Regional Difference in Retail Product Association of Market Basket Analysis in US*

Byong-Kook YOO[1], Soon-Hong KIM[2]

Abstract

Purpose: Market basket analysis is one of the most frequently used methods in the retail industry today as a technique to discover the product association. It is empirically analyzed how these product associations differ regionally in the case of the United States. **Research design, data, and methodology**: Based on the purchasing data of consumer panels collected from 49 US states, the association rules for each state was extracted with the corresponding lift values indicating product association. The difference in lift values in 49 states by the association rule was compared and tested for 49 states and for 4 census regions (Northeast, Midwest, South, West). **Results**: The association rules of 3/4 of the same association rules show positive associations or negative associations depending on the lift values of the states. There were significant differences in the lift values for 49 states, and for 4 census regions. These significant differences in the lift values were found to be related to the distance between states and whether states belong to the same census region. **Conclusions:** Retail product associations shown by market basket analysis may vary depending on regional distance or regional heterogeneity. It is necessary to pay attention to these points in multi-store environment.

Keywords : Market Basket Analysis, Association Rule, Retail Product Association, Regional Difference

JEL Classification Code: L81, C81, R58

1. Introduction

Market basket analysis (also known as association rule mining) is a technique for discovering the association of consumer purchasing patterns from large-scale transaction data of stores (Aguinis et al., 2013). For example, if supermarket consumers tend to buy milk, bread, and cheese together, or if bank customers tend to use certain services together, this will help the company's marketing strategy (store layout, product recommendation, product mixing and bundling, etc.). These consumer purchasing patterns derived from market basket analysis in the form of rules (if A is B) are called association rules.

Market basket analysis is useful and easy to understand, and has recently been used to identify the relationship between complex scientific phenomena occurring simultaneously in various fields such as bioinformatics, nuclear science, pharmacodynamics, immunology, and geophysics (Aguinis et al., 2013; Boratto et al., 2020; Kanagawa et al., 2009; Koperski & Han, 1995; Reddy & Reddy, 2021; Szymkowiak et al., 2018).

Today, it is common for many retail businesses to have various subsidiaries, branches, dealers, or franchises in different locations. For example, the supermarket chain Walmart, has the largest number of stores worldwide. For retail companies with multiple stores, the discovery of

1 First Author. Professor, Division of International Trade, Incheon National University, South Korea. Email: bkyoo@inu.ac.kr
2 Corresponding Author. Professor, Division of International Trade, Incheon National University, South Korea.

Email: snow8817@inu.ac.kr

purchasing patterns that can vary depending on the individual stores' region can be useful in forming marketing, sales, service, and operation strategies at company, local, and store levels (Dangerfield et al., 2021; Wang et al., 2022).

One of the important issues to consider when performing market basket analysis in a multi-store environment is a question of whether the association rules derived from one store can be valid for other stores as well. If the association rule for one store can be applied to other similar stores, the company's marketing strategy can be implemented at a lower cost. However, in a multi-store environment, we have to consider the possibility that the retail product associations (or complementarity) may vary between stores that are geographically separated. In other words, consumer purchase spending patterns may vary among regions because of many factors: prices, income, population characteristics, climate, consumer tastes, and so on. More specifically, factors such as regional inequalities (Bono et al., 2007) and regional variation in personality (Rentfrow et al., 2013), and preference for local product (Aprile et al., 2016; Skallerud & Wien, 2019) have been studied.

In the meantime, major studies on market basket analysis have mainly focused on its application to various fields other than retail products or the improvement of rule mining algorithms (Aguinis et al., 2013; Kamakura, 2012; Martinez et al., 2021; Tang et al., 2008). But, there were a few empirical studies of market basket analysis on regional heterogeneity (Szymkowiak et al., 2018; Williams et al., 2007). Szymkowiak et al. (2018) used market basket analysis to analyze differences in marital status by region using the Polish socio-demographic census data. Williams et al. (2007) presented a trace element survey of U.S.-grown rice purchased in U.S. supermarkets, comparing and contrasting grain from California and the South Central U.S., to investigate variation in As (Arsenic) contamination between these 2 regions.

In order to address the above issue, we used data on retail purchases made by households in 49 states, US. In order to examine how the associations of retail products differ by state, we try to compare lift values of association rules obtained through market basket analysis of the retail purchase data. To this end, we first examine whether there is a significant difference in lift values by 49 states and 4 census regions for the same association rule in 49 states. Then we look at how these differences are affected by distance between states or census regions.

2. Literature Review

2.1. Market basket analysis

Market basket analysis (MBA) is a data mining technique that started in the marketing field to investigate the relationship between groups such as products, items, and categories, and is known as association rule mining or affinity analysis. Agrawal et al. (1993) attempted to discover association rules in a large repository of previously collected customer transaction data using market basket analysis for the first time. The Apriori algorithm that they developed is widely applied to the marketing field and can be utilized for recommendation of related products (Kaur & Kang, 2016; Ünvan, 2021; Zamil et al., 2020).

Market basket analysis aims to explore patterns such as one or more products that are often purchased together based on large-scale consumer purchase data. Association (or complementarity) between products can be investigated by such a search and these results can be used to make decisions such as stocking the two items near each other, thereby increasing the likelihood that customers will easily find and purchase both products instead of just one of them (Griva et al., 2018; Nurmayanti et al., 2021; Rana & Mondal, 2021; Russell & Petersen, 2000; Ünvan, 2021).

2.2. Temporal Studies

In general, it is assumed that the market basket analysis targets products that are commonly provided in all stores. Therefore, it has limitations in temporal or spatial analysis in that it must happen at the same time (joint buying) and in the same space (Kamakura, 2012; Rana & Mondal, 2021). For example, an association rule that is valid at one store (or at one point in time) of a chain store may not be valid at another store (or at another time) in the same chain. Kamakura (2012) tried to overcome the temporal limitation of market basket analysis by introducing the incorporation of a longitudinal component (sequential buying) into market basket analysis. Rana and Mondal (2021) proposed a methodology for mining seasonally frequent patterns and association rules with multilevel data environments.

2.3. Spatial Studies

Market basket analysis studies that overcome the limitations of spatial analysis mainly focus on improving the existing Apriori algorithm of market basket analysis (Chen et al., 2005; Koperski & Han, 1995; Tang et al., 2008). Chen et al. (2005) presented a new algorithmic method capable of various product-mix strategies under multi-store environments. Koperski and Han (1995) proposed an efficient method for mining strong spatial association rules in geographic information databases when one of the association rules is spatial information. Tang et al. (2008) proposed an approach for extracting association rules from transactional records in a multiple-store and multiple-period environment. Nevertheless, it cannot be denied that the

Apriori algorithm is still the representative method of market basket analysis in marketing and other fields (Aguinis et al., 2013; Kaur & Kang, 2016).

Recently, Ludwig et al. (2021) explored how association rules derived from Open Street Map data vary across geographic regions and depend on different context variables. Using geo-location data shared by tourists on tourism platforms, Vavpotič et al. (2021) applied the market basket analysis to tourism in order to determine the set of tourist experiences that are consumed by tourists during their visit to a certain tourist destination. However, studies comparing the degree of product association derived from data on retail purchase by region are very rare.

In addition to the regional factors mentioned above, the association rules of market basket analysis may vary in the degree of association according to various groups (Pradhan et al., 2022). That is, the degree of association between specific products may vary depending on the group of the research target. Pradhan et al. (2022) aimed to find out whether the principle of market basket analysis was applicable in a segmented market comprising highly efficient customers (i.e. customers having higher CLV) and less efficient customers (i.e. customers having lower CLV), where customer lifetime value (CLV) was the relative worth of the customer to the firm.

3. Methodology

3.1. Data

This study was based on the US Consumer Panel Data (the consumer records purchases by scanning UPC product codes at home) by the AC Nielsen in 2009. These data track a panel of nation-wide US households every year and their purchases of fast-moving consumer goods from a wide range of retail outlets across all US markets. Our data are based on 3,445,787 transaction data of 60,473 panel households residing in 49 states in 2009. Summary statistics (mean, minimum, maximum) for 49 states for the number of households, number of transactions, number of products, and number of stores are presented in Table 1, where the average number of participating households was 1,234, the number of transactions was 70,322, the number of products was 128, and the number of retail stores was 711.

Table 1: Summary statistics

Number of Samples	Mean	Min	Max	Total
Household	1,234	78	5,039	60,473
Transaction	70,322	3,483	250,838	3,445,787
Product	128	90	166	6,232
Retail Store	711	58	3,628	34,841

3.2. Association Rules

An important outcome of market basket analysis is the discovery of association rules in the form of if-then statements. Typically, the association rule has the following form: $\{A, B\} \Rightarrow \{C\}$. It means that if A and B are purchased, then there is a high probability that C will be purchased. In this association rule, $\{A, B\}$ is an antecedent and $\{C\}$ is a consequent.

Three indexes (support, confidence, lift) are used to evaluate the association rule (Aguinis et al., 2013; Pillai, & Jolhe, 2021). When event A and event B are events where product A and product B are purchased, respectively, the three indexes can be defined as follows.

Support is defined as the probability that event A and event B occur simultaneously and can be expressed as P(A∩B). In other words, it refers to the probability (ratio) of a transaction in which product A and product B are simultaneously purchased. A rule that has low support may occur simply by chance. A low support rule may also be uninteresting from a business perspective because it may not be profitable to promote items that are seldom bought together. For these reasons, support is often used to eliminate uninteresting rules.

Confidence is the conditional probability ($P(A) = \frac{P(A \cap B)}{P(A)}$) that event B will occur when event A occurs, and is the probability that product B is included in all transactions involving product A. Confidence and support measure the strength of an association rule. Since the transactional database is quite large, there is a higher risk of getting too many unimportant and rules which may not be of our interest. To avoid these kinds of errors we commonly define a threshold of support and confidence prior to the analysis, so that only useful and interesting rules are generated in our result.

Lift is defined as $\frac{P(A \cap B)}{P(A)P(B)}$, which is the ratio of the conditional probability($\frac{P(A \cap B)}{P(A)}$) that event B will occur when event A occurs to the probability($P(B)$) that event B will occur. Lift provides information on whether an association exists or not, or if the association is positive or negative. If the probability of including product B in all transactions involving product A is greater than the probability of a transaction in which product B is purchased (Lift > 1), it means that the purchase of product A is positively associated with the purchase of product B. On the other hand, if the probability of including product B in all transactions involving product A is less than the probability of a transaction in which only product B is purchased (Lift < 1), then it means that the purchase of product A is negatively associated with the purchase of product B. Also, if the purchase of product A is not associated with the purchase of product B, then P(A∩B) = P(A)P(B), and lift value becomes

1. Therefore, as the lift value is greater than 1, the purchase of two products can be interpreted as having strong association. Hereinafter, we will examine how the lift value of the association rule can differ by region (or state) despite the same association rule.

The Apriori algorithm used to derive the association rule in this paper is an iterative algorithm which looks for so-called frequent itemsets (Patel, 2022). When performing the Apriori algorithm, support and confidence of a frequent itemset are assumed to be greater than or less than a certain minimum threshold, respectively (Szymkowiak et al., 2018).

Table 2: Number of associate rules

St.	n	St.	n	St.	n	St.	n	St.	n	St.	n	St.	n
AL	871	AR	1,154	AZ	606	CA	569	CO	956	CT	1,044	DC	889
DE	830	FL	667	GA	699	IA	655	ID	725	IL	837	IN	901
KS	828	KY	971	LA	849	MA	1,148	MD	759	ME	956	MI	794
MN	831	MO	850	MS	807	MT	642	NC	692	ND	1816	NE	840
NH	1,373	NJ	1,109	NM	875	NV	900	NY	658	OH	886	OK	1,708
OR	516	PA	856	RI	986	SC	730	SD	748	TN	824	TX	891
UT	857	VA	886	VT	1,904	WA	553	WI	965	WV	885	WY	810

4. Results

For the above purchase data, association rules for each of 49 states were generated using the arules package of the R software. When the association rule was generated using the apriori function of the arules package, the minimum values of support and confidence were set to 0.001 and 0.002, respectively. This relatively low number reflects the intention to generate as many rules with high relevance as possible, even for rules with low frequency. A total of 44,106 association rules were extracted from 49 states. The number of association rules extracted from each state is shown in Table 2, respectively Among these association rules, there are unique association rules according to the characteristics of each state, while the same association rules exist for several states. Here, the same association rules refer to the rules in which A = A', B = B' when the association rules of some two states are A => B and A' => B'. The number of states having the same specific association rule can be from a minimum of 2 to a maximum of 49. In this study, for comparison of all states, the same association rules for 49 states were considered.

The same association rules for all states were found In 159 out of 44,106 association rules, or 0.4%. That is, as shown in Table 3, each state has the same 159 association rules with a unique lift value for each state. These 159 association rules and 49 states can be considered as subjects and treatments in the experimental study, That is, It can be interpreted as performing 49 regionally different treatments on 159 subjects.

4.1. Lift Difference

Lift difference by state can be divided into a quantitative Lift difference by state can be divided into a quantitative difference and a qualitative difference.

First, the quantitative difference can be summarized as the difference between the maximum and minimum lift values. The average of these maximum values was 1.885 and the average of the minimum values was 0.784, so the difference was about 1.100.

Table 3: The Same Association Rules for All States

NO	Association Rule	AL lift value	………………	WY lift value
1	{BAKERY BREAKFAST CAKES/SWEET ROLLS FRESH, NUTS BAGS} => {SOFT DRINKS CARBONATED}	1.351	……………..	1.345
2	{BAKERY BREAKFAST CAKES/SWEET ROLLS FRESH,SOFT DRINKS CARBONATED} => {NUTS BAGS}	1.625	……………..	2.649
⋮		⋮	⋮	⋮
158	{SOFT DRINKS LOW CALORIE} => {NUTS BAGS}	1.139	……………..	1.121
159	{SOFT DRINKS LOW CALORIE} => {SOFT DRINKS CARBONATED}	1.141	……………..	1.112

Figure 1: Maximum and Minimum Lift Values

Specifically, in Table 4, the association rule with the largest difference was {FRUIT DRINKS & JUICES CRANBERRY} => {CANNED FRUIT FRUIT COCKTAIL}. That is, in North Dakota (ND), the maximum was 5.853, but in New Mexico (NM), the minimum was 1.183, showing a difference of 4.680. On the other hand, the association rule with the smallest difference was {CANDY CHOCOLATE} => {NUTS BAGS}, with a maximum of 1.045 in North Dakota (ND) and a minimum of .0.683 in Rhode Island (RI), showing a difference of 0.362.

Table 4: The Same Association Rules for All States

No	Association Rule	Max (state)	Min (state)	Max-Min
96	{FRUIT DRINKS & JUICES CRANBERRY} => {CANNED FRUIT FRUIT COCKTAIL}	5.853 (ND)	1.183 (NM)	4.680
27	{CANDY CHOCOLATE} => {NUTS BAGS}	1.045 (ND)	0.683 (RI)	0.362

Second, the qualitative difference in lift can be viewed by the association based on the lift value of 1(see Figure 1). For the 159 association rules, on average, the maximum value is greater than 1 (1.885), but the minimum value is 0.784, which is less than 1. This means that even the same association rule may show a positive association or a negative association depending on the state. Specifically, in the case of {CANDY CHOCOLATE} => {NUTS BAGS} in Table 4, it showed a positive association (1.045) in North Dakota (ND), but a negative association (0.683) in Rhode Island (RI).

Among the total 159 association rules, 38 (23.9%) association rules with lift greater than 1 in all states were found. On the other hand, there were 2 (1.2%) association rules with a lift smaller than 1 in all states. In the remaining

119 association rules (74.8%), the lift values may be greater than or less than 1 depending on the state. This means that about three-quarters of the association rules may show the opposite associations (positive or negative) depending on the state. In this way, the proportion of these opposite associations in each state varies depending on the association rule. Figure 2 shows states showing the positive associations or the negative associations when the association rule is "{BAKRY BREAKFAST CAKES/SWEET ROLLS FRESH} => {SOFT DRINKS CARBONATED}".

Note: For spatial coordinate data US states polygons of spData R were used

Figure 2: Positive and Negative Associations by State

4.2. Test of Lift Difference

To test the regional differences in lift, the following methods are used. First, the difference in lift is tested for all 49 states. Second, 49 states are divided into 4 census regions and the difference in lift is tested for 4 regions.

Table 5: Census Regions and Lift Statistics

Census Region	State	Mean	Min	Max
Northeast	CT, MA, ME, NH, NJ, NY, PA, RI, VT	1.24	0.59	2.36
Midwest	IA, IL, IN, KS, MI, MN, MO, ND, NE, OH, SD, WI	1.29	0.41	2.67
South	AL, AR, DC, DE, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV	1.18	0.58	2.51
West	AZ, CA, CO, ID, MT, NM, NV, OR, UT, WA, WY	1.15	0.67	2.41

Table 5 shows the 4 census regions (Northeast, Midwest, South, West) that are defined by the United States Census Bureau and their lift statistics As a result of performing a

Shapiro-Wilk's test on the lift values of 49 states, none of the states satisfy the normality test.

Therefore, we intend to use the Friedman test, a non-parametric approach, as a test method. The Friedman test is a non-parametric alternative to the Repeated Measures ANOVA. The Friedman test was conducted in 49 states and 4 regions, respectively. As a result of Friedman Test, p-values were found to be close to 0.0 (Table 6). Therefore, the null hypothesis that the lift does not differ by state or region is rejected.

Table 6: Test of Lift Difference between States and Census Regions

	Degree of Freedom	N	Chi-Square Statistic	p-value
State	48	159	1,920	0.00
Region	3	159	204.49	0.00

From the output of the Friedman test, we know that there is a significant difference between states, but we don't know which pairs of states are different. A significant Friedman test can be followed up by paired Wilcoxon signed-rank tests for identifying which pairs are different. Paired Wilcoxon signed-rank test is the non-parametric equivalent of the paired t-test.

Taking two from 49 states, a total of 1,176 paired Wilcoxon signed-rank tests (49 C_2) can be performed. Table 7 classifies the paired Wilcoxon signed-rank test results according to significance. Among a total of 1,176 paired Wilcoxon signed-rank tests, 56% showed a significant difference at a significance level of 0.05 or less, and 44% showed a non-significant difference.

**Table 7:** Paired Wilcoxon Signed-rank Test Results

Group			N	Proportion
Significant	p-value	> 0.001	530	45%
		> 0.01	65	6%
		> 0.05	57	5%
		Total	652	56%
Non-significant			524	44%
Total			1,176	100%

4.3. Inter-state Distance and Census Region

The factors contributing to these regional lift differences can be very diverse. In this study, we focus on the two factors, the physical distance between two states (inter-state distance) and whether states belong to the same census region. In the following, we divide the above paired Wilcoxon signed-rank test results into two groups (Significant group, Non-significant group) according to the significant results (p-value < 0.05) and non-significant results (p-value ≥ 0.05).

4.3.1. Inter-state Distance

We can use the st_distance function of R's sf package to calculate the inter-state distance between two states. This st_distance function calculates the shortest distance between two states. In the st_distance function, the distance between adjacent states is set to 0. For example, the inter-state distance between New York and California is calculated as approximately 3,783 km, but the inter-state distance between adjacent New York and Pennsylvania is calculated as 0. Figure 3 shows the significant and non-significant groups on the map based on New York (NY). For example, there was a significant difference in lift values between New York and California (CA), but the difference in lift values was not significant when comparing New York and Pennsylvania (PA).
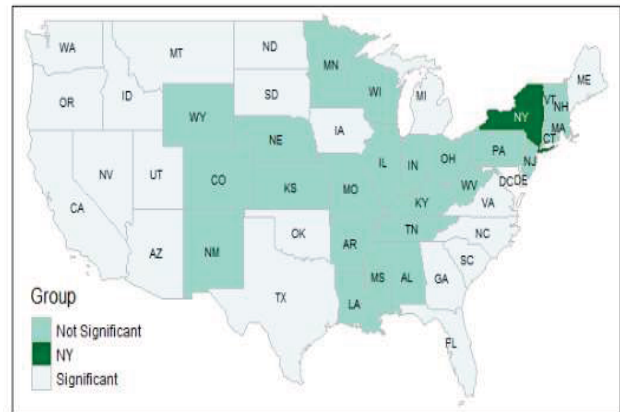
Note: For spatial coordinate data US states polygons of spData package in R were used

Figure 3: The Significant and Non-significant Groups

In Table 8, the mean of the inter-state distance of the significant group is higher than the mean of the non-significant group. Wilcoxon rank-sum test showed that there was a significant difference in inter-state distance between the two groups (Significant, Non-significant) in lift. In other words, the inter-state distance of the significant group is farther than the inter-state distance of the non-significant group.

Table 8: Wilcoxon Rank-sum Test Result

Index	Group	N	Mean (Inter-State Distance, km)
lift	Non-Significant	524	1,260.0
	Significant	652	1,395.9
	Wilcoxon rank-sum test statistics = 157.462, p-value = 0.021		

4.3.2. Census Region

Comparisons between two states can be divided into

intra-region (or within the same census region) comparisons and inter-region (or between census regions) comparisons. For example, in the above Table 5, Connecticut (CT) and Massachusetts (MA) belong to the same census region (Northeast), so the comparison between them belongs to the intra-region comparison. On the other hand, Connecticut (CT) and Iowa (IA) belong to the Northeast and Midwest, respectively, so their comparison belongs to the inter-region comparison. The results of 1,176 paired Wilcoxon signed-rank tests were classified into significance (Significant, Non-significant) and regions (intra-region, inter-region) as shown in the contingency table below (Table 9).

Table 9: Contingency Table

Comparison (Intra or Inter)		Significant?		
		Yes	No	Total
Intra-region	Count	147	146	293
	Expected Count	162	131	293
	% Intra-region	50.2%	49.8%	100.0%
Inter-region	Count	505	378	883
	Expected Count	490	393	883
	% Inter- region	57.2%	42.8%	100.0%
Total				1,176

There were 147 cases (50.2%) in which the lift difference was significant in the intra-region comparisons. However, non-significant cases were 146 (49.8%), indicating that the difference between significant cases and non-significant cases was relatively small. On the other hand, 505 cases (57.2%) were significant in inter-region comparisons while 378 cases (42.8%) were non-significant, showing a greater difference than intra-region comparisons ($\chi 2$ (1) = 4.11, p-value = 0.043). Therefore, it was found that the significant difference in lift occurred more in the inter-region comparisons than in the intra-region comparisons.

5. Conclusions

5.1. Discussions and Implications

Market basket analysis is one of the most used primary methods in the retail industry today as a technique to discover the associations of consumer purchasing patterns from large-scale transaction data of stores. In particular, the lift value for each association rule derived from the market basket analysis is an important indicator showing the association between the retail products constituting the association rule. That is, in the case of a positive association (lift > 1), there is strong complementarity between products, and the purchase of one product plays a role in promoting the purchase of another product. In this case, various marketing activities such as bundling, purchase recommendation, and display layout are possible to increase sales. On the other hand, in the case of a negative association (lift<1), the above marketing activities may have adverse effects, so more differentiated management is required.

In this study, we focused on empirically analyzing how the lift values of the same association rule may differ by region. For this purpose, each market basket analysis was conducted using the consumer purchasing data collected in 49 US states. The lift values of the same 159 association rules obtained through 49 market basket analyses were compared. The conclusions drawn from the analysis results are as follows.

First, the lift value of the same association rule showed a significant difference depending on the state. That is, the average of the maximum values of lifts was 1.885, while the average of the minimum values was 0.784. Moreover, even though the association rules are the same, 3/4 of the cases had the opposite associations depending on the state.

Second, in order to statistically verify this difference, as a result of the Friedman test conducted for 49 states and 4 regions, it was found that there is a significant regional difference in lift values.

Third, two groups (significant group and non-significant group) were compared by calculating the inter-state distance between the two states. As a result, there was a significant difference in the inter-state distance between the two groups. In other words, it was found that the significant group had a longer inter-state distance than the non-significant group. This fact can be seen to mean that the association between products appears similarly in a nearby region rather than in a distant region.

Furthermore, two groups (significant group and non-significant group) showed a significant correlation with whether states belonged to the same census region. In other words, it can be said that the association between products appears more similar in the same census region.

The above results show that the association between retail products by market basket analysis can vary depending on distance or regional heterogeneity. In particular, when multiple stores are operated in various geographic locations in a multi-store environment, it is necessary to pay attention to these points when considering whether the association between retail products in one store are utilized in other stores.

5.2. Limitations and Future Directions for Research

In this study, there were inevitably several limitations, and are as follows, as well as future research tasks.

First, various factors such as socio-cultural factors or characteristics of resident consumers can be considered in addition to the distance between regions and the factor of the

census region discussed in this study, for regional factors affecting the association between products. Further studies on these various heterogeneities are needed.

Second, this study attempted to compare the spatial difference of association using the purchasing information in 49 states of the United States. If sufficient data are available, such studies need to be conducted at the regional level in various countries other than the United States.

Third, from the point of view of a company with multiple stores, comparisons between   stores as well as regional comparisons may be of more interest. In this regard, it is necessary to study the comparison of association rules in more diverse stores in the future.

Fourth, in general, market basket analysis has intrinsic limitations not only in spatial analysis but also in temporal analysis. In other words, in order to solve the limitations of the existing market basket analysis, it can be said that an additional empirical study on the temporal difference according to factors such as the day of the week or the season is needed in addition to the regional and spatial differences in this study.

References

Aguinis, H., Forcum, L. E., & Joo, H. (2013). Using market basket analysis in management research. *Journal of Management, 39*(7), 1799-1824.

Aprile, M. C., Caputo, V., & Nayga Jr, R. M. (2016). Consumers' preferences and attitudes toward local food products. *Journal of food products marketing, 22*(1), 19-42.

Bono, F., Cuffaro, M., & Giaimo, R. (2007). Regional inequalities in consumption patterns: A multilevel approach to the case of Italy. *International Statistical Review, 75*(1), 44-57.

Boratto, L., Manca, M., Lugano, G., & Gogola, M. (2020). Characterizing user behavior in journey planning. *Computing, 102*(5), 1245-1258.

Chen, Y.-L., Tang, K., Shen, R.-J., & Hu, Y.-H. (2005). Market basket analysis in a multiple store environment. *Decision Support Systems, 40*(2), 339-354.

Dangerfield, F., Lamb, K. E., Oostenbach, L. H., Ball, K., & Thornton, L. E. (2021). Urban-regional patterns of food purchasing behaviour: A cross-sectional analysis of the 2015–2016 Australian Household Expenditure Survey. *European Journal of Clinical Nutrition, 75*(4), 697-707.

Griva, A., Bardaki, C., Pramatari, K., & Papakiriakopoulos, D. (2018). Retail business analytics: Customer visit segmentation using market basket data. *Expert Systems with Applications, 100*(15 June), 1-16.

Kamakura, W. (2012). Sequential market basket analysis. *Marketing Letters, 23*(3), 505-516.

Kanagawa, Y., Matsumoto, S., Koike, S., & Imamura, T. (2009). Association analysis of food allergens. *Pediatric Allergy and Immunology, 20*(4), 347-352.

Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the changing trends of market data using association rule mining.

Procedia Computer Science, 85, 78-85.

Koperski, K., & Han, J. (1995). Discovery of spatial association rules in geographic information databases. In M. J. Egenhofer, & J. R. Herring (Eds.), *Lecture Notes in Computer Science: Vol. 951. Advances in Spatial Databases*. Springer, Berlin, Heidelberg.

Ludwig, C., Fendrich, S., & Zipf, A. (2021). Regional variations of context-based association rules in OpenStreetMap. *Transactions in GIS, 25*(2), 602-621.

Martinez, M., Escobar, B., Garcıa-Dıaz, M. E., & Pinto-Roa, D. P. (2021). Market basket analysis with association rules in the retail sector using Orange. Case Study: Appliances Sales Company. *CLEI Electronic Journal, 24*(2), Paper 12.

Nurmayanti, W. P., Sastriana, H. M., Rahim, A., Gazali, M., Hirzi, R. H., Ramdani, Z., & Malthuf, M. (2021). Market basket analysis with apriori algorithm and frequent pattern growth (Fp-Growth) on outdoor product sales data. *International Journal of Educational Research & Social Sciences, 2*(1), 132-139.

Patel, H. K. (2022). Association rule mining using retail market basket dataset by Apriori and FP growth algorithms. *Journal of Algebraic Statistics, 13*(3), 798-803.

Pillai, A. R., & Jolhe, D. A. (2021). Market basket analysis: Case study of a supermarket. In *Advances in Mechanical Engineering* (pp. 727-734). Springer, Singapore.

Pradhan, S., Priya, P., & Patel, G. (2022). Product bundling for 'Efficient' vs 'Non-Efficient' customers: Market basket analysis employing genetic algorithm. *The International Review of Retail, Distribution and Consumer Research, 32*(3), 293-310.

Rana, S., & Mondal, M. N. I. (2021). A seasonal and multilevel ssociation based approach for market basket analysis in retail supermarket. *European Journal of Information Technologies and Computer Science, 1*(4), 9-15.

Reddy, V. N., & Reddy, P. S. S. (2021). Market basket analysis using machine learning algorithms. *International Research Journal of Engineering and Technology, 8*(7), 2570-2572.

Rentfrow, P. J., Gosling, S. D., Jokela, M., Stillwell, D. J., Kosinski, M., & Potter, J. (2013). Divided we stand: Three psychological regions of the United States and their political, economic, social, and health correlates. *Journal of personality and social psychology, 105*(6), 996-1012.

Russell, G. J., & Petersen, A. (2000). Analysis of cross category dependence in market basket selection. *Journal of Retailing, 76*(3), 367-392.

Skallerud, K., & Wien, A. H. (2019). Preference for local food as a matter of helping behaviour: Insights from Norway. *Journal of Rural Studies, 67*(April), 79-88.

Szymkowiak, M., Klimanek, T., & Jozefowski, T. (2018). Applying market basket analysis to official statistical data. *Econometrics, 22*(1), 39-57.

Tang, K., Chen, Y.-L., & Hu, H.-W. (2008). Context-based market basket analysis in a multiple store environment. *Decision Support Systems, 45*(1), 150-163.

Ünvan, Y. A. (2021). Market basket analysis with association rules. *Communications in Statistics-Theory and Methods, 50*(7), 1615-1628.

Vavpotič, D., Knavs, K., & Cvelbar, L. K. (2021). Using a market basket analysis in tourism studies. *Tourism Economics, 27*(8),

1801-1819.

Wang, W., Wang, L., Wang, X., & Wang, Y. (2022). Geographical determinants of regional retail Sales: Evidence from 12,500 retail shops in Qiannan County, China. *ISPRS International Journal of Geo-Information, 11*(5), 302.

Williams, P. N., Raab, A., Feldmann, J., & Meharg, A. A. (2007). Market basket survey shows elevated levels of As in South Central US processed rice compared to California:

consequences for human dietary exposure. *Environmental science & technology, 41*(7), 2178-2183.

Zamil, A. M. A., Al Adwan, A., & Vasista, T. G. (2020). Enhancing customer loyalty with market basket analysis using innovative methods: A python implementation approach. *International Journal of Innovation, Creativity and Change, 14*(2), 1351-1368.