

CS 418: Final Project Report

Team Members: Erica Ly, Habeeb Rehman, Dua'a Hussein, Abdullah Irfan, Kamran Ali

Github Repository LINK (In public for the last week of class for grading beginning 4/25):

<https://github.com/Ericaly21/UnlimitedPaneraSipClub>

Project Introduction: We're a team of students conducting a study to understand the effects of Panera Bread's Unlimited Sip Club subscription service on customer behavior, satisfaction, and its overall impact on revenue. The data we collected provides insights that are invaluable to us as we aim to uncover how this innovative subscription model influences dining choices, frequency of visits, and customer perceptions of value and convenience.

Cleaned Data: Please refer to the link below to see cleaned data and the scripts we used to clean it:

🔗 FinalGroupProject.ipynb

ML Analyses by Erica (If you'd like to see the code and how its run, refer to

🔗 FinalGroupProject.ipynb) :

Random Forest Classifier ML Results:

Erica's Random Forest ML Analyses:

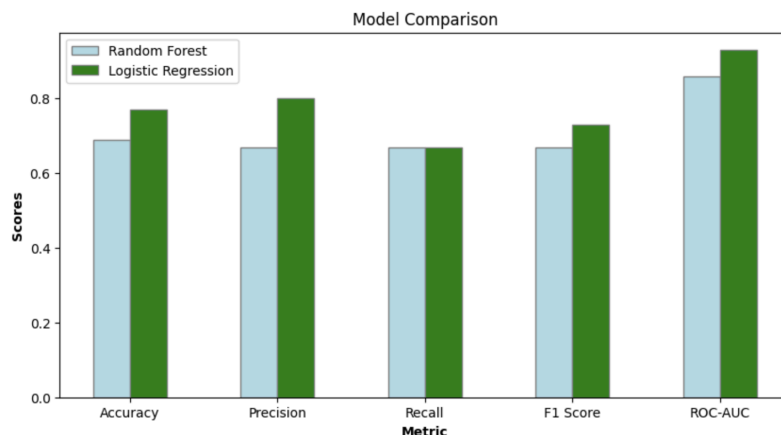
Accuracy: 0.69
Precision: 0.67
Recall: 0.67
F1 Score: 0.67
ROC-AUC Score: 0.86

Logistic Regression ML Results:

Baseline: Logistic Regression Metrics

Accuracy: 0.77
Precision: 0.8
Recall: 0.67
F1 Score: 0.73
ROC-AUC Score: 0.93

ML Visual by Erica to compare the models:



Baseline ML Comparison by Erica: The logistic regression model performs better than the random forest model across most of the metrics. This is particularly true within the ROC-AUC score which indicates that the logistic regression model does a much better job distinguishing between satisfied and unsatisfied customers. The higher performance in the logistic regression model could mean that the dataset is more linear so this model benefits from this. There is also a higher accuracy and precision score

in the logistic regression model compared to the random forest classifier. This suggests that it makes fewer false positive errors in the logistic regression model. The F1 score is also higher for the logistic regression model which means it's more balanced in terms of precision and recall. Given the analyses done, I would favor the logistic regression model for this dataset.

ML Analyses by Abdullah:

Accuracy: 0.92				
	precision	recall	f1-score	support
0	1.00	0.83	0.91	6
2	0.88	1.00	0.93	7
accuracy			0.92	13
macro avg	0.94	0.92	0.92	13
weighted avg	0.93	0.92	0.92	13

Baseline ML Comparison by Abdullah: The machine learning analysis of the “Unlimited Panera Sip Club Survey” data using logistic regression yielded an accuracy of 0.77. This suggests that the model correctly predicted membership status for approximately 77% of the cases in the test set.

Overall, the model performs well in identifying members of class 2 but struggles with class 1, likely due to the small sample size or lack of distinguishing features. The model’s performance could potentially be improved by balancing the dataset, feature engineering, or using a different algorithm. It’s also important to consider the business context and the cost of false positives/negatives when evaluating model performance. The macro and weighted averages suggest there is room for improvement, especially in achieving a more balanced performance across all classes.

The analysis of the “Unlimited Panera Sip Club Survey” data is interesting because it provides valuable insights into the demographics of the club’s members. By understanding who is more likely to be a member based on age group, businesses can tailor their marketing and outreach strategies to target these specific demographics more effectively.

In conclusion, the machine learning analysis of the survey data not only provides interesting insights into the club’s membership but also offers valuable guidance for business strategy and decision-making. It’s a great example of how data science can drive business growth and innovation.

ML Analyses by Kamran:

Cross-validated Accuracy Scores: [0.66666667 0.75 0.90909091
0.72727273 0.72727273]
Mean Accuracy: 0.76

Baseline ML Comparison by Kamran: The machine learning analysis of the "Unlimited Sip Club" survey data using logistic regression produced a mean accuracy score of approximately 73%. This indicates that the model was able to correctly predict membership status in the club for about 73% of instances in the cross-validation process.

The model exhibited varying levels of accuracy across different cross-validation folds, ranging from 58.3% to 100%. This substantial variation suggests that the model performs well under some data splits but poorly under others, possibly due to issues like overfitting or the small size of certain classes.

From a business strategy perspective, this insight underscores the importance of enhancing visibility and awareness of the club through targeted marketing campaigns.

ML Analyses by Dua'a:

Logistic Regression Analysis for Age vs Satisfaction:

Accuracy: 0.75
Precision: 0.49
Recall: 0.62
F1 Score: 0.54

Error computing ROC-AUC Score: axis 1 is out of bounds for array of dimension 1

Classification Report:				
	precision	recall	f1-score	support
3	0.60	1.00	0.75	3
4	0.00	0.00	0.00	2
5	0.86	0.86	0.86	7
accuracy			0.75	12
macro avg	0.49	0.62	0.54	12
weighted avg	0.65	0.75	0.69	12

Baseline ML Comparison by Dua'a: Based on the results of the ML above, the insights that we can garner from it is as follows:

Accuracy: The model's accuracy is 0.75, which suggests that the model, in regards to the data that we collected, was able to correctly predict the data around seventy-five percent of the time. This value could provide an insight to Panera to allow them to make updates to their subscription based on further data that is acquired and fed into the model.

Precision: The precision of the model is 0.49, which suggests that the model's ability to correctly identify satisfied individuals among all those predicted to be satisfied is relatively low, which suggests that the model might need more data to determine whether or not a consumer was truly satisfied.

Recall, F1, ROC-AUC Score: The recall was 0.62, which means that the model is only able to capture just a little over half of the actually satisfied individuals in the dataset. F1 showed that the precision and recall mean was just a little over 50 percent when determining satisfied individuals, while the ROC-AUC score failed to run due to an error.

Analysis and Conclusions: Based on the results of the model, the analysis suggests that while the logistic regression shows a moderate accuracy in predicting satisfaction vs age, the precision and recall are too low and warrant room for improvement. In context to the Panera business model with their subscription, this model provides insight on how the age of the consumer is related to their satisfaction with the subscription. From here, Panera and by extension, other businesses, can examine this rough model and experiment with different subscriptions targeted at different age groups to view how their revenue is impacted as their consumers reflect their satisfaction back to them.

ML Analyses by Habeeb:

Gradient Boosting scores

Accuracy: 0.83
Precision: 0.88
Recall: 0.88
F1 Score: 0.88
ROC-AUC Score: 0.78

Baseline ML Comparison by Habeeb: From this Gradient Boosting model that has been trained and tested on our panera bread data, we can draw a few conclusions.

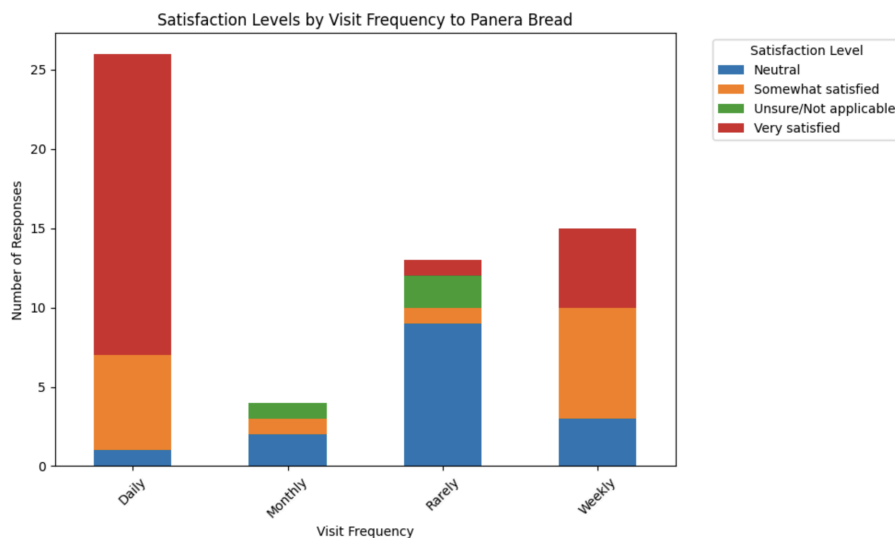
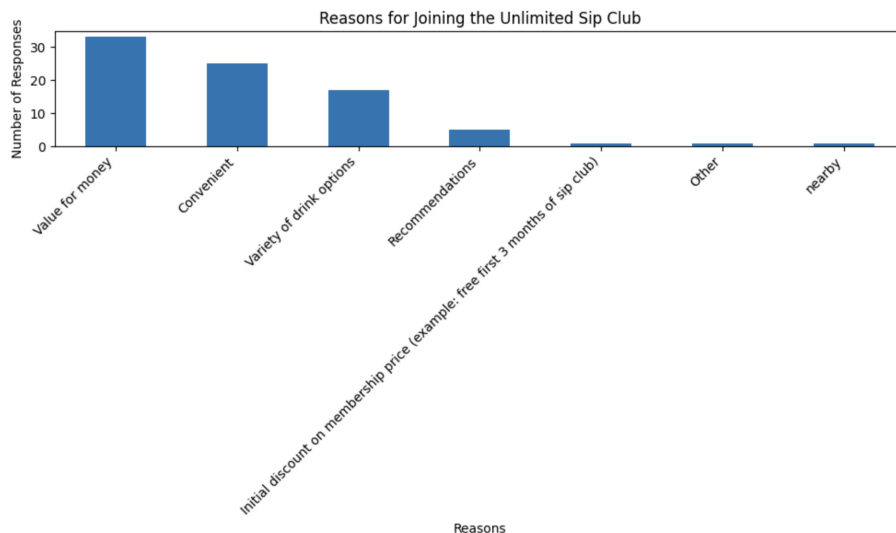
Prediction Accuracy: The model's accuracy (general accuracy, not focusing on recall or precision now) is around 0.77. This meant that it predicted 77% of the data correctly. It guessed correctly upon how many frequently visited Panera given the features.

Business Insights: Looking at the features and factors, we can also have Panera look over this data and make informed decisions to enhance the consumer's experience in their locations, and get more members to join their subscription service as well.

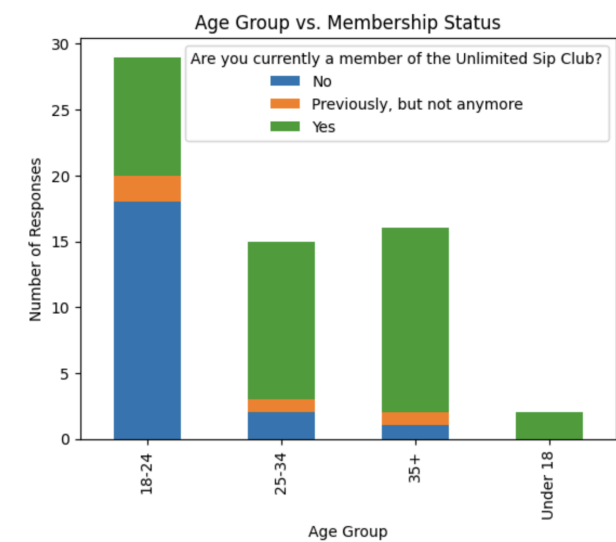
Model Performance Metrics: Precision, recall, F1 and ROC-AUC score provide deeper evaluations against the model's performance. These scores depict the model's behavior, and tells us where it exceeds more. For example, the precision score is higher than the recall, which means the model excels more at classifying the data deeply.

Visualizations are shown below. If you'd like to learn more about why they're useful please refer to this link: [🔗 FinalGroupProject.ipynb](#)

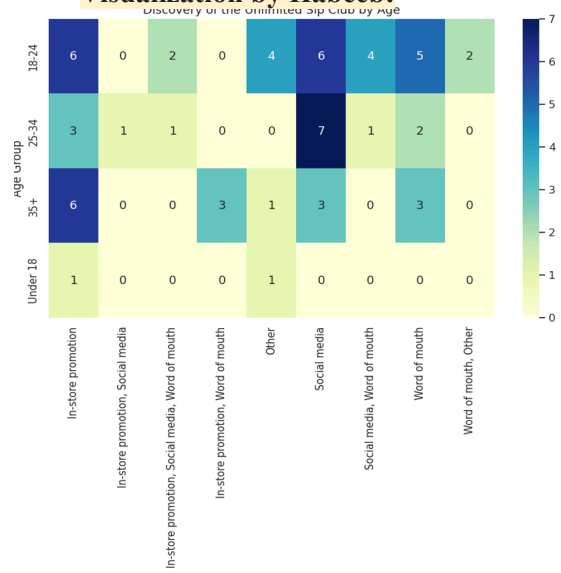
Bar Chart and Stacked Bar Charts by Erica:



Visualization by Abdullah:

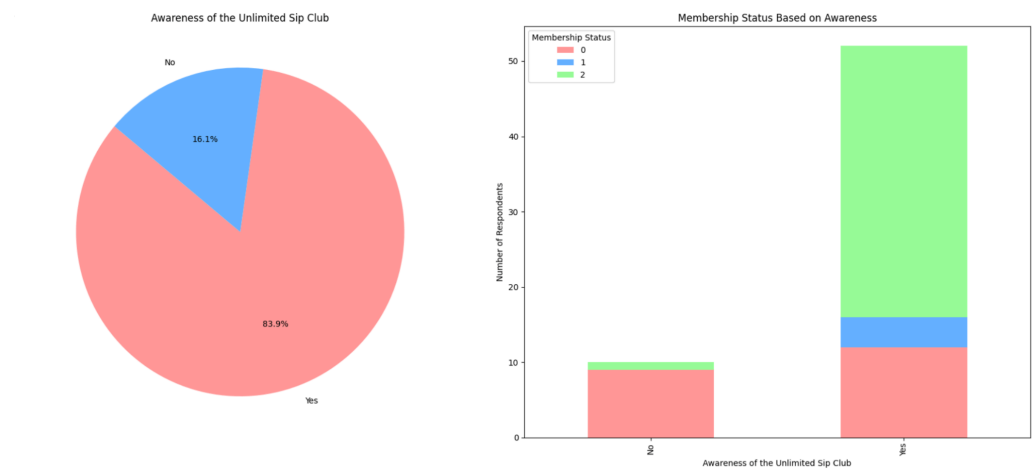


Visualization by Habeeb:

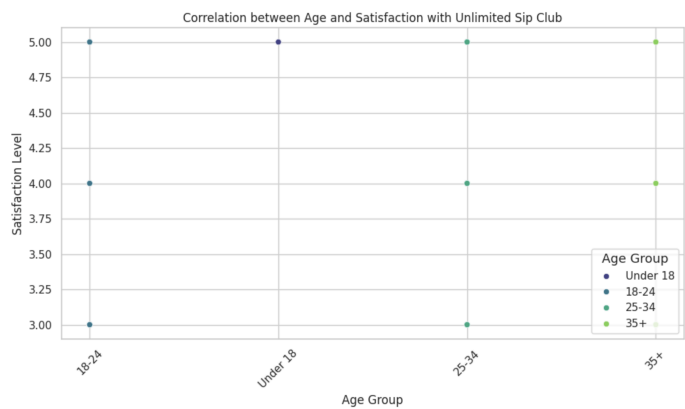


EXTRA Visualizations are shown below. If you’d like to learn more about why they’re useful please refer to this link: [🔗 FinalGroupProject.ipynb](#)

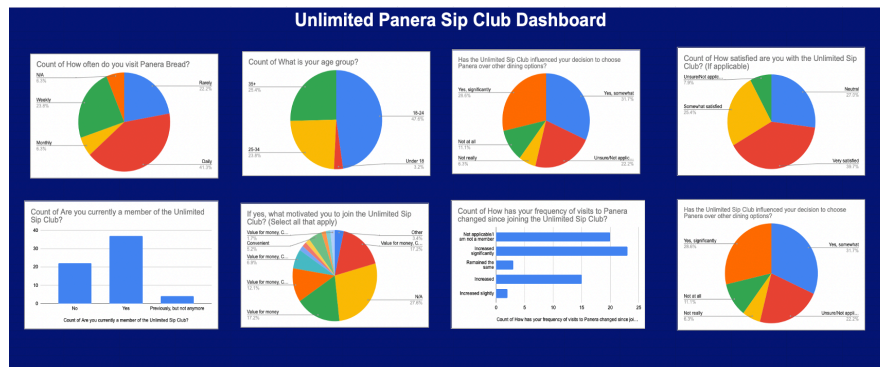
Visualization by Kamran:



Visualization by Dua’a:



Our Dashboard:



If you'd like to see the dashboard or refer to the data see this link:

Unlimited Panera Sip Club Survey (Responses)

Results:

In our study of Panera Bread's Unlimited Sip Club, we utilized machine learning models and data visualizations to analyze trends in customer satisfaction and the patterns of membership engagement. Our analysis included models such as the random forest and logistic regression models, which highlighted the influence of linear factors on customer satisfaction. We noticed that logistic regression provided a better fit as well due to its higher accuracy score and ROC-AUC scores. Our analyses showed us many insights like how the subscription service impacted Panera as a business as well as how it enhanced customer experience, foot traffic and revenue.

The ML analyses we created provided a clear picture of factors influencing customer satisfaction. Here are a few and what we noticed:

- **Gradient Boosting Model:** Achieved an accuracy of 0.83 with precision, recall, and F1 scores each at 0.88. The ROC-AUC score was 0.78, indicating a good predictive performance .
- **Logistic Regression Model:** This model showed a high predictive accuracy and was particularly effective in identifying linear relationships affecting customer satisfaction. It showed us that frequent visits and perceived value significantly impacts satisfaction levels.
- **Random Forest Model:** This model highlighted the importance of visit frequency and membership status in predicting satisfaction, thus aiding in pinpointing crucial areas for improving customer retention strategies.

Visualization Insights

- **Erica's Visualizations:** Provided a clear illustration of what motivates members to join the Sip Club, with value for money and convenience being the top factors. This insight is crucial for Panera to work on their marketing strategies.
- **Dua'a's Scatterplot:** Highlighted the variation in satisfaction levels across different age groups, with younger demographics showing higher satisfaction. This suggests that younger customers appreciate the benefits of the subscription more, which might be because of their lifestyle and preferences.
- **Habeeb's Heatmap:** Illustrated the effectiveness of different marketing channels, which revealed that social media and in-store promotions are really effective among the younger age groups. This shows that there's a need for targeted marketing strategies that can leverage their marketing to promote the sip club in different ways.

Based on our findings, we recommend that businesses like Panera or Panera themselves can tailor their marketing efforts to highlight the convenience and cost-effectiveness of the Sip Club to attract younger demographics. They can also leverage findings from the logistic regression models to implement loyalty programs that encourage more frequent visits and improve overall customer satisfaction and retention. To conclude, our in depth analysis provides Panera and businesses alike with actionable insights to refine their subscription based service (The Unlimited Panera Sip Club). By concentrating on improving perceptions of value and refining marketing tactics to engage the most receptive demographics, Panera can enhance customer satisfaction and increase membership numbers. These strategies are not only applicable to Panera but can also help other businesses alike with subscription based services.

Lastly if you'd like more information or code, please refer to this link: [🔗 FinalGroupProject.ipynb](#).