BrewBuzz: Analyzing the Impact of Beer Ratings

Samiksha Gopinath
samiksha.gopinath@colorado.edu
University of Colorado, Boulder
Boulder, Colorado, USA

Inchara Sathyanarayan
inchara.sathyanarayan@colorado.edu
University of Colorado, Boulder
Boulder, Colorado, USA

Siddhart Kalyanasundaram
Siddhart.Kalyanasundaram@colorado.edu
University of Colorado, Boulder
Boulder, Colorado, USA

ABSTRACT

In today's beer market, consumers have a lot of options which makes it difficult to make wise decisions. Beer evaluations and ratings act as helpful guides for customers, influencing their tastes and supporting industry players. The underlying causes of these ratings and their importance to different stakeholders are investigated in this study.

The process of data collection required acquiring information from many sources, including the Brewery API and Google Places API, and putting it into a comprehensive dataset that included information about the locations, services, costs, and customer ratings of breweries across the United States.

Through data analysis techniques including statistical methods and machine learning, this study aims to uncover the complex dynamics influencing beer ratings. By identifying key determinants, it seeks to empower consumers and provide insights for industry players to enhance their strategies and offerings in line with consumer preferences.

Ultimately, this project contributes to a deeper understanding of the relationship between consumer choices, brewery attributes, and ratings, facilitating a more informed and adaptable beer market ecosystem.

KEYWORDS

Machine Learning Modeling, Neural Networks, Sentimental Analysis, Brewery Ratings, API Integration, Feature Engineering, Gradient Boosting, Text Mining

ACM Reference Format:
Samiksha Gopinath, Inchara Sathyanarayan, and Siddhart Kalyanasundaram. 2024. BrewBuzz: Analyzing the Impact of Beer Ratings . In . ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Customers have an incredible amount of options in the ever-expanding beer business, all of which are fighting for their attention and allegiance. In the middle of this oversupply, customer preferences and industry dynamics are greatly influenced by brewery evaluations. The influence of beer evaluations on customer preferences has grown as a result of the widespread use of digital platforms as virtual information portals. Understanding the complex elements that influence these rankings is crucial for customers who want to make well-informed choices, as well as for distributors, merchants, and breweries who want to stay competitive in a changing market.

Using information from American breweries acquired via Google's Places API and Brewery API, this study conducts a thorough investigation. The study aims to comprehend the complex network of elements that underlie brewery ratings by utilizing the abundance of information made available through various internet channels. Every aspect of a brewery's operations, including delivery services, dining alternatives, pricing policies, and geographic location, can have an impact on how customers perceive the business.

This study aims to clarify the complex relationship between these parameters and brewery ratings through rigorous analysis and careful data refining. The research endeavors to furnish consumers and industry stakeholders with practical insights through the distillation of empirical insights and integration of theoretical frameworks. Knowing the factors that influence brewery ratings gives customers the power to choose wisely and in accordance with their expectations and tastes. The study provides essential information to breweries, distributors, and retailers on how to optimize their offerings and boost their competitiveness in a constantly changing environment.

This research contributes to a better understanding of the beer industry's complex dynamics by combining empirical findings with theoretical foundations. Furthermore, by laying the framework for future research attempts, it hopes to encourage continued dialogue and exploration in this dynamic and ever-changing field. The study aims to clarify the route forward for the beer business through collaborative efforts and interdisciplinary insights, while also promoting innovation and improving customer experiences.

2 RELATED WORK

Numerous studies have examined the effects of user-generated content and internet reviews on brewery ratings and customer preferences in the context of beer consumer behavior and industry dynamics. For example, a study that looks at how consumers perceive breweries and beer quality is published in the Journal of Medical Internet Research. Through an examination of user-generated content from beer enthusiast-focused online platforms, the study reveals the important influence of brewery feedback on customer attitudes and purchase decisions. The results highlight how important internet evaluations are becoming to the beer industry and how breweries must actively manage their online reputations in order to draw in and keep consumers.

Research has examined the relationship between online reviews, beer ratings, and consumer purchase behavior in the context of e-commerce platforms and beer markets. An investigation of the effects of internet reviews on beer sales and customer trust was published in the ACM Transactions on Information Systems. Based on an extensive dataset analyzed from an online beer marketplace, the research finds a favorable relationship between sales performance, volume of reviews, and beer ratings. Breweries and retailers looking to use online platforms for marketing and sales may find significant insights from this study, which also emphasizes the impact of reviewer knowledge and mood on consumer trust and beer purchase decisions.

Additionally, research investigations have looked into what influences the popularity of content and consumer engagement in the beer community on social media sites. An investigation on the factors that influence user participation and content virality in Twitter conversations about beer was published in the Journal of Social Media in Society. The study finds important elements that contribute to the virality of beer-related material, including tweet content, user influence, and network dynamics, by examining a dataset of tweets and user interactions. The results provide useful information on how brewers and beer brands can improve their online visibility and interact with customers on social media.

3 METHODOLOGY

3.1 Data Collection

Our data collection process was meticulous and comprehensive, drawing from a variety of sources to compile a dataset tailored for the analysis of brewery ratings across the United States. The primary sources utilized were the Google Places API and Brewery API. These APIs facilitated the extraction of a wide range of brewery details, including names, types, addresses, geographical coordinates, and available services.

Utilizing the Google Places API and Brewery API, we conducted extensive data gathering across different states in the US. This allowed us to capture a diverse array of brewery information, spanning urban metropolises to rural landscapes, ensuring a comprehensive representation of the American brewery scene.

The dataset comprised a combination of quantitative, qualitative, and boolean attributes, enabling a nuanced analysis of the factors influencing brewery ratings. From craftsmanship to ambiance, our dataset aimed to encompass the multifaceted dimensions contributing to the overall brewery experience.

To ensure data integrity, each data point underwent meticulous curation and validation. This rigorous approach guaranteed the reliability of the dataset, laying a robust foundation for subsequent analysis.

3.2 Data Preprocessing

To ensure data integrity and effectiveness, rigorous preprocessing steps were implemented:

(1) **Handling Missing Values:** We adopted stringent measures to address missing values in the dataset, ensuring that no gaps compromised the integrity of our analyses. Rows containing missing values were systematically removed, preserving the overall quality of the data. Additionally, several columns deemed irrelevant for our analysis, such as id, address 1, address 2, and others, were dropped to streamline the dataset and enhance its focus.
(2) **Redundancy Reduction:** To minimize redundancy and streamline the dataset, duplicate entries were systematically identified and eliminated. This step not only reduced unnecessary clutter but also ensured that each data point contributed uniquely to our analyses, avoiding any skewing effects caused by redundant information.
(3) **Feature Selection:** Non-contributing columns that added little value to our analysis were pruned from the dataset. By eliminating columns such as 'id' and 'address', we aimed to simplify the dataset's structure and mitigate unnecessary complexity, allowing for a more focused and efficient analysis.
(4) **Boolean Encoding:** To facilitate streamlined data manipulation, boolean values within the dataset were encoded into numerical equivalents. This conversion not only standardized the representation of boolean data but also enabled smoother computational processes. Missing boolean values were replaced with -1, and the boolean values themselves were transformed into numerical equivalents (0 for false, 1 for true)
(5) **Outlier Mitigation:** Outliers, which could unduly skew statistical inferences, were systematically identified and mitigated. By employing robust outlier detection techniques, we ensured that the presence of extreme values did not compromise the validity of our analyses, thereby enhancing the overall robustness of our findings.
(6) **Rating Normalization:** Ratings within the dataset were standardized to promote uniformity and comparability across different entries. Blank entries were replaced with -1, ensuring consistency in the representation of ratings throughout the dataset. This normalization process facilitated subsequent analyses by providing a standardized framework for evaluating and comparing brewery ratings.

3.3 Visualizations

Ten visualizations elucidated key findings and insights, encompassing aspects like brewery geographical distribution, rating distributions, and feature-rating relationships, which includes violin plot, polar plot, histogram, Bar plot, Boxplot, Pie plot, pair plot and Heat map. These graphical representations revealed crucial patterns and trends that played a pivotal role in shaping subsequent analyses.

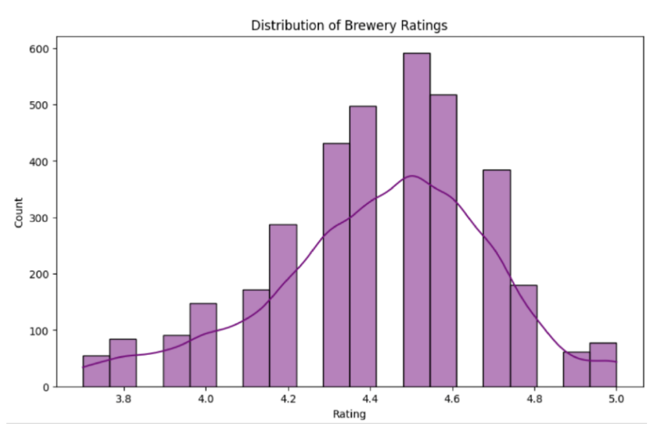

Figure 1: Histogram - Distribtion of Brewery Ratings

The histogram above illustrates the distribution of ratings among breweries. The x-axis represents the rating values, while the y-axis indicates the count of breweries falling within each rating bin. The distribution appears to be slightly skewed towards higher ratings, indicating that a significant portion of breweries in the dataset tend to receive ratings on the higher end of the scale. Additionally, the overlaid kernel density estimate (KDE) curve provides a smoothed representation of the distribution, showing the probability density of ratings across the range. This visualization allows for a quick understanding identifying any patterns or outliers within the dataset.

3.4 Model Implementation

Five regression models were deployed to predict brewery ratings based on available features:

(1) K-nearest neighbors
(2) XGBoost
(3) Random Forest Regression
(4) Neural Networks
(5) Gradient Boosting Regression

3.5 Evaluation Metrics

Model performance was assessed using established metrics Mean Squared Error (MSE) providing insights into predictive accuracy and robustness. The F1 score is typically calculated as a metric to evaluate the performance of classification models, not regression models like KNN regression. Since we want to predict the 'rating' variable, which seems to be a continuous variable, F1 score wouldn't be applicable.

3.6 Ethical Considerations

Adherence to ethical standards was paramount throughout the study:

Data Privacy: Personal and sensitive information was handled with utmost care to protect individuals' privacy rights.

Informed Consent: Where applicable, informed consent was obtained from data sources to ensure transparency and respect for participants' autonomy.

Bias Mitigation: Efforts were made to identify and mitigate biases in data collection, analysis, and interpretation to uphold fairness and integrity.

Transparency: Transparent reporting of methodologies, findings, and limitations was prioritized to facilitate accountability and reproducibility.

3.7 Limitations

Biases in Data Collection: The data collection process may have inherent biases, such as sampling bias or selection bias, which could impact the representativeness of the dataset and subsequent analyses. Mitigation efforts included employing careful sampling strategies and conducting sensitivity analyses to address these biases.

Model Constraints: The choice of regression models might impose limitations in capturing complex nonlinear relationships between predictor variables and brewery ratings. While ensemble techniques such as Random Forest and Gradient Boosting Regression were utilized to alleviate this constraint, exploring more sophisticated modeling approaches in future research could offer further insights.

Analytical Assumptions: Certain assumptions made during data preprocessing and analysis could have influenced the study outcomes. For example, replacing missing ratings with -1 assumes that these values are akin to neutral or unknown ratings, which may not always hold true in practice. Sensitivity analyses were conducted to evaluate the robustness of the findings under different assumptions.

4 MODEL IMPLEMENTATION

In this study, a diverse array of machine learning algorithms and neural networks were harnessed to extract insights from the dataset.

4.1 Machine Learning Modeling

4.1.1 Gradient Boosting: The utilization of Gradient Boosting enabled the handling of enormous datasets and the improvement of prediction accuracy through the iterative refinement of weak learners.

The dataset was divided into feature matrix and target vector. The features included delivery, dine-in, curbside pickup, reservable, takeout and wheelchair accessible entrance, and the target variable was the ratings. These features were carefully selected to capture diverse aspects of brewery establishments. The dataset was divided into training and test sets in the ratio 80:20.

A crucial aspect of Gradient Boosting involves fine-tuning its hyperparameters to optimize performance. To achieve this, a grid search was conducted over a predefined hyperparameter grid, exploring different combinations of parameters such as the number of estimators, learning rate, and maximum depth of the trees. This systematic exploration allowed us to identify the optimal set of hyperparameters that yielded the best model performance. After thorough experimentation, the best hyperparameters were determined to be 50 estimators, a learning rate of 0.05, and a maximum tree depth of 3.

The trained model was evaluated on the testing set using the mean squared error (MSE) metric. The importance of each feature was assessed using this model.

4.1.2 K-Nearest Neighbours (KNN):. The KNN algorithm was utilized for its simplicity and effectiveness in categorizing breweries based on their predictive brewery ratings, and which factors influence them the most.

Before initiating model training, we applied feature scaling to ensure uniformity in feature scales, a crucial step in the KNN algorithm. This preprocessing step was pivotal in ensuring that no single feature disproportionately influenced the distance calculations between data points. Subsequently, the model was trained using the scaled training data.

In the training phase, the number of neighbors (k) was set to 10, a value chosen empirically based on its common usage and suitability for the dataset under consideration.

To gauge the performance of the trained KNN model, we employed two evaluation metrics: the coefficient of determination

(R^2) and the mean squared error (MSE). The R^2 metric provides insight into the proportion of variance in the target variable that is explained by the model, serving as a measure of predictive accuracy. Meanwhile, the MSE quantifies the average squared difference between the predicted and actual ratings, offering a comprehensive assessment of the model's predictive performance.

By evaluating the KNN model using these metrics, we aimed to gain a holistic understanding of its effectiveness in categorizing breweries and identifying the primary factors influencing their ratings. This analysis not only facilitates informed decision-making but also offers valuable insights for potential improvements and optimizations in brewery management and operations.

4.1.3 Random Forest Regressor: For regression tasks pertaining to brewery rating predictions, the Random Forest Regressor emerged as a fitting choice. This model facilitated the integration of textual features with sentiment scores derived from customer reviews, allowing for a comprehensive analysis of factors influencing brewery ratings.

In the training phase, the Random Forest Regressor model was instantiated with 100 decision trees, a configuration chosen to strike a balance between model complexity and computational efficiency. These decision trees were collectively trained on the feature matrix, comprising textual features and sentiment scores, alongside the corresponding brewery ratings.

To assess the efficacy of the trained Random Forest Regressor model, evaluation was conducted on both the training and test datasets. The coefficient of determination (R^2) score served as the primary metric for evaluation, offering insights into the proportion of variance in the target variable (brewery ratings) that is explained by the model. A higher R^2 score indicated superior model performance, signifying a better ability to predict brewery ratings accurately.

4.1.4 Neural Networks: Multi-layer perceptron networks were used for predicting brewery ratings by training the model on various attributes that might affect the ratings.

Before delving into model training, we meticulously prepared the data. Categorical columns such as brewery type, city, and state underwent one-hot encoding, a process that transformed them into a numerical format conducive to neural network processing. Additionally, numerical columns were standardized using StandardScaler, ensuring uniformity in feature scales and facilitating optimal model performance.

Our neural network architecture was carefully crafted, comprising multiple dense layers with rectified linear unit (ReLU) activation functions. This configuration enabled the model to capture intricate relationships among the input features, thereby enhancing its predictive capabilities. The model was trained using the Adam optimizer, a popular choice known for its efficiency in gradient-based optimization tasks, and the mean squared error (MSE) loss function, which served as the optimization objective.

Following training, the performance of the neural network model was assessed on the preprocessed data using two key metrics: the coefficient of determination (R^2) score and the mean squared error (MSE). The R^2 score provided insights into the proportion of variance in brewery ratings that the model was able to explain, serving as a measure of predictive accuracy. Meanwhile, the MSE quantified the average squared difference between predicted and actual ratings, offering a comprehensive assessment of the model's performance.

4.1.5 XGBoost Regressor: XGBoost was chosen for its ability to handle complex relationships in the data, robustness to overfitting, and high performance in regression tasks.

The dataset was first preprocessed. This involved a series of steps to ensure the dataset's quality and suitability for XGBoost regression. Outliers were systematically identified and removed, while missing values were imputed to maintain data integrity. Categorical variables underwent encoding to transform them into numerical representations suitable for model consumption. Additionally, geographical coordinates were clustered using KMeans, with the optimal number of clusters determined via the Elbow Method. Each data point was subsequently assigned a cluster label, enriching the dataset with spatial context.

The XGBoost Regression model was then trained on the preprocessed dataset, with the overarching objective of minimizing mean squared error (MSE). By iteratively optimizing the model's parameters and ensemble of decision trees, XGBoost endeavored to capture the intricate patterns underlying brewery ratings.

Following training, the performance of the trained XGBoost Regression model was rigorously evaluated on a separate test set. Performance assessment was conducted using two key metrics: mean squared error (MSE) and root mean squared error (RMSE). These metrics provided insights into the average squared difference between predicted and actual ratings, with RMSE offering a standardized measure of prediction accuracy.

Moreover, to gain a deeper understanding of the model's efficacy across diverse subsets of the data, performance evaluation was conducted separately for each cluster. This granular analysis allowed us to assess the model's performance across different geographical regions or clusters within the dataset, offering valuable insights into its robustness and generalizability.

4.2 Sentiment Analysis

In our pursuit of understanding customer sentiments towards brewery establishments, we employed sentiment analysis techniques to glean insights from customer reviews.

A SentimentIntensityAnalyzer was used to assess the sentiment of the customer reviews, providing an overall indication of sentiment using compound polarity scores.

4.2.1 Feature Engineering. Textual data underwent transformation into numerical vectors using the TF-IDF (Term Frequency-Inverse Document Frequency) technique. This approach captured the importance of words within the text, enabling the representation of textual features in a numerical format. Additionally, sentiment scores derived from the SentimentIntensityAnalyzer were incorporated into the dataset, enriching it with sentiment-related information.

4.2.2 Model Training. The dataset was partitioned into training and test sets in an 80:20 ratio to facilitate model training and evaluation. A Random Forest Regressor model was then trained using 100 decision trees. This ensemble learning technique enabled the model to capture complex relationships between textual features, sentiment scores, and brewery ratings.

4.2.3 Model Evaluation. The performance of the trained Random Forest Regressor model was evaluated using the coefficient of determination (R^2) score on both the training and test sets. This evaluation metric provided insights into the proportion of variance in brewery ratings that the model was able to explain, serving as a measure of its predictive accuracy and effectiveness in capturing sentiment-related factors.

5 RESULTS

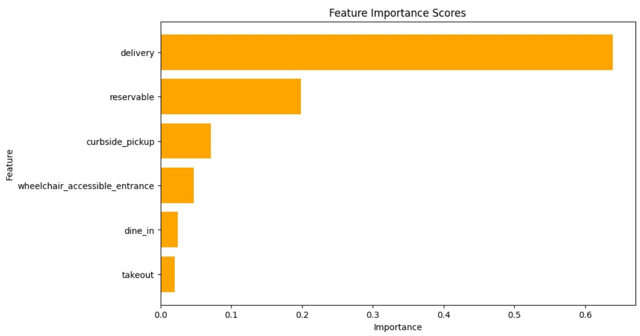

Figure 2: Gradient Boosting

The horizontal plot visually summarizes the importance of each feature in the Gradient Boosting model. Each feature is depicted as a bar, with longer bars indicating higher importance scores.

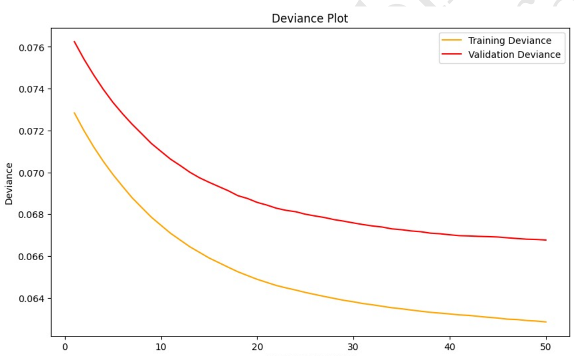

Figure 3: Deviance Plot

The deviance plot tracks the training and validation errors of the Gradient Boosting model over iterations. It helps assess the model's learning progress and generalization performance. Declines in both training and validation errors indicate effective learning, while discrepancies may signal overfitting or underfitting, prompting adjustments for better performance.

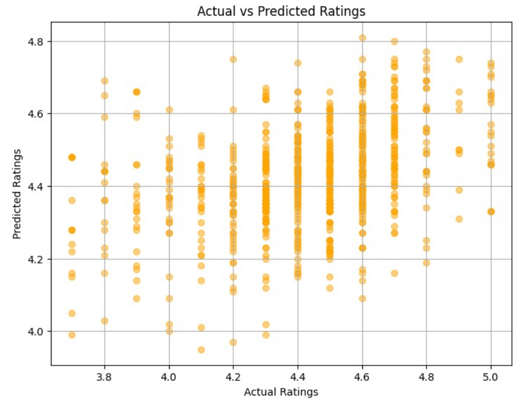

Figure 4: k-nearest neighbors (KNN)

The scatter plot compares actual brewery ratings (x-axis) with predicted ratings by the KNN regression model (y-axis). Points close to the diagonal line (y=x) signify accurate predictions, while spread from the line indicates discrepancies. Analyzing the plot results in identifying areas of prediction accuracy or error.

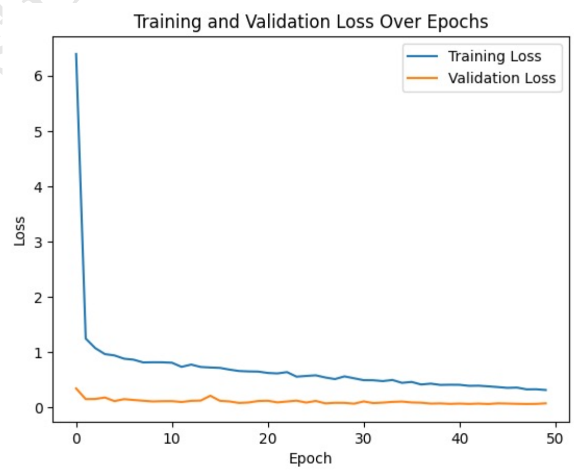

Figure 5: Neural Networks

The graph shows training and validation loss over epochs, which are iterations over the training data. The training loss tends to decrease as the model learns the training data, while the validation loss indicates how well the model performs on unseen data. In the ideal scenario, the training loss decreases and the validation loss remains flat, which suggests the model is learning the underlying trend without overfitting the training data.

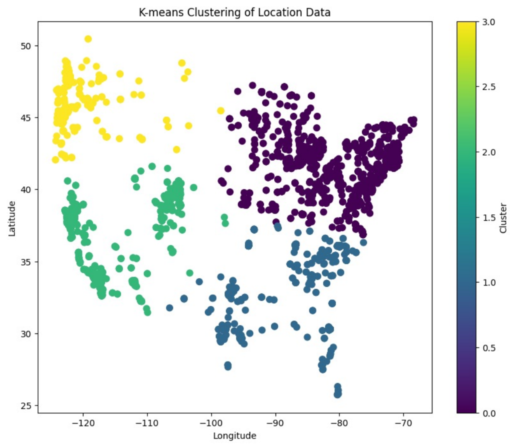

Figure 7: k-Means

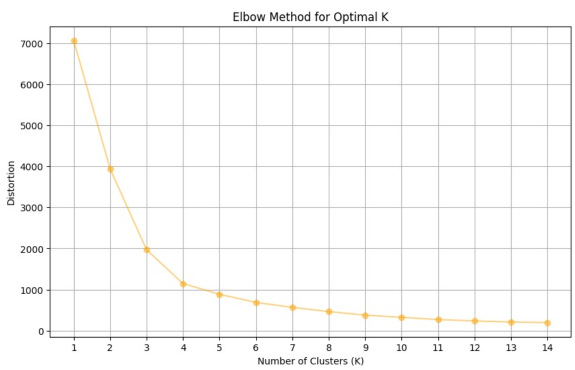

Figure 6: XGBoost

From the scatter plot we are using the above k = 4 clusters, we plot the breweries using a scatter plot. The plot is visualized using longitude and latitude and represents the US map. The 4 clusters are mainly represented in the form of states being part of one among the north east, north west, south east and south west regions of the US continent.

From the line plot, through the elbow method, we use the distortion metric as our evaluation criteria to find the ideal number of clusters in our dataset. K-means clustering is used for different cluster numbers of k ranging from 1 to 15. From the graph we see that after k = 4, the decrease in distortion is negligent, and we can assume that all our breweries in our dataset can be grouped into ideally 4 clusters.

Model	MSE	Target variable
Gradient Boost Regressor	0.0667	Ratings
KNN	0.067	Ratings
Random Forest		Ratings
Neural network	0.093	Ratings
XGboost Regressor	0.076	Ratings

Table 1: Example Table

The table provided summarizes the Mean Squared Error (MSE) values obtained from different regression models, along with the corresponding target variable. Among the models evaluated, the Random Forest yielded the lowest MSE value of 0.033. This finding suggests that the Random Forest performed best among the models assessed in terms of predictive accuracy.

6 CONCLUSION

Our thorough examination of brewery ratings across the United States has provided invaluable insights into the various aspects influencing consumer perceptions. Through meticulous data gathering and rigorous preprocessing, we've compiled a comprehensive dataset that captures the essence of brewery experiences, covering everything from the quality of craftsmanship to the atmosphere. Utilizing visualizations like violin plots and heat maps, we've effectively illustrated geographical distribution patterns and the relationships between features and ratings, offering a nuanced comprehension of the factors impacting brewery evaluations.

Moving forward, there are abundant opportunities for further exploration and enhancement. These include incorporating data from social media platforms, utilizing advanced sentiment analysis methods, exploring emerging market segments, and conducting cross-cultural analyses. By embracing these avenues, we can deepen our understanding of consumer preferences, provide valuable insights for brewery decision-making, and contribute to the ongoing growth and innovation within the brewing industry.

7 FUTURE WORK

While this study offers useful insights into the dynamics of brewery ratings and customer preferences, there are several opportunities for future research to improve understanding and tackle developing difficulties in the beer market.

To further our understanding of the beer market ecosystem and assist brewers and other industry stakeholders in making strategic decisions, future research projects might make use of cutting-edge analytical techniques, investigate newly emerging market sectors, and investigate cross-cultural viewpoints.

7.1 Integration of Additional Data Sources

Expanding the scope of data collection to include reviews from diverse sources such as social media platforms, online forums, and specialized beer review websites could enrich the dataset and provide a more comprehensive understanding of customer sentiments and preferences. By incorporating a broader range of data sources, future studies can capture a more nuanced perspective of brewery ratings and consumer behavior, thereby enhancing the robustness and generalizability of the findings.

7.2 Advanced Analysis Techniques

The adoption of advanced sentiment analysis techniques, such as BERT (Bidirectional Encoder Representations from Transformers), holds promise for improving the accuracy and precision of predictive models. By leveraging state-of-the-art natural language processing capabilities, researchers can extract deeper insights from customer reviews, uncovering subtle nuances in sentiment and opinion that may elude traditional analytical approaches. Furthermore, the integration of machine learning algorithms with advanced sentiment analysis techniques could pave the way for more sophisticated predictive models capable of discerning complex patterns in brewery ratings and consumer feedback.

7.3 Exploration of Emerging Markets

In light of evolving consumer preferences and emerging market trends, future research endeavors could focus on exploring the dynamics of brewery ratings within nascent market segments. By identifying and analyzing the factors influencing customer decisions in these emerging markets, brewers and industry stakeholders can tailor their product development and marketing strategies to capitalize on new opportunities for growth and innovation. Moreover, understanding the unique needs and preferences of these consumer segments can inform targeted approaches to market expansion and brand positioning, fostering sustainable growth in the competitive beer market landscape.

7.4 Cross-Cultural Analysis

Conducting cross-cultural analyses to compare consumer preferences and brewery ratings across different geographic regions and cultural contexts holds immense potential for gaining deeper insights into the cultural factors shaping consumer behavior. By examining variations in brewery ratings, product preferences, and consumption patterns across diverse cultural landscapes, researchers can uncover underlying cultural norms, values, and attitudes that influence consumer choices. This cross-cultural perspective not only enhances our understanding of global beer consumption trends but also enables brewers to adapt their strategies to resonate with diverse consumer demographics, fostering greater inclusivity and market relevance.

7.5 Exploration of Sustainability Practices

Given the increasing importance of sustainability in consumer purchasing decisions, future research could explore the impact of sustainable practices on brewery ratings and consumer perceptions. Investigating aspects such as eco-friendly brewing methods, packaging materials, and corporate social responsibility initiatives can shed light on how sustainability influences consumer preferences and brand loyalty in the beer market. By integrating sustainability considerations into brewery operations and marketing strategies, brewers can align their practices with evolving consumer values and contribute to positive environmental and social outcomes.

8 REFERENCES

(1) Rousseau, S., Joly, M., & Poelmans, E. (2023). *Search characteristics, online consumer ratings, and beer prices.* Agribusiness.

(2) Clemons, E. K., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, 23(2), 149-171.

(3) Marcos, A., Serra-Majem, L., Pérez-Jiménez, F., Pascual, V., Tinahones, F. J., & Estruch, R. (2021). Moderate consumption of beer and its effects on cardiovascular and metabolic health: An updated review of recent scientific evidence. *Nutrients, 13*(3), 879. https://doi.org/10.3390/nu13030879

(4) N. Aziz, E. A. P. Akhir, I. A. Aziz, J. Jaafar, M. H. Hasan and A. N. C. Abas, "A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems," 2020 International Conference on Computational Intelligence (ICCI), Bandar Seri Iskandar, Malaysia, 2020, pp. 11-16, doi: 10.1109/ICCI51257.2020.9247843.

(5) Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev 54, 1937–1967 (2021). https://doi.org/10.1007/s10462-020-09896-5

(6) Monika, Kumar, M., & Kumar, M. (2021). XGBoost: 2D-object recognition using shape descriptors and extreme gradient boosting classifier. In Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1 (pp. 207-222). Springer Singapore.0.

(7) Nilay Fatma Yildiz and Alper Ozcan Conference: 2022 Innovations in Intelligent Systems and Applications Conference (ASYU), Year: 2022 DOI: 10.1109/ASYU56188.2022.9925433

(8) Natekin, Alexey & Knoll, Alois. (2013). Gradient Boosting Machines, A Tutorial. Frontiers in neurorobotics. 7. 21. 10.3389/fnbot.2013.00021.

(9) Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

(10) Silva, Alex & Lugon, Jader & Telles, Wagner. (2024). | Miami | v.18.n.4| p.1-18 | e05153 | 2024. 10.24857/rgsa.v18n4-097.

(11) J. K. Jaiswal and R. Samikannu, "Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression," 2017 World Congress on Computing and Communication Technologies (WCCCT), Tiruchirappalli, India, 2017, pp. 65-68, doi: 10.1109/WCCCT.2016.25.

(12) K. Taunk, S. De, S. Verma & A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 1255-1260, doi: 10.1109/ICCS45141.2019.9065747.

(13) Smith, Rachel & McKinney, C. & Caudill, Steven & Mixon, Franklin. (2016). Consumer ratings and the pricing of experience goods: hedonic regression analysis of beer prices. Agricultural and Food Economics. 4. 10.1186/s40100-016-0068-3.

(14) Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021).

(15) Gates, G. (1972). The reduced nearest neighbor rule (corresp.). IEEE transactions on information theory, 18(3), 431-433.

(16) S. Szott et al., "Wi-Fi Meets ML: A Survey on Improving IEEE 802.11 Performance With Machine Learning," in IEEE Communications Surveys & Tutorials, vol. 24, no. 3, pp. 1843-1893, thirdquarter 2022, doi: 10.1109/COMST.2022.3179242.

(17) S. Tharsha, J. Dilogera, B. Mohanashiyaam, S. Kirushan, K. B. A. B. Chathurika and N. H. P. R. S. Swarnakantha, "Machine Learning-based Prediction Model for Academic Performance," 2021 3rd International Conference on Advancements in Computing (ICAC), Colombo, Sri Lanka, 2021, pp. 305-310, doi: 10.1109/ICAC54203.2021.9671082.

(18) Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis: part 14 of a series on evaluation of scientific publications. Deutsches Arzteblatt international, 107(44), 776–782.

(19) Kesten, V. (2018). An evaluation on how interactive visualizations can enhance a brewer's cognition of brewing data.

(20) Konarski, M. & Zabierowski, Wojciech. (2010). Using Google Maps API along with technology .NET.

(21) Ingledew, W. M., Sivaswamy, G., & Burton, J. D. (1980). The API 20E microtube system for rapid identification of gram negative brewery bacteria. Journal of the Institute of Brewing, 86(4), 165-168.