OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

What Drives Sales Concentration? Investigation of Pareto's Principle in Digital Food Delivery Ecosystems

Daniel Lee
daniellee63540@gmail.com

ABSTRACT

Digital food delivery platforms have transformed consumer purchasing behavior by expanding menu accessibility and reducing search costs. Although understanding customer purchasing behavior has high economic value, only a limited number of studies have investigated sale concentration across online food delivery platforms. To address this question, customer review data (N = 19,330) were collected from 80 stores of five major fast-food franchises across six regions of South Korea using web crawling and text parsing. Sales distributions were estimated based on menu-specific keywords. The resulting Pareto effects were then analyzed in relation to internal menu structures and external market conditions. As a result, the strength of the Pareto effect varied across franchises, with the top 20% of popular menu items accounting for 55.4%-82.4% of total estimated sales. In addition, stronger sales concentration tended to occur when top-selling menu items were less heterogeneous and menu sizes were smaller. By observing the key determinant of sales concentration, this study provides practical implications for menu design and business strategies in digital food-service environments.

INTRODUCTION

Digital technology has dramatically changed the way people purchase food. Online delivery platforms allow consumers to order from restaurants without distance or location constraints, offering a far wider variety of food than traditional dining environments. These platforms also use recommendation systems and ranking algorithms that highlight popular menu items based on previous customer behavior, influencing consumers' purchasing decisions (Sánchez et al., 2023). Traditionally, sales in many retail industries have followed Pareto's principle, which states that a small proportion of products generates most of the total sales (Brynjolfsson et al., 2006). In the restaurant industry, this pattern has played an important role in menu engineering, pricing decisions, and inventory management (Kimes, 1999). However, digital marketplaces operate differently from physical stores. By removing many of the time and location barriers associated with traditional shopping, online platforms make it easier for consumers to explore a broader range of products (Bentley et al., 2009). This shift has led to increased interest in the long-tail effect, where less popular products collectively account for a larger share of overall sales than in conventional markets (Bentley et al., 2009).

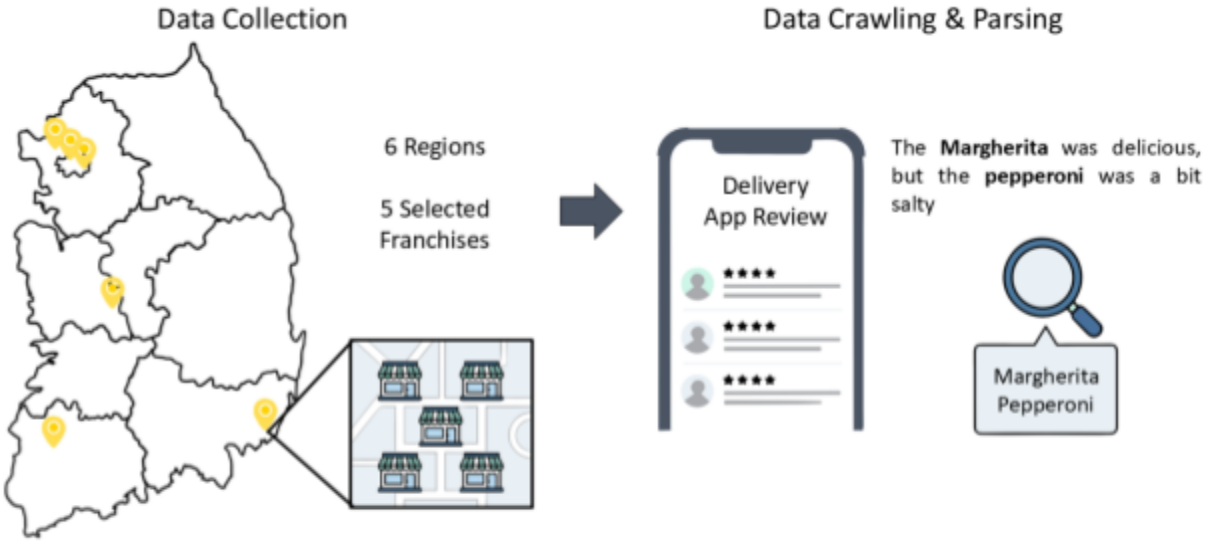


Figure 1: Workflow of data collection and text parsing.

Since food is among the most frequently purchased consumer goods, understanding how consumers choose menu items has important economic value (Kearney, 2010). Although food delivery services have grown rapidly, only a limited number of studies have examined whether the strong Pareto-like sales concentration commonly observed in traditional restaurants is maintained in online delivery environments, where consumers have access to larger menus and algorithm-driven recommendations. The present study aimed to investigate how product sales are distributed across menu items on a modern food delivery platform and to determine whether digitalization promotes a shift toward long-tail consumption. It was hypothesized that the greater accessibility of menu items in online delivery services would weaken the traditional Pareto effect by increasing the contribution of lower-ranked products to total sales.

To address this question, customer review data were collected from five major fast-food franchises across six regions of South Korea. Using web crawling and text-parsing techniques, menu-specific keywords were extracted from customer reviews to estimate product-level sales distributions. The study also examined whether differences in menu diversity, brand popularity, market competitiveness, and the consistency of best-selling menu items across stores were associated with the strength of the Pareto effect. By combining large-scale consumer review data with quantitative analyses of sales concentration, the study provides new insight into how digital food delivery platforms may influence consumer purchasing behavior. The findings also provide practical implications for menu engineering, pricing strategies, inventory management, and marketing optimization in increasingly digital food-service environments.

METHODS

Data Collection and Preprocessing

To investigate the applicability of Pareto's principle to the online food delivery sector, five major fast-food franchises in South Korea were selected: KFC, McDonalds, Papa Johns, Dominos, and Burger King. Data were collected from six regions nationwide where delivery services for these franchises are actively available (Figure 1). Customer reviews from South Korea's leading food delivery platform, Baemin, were used as a proxy for product sales volume starting from 2022 to June 2026.

Data crawling was performed using Selenium (v4.44.0) in Python (v3.8.8) to extract text-based reviews (Chapagain 2019). A total of 19,330 reviews were collected from 80 distinct store locations across the five selected franchises. To estimate the sales proportion of each product, menu-specific keywords were extracted from the raw review texts based on a predefined menu catalog (Figure 1). The number of mentions for each menu item was then multiplied by its corresponding price, assuming that the proportion of customers who leave reviews is consistent across menu items. Side dishes and beverages were excluded from the analysis to focus exclusively on main menu items.

Quantification of Variables Affecting the Pareto's Principle

The cumulative sales share accounted for by the top 20% most popular products was extracted from cumulative sales curves and used as the primary dependent variable representing the strength of the Pareto effect. To identify the internal factors driving sales concentration, the variability of the best-selling item across stores was evaluated based on the number of mentions of each menu item in customer reviews multiplied by its corresponding price. The heterogeneity of the top-ranked menu items across different stores was quantified using Shannon entropy. The Shannon entropy H(X) for each franchise was calculated as follows (Lesne 2014):

$$H(X) = \quad - \sum_{i=1}^{n} P\left(x_i\right) P(x_i)$$

where $P(x_i)$ denotes the probability that a specific menu item x_i is the top-ranked product within a single store. If the same menu item is ranked first across all stores, the Shannon entropy converges to zero. In contrast, greater heterogeneity in top-ranked menu items across stores results in a higher entropy value. The entropy analysis was conducted at a nationwide level for each brand, yielding five data points corresponding to the five selected franchises.

Furthermore, to investigate the structural and environmental determinants of the Pareto effect, three independent variables were measured at the regional franchise level. The first was product diversity, defined as the number of main meal items listed on the menu. The second was brand popularity, indicated by the total number of reviews. The third was market competitiveness, represented by the number of same-category stores. Unlike the entropy analysis, this analysis generated a total of 30 observation points, corresponding to 5 franchises across 6 regions.

RESULTS

Sales Distributions Across Food Franchises

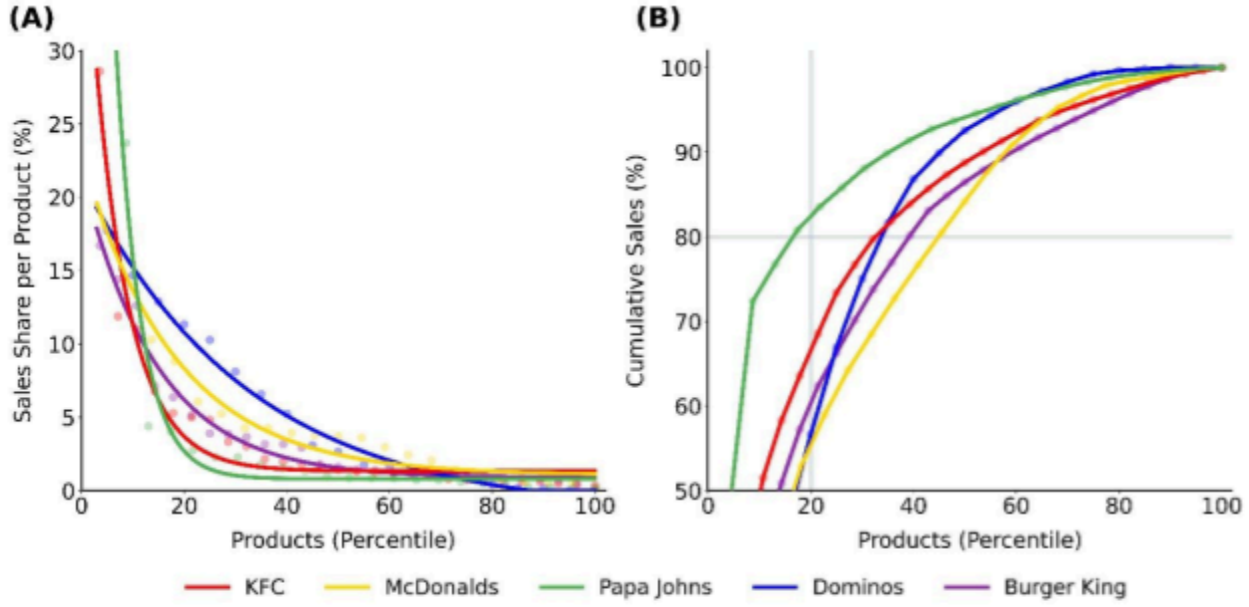


Figure 2: Estimated sales distribution of products across five food franchises. (A) Sales share per product by popularity percentile. Dots represent empirical data and lines indicate the fitted exponential decay models. (B) Cumulative sales as a function of product popularity percentile, constructed using linear interpolation. The light gray lines mark the 20% product and 80% cumulative sales thresholds to illustrate Pareto's principle.

The estimated distribution of sales differed noticeably among the five food franchises (Figure 2). The proportion of total sales generated by the top 20% most popular menu items varied across brands, indicating different levels of agreement with Pareto's principle. Papa John's showed the strongest Pareto-like pattern, with the top 20% of products accounting for 82.4% of total sales, closely matching the classic 80/20 rule (Brynjolfsson et al., 2006). The corresponding values were lower for KFC (66.6%), Burger King (60.3%), Domino's Pizza (56.7%), and McDonald's (55.4%).

ABC classification (Kheybari et al., 2019) further highlighted differences in how sales were distributed across menu items. Products ranked between the top 20% and 50% contributed only 11.6% of total sales at Papa John's, compared with 24.6% at KFC, 26.1% at Burger King, 35.9% at Domino's Pizza, and 28.8% at McDonald's. These findings indicate that sales at Papa John's were concentrated on a much smaller set of menu items than at the other franchises.

The contribution of the highest-selling menu item also varied across brands. In Papa John's, the top-ranked product alone accounted for 48.7% of total sales. The corresponding values were much lower in KFC (28.6%), McDonald's (18.1%), Domino's Pizza (17.7%), and Burger King (16.7%). Overall, franchises with stronger Pareto effects also showed greater reliance on a single flagship menu item.

Relationship Between Top Menu Heterogeneity and Sales Concentration

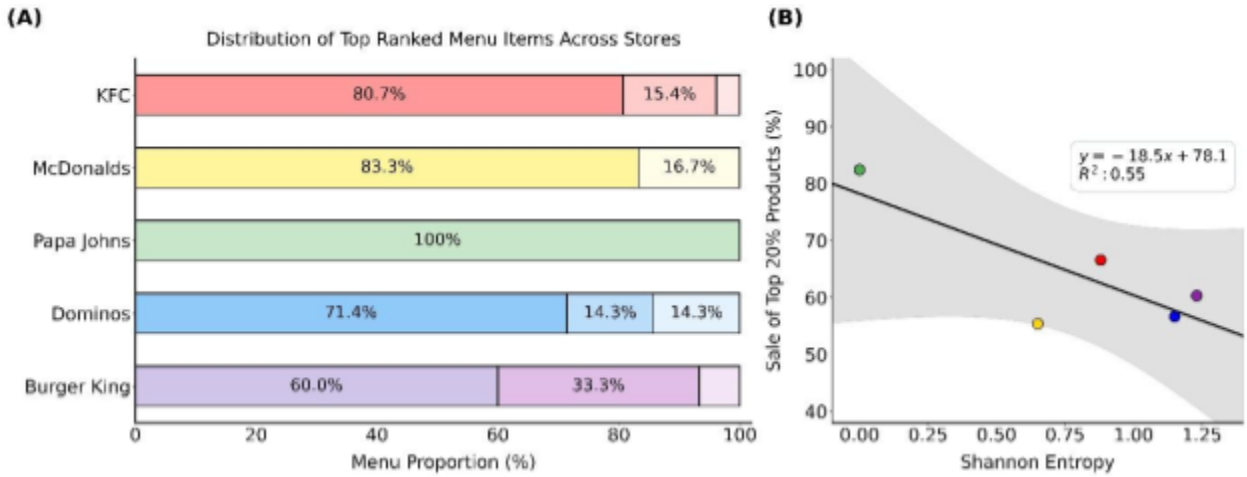


Figure 3: Relationship between the dominance of the top-ranked menu item and Pareto's principle. (A) Distribution of top-ranked menu items across stores within each franchise. (B) Heterogeneity of top-ranked menu items, measured by Shannon Entropy, versus the cumulative sales of the top 20% products. The solid line represents the linear regression fit with the gray shade indicating the confidence interval.

The consistency of the best-selling menu item across stores was measured using Shannon entropy (Figure 3). Papa John's had a Shannon entropy of 0, indicating that the same menu item ranked first in every store included in the analysis. Higher entropy values were observed for McDonald's (0.65), KFC (0.88), Domino's Pizza (1.15), and Burger King (1.23), reflecting greater variation in the top-selling menu item across stores.

Linear regression analysis revealed a negative relationship between Shannon entropy and the cumulative sales share of the top 20% products with a slope of -18.5 (R^2 = 0.55). Franchises with more consistent best-selling products generally showed stronger Pareto effects, whereas those with greater variation in store-level best-selling products exhibited weaker sales concentration. Papa John's had both the lowest entropy and the strongest Pareto effect, while Burger King had the highest entropy and one of the weakest Pareto-like sales distributions.

Investigation of Variables Affecting the Pareto Effect

The relationships between the Pareto effect and three regional franchise-level variables were examined using linear regression analyses (Figure 4). The analysis included 30 observations from five franchises across six regions. Product diversity was measured by the number of main-course items on the menu, brand popularity by the total number of customer reviews, and market competitiveness by the number of same-category stores in each region.

Among the three variables, product diversity showed the clearest relationship with the Pareto effect (Figure 4). Franchises with fewer main meal items generally had a higher cumulative sales share among the top 20% of products, suggesting a modest negative relationship between menu size and sales concentration. In contrast, brand popularity and market competitiveness showed little relationship with

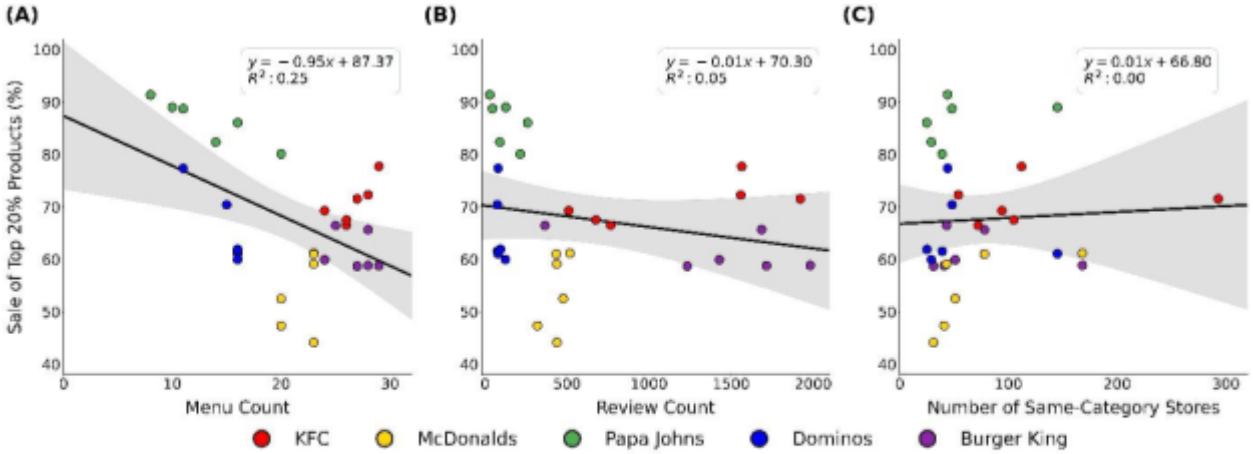


Figure 4: Investigation of variables affecting Pareto's principle. The cumulative sales of the top 20% products are plotted against three regional franchise-level variables: (A) product diversity (measured by the number of items listed in a menu), (B) brand popularity (measured by the review count), and (C) market competitiveness (measured by the number of same-category stores). Solid lines and shaded areas represent linear regression fits and their corresponding confidence intervals.

the Pareto effect. These findings suggest that menu structure is more closely associated with Pareto-like sales distributions than external market factors such as brand popularity or competitive density.

DISCUSSION

The present findings show the sales concentration on online food delivery platforms. Traditionally, restaurant sales have followed Pareto's principle, in which a small number of flagship menu items account for most of the total sales (Brynjolfsson et al., 2006). However, the rise of online delivery platforms gave consumers access to a wider range of menu items than traditional restaurants. As a result, underperforming products may attract greater attention, resulting in a more even distribution of sales (Brynjolfsson et al., 2011). Consistent with this idea, the top 20% of products accounted for only 55.4%-66.6% of total sales in McDonald's, Domino's Pizza, Burger King, and KFC, showing weaker sales concentration than expected under the traditional Pareto principle (Figure 2). However, the long-tail effect was not observed, as the bottom 80% of products by sales rank did not collectively outperform the top 20%. Food category also did not appear to determine the strength of the Pareto effect. For example, although Papa John's and Domino's Pizza are both pizza franchises, Papa John's showed the strongest sales concentration (82.4%), whereas Domino's Pizza showed one of the weakest (56.7%). These results

suggest that digitalization alone does not fully account for sales concentration on online delivery platforms.

Figures 3 and 4 show that sales concentration was more closely related to internal menu structures than to external market conditions. Franchises in which the same signature menu item consistently ranked first across stores showed stronger Pareto effects, supported by the negative relationship between Shannon entropy and the cumulative sales share of the top 20% products. In contrast, brand popularity and regional market competitiveness showed little relationship with the strength of the Pareto effect. These findings suggest that menu structure remains an important factor in consumer behavior (Skov et al., 2013), even in online food delivery platforms. Franchises with more menu items also tended to have more evenly distributed sales.

From a practical perspective, this study may help restaurants improve menu design, pricing strategies, inventory management, and marketing. However, several limitations should be considered. First, product sales were estimated from customer review data rather than actual sales records. Although customer reviews provide a useful measure of purchasing behavior, they may introduce bias because customers are not equally likely to review every product. Popular menu items may receive more reviews due to social influence, while niche products may be reviewed more often by highly motivated customers (Muchnik et al., 2013; Fleder & Hosanager, 2009). Second, the analysis included only major fast-food franchises across six metropolitan regions in South Korea. Although these regions have high use of food delivery services, the dataset may not fully represent purchasing patterns across the country or other restaurant categories. Future studies using larger datasets, direct sales, records, and a wider range of regions and restaurant types would help validate and expand these findings.

CONCLUSION

In conclusion, the present study suggests that online food delivery platforms weaken the traditional Pareto pattern by increasing the contribution of lower-ranked menu items, although the long-tail effect was not observed. At the same time, internal menu structures remained important in determining the strength of the Pareto effect. Franchises that focused on signature dishes and steady sellers exhibited sales distributions closely following Pareto's principle, whereas franchises with more diverse menus showed more evenly distributed sales. These findings may provide insight into how digitalization influences consumer behavior in the food-service industry and offer practical guidance for business strategies on online food delivery platforms.

REFERENCES

Bentley, R. Alexander, Mark E. Madsen, and Paul Ormerod. "Physical space and long-tail markets." *Physica A: Statistical Mechanics and its Applications* 388.5 (2009): 691-696.

Brynjolfsson, Erik, Yu Jeffrey Hu, and Michael D. Smith. "From niches to riches: Anatomy of the long tail." *Sloan management review* 47.4 (2006): 67-71.

Brynjolfsson, Erik, Yu Hu, and Duncan Simester. "Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales." *Management science* 57.8 (2011): 1373-1386.

Chapagain, Anish. *Hands-On Web Scraping with Python: Perform advanced scraping operations using various Python libraries and tools such as Selenium, Regex, and others*. Packt Publishing Ltd, 2019.

Fleder, Daniel, and Kartik Hosanagar. "Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity." *Management science* 55.5 (2009): 697-712.

Kearney, John. "Food consumption trends and drivers." *Philosophical transactions of the royal society B: biological sciences* 365.1554 (2010): 2793.

Kheybari, Siamak, et al. "ABC classification according to Pareto's principle: a hybrid methodology." *Opsearch* 56.2 (2019): 539-562.

Kimes, Sheryl E. "Implementing restaurant revenue management: A five-step approach." *Cornell Hotel and Restaurant Administration Quarterly* 40.3 (1999): 16-21.

Lesne, Annick. "Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics." *Mathematical Structures in Computer Science* 24.3 (2014): e240311.

Muchnik, Lev, Sinan Aral, and Sean J. Taylor. "Social influence bias: A randomized experiment." *Science* 341.6146 (2013): 647-651.

Sánchez, Claudia N., et al. "Recommendation system for a delivery food application based on number of orders." *Applied Sciences* 13.4 (2023): 2299.

Skov, Laurits Rohden, et al. "Choice architecture as a means to change eating behaviour in self‐service settings: a systematic review." *Obesity Reviews* 14.3 (2013): 187-196.