. Their research aligns with our project's focus on trend analysis using big data techniques. They also suggested that breweries might benefit from producing higher ABV beers to meet consumer preferences.

Aquilani et al. [14] investigated consumer preferences in the craft beer industry, identifying aroma, perceived quality, and draft beer preference as key factors influencing consumption. Craft beer enthusiasts tend to prefer distinct flavor profiles, including amber, red, and dark beers, as well as unique flavors like honey and chestnut. The research by Manikas, Godfrey, and Woldt [13] and Aquilani et al. [14] provided a valuable framework for analyzing beer ratings and preferences, highlighting the relevance of ABV, review characteristics, and craft beer preferences to our project's focus on visualizing beer review trends.

The integration of geospatial analysis in consumer studies is an emerging field that adds context to the study of consumer behavior and market trends. Zhang and Shi [15], highlighted the potential of web mapping services like Google Maps to visualize complex data, including geospatial information. Their case study on conference presenters shows how geocoding and GeoRSS can be used for location-based decision-making, offering a new perspective on data presentation and analysis.

Applying this approach to the craft beer market offers an opportunity to uncover regional preferences, distribution patterns, and how location influences consumer trends. By employing the methods demonstrated by Zhang and Shi [15], our project has successfully visualized and analyzed the geographical distribution of breweries and consumer ratings.

IV. PROPOSED METHOD

A. Rationale and Methodological Foundation

This project's multi-faceted analytical approach integrated diverse data types, including sensory attributes, review scores, and geospatial information. Key advantages include:

- Combining sensory reviews with geospatial data captured both consumer sentiment and spatial trends.
- Iterative K-Means clustering, fine-tuned via the Elbow Method, identified patterns within consumer feedback.
- Mapping supported by the Google Maps API revealed regional market dynamics and consumer preference landscapes.

B. Methodology Elaboration

1. **Dataset Preparation and Refinement:**
    - Data cleaning and transformation were performed using **Pandas** within Python, to ensure data integrity by resolving issues such as missing values and outliers.
    - The **Google Maps Geocoding API** enriched the dataset with latitude and longitude coordinates, enabling spatial analysis and visualization of brewery locations.
2. **Exploratory Data Analysis (EDA):**
    - Interactive visualizations (**Plotly**, **Seaborn)** facilitated interpretation of complex data relationships.
    - Statistical measures were used to examine the distribution of review scores.
3. **Clustering and Segmentation:**
    - **Scikit-learn's StandardScaler** was used to standardize feature scales before clustering.
    - K-Means clustering (**scikit-learn**) segmented reviews into clusters based on similarity metrics [5], with seven optimal clusters identified. Visualizations (heatmaps, bar plots) helped understand consumer preferences within each cluster [9].

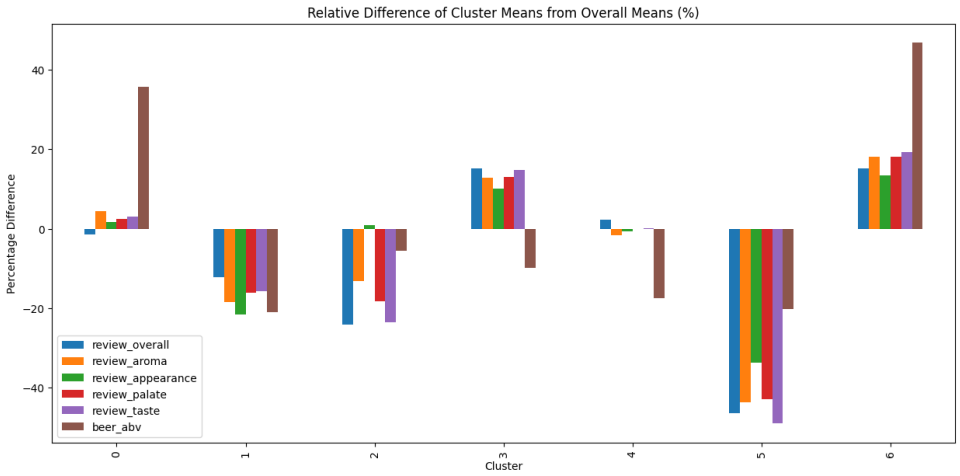

Figure 1- Relative Difference of Cluster Means from Overall Means (%)

4. **Geospatial Analysis and Data Integration:**
    - A batch geocoding (**googlemaps Python Client Library)** automated the extraction of geographic coordinates for breweries, refining the dataset to include spatial attributes, facilitating regional trend mapping [6].
    - Exception handling managed limitations and quotas associated with Google Maps services.
5. **Visualization and Reporting:**
    - An interactive Tableau dashboard was developed, featuring top-rated beers, popularity charts by beer style, brewery performance metrics, and comparative analysis of beer styles by review scores.
    - Geospatial mapping plotted brewery locations against demographics and market segments.
    - A clustering filter derived from the data segmentation allowed for granular examination of consumer preferences.

C. Contributions

This project's methodology offers a template for market analysis, blending consumer research with spatial analytics. The outcomes have both academic and commercial utility, emphasizing the role of evidence-based strategies in product innovation and market positioning.

V. EXPERIMENTS / EVALUATIONS

Testing Overview

We conducted a controlled evaluation of our clustering model and Tableau visualizations. Our dataset included over 1.5 million beer reviews with sensory attributes and brewery geolocations. We utilized Python libraries (Pandas, scikit-learn, Plotly, Seaborn), the Google Maps API, and Tableau for analysis, geospatial processing, and dashboard creation. The Davies-Bouldin Index was used to assess clustering quality.

Key Experimental Questions

1. How does clustering configuration impact the quality of review segmentation?
2. Does geospatial integration reveal meaningful regional preferences and trends?
3. Do Tableau visualizations enhance stakeholder engagement with the data?

Findings and Implications

- **Data Preprocessing:** data cleaning and handling of missing values/outliers were needed for a reliable analysis. Our correlation matrix revealed a positive relationship between ABV and review scores, aligning with prior findings [13]. Here, we observed most ratings falling between 3.5 and 4.5, indicating a general trend towards positive evaluations, with some beers receiving exceptionally high scores. Additionally, the EDA identified American IPA, Russian Imperial Stout, and American Double IPA as the most reviewed beer styles.

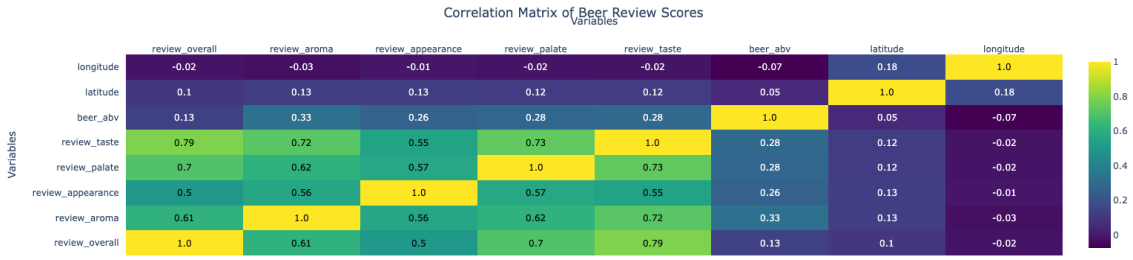

Figure 2 – Correlation Matrix of Beer Review Scores

- **Clustering:** the Elbow Method indicated seven optimal clusters. Analysis of cluster centroids disclosed distinct characteristics [5] with some clusters gravitating towards high ABV and sensory scores and others indicating more moderate attributes. These preferences ranged from high-rated flavorful beers (Clusters 3 and 6) to more standard or less favored beers (Clusters 1, 4, and 5).
- **Geospatial Trends**: geospatial mapping showed top-rated breweries concentrated in established craft beer scenes (Pacific Northwest, Northeast US, Germany, Belgium), aligning with industry knowledge. This supports geographically informed marketing and distribution strategies.

Limitations and Future Work

We recognize limitations in our approach, including potential biases in self-reported data and the dynamic nature of consumer preferences that may evolve beyond the scope of our dataset. Future work could include real-time data analysis and the exploration of alternative clustering algorithms.

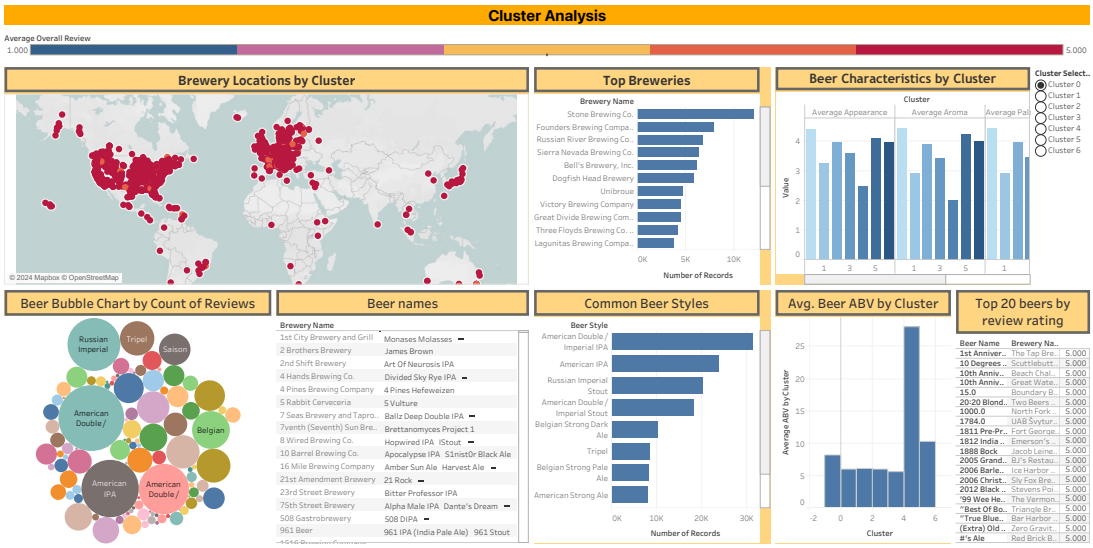

Figure 3 -Interactive Dashboard illustrating the craft beer market through consumer ratings, preferred styles and regional patterns.

VI. CONCLUSIONS AND DISCUSSION

Our analysis revealed seven distinct consumer segments within the craft beer market, supporting findings by Aquilani et al. [14] on the importance of aroma, quality, and draft availability. Clusters with exceptional ratings and high ABV (like Cluster 6, 10.38%) likely represent enthusiasts, while more moderate preferences (like Cluster 1, 5.59% ABV) suggest casual drinkers. This segmentation gives breweries tools to tailor offerings to specific market segments [12].

Our integration of geospatial data uncovered regional differences in beer preferences and brewery popularity. Top-rated breweries were concentrated in areas with established craft beer scenes (West Coast and Northeast US), and traditional brewing countries like Germany and Belgium. This information can guide breweries' marketing and distribution strategies [7] [10].

We confirmed a positive correlation between ABV and overall review scores (Pearson's $r = 0.23$, $p < 0.001$), supporting the research by Manikas et al. [13]. Additionally, the most reviewed beer styles were American IPA (187,255 reviews), Russian Imperial Stout (116,739 reviews), and American Double IPA (90,306 reviews).

These findings offer breweries a guide for product development, marketing, and distribution strategies tailored to consumer segments and preferences. This approach based on evidence, can enhance customer satisfaction, brand loyalty, and business success. Our study contributes to ongoing research on consumer behavior in the craft beer market.

Under limitations, although extensive, our dataset relies on self-reported reviews on a single platform and the potential for preferences to evolve over time. Future research could integrate real-time data and qualitative methods, such as consumer interviews or focus groups, which could offer richer information into the motivations behind beer preferences.

Our findings benefit breweries seeking to align offerings with consumer segments and regional preferences [7] [11]. Marketers and researchers can use our framework to further investigate this dynamic market.

In conclusion, this project successfully revealed key consumer preferences in the craft beer market. By identifying segments, regional variations, and style trends, we provide breweries with valuable information for decision-making in a competitive landscape. Strategies based on data are important for breweries seeking to adapt and thrive in the evolving craft beer market.

VII. REFERENCES

1. Lerro, M., Marotta, G. & Nazzaro, C. *Measuring consumers' preferences for craft beer attributes through Best-Worst Scaling. Agric Econ* 8, 1 (2020). https://doi.org/10.1186/s40100-019-0138-4
2. Pilone, V.; Di Pasquale, A.; Stasi, A. *Consumer Preferences for Craft Beer by Means of Artificial Intelligence: Are Italian Producers Doing Well?* Beverages 2023, *9*, 26. https://doi.org/10.3390/beverages9010026
3. R. Nieto-Villegas, R. Bernabéu, A. Rabadan, *Effects on beer attribute preferences of consumers' attitudes towards sustainability: The case of craft beer and beer packaging*, Journal of Agriculture and Food Research, Volume 15, 2024, 101050, ISSN 2666-1543, https://doi.org/10.1016/j.jafr.2024.101050.
4. 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
5. Scott Taylor Jr. & Robin B. DiPietro (2019) *Segmenting craft beer drinkers: An analysis of motivations, willingness to pay, and repeat patronage intentions*, International Journal of Hospitality & Tourism Administration, 20:4, 423-448, https://doi.org/10.1080/15256480.2017.1397585
6. Ryan M. Hynes & Dieter F. Kogler (2023) G*eography and branding in the craft beer industry, Regional Studies,* https://doi.org/10.1080/00343404.2023.2255618
7. TREMBLAY, VICTOR J., et al. "*The Dynamics of Industry Concentration for U.S. Micro and Macro Brewers."* Review of Industrial Organization, vol. 26, no. 3, 2005, pp. 307–24. *JSTOR*, http://www.jstor.org/stable/41799237.
8. Blin, Jean-Marie, and Claude Cohen. *"Technological Similarity and Aggregation in Input-Output Systems: A Cluster-Analytic Approach." The Review of Economics and Statistics*, vol. 59, no. 1, 1977, pp. 82–91. *JSTOR*, https://doi.org/10.2307/1924907.
9. Calvo-Porral, Cristina & Orosa-González, Javier & Blázquez-Lozano, Félix. (2018). *A clustered-based segmentation of beer consumers: from "beer lovers" to "beer to fuddle"*. British Food Journal. 120. 1280-1294. 10.1108/BFJ-11-2017-0628. 10.1108/BFJ-11-2017-0628
10. Garavaglia, Christian, and Johan Swinnen. "*The Craft Beer Revolution: An International Perspective." Choices*, vol. 32, no. 3, 2017, pp. 1–8. *JSTOR*, http://www.jstor.org/stable/90015005.
11. Wedel, Michel, and P. K. Kannan. *"Marketing Analytics for Data-Rich Environments."* Journal of Marketing, vol. 80, no. 6, 2016, pp. 97–121. *JSTOR*, http://www.jstor.org/stable/44134975.
12. Alali, A. (2023) *Analysing Craft Beer Preferences & Market Segmentation in Västerbotten, Sweden. A Quantitative Exploration of Key Attributes*. Thesis. Available at: https://umu.diva-portal.org/smash/get/diva2:1833378/FULLTEXT03.pdf (Accessed: 25 March 2024).
13. Manikas, A., Godfrey, M., & Woldt, J. (2022). *What Drives Higher Beer Ratings? Evidence from Big Data. International Journal of Management and Marketing Research*, 15(1), 1-13. https://ideas.repec.org/a/ibf/ijmmre/v15y2022i1p1-13.html
14. Aquilani, B., Laureti, T., Poponi, S., & Secondi, L. (2015). *Beer choice and consumption determinants when craft beers are tasted: An exploratory study of consumer preferences*. Food Quality and Preference, 41, 214-224. https://doi.org/10.1016/j.foodqual.2014.12.005

15. J. Zhang and H. Shi, "*Geospatial visualization using google maps: a case study on conference presenters,*" Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007), Iowa City, IA, USA, 2007, pp. 472-476, **DOI:** 10.1109/IMSCCS.2007.9