To further illustrate the importance of analyzing sales data in bakery settings and the impact of ebbs and flows of business throughout the year, our analysis also focused on the intra-relatedness of assorted items within single transactions in order to optimize seasonal performance. The ability to algorithmically predict item groupings that are frequently purchased together offers bakeries a modern avenue to optimize profits and streamline their operations. By leveraging these insights, bakeries can implement targeted promotions and sales strategies aimed at companion items that are frequently bought in the same transaction. For instance, identifying that cakes and coffee are frequently purchased together could prompt the bakery to offer a bundled discount on this pairing, enticing customers who might initially come in with the sole intent of purchasing a pre-ordered cake to make a spontaneous coffee purchase and boost overall sales revenue.

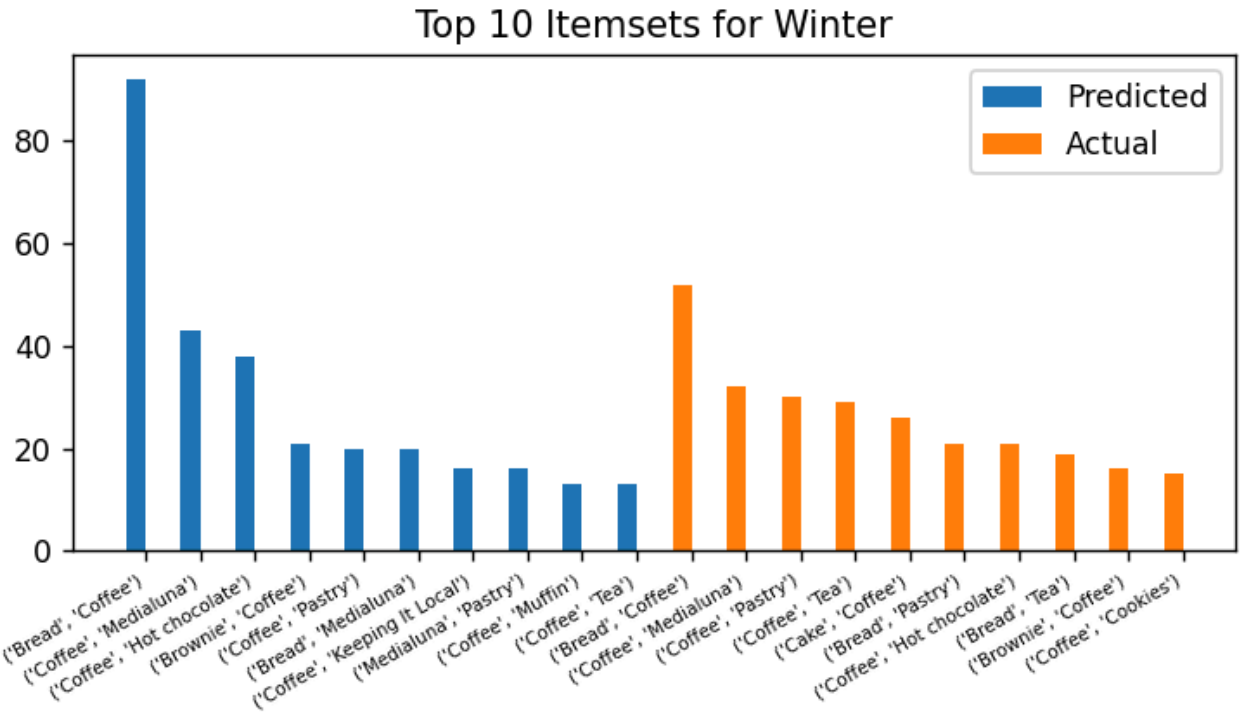
Additionally, understanding the seasonal dynamics of popular item pairings enables bakeries to make informed decisions about inventory management. By anticipating which materials will be in high demand during specific seasons or promotional periods, bakeries can optimize their ingredient purchasing schedules, minimize overstocking, and reduce potential losses due to excess inventory.

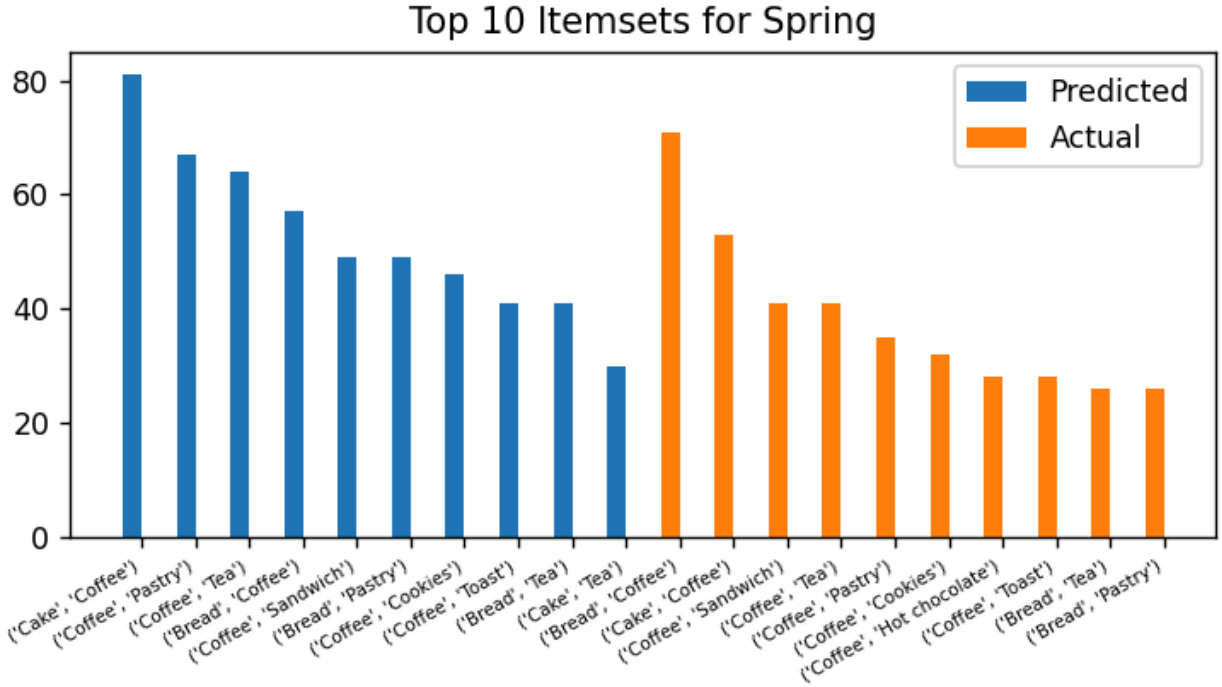
In terms of marketing, predictive insights derived from transactional data can inform strategic decisions on when and how to prioritize promotional campaigns. For instance, through predicting which item groupings will be popular during specific seasons, bakeries can align advertising efforts to capitalize on these trends, maximizing the impact of marketing investments.

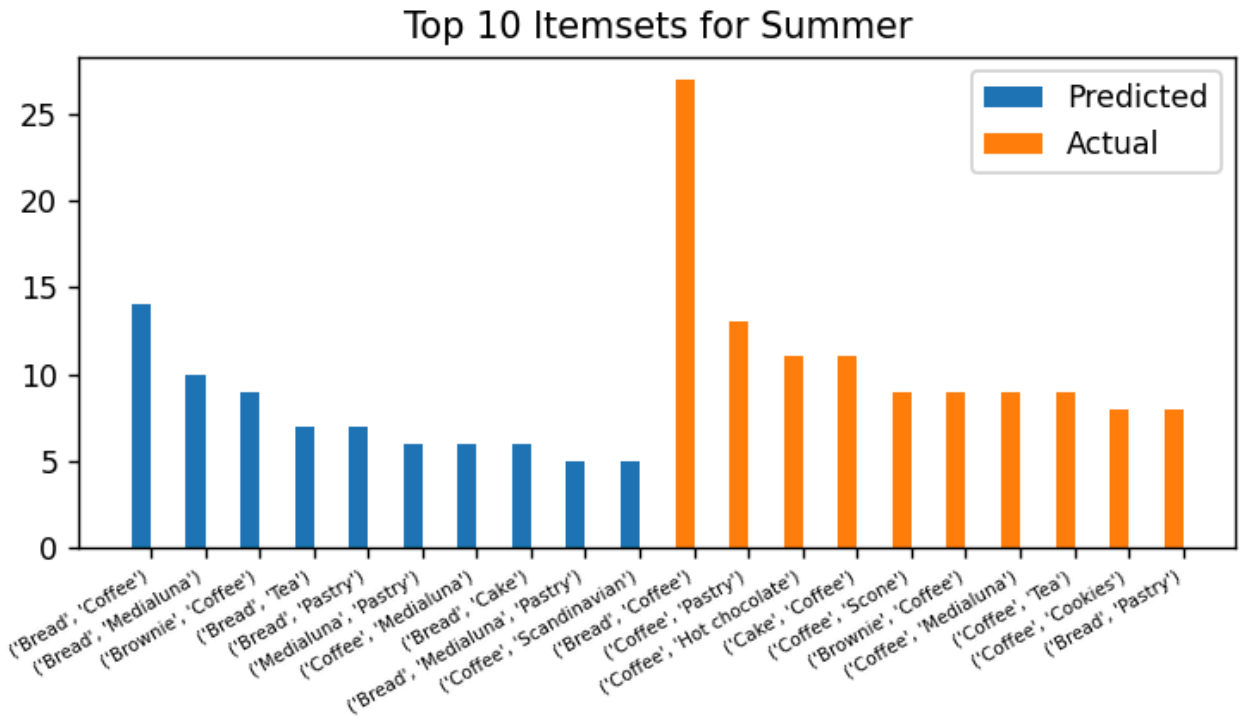
When it comes to predictions made during this analysis, the optimal strategy that emerged was to employ a decision tree classifier trained via attributes refined through recursive feature elimination. This approach excelled in uncovering and using correlations within transactional data that had hindered other classification methods, but resulted in more accurate predictions of consumer preferences based on the records of previous purchases.

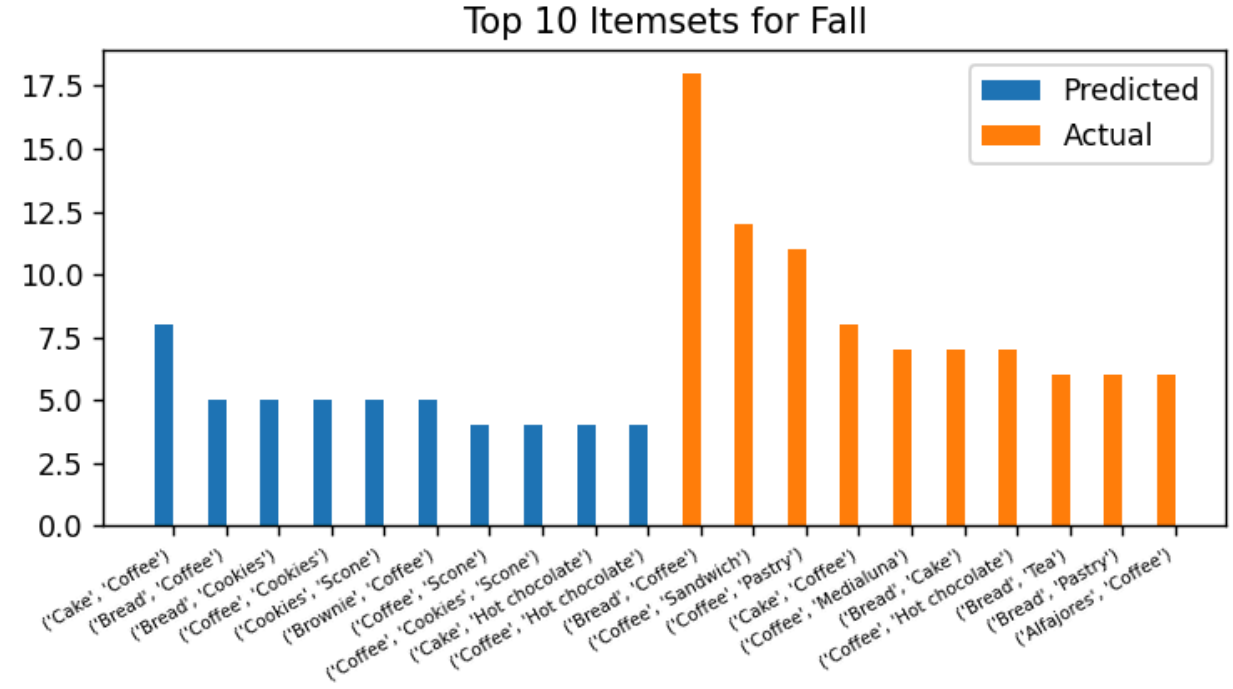
By contrast, the Multinomial Naive Bayes classifier faced even greater challenges in predicting the season in which transactions took place. Because Naive Bayes models rely on the assumption of conditional independence of features, they often fail to properly represent the complex and correlated nature of customer behavior. This faulty reliance underscores the need for more adaptable classification techniques in this context, such as the decision tree classifier.

The following graphs show the areas where the model worked well and where it struggled due to a lack of data for the summer and fall seasons:

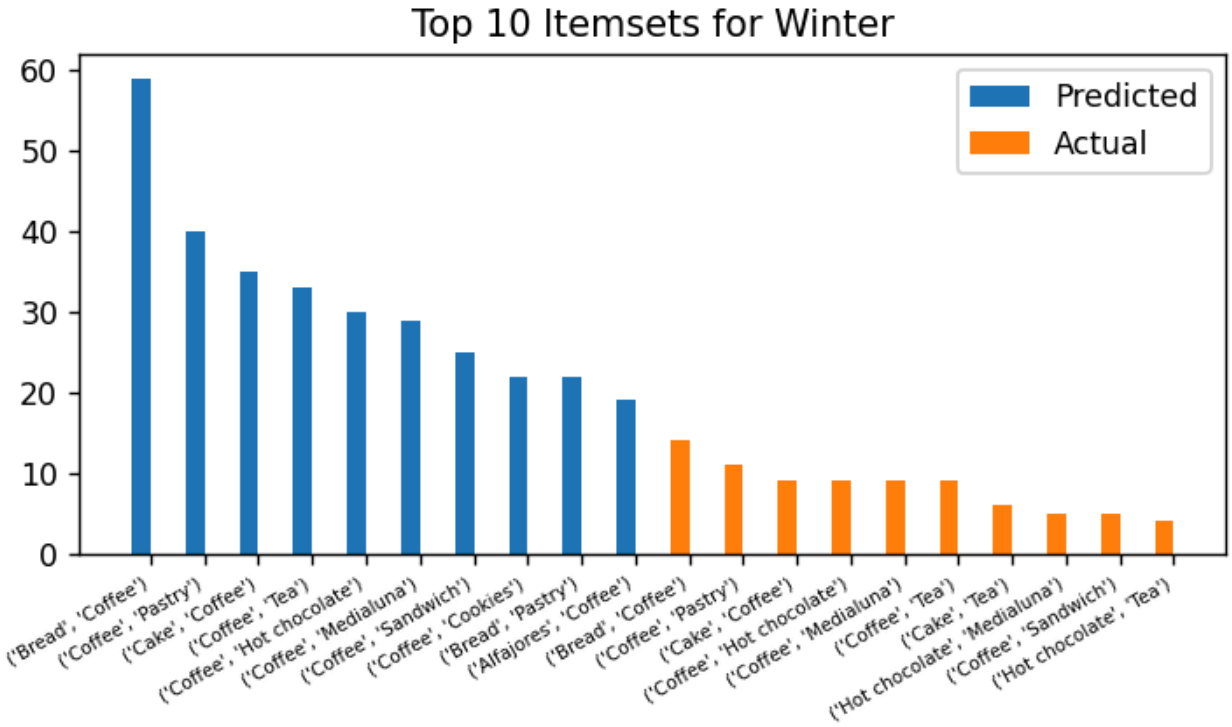

Top 10 Itemsets for Winter
Predicted
Actual
0
20
40
60
80
('Bread', 'Coffee')
('Coffee', 'Medialuna')
('Coffee', 'Hot chocolate')
('Brownie', 'Coffee')
('Coffee', 'Pastry')
('Bread', 'Medialuna')
('Coffee', 'Keeping It Local')
('Medialuna', 'Pastry')
('Coffee', 'Muffin')
('Coffee', 'Tea')
('Bread', 'Coffee')
('Coffee', 'Medialuna')
('Coffee', 'Pastry')
('Coffee', 'Tea')
('Cake', 'Coffee')
('Bread', 'Pastry')
('Coffee', 'Hot chocolate')
('Bread', 'Tea')
('Brownie', 'Coffee')
('Coffee', 'Cookies')

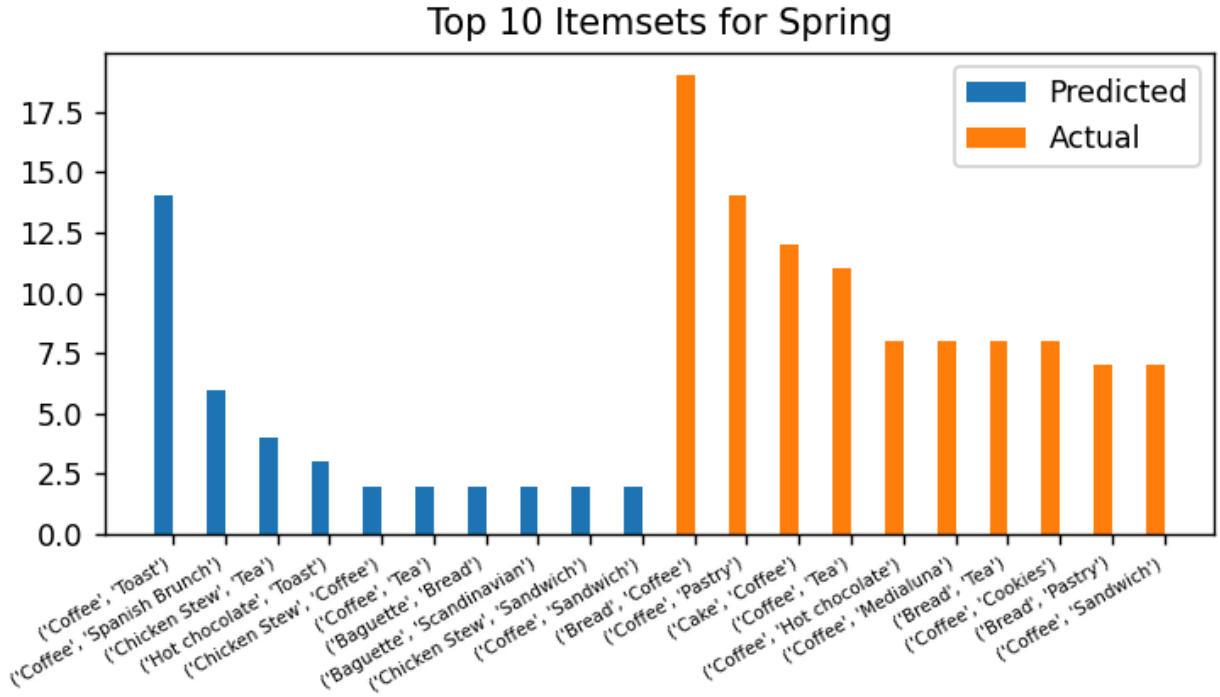

Top 10 Itemsets for Spring
Predicted
Actual
0
20
40
60
80
('Cake', 'Coffee')
('Coffee', 'Pastry')
('Coffee', 'Tea')
('Bread', 'Coffee')
('Coffee', 'Sandwich')
('Bread', 'Pastry')
('Coffee', 'Cookies')
('Coffee', 'Toast')
('Bread', 'Tea')
('Cake', 'Tea')
('Bread', 'Coffee')
('Cake', 'Coffee')
('Coffee', 'Sandwich')
('Coffee', 'Tea')
('Coffee', 'Pastry')
('Coffee', 'Cookies')
('Coffee', 'Hot chocolate')
('Coffee', 'Toast')
('Bread', 'Tea')
('Bread', 'Pastry')

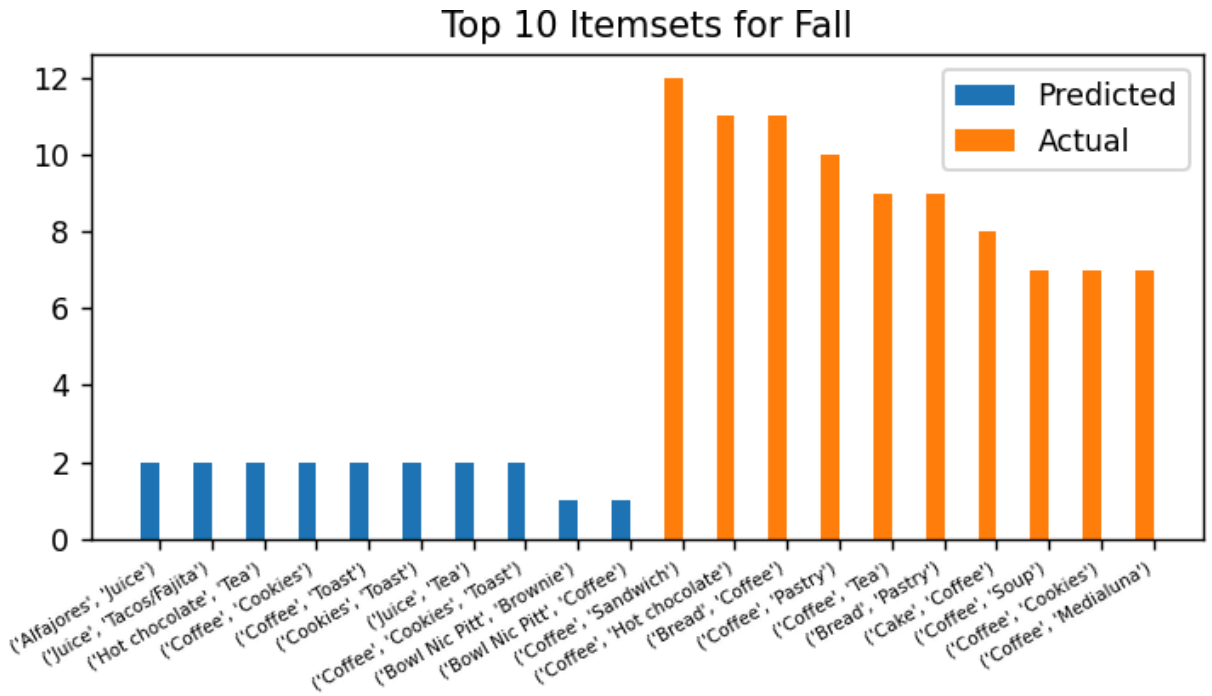

Top 10 Itemsets for Summer
Predicted
Actual
25
20
15
10
5
0
('Bread', 'Coffee')
('Bread', 'Medialuna')
('Brownie', 'Coffee')
('Bread', 'Tea')
('Bread', 'Pastry')
('Medialuna', 'Pastry')
('Coffee', 'Medialuna')
('Bread', 'Cake')
('Bread', 'Medialuna', 'Pastry')
('Coffee', 'Scandinavian')
('Bread', 'Coffee')
('Coffee', 'Pastry')
('Coffee', 'Hot chocolate')
('Cake', 'Coffee')
('Coffee', 'Scone')
('Brownie', 'Coffee')
('Coffee', 'Medialuna')
('Coffee', 'Tea')
('Coffee', 'Cookies')
('Bread', 'Pastry')

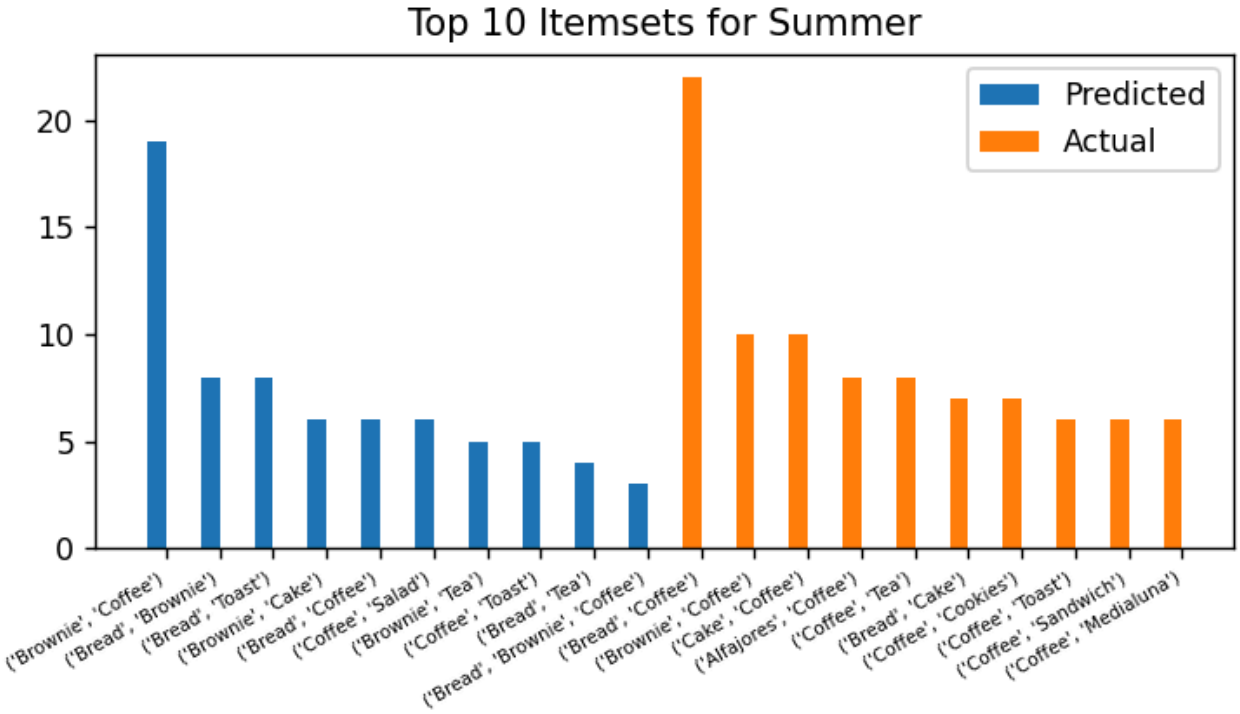

Top 10 Itemsets for Fall
Predicted
Actual
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
('Cake', 'Coffee')
('Bread', 'Coffee')
('Bread', 'Cookies')
('Coffee', 'Cookies')
('Cookies', 'Scone')
('Brownie', 'Coffee')
('Coffee', 'Scone')
('Coffee', 'Cookies', 'Scone')
('Cake', 'Hot chocolate')
('Coffee', 'Hot chocolate')
('Bread', 'Coffee')
('Coffee', 'Sandwich')
('Coffee', 'Pastry')
('Cake', 'Coffee')
('Coffee', 'Medialuna')
('Bread', 'Cake')
('Coffee', 'Hot chocolate')
('Bread', 'Tea')
('Bread', 'Pastry')
('Alfajores', 'Coffee')

It is clear from reviewing these graphs that while the predicted counts of transactions may be slightly off during winter and spring, the itemsets which were predicted to be most popular were fairly close to the actual itemsets. As mentioned previously, the uneven distribution of our data between winter/spring and summer/fall created difficulties when it came to predicting itemsets during the latter seasons. When we attempted to normalize this imbalance we ended up with even lower accuracy than before normalization, which can be seen in the following graphs:









We believe that our decision tree model requires more transactions than our dataset contained for the summer and fall seasons in order to properly classify them, as even though the pre-normalized predictions for the winter and spring seasons were not perfect, they vastly outperformed the summer and fall seasons. Also, because bread and coffee are one of the most popular combinations for each season, it is likely that too many of these transactions were classified as occurring in the winter due to overfitting in

our tree. If we had access to more transactions for summer/fall, our predictions would likely be more accurate and therefore more useful for bakery businesses.