






ARTICLE



<https://doi.org/10.1057/s41599-023-02183-y>

OPEN

# Evidence of the time-varying impacts of the COVID-19 pandemic on online search activities relating to shopping products in South Korea

Jiam Song <sup>1</sup>, Kwangmin Jung <sup>2</sup> & Jonghun Kam <sup>1</sup>✉

The COVID-19 pandemic has changed the level of the received risk of the public and their social behavior patterns since 2020. This study aims to investigate temporal changes of online search activities of the public about shopping products, harnessing the NAVER DataLab Shopping Insight (NDLSI) data (weekly online search activity volumes about +1,800 shopping products) over 2017–2021. This study conducts the singular value decomposition (SVD) analysis of the NDLSI data to extract the major principal components of online search activity volumes about shopping products. Before the pandemic, the NDLSI data shows that the first principal mode (15% of variance explained) is strongly associated with an increasing trend of search activity volumes relating to shopping products. The second principal mode (10%) is strongly associated with the seasonality of monthly temperature, but in advance of four weeks. After removing the increasing trend and seasonality in the NDLSI data, the first major mode (27%) is related to the multiple waves of the new confirm cases of corona virus variants. Generally, life/health, digital/home appliance, food, childbirth/childcare shopping products are associated with the waves of the COVID-19 pandemic. While search activities for 241 shopping products are associated with the new confirmed cases of corona virus variants after the first wave, 124 and 190 shopping products are associated after the second and third waves. These changes of the public interest in online shopping products are strongly associated with changes in the COVID-19 prevention policies and risk of being exposed to the corona virus variants. This study highlights the need to better understand changes in social behavior patterns, including but not limited to e-commerce activities, for the next pandemic preparation.

<sup>1</sup>Division of Environmental Science and Engineering, POSTECH, Pohang 37673, South Korea. <sup>2</sup>Department of Industrial and Management Engineering, POSTECH, Pohang 37673, South Korea. ✉email: [jhkam@postech.ac.kr](mailto:jhkam@postech.ac.kr)

## Introduction

The COVID-19 pandemic has spread out since early 2020. The corona virus is contagious and has transformed to variants. The global community has been suffered with multiple waves of new confirmed cases of the corona virus variants. The herd community of the corona virus consists of the natural infected groups and vaccinated groups. However, the herd immunity for COVID-19 is required to prevent multiple variants from Alpha through Omicron (Moghnieh et al. 2022). Particularly, the absence of vaccines for corona virus and its variants has exacerbated the pandemic over the world. The variants have decreased the probability to form the herd immunity.

The spread of corona virus variants has significantly increased public risk perception, thereby leading people to avoid in-person activities (Dryhurst et al. 2020). Markets have responded to such changes in socioeconomic landscape by rapidly adapting digital transformations, which consequently boosted online platforms relating to shopping. The public have become preferred to online shopping, rather than in-person shopping, particularly when the number of infected people increases (Grashuis et al. 2020; Li et al. 2020; Mouratidis and Papagiannakis, 2021; Pham et al. 2020). This shift of the public's lifestyle provides an opportunity to understand the impact of the COVID-19 pandemic on socioeconomic change via big social monitoring data relating to online information seeking activities.

The impact of the COVID-19 pandemic can be examined by comparing socioeconomic activities before and after COVID-19 pandemic. However, the long-lasting pandemic crisis makes it difficult to investigate the time-varying impact of the COVID-19 pandemic. Few literature has considered temporal changes of the impact of COVID-19 through its multiple waves due to the cost of collecting relevant data and the time-consuming data preprocessing. Online social monitoring data enables us to investigate the impact of the multiple waves of the corona virus variants and relevant prevention policies on online socioeconomic activities, which are costly-efficient and real-time monitoring. Recent studies have investigated changes in online activity patterns during the COVID-19 pandemic (Gu et al. 2021; Lampos et al. 2021; Nasser et al. 2021). However, the socioeconomic impact of the multiple waves of the corona virus variants remains unknown.

During the COVID-19 pandemic, online shopping patterns has been investigated in various ways. A previous study discussed a chance to die or modify old purchasing habits from in-person activities and to create new habits (Sheth, 2020). The new habits are likely to be influenced by socioeconomic constraints, such as public policy, technology and changing demographics. Another study proposed this behavior pattern change during the COVID-19 pandemic introducing the "react", "cope", and "adapt" phases of the Reacting Coping Adapt (RCA) framework (Kirk and Rifkin, 2020). At the "react" phase, the public change their purchasing behavior based on pandemic risk perception as a social response to dynamic social distancing policies. At the "coping" phase, they start adopting new purchasing pattern based on the public policy level. At "adapt" phase, they establish/stabilize new purchasing pattern and less reactive to the pandemic situation (Guthrie et al. 2021; Kirk and Rifkin, 2020). The RCA framework has been validated by the online shopping patterns in France before, during, and after the COVID-19 pandemic (Guthrie et al. 2021). The application of this RCA framework to other countries and social behaviors is still lacking.

Nowadays, the Internet service providers monitor and record online search activities through data logging and analyze these online search activity data to detect changes in the user's interest and optimize the search algorithm for most relevant information to their interest in a timely manner. For example, increased online search activities about a specific shopping product hint an

emerging demand of the shopping product, which is a practical information for inventory and supply chain management.

Online social network data, such as Twitter, have been already used to predict stock market price change (Almehmadi, 2021). Online information search activity data, such as Google Trends, have been used to forecast the near-term values of economic indicators (Carrière-Swallow and Labbé, 2013; Choi and Varian, 2012), private consumption (Vosen and Schmidt, 2011), and epidemics (Carneiro and Mylonakis, 2009; Teng et al. 2017). Recently, the utility of these data has been examined in investigating spatiotemporal changes of social response to natural disasters, such as earthquakes and droughts (Gizzi et al. 2020; Kam et al. 2021; Kam et al. 2019; Kim et al. 2019). However, These social monitoring big data have been underutilized to investigate the changes of socioeconomic activities during the multiple waves of the corona virus variants.

The NAVER Shopping website is the most popular online shopping platform among the citizens of the Republic of Korea with online sales valued at about 2.7 billion KRW in the third quarter of 2021 (2.3 million USD) (<https://www.wiseapp.co.kr/insight/detail/89>). Online shopping activities via the NAVER shopping website can capture major modes of online shopping activities of the Koreans. For example, increased online search activities relating to a specific shopping product hint an emerging demand of the NAVER's users relating to the shopping product (Woo and Owen, 2019). Rumors about an emerging topic can affect the public's social behavior patterns via social media (Alkhodair et al. 2020). However, the quality of social monitoring data determines an appropriate analysis spatial scale, and a careful design of data preprocesses is necessary for quality control (Wilcoxson et al. 2020). Recently, it has been found that the public interest in nationwide natural disasters and global pandemics can reduce the impact of rumors on social media and online seeking activities because the rumors can be verified by the direct and indirect experience of the public from the disaster or pandemic (Park et al. 2022; Kam et al. 2021).

Recent studies found a relationship between decision making and consumer behavior patterns at the individual level during the COVID-19 pandemic (Birtus and Lăzăroiu, 2021; Smith and Machova, 2021; Vătămănescu et al. 2021). Statewide sentimental alterations have been also found from the public's complaints about water pollution during the COVID-19 pandemic (Liu et al. 2023). However, the impact of the COVID-19 pandemic and associated prevention policies on national-level social behavior pattern remains unknown. Online social monitoring data provides a unique opportunity to examine the relationship between decision making and consumer behaviors as response to changes of the COVID-19 pandemic prevention policies.

This study aims to investigate the impact of multi-year COVID-19 pandemic, using the NAVER DataLab Shopping Insight (NDLSI) data that provided by the NAVER Corporation. The data provides online search activity volumes relating to +1,800 shopping products at the nation level, which can detect an emerging change of online purchasing activities of the Koreans. The NAVER Corporation has operated the online search engine since 1999 and is the most popular internet search engine platform in South Korea. It had 1.2 billion visits from August through October 2022, and 94% of these visits solely from the Republic of Korea (<https://www.similarweb.com/website/naver.com/#traffic>). The NAVER Coporation provides weekly online search activity volume data of 1,800 shopping times since 2017 via the NDLSI platform. Such big social monitoring data provide a unique research opportunity to examine the COVID-19 impact on online shopping activities of the Koreans within the RCA framework by answering the following questions:

1. What are the major components of the dynamic patterns of online search activities before and after COVID-19?
2. How have the social behavior patterns related to online shopping search activities changed along multiple waves of corona virus variants?
3. Which prevention policies are key factors of the temporal changes of online shopping search activities during the COVID-19 pandemic?

To answer these questions, this study extracts the major modes of information seeking behavior patterns relating to shopping products from the NDLSI data (2017–2021) via the singular value decomposition algorithm-based Principal Component Analysis (PCA). Furthermore, the RCA framework is validated by the major modes of the NDLSI data during the multiple waves of the COVID-19 pandemic. The PCA analysis of the NDLSI data will advance the current understanding about changes in e-commerce before and after the two-year long COVID-19 pandemic.

### Data and methods

**NAVER DataLab Shopping Insight (NDLSI) data.** The NDLSI data includes the number of clicks on 1,837 shopping products from the NAVER Shopping platform. This study uses the NDLSI data that provide 214-week online search activities relating to 1,837 shopping products (July 31, 2017 through August 30, 2021). Weekly relative search activity volumes of the NDLSI data range from 0 to 100 (normalized by the maximum number of clicks during the search period and multiplied by 100). The NDLSI data is classified at the three levels: 11 categories for the first level, 204 categories for the second level, and 1,837 items for the third level (see Table S1. in Supplementary Material). These categories of shopping products are provided from NAVER shopping platform, which are based on the merchant category codes (MCCs) that a credit card issuer uses to categorize the transactions consumers complete using a particular card. The MCCs are used to classify merchants and businesses by the type of goods or services provided in order to keep a track of transactions. Recently, changes in credit/debit card spending in the MCCs have been analyzed during the COVID-19 pandemic (Darougeh, 2021; Dunphy et al. 2022). The first level categories include Fashion clothing, Fashion Miscellaneous Goods, Cosmetics/Beauty, Digital/Home Appliance, Furniture/Interior, Childbirth/Childcare, Food, Sports/Leisure, Life/Health, Leisure/Life convenience, and Duty-free shops. The category and product names are provided in Korean. In this study, the category and product names are translated in English via the Google Translator.

**Six COVID-19 metrics.** This study uses the six COVID-19 metrics from the Center for Systems Science and Engineering at Johns Hopkins University (JHU CSSE) COVID-19 dataset (Dong et al. 2020). The six COVID-19 metrics include new confirmed cases, stringency index, residential index, vaccination index, new death cases, and fatality. New confirmed/death cases are the number of the corresponding case of the Koreans over the study period. The stringency index is estimated based on the nine metrics: school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements, and international travel controls. The stringency index shows the strictness of the government prevention policies in quantitative method (Dong et al. 2020). The value ranges from 0 (lowest stringency) to 100 (highest stringency). Higher stringency values represent more strict prevention policies. The residential index shows the number of people who spend more time at home after the COVID-19

pandemic than before. The vaccination index is a partial vaccinated index that represents the percent of who have vaccinated at least once. The fatality index is the ratio of the number of the number of new death cases to the number of new confirmed cases. While these daily six metrics are available, this study computes and uses the weekly sums of new confirmed cases and the weekly averages of the other five COVID-19 metrics, which is a consistent temporal scale with the NDLSI data analysis. The Korea Meteorological Administration (KMA) provides the historical meteorological data of the Republic of Korea through the Open MET Data Portal platform (<https://data.kma.go.kr/cmmn/main.do>). In this study, weekly temperature averages of the 95 stations in the Republic of Korea are computed to extract the seasonality of the regional climate system.

**Singular value decomposition (SVD)-based principal component analysis.** In the machine learning field, the principal component analysis (PCA) is a popular unsupervised learning method. The PCA technique is known as a data compressing technique to extract key features of the high dimension data. Singular value decomposition (SVD) algorithm can be used to extract the PCA major modes (Vosen and Schmidt, 2011; Wilks, 2011). The SVD algorithm-based PCA decomposes a covariance matrix into three matrixes if the A matrix has  $m \times n$  dimension ( $n < m$ ; Eq. 1). These matrixes include the U matrix (an  $m$  by  $m$  matrix), the  $\Sigma$  matrix (a  $m$  diagonal matrix) and the V transpose matrix (an  $n$  by  $n$  matrix). The  $\Sigma$  matrix is a diagonal matrix which have one to one correspondence with the U matrix. The U matrix shows the orthogonal eigenvectors, which are known as the principal components (PCs).

$$A = U\Sigma V^T \quad (1)$$

In this study, the SVD algorithm is employed to the covariance matrix of the NDLSI data over the five different periods. The five periods include the period before the COVID-19 pandemic (July 31, 2017–December 31, 2019), Wave 1 (July 31, 2017–May 25, 2020), Wave 2 (July 31, 2017–October 19, 2020), Wave 3 (July 31, 2017–March 1, 2021), and Wave 4 (July 31, 2017–August 31, 2021) of the corona virus variants. Here, the waves are defined based on the surges of the new confirm cases. To explore shopping products with an increasing/decreasing interest of the public during each wave of the COVID-19 pandemic, the SVD analysis period for the wave of interest covers before the emergence of the next wave, which includes the overlapped analysis period of the previous wave. It enables us to investigate the impact of the wave of interest on the public interest relating to shopping products compared to that of the previous wave.

Two major modes are found before the COVID-19 pandemic: the increasing line trend and the seasonality pattern of the online search activities. Before employing the SVD-based PCA analysis to the NDLSI data, the linear trend and the seasonality are removed from the NDLSI matrixes over the period of Wave 1, 2, 3 and 4. The detrended NDLSI data over the different periods enable us to investigate changes of online search activities relating to shopping products over the multiple waves of COVID-19. Not available values in the NDLSI data were replaced with zeros. The U and V matrixes are the same eigenvectors of the covariance matrix and the  $\Sigma$  matrix includes the eigenvalues. The  $\Sigma$  matrix's diagonal values show the quantitative contribution of the corresponding vector to the total variance of the covariance matrix.

**Spearman's rank correlation.** Spearman's rank correlation is a non-parametric metric to find a relationship between two variables based on their ranks (Spearman, 1904). This study uses

Spearman's rank correlation because online search activities of most items in the NDLSI data do not have normal distribution. Spearman's rank correlation efficiency ranges from  $-1$  (negative perfect relation between two variables) to  $+1$  (positive perfect). In this study, Spearman's rank correlation is used to trace the user's interest in shopping products that are associated with the wave of corona virus variants. Furthermore, Spearman's rank correlation is computed between the COVID-related metrics and NDLSI data to examine which socio-economic factors associated on e-commerce search activities. First, Spearman's rank correlation coefficients are computed between the first principal component (PC1; one time series) and the search activities of  $+1,800$  shopping products ( $>1,800$  time series) during the periods of Wave 1 through 4. Furthermore, the distribution of Spearman's rank correlation coefficient is constructed by the kernel density estimate (KDE) method from the Joyplot python package (<https://github.com/leotac/joyppy>). Shopping products with a high coefficient have increased from Wave 1 through 4 (Fig. S1). In this study, 0.45 of Spearman's coefficient is a threshold value to detect up to 20% of associated item with the PC1 mode with the COVID-19 pandemic.

**Quantile-Quantile plot (QQ plot).** The number of the PC-associated shopping products affect the construction of the reliable correlation distributions with the COVID-19 metrics. A Quantile-Quantile (QQ) plot is a common visualization method to determine whether two data sets are came from same distributions or not. Despite different numbers of the COVID-19 associated shopping products during each wave period, the QQ plot can detect the stability of the correlation distribution shape. The QQ plot is based on the ranks of each data, which gives an advantage that the two dataset still can be compared in the QQ plot even though the sample sizes of the two datasets are different. The one-to-one line is a reference line of the QQ plot. When the quantile line of the two data is close to the reference line, the two sample data are from the same distribution (Nist, 2006). In this study, the QQ plots are constructed for the sensitive analysis of the stability of the correlation distribution shape to shopping products numbers. This analysis can determine how many shopping products are needed to generate the reliable distributions of its correlation with the waves of corona virus variants (Figs. S2 and S3).

## Results

**Principal components of NDLSI data.** Before the COVID-19 pandemic (hereafter, Wave 0), the first and second Principal Component (PC) modes (PC1 and PC2, respectively) explained around 15% and 10% of the total variance, respectively. PC1 was a monotonic increasing trend of online search activities for shopping products. PC2 was strongly associated with the seasonality of weekly mean temperature, however the seasonal cycle of online search activities relating to shopping products was four weeks ahead of the seasonality of the temperature (Fig. 1A, B). Based on Spearman's rank correlation coefficients with the PC1 and PC2, the top 10 shopping products showed that these two major modes captured well an increasing trend of shopping product-specific e-commerce and the seasonality of online search relating to shopping items during Wave 0 (Fig. 1C-F).

Results showed that the top 10 PC1-related shopping products included toothpaste, table tennis shoes, and cleaning tissue, packed lunch, and hair spray. Shopping products with a negative correlation coefficient with the PC1 mode included (car) hands free, sea fishing, Random Access Memory (RAM), and Network Attached Storage (NAS). Shopping products with a positive (negative) correlation coefficient with the PC2 mode (the

seasonality of temperature in advance of four weeks) were summer (winter) season shopping products. Based on the correlation coefficients with PC2, summer season shopping products included fan, parasol, yeolmu kimchi (a type of kimchi for summer), and tarp. Winter season shopping products included brooch, beanie, and neck cape. These PC2-based items were the well-known popular shopping products for summer and winter, respectively, confirming that the PCA technique is useful to extract and interpret key features in the NDLSI data when the principal major mode is associated with a certain temporal pattern (herein, the seasonality of temperature).

## Flow of PC1 related items during the COVID-19 pandemic.

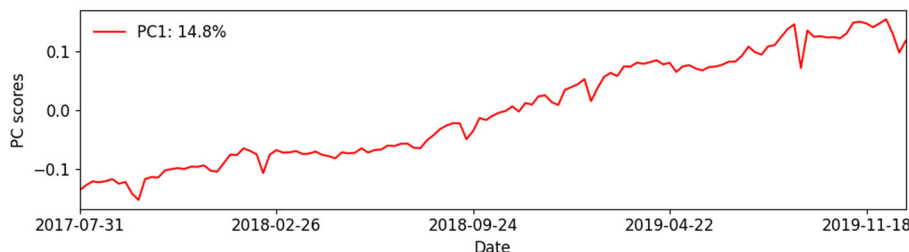
Results from the PCA analysis of the detrended NDLSI data showed that PC1 resembled the new confirm cases of COVID-19 over the four waves of the corona virus variants (Fig. 2). The percent of explained variance by the PC1 mode increased from the first wave (20%) through the fourth wave (27%), which means that associated shopping products with the corona virus variants increased during the COVID-19 pandemic. The first-level category shopping products associated with the PC1 mode showed temporal changes from Wave 1 through 4 (Fig. 3). For visualization, the Sankey diagram was constructed, which has been often used as an efficient visualization for changes of the flow/volume of the data (Lupton and Allwood, 2017).

Based on the result of the explained variance by the PC1 mode (around 20% of the total variance), changes in online search activities relating to shopping products with the correlation coefficient, 0.45, or higher (close to 20% of total items) were analyzed. Overall, life/health, digital/home appliance items showed a large percentage during the study periods). Outdoor activity-related category items, including cosmetics/beauty, fashion clothing and fashion miscellaneous goods, account for small portions than other category items. Associated items with the corona virus variants have increased from Wave 1 through Wave 4 by more than twice (from 327 to 714). After the first wave, new 241 shopping products showed the correlation coefficient, 0.45 or above. This inflow of online search activities were associated with shopping items in the categories of life/health (25%), digital/home appliances (15%), and food (15%) (Fig. 4).

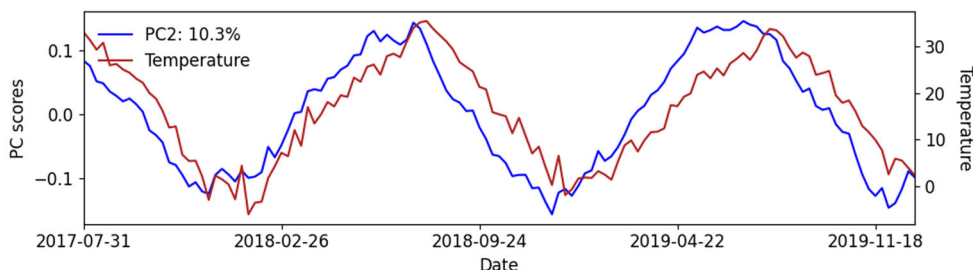
After Wave 2, the inflow of the 125 items included life/health (29%), digital/home appliance (19%), and childbirth/childcare (12%) items with decreased item numbers (125 items). After Wave 3, the inflow of 190 items included life/health (22%), digital/home appliance (17%), and childbirth/childcare (19%) items. Interestingly, duty-free shopping products and leisure/life convenience items first appeared after the Wave 2 and 4, respectively. The leisure/life convenience category items included work out class (fitness/personal training and Pilates) abroad travel items (abroad travel package, airline ticket, Wi-Fi/Universal Subscriber Identity Module (USIM)). Increasing online search activities relating to work out class may be come from a concern about health due to a restrict quarantine policy. Increased interest in abroad travel cases after Wave 4 suggests that the public in South Korea might have a low perceived risk of the COVID-19 pandemic and begin to consider that the pandemic is over.

To investigate the temporal change of the third-level (product-specific) category shopping products associated with the waves of the corona virus variants, changes in the correlation coefficients of the top 10 items were selected for each waves (Fig. 5). The results showed that 31 shopping products were associated with the PC1 component throughout the four waves. More than 32% shopping products were in the category of life/health shopping products. These 31 items can be classified into two groups: the

(A) PC1 mode before COVID-19



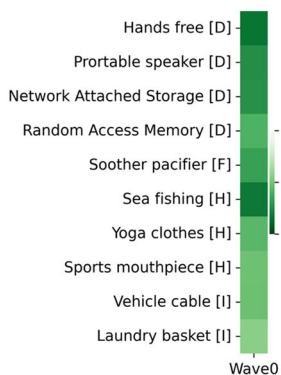
(B) PC2 mode before COVID-19



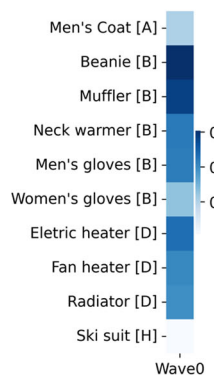
(C) PC1 positive



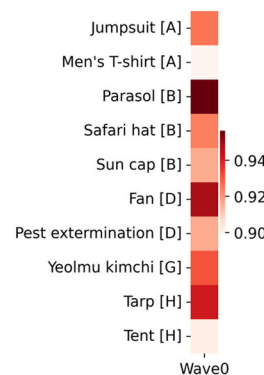
(D) PC1 negative



(E) Summer



(F) Winter



- [A]: Fashion clothing
- [B]: Fashion Miscellaneous goods
- [C]: Cosmetics/Beauty
- [D]: Digital/Home Appliance
- [E]: Furniture/Interior
- [F]: Childbirth/Childcare
- [G]: Food
- [H]: Sports/Leisure
- [I]: Life/Health
- [J]: Leisure/Life convenience
- [K]: Duty-free shops

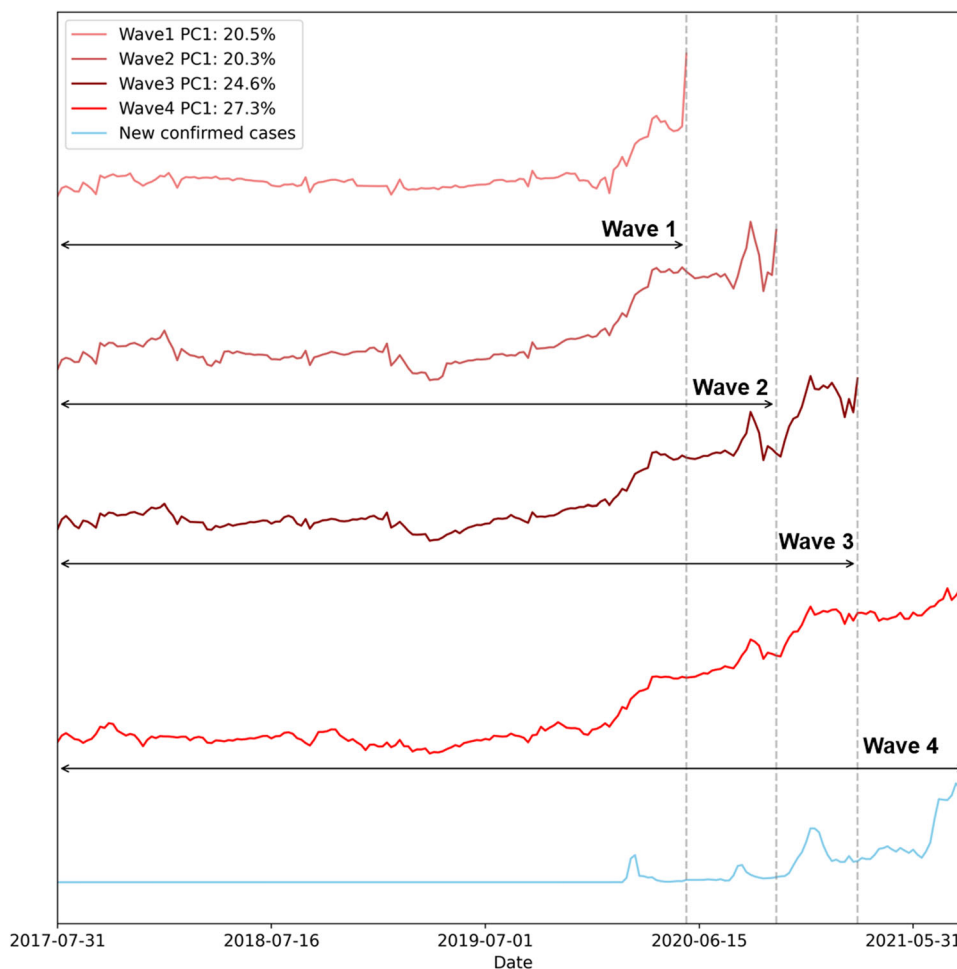
**Fig. 1 Two major modes of online search activities relating shopping products of the Koreans.** Weekly time series of Principal Component 1 (PC1) of NDLSI data before COVID-19 (A) and PC2 (B) with heatmap of correlation coefficients of top 10 correlated items. Associated online search activities of top 10 shopping products with the PC1 time series (A): Positive (C) and negative correlation (D). Associated online search activities of top 10 shopping products with seasonality (B): Summer- and winter-related shopping products in (E) and (F), respectively.

items with a higher and lower correlation coefficient over time. The first group items included minidisc player monitor arms, webcam, interphone box, fabric, handicraft supplies/subsidiary materials, character card/ticket, processed snacks, cooking oil/oil, bread, tuning supplies, craft, feed, seeds/seedlings, water aperture, gravel/sands/soil, landscape tree/sapling. These first group shopping products showed a persistent increase in the correlation coefficient through the multiple waves. The second group items included gas range, microwave, toothbrush, hula hoop. These second group shopping products showed a decrease in the correlation coefficient (Fig. 5).

These two shopping product groups might originate from the different social response to the strictness of prevention policies. During the first wave, the government forced the public to stay at

home to minimize the risk of being exposed to the corona virus. However, the prevention policies became less strict at Wave 4 to account for the fatigue of the public from the multi-year pandemic and revive local business and industry sectors. While the first group items have become more associated with the waves of the corona virus variants, the second group items no longer showed a high correlation coefficient with the corona virus variants.

**Association with the six COVID-19 metrics.** A surge of new confirmed cases of corona virus variants can influence social behavior patterns relating to e-commerce in a different way due to a different level of the COVID-19 prevention policy and the easy access of online shopping activities. In this study, Spearman's rank correlation coefficients between the six COVID-19 metrics



**Fig. 2** First major modes of the detrend NDLSI data of the four waves of the corona virus variants. Weekly time series of the PC1 mode of the detrended NDLSI data up to Wave 1 through Wave 4 (gray dash lines) along South Korea's COVID-19 new confirmed cases (a sky line).

and the NDLSI data are computed to investigate potential causes of changes in online search activity volumes of shopping products (Fig. 6).

The six COVID-19 metrics showed different correlation distributions with the six COVID-19 metrics (Fig. 7). As the sensitivity test of the correlation distribution shape to the number of shopping products, the Quantile-Quantile (QQ) plots have been made along the different shopping times (see Figs. S2 and S3). According to the QQ plots, the top 50 items were chosen to construct the correlation distributions of the top 50 shopping products with the vaccination index. The correlation coefficients were widely distributed, indicating a relatively weak association with online search activities relating to the shopping products (Fig. 7A). The correlation distributions with the stringency and fatality indices showed a low variance with high correlation coefficients above 0.8. The correlation distribution with the residential index showed a relatively low correlation coefficients than those with the stringency and fatality indices. New confirmed and death cases showed a relatively high variance than the correlation distributions with the fatality and stringency data. The categories of the top 50 shopping products included life/health (20%), digital/home appliance (16%) and food (16%), shopping products (Fig. 7B).

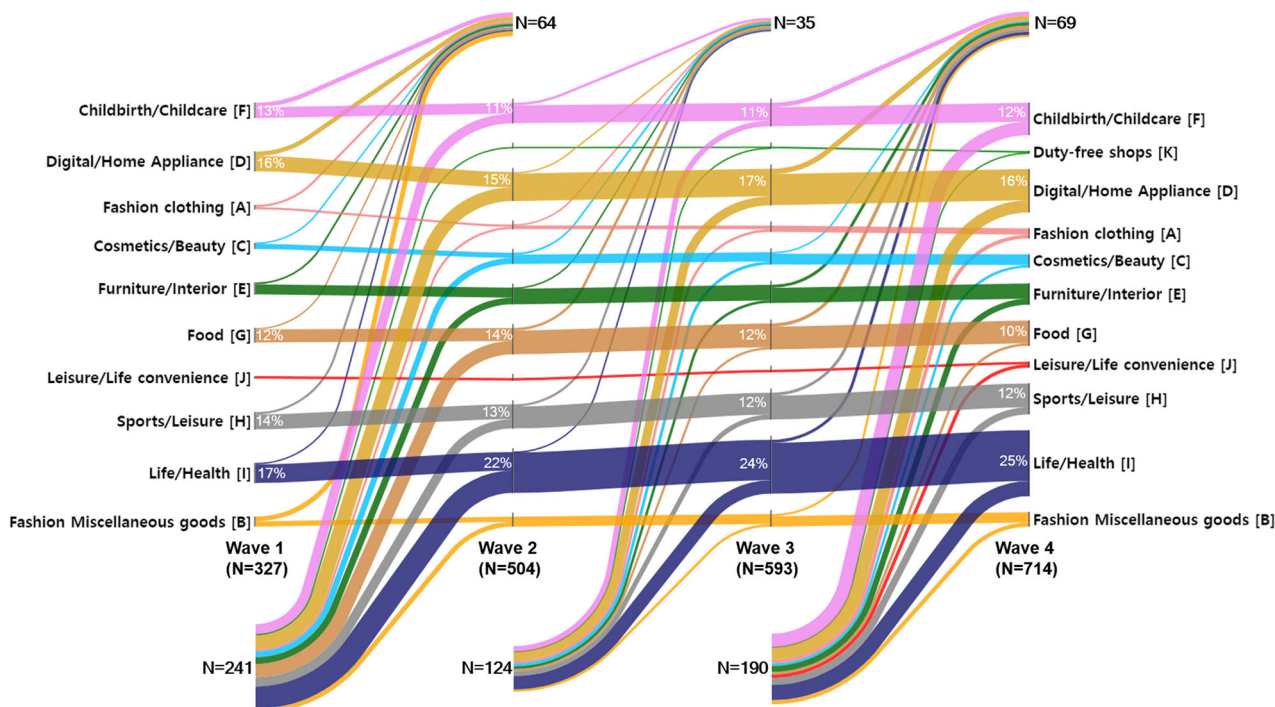
To investigate associations of online search activities relating shopping products with the six COVID-19 metrics, the Spearman's rank correlation coefficients with the 31 PC1-associated items associated with the COVID-19 pandemic were computed (Fig. 8). New confirmed cases, stringency, residential index, new

death cases and fatality showed a high correlation coefficient with the most of top 10 shopping products. The vaccination index showed no significant correlation coefficient with the top 10 shopping products. Gas range, baby walker and toothbrush items showed a relatively low correlation with the COVID-19 metrics than other shopping products. Online search activities relating to these shopping items showed a decreasing correlation during COVID-19 pandemic (see Fig. 5), that is, these items no longer show a significant effect of the COVID-19 pandemic after the Wave 4.

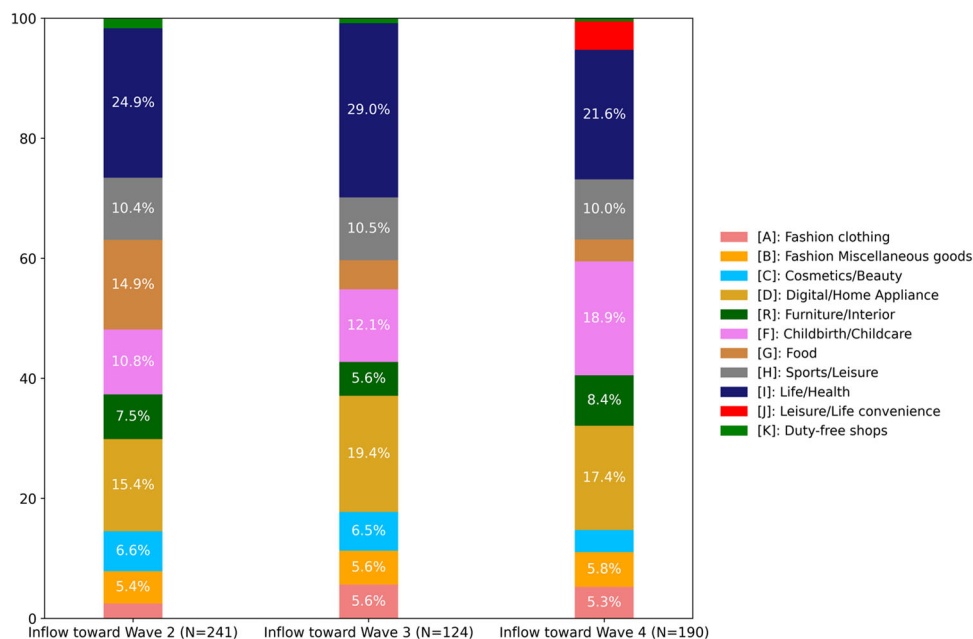
Overall, the stringency and fatality metrics generally have high association with the changes in online search activity patterns for shopping product. Stringency can be regarded as how government control public strictly. Fatality shows seriousness of pandemic. The results indicate that consumer behavior response sensitively to extent of restriction policies and seriousness of pandemic.

## Discussion

This study used the NDLSI data about the online search activity volumes for shopping products, not real purchasing data. Using the data of online search activities can provide an evidence on emerging purchasing patterns of the public in the next regime, implying that the public might tend to purchase items that have been most searched in the previous timeframe (Chen et al. 2017). Lately, credit card data<sup>1</sup> and bar cord data<sup>2</sup> include the records of actual purchase activities. Integrating the actual purchase data



**Fig. 3 Changes in the first-level category shopping products associated with the corona virus variants during the four waves.** Sankey diagram of COVID-19 associated shopping products during the four waves.

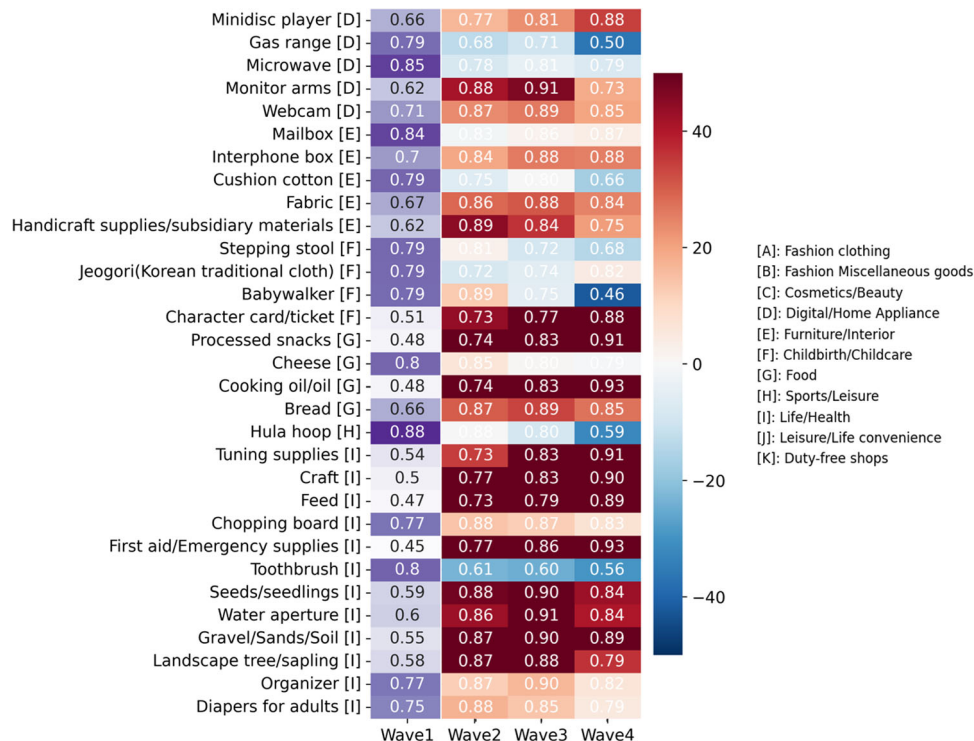


**Fig. 4 Changes in new shopping products associated with the corona virus variants during the period of Wave 2 through 4.** Percentages of the first-level shopping product categories of inflow after Wave 2, 3, and 4.

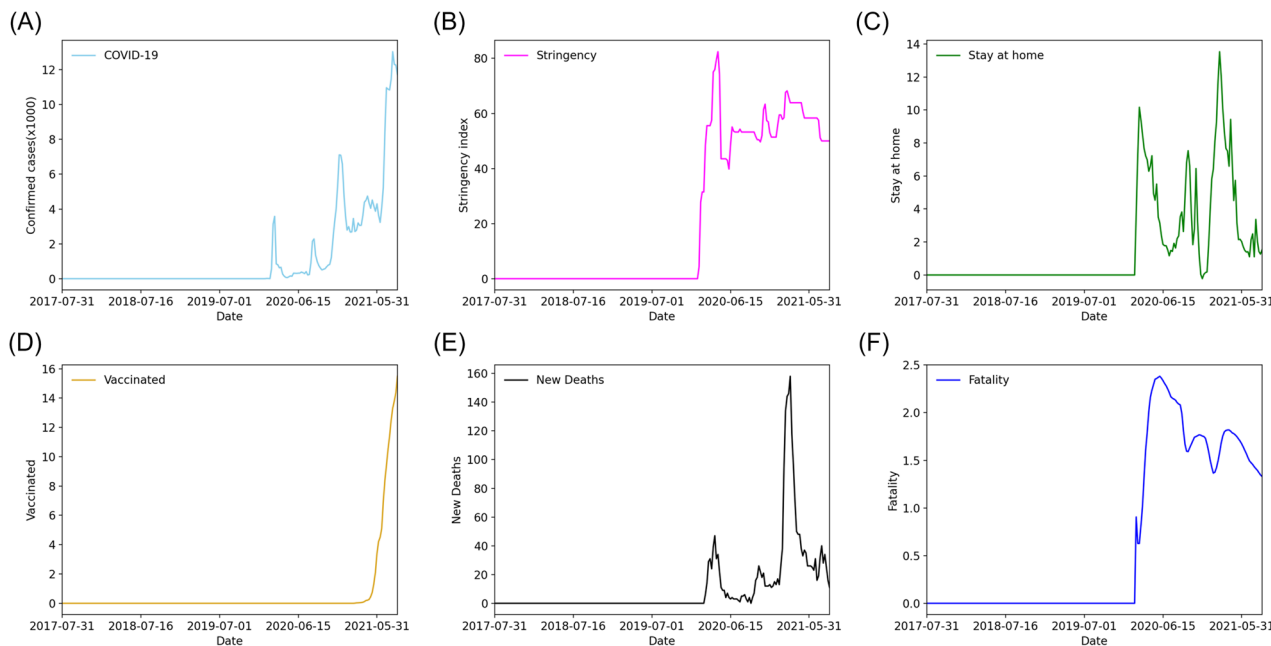
and online search activity data can provide more practical guidelines and plans for socio-economic changes not only during the COVID-19 pandemic, but also the post pandemic period.

This study revealed that the public interest in online shopping products had been changed not only after the first wave of the COVID-19 pandemic but also during the following three waves. These dynamic patterns of the public interest in online shopping products were possibly explained by the RCA framework (Kirk and Rifkin, 2020). The RCA framework consists of reacting,

copied, and adapting phases, and significant changes in social behavior patterns are expected during a transition period from one to another phase. The first wave was a typical 'react' phase because people responded to the pandemic situation. A large inflow volume after Wave 2 (241 items) indicated a coping phase. The new confirmed cases were relatively low during Wave 2 (see a line colored in sky in Fig. 2) compared with those during other waves. Inflow of online search activities relating shopping products was the minimum after Wave 2. This finding suggests that a



**Fig. 5 Heatmap of Spearman's rank correlation coefficient between 31 items with PC1 from Wave 1 to 4.** The numbers of the Wave 1 through 4 heatmaps are Spearman's rank correlation coefficients of the shopping products with the PC1 mode. The Wave 2 to 4 heatmap depict the percent changes of Spearman's rank correlation coefficients compared with the correlation coefficients after Wave 1  $((Corr_X - Corr_1) / Corr_1) * 100$ , where X depicts the wave occurrence order (X = 2, 3, and 4).

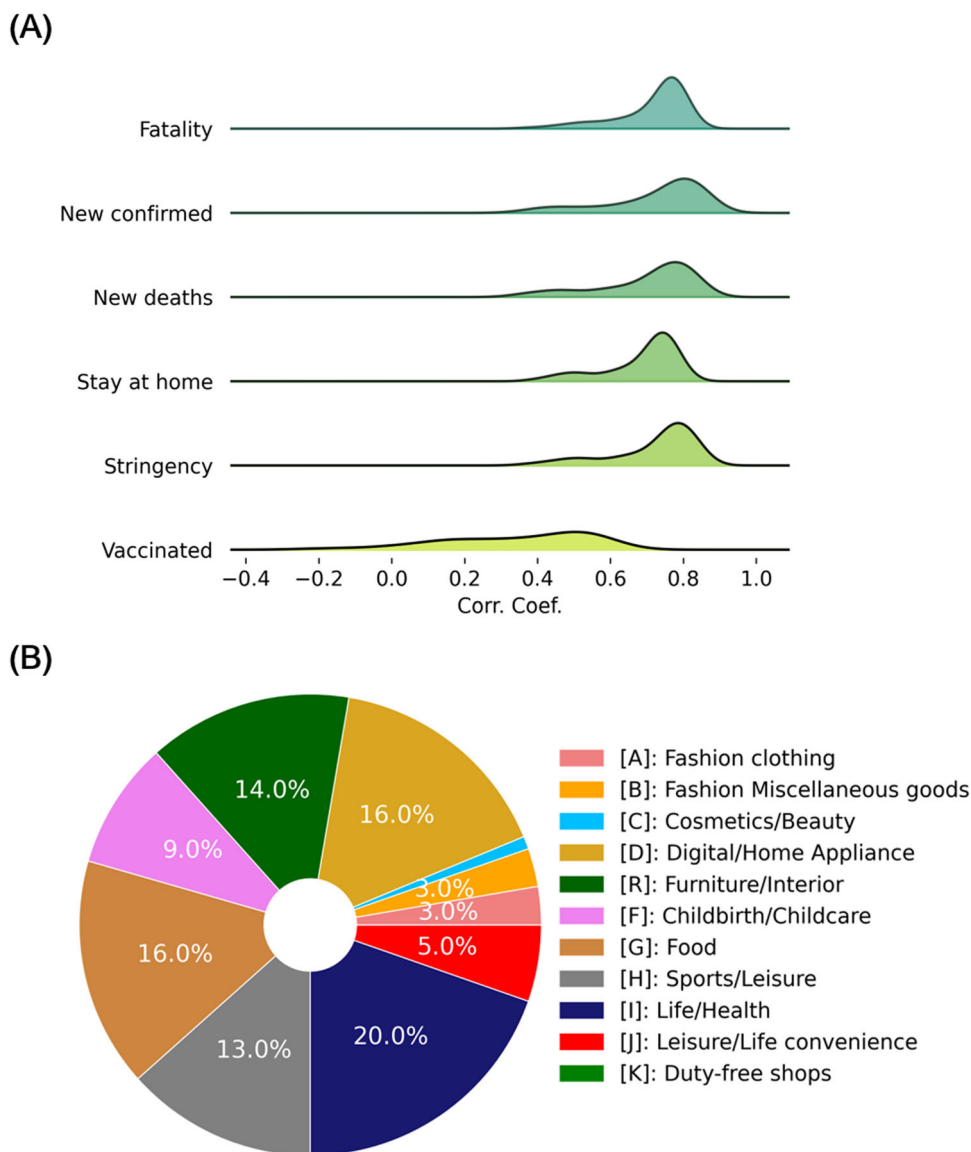


**Fig. 6 Changes in the six COVID-19 metrics over the four waves.** Weekly time series of the COVID-19 new confirmed cases (A), the stringency (B), residential index (C), vaccinated rate (D), new deaths by the corona virus (E), and fatality (F).

transition from a 'react' to 'coping' phase might occur between Wave 2 and Wave 3. After Wave 3, the public coped with the long-term pandemic. During Wave 4, the categories related with outdoor activities show a low percentage, indicating a low level of the public interest in outdoor activities due to the COVID-19 quarantine policy. The result that the inflow of online search

activities relating to leisure/life convenience items (workout class, abroad travel) at Wave 4 indicates that the public became less reactive to the wave of the corona virus variants, which hints an emerging signal of a low perceived risk of the COVID-19 pandemic after Wave 4. Therefore, the 'adapt' phase transition is expected after Wave 4.

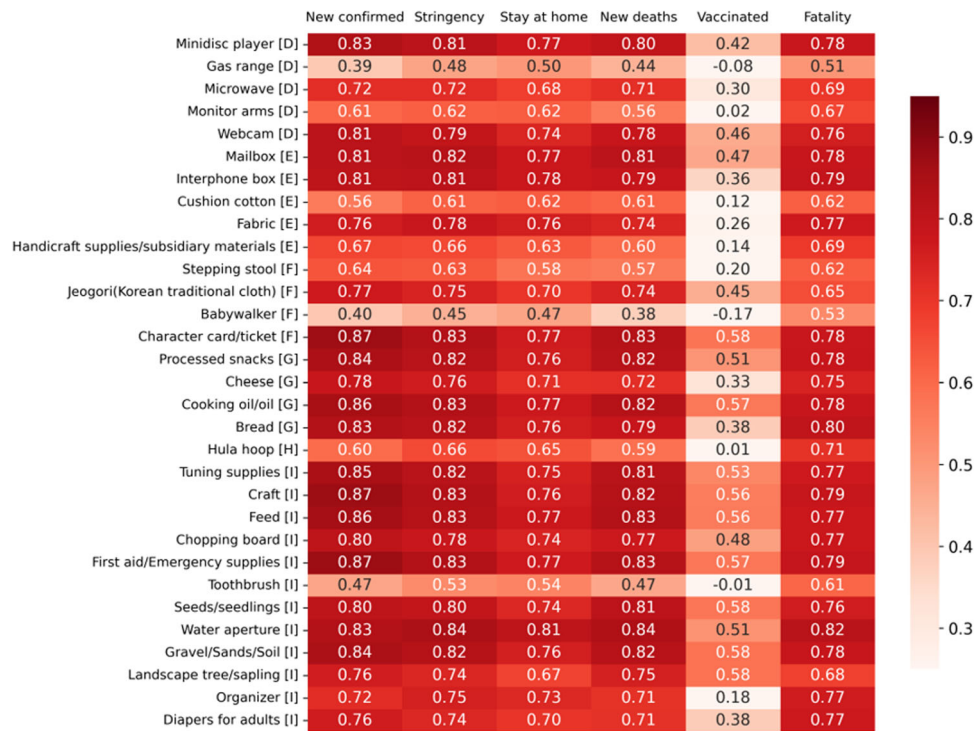




**Fig. 7 Correlation distributions of the six COVID-19 metrics with the first mode of online search activities relating to shopping products.** Distributions of Spearman's rank correlation coefficient of top 50 items related to the COVID-19 pandemic with six COVID-19 metrics (A), and the pie chart of first category percentage of items of top 50 items (B).

Understanding the public's purchasing patterns amid a global crisis via big social monitoring data is critical from the risk management perspective. Risk control (e.g., self-protection) and financing (insurance) strategies can be improved for the next global crisis by understanding and predicting changes in social behaviors. This study found that the shopping products with an increased interest of the public have been changed during the two year-long COVID-19 pandemic, which can be explained by different stages of the RCA framework. The social behavior patterns found by this study had been also reported from the observed reacting and coping consumer behaviors in mass media and online and reacting public behavior to social distance during the COVID-19 pandemic (Guthrie et al. 2021; Kirk and Rifkin, 2020; Tintori et al. 2020). Specifically, better understanding and predicting of which products can help markets manage inventory of shopping products that are in an emerging high/low demand throughout different regimes of the crisis. This study found that associations of these products were more clear when they were used for self-protection measures (e.g., facial masks in the COVID-19 pandemic).

Governments and authorities can accordingly implement changes in the public's actions to prevent potential market failures that, for example, self-protection measures may not be sufficiently supplied, or big market players use their power to dominate necessity markets (Stiglitz, 2021). These responses from the public and private sectors can be optimized with prevention plans in a timely manner of different waves of the crisis by analyzing big social monitoring data. This study found changes in the interest and demand of the shopping products related to self-protection measures during the COVID-19 pandemic, which hints how to facilitate big social monitoring data to mitigate the adverse effects of daily infections. Furthermore, this information can help insurance industries manage systematic risks that cannot be fully controlled by individuals or other industry sectors, which can offer risk transfer measures (Alonso et al. 2020; Harris et al. 2021; Peiffer-Smadja et al. 2020; Rita et al. 2019). This study also found a strong association between changes in online search activities of the public relating to shopping items and perceived risk, which was previously found in the travel insurance purchasing patterns (Al Mamun et al. 2022; Tan and Caponecchia,



**Fig. 8 Associations of online search activities relating to shopping items with the six COVID-19 metrics.** Heatmap of Spearman's rank correlation coefficient between COVID-19 metrics and the 31 shopping products.

2021). This information can give an insight for how to increase the public's willingness to prepare for the next pandemic.

Search engine optimization (SEO) algorithms for searching items have been developed, particularly in the e-commerce sector to increase the customer's satisfaction and loyalty (Husain et al. 2020; Liu et al. 2008; Pratminingsih et al. 2013). Some online search engine platforms collect the data of the user's online activities and optimize the customized recommendation algorithm that could give more relevant result of searching. Especially, e-commerce sites, such as Amazon, have developed this customized SEO algorithm to increase a chance to purchase the products (Heng et al. 2018; Linden et al. 2003). In this study, the observational evidence of the COVID-19 impact on online search activities about shopping products was reported, which was also found in online shopping pattern for apparel (Watanabe et al. 2021). The SEO algorithms developed by the data before the COVID-19 pandemic increased the user's complaint by three times (Dahiya et al. 2021), implying that the COVID-19 pandemic was an unprecedented event since the advent of Internet that supposedly cause a drastic context difference. Therefore, the SEO algorithms are needed to update until the data after the pandemic is sufficient. Furthermore, the expected continued growth of online commerce industries requires the coping strategies to adapt an increasing trend of not only pandemics but also other disasters such as climatic extremes, pandemic, war, and terror.

This study provides an insight about how social big monitoring data can help authorities to better understand the social response to COVID-19 via near real-time social monitoring data. In this study, the NDLSI data about the online search activity volumes relating to shopping products, not real purchasing data, were used. The NDLSI data analysis provided a possible evidence on an emerging change in the public's purchasing patterns at the shopping product level. Previously, it was found that the public tended to purchase shopping products that have been most searched in the previous timeframe (Chen et al. 2017).

Associations between the public interest in shopping products and purchase records can be explored using credit card data<sup>3</sup> and barcode data<sup>4</sup>. These data have been used to investigate changes in spending associated with stringent nonpharmaceutical interventions during the COVID-19 pandemic (Horvath et al. 2023). Integrating the actual purchase data and online search activity data can give more practical guidelines and plans for socio-economic changes during not only the COVID-19 pandemic, but also the post pandemic period. Furthermore, the e-commerce sector can harness social big monitoring data to develop their strategic plans for supply chain management for the next pandemic.

This study also explored associations of changes of online search activity patterns with the COVID-19 metrics. The results showed that the COVID-19 metrics, except for vaccination, were strongly associated with changes in online search activity patterns relating to shopping products. The stringency index was a reliable indicator of the strictness of the government's response to the COVID-19 pandemic and had a significant impact on social behavior patterns, which is in line with the findings of Makki et al. (2020) that the timing and duration of the stringency implementation are key factors to prevent the spread of the corona virus variants. Furthermore, a recent study found that policy perceptions affect the practice of volunteered prevention behaviors, such as mask wearing and social distancing (Lee et al. 2021). They found that the perceived policy stringency was associated with actual risk and political ideology, causing non-compliance in communities during the COVID-19 pandemic.

The proposed methods in this study have some limitations. For example, the results based on the correlation analysis provide potential, not actual, triggers of changes in the social behavior patterns during the COVID-19 pandemic, which have previously known as the caveat of the correlation analysis (Haley and Drazen, 1998; Stigler, 2005). The findings of this study about potential triggers however can help design more effective and efficient interview and survey questionnaires to investigate true

triggers of changes in the public interest in shopping products. Combined information from big social monitoring and survey/interview data will create new knowledge about the dynamics of social behavior patterns and help develop a reliable social behavior prediction modeling.

## Conclusions

This study succeeded to extract the major modes of the public's interest in shopping products and investigate changes in online search activities relating to associated shopping products with the COVID-19 pandemic. The SVD algorithm-based PCA analysis of the NDLSI data showed the dynamic patterns of online search activities relating to shopping products during the two year-long COVID-19 pandemic. Before the COVID-19 pandemic, an increasing trend and seasonality of online search activity volumes about shopping products are the major mode of the NDLSI data. After the COVID-19 pandemic, the impact of COVID-19 on online search activities relating to shopping products were various during the four waves of the corona virus variants, particularly when the objective risk was dramatically increased. Changes of the online search activity patterns were associated with the change of the COVID-19 prevention policy and objective risk of being exposed to the corona virus variants. This study attempted to explain the changes of these online search activity patterns within the RCA framework by identifying the react, coping, and adapt phases.

This study highlights the utility of online social monitoring data in developing strategic plans for preparation, mitigation, and recovery policies for the next pandemic. Furthermore, the findings of this study can guide how to design interview and survey questionnaires to investigate actual drivers of social behavior changes during the COVID-19 pandemic. Integrated studies using online social monitoring data and survey and interview data will advance the current knowledge and prediction skill of social behavior changes, which can provide actionable information to mitigate its adverse effects for the sustainable development of our communities Kim et al. (2019), Spearman (1904).

## Data availability

The data used in this study are available at Harvard Dataverse: <https://doi.org/10.7910/DVN/JT8RCK>.

Received: 3 December 2022; Accepted: 20 September 2023;

Published online: 09 October 2023

## Notes

- <https://www.bccard.com/card/html/company/en/index.jsp>
- <https://www.chicagobooth.edu/research/kilts/datasets/nielsenIQ-nielsen>
- <https://www.bccard.com/card/html/company/en/index.jsp>
- <https://www.chicagobooth.edu/research/kilts/datasets/nielsenIQ-nielsen>

## References

- Al Mamun A, Rahman MK, Yang Q, Jannat T, Salameh AA, Fazal SA (2022) Predicting the willingness and purchase of travel insurance during the COVID-19 pandemic. *Front Public Health* 10:907005
- Alkhodair SA, Ding SH, Fung BC, Liu J (2020) Detecting breaking news rumors of emerging topics in social media. *Inform Process Manag* 57(2):102018
- Almehmadi A (2021) COVID-19 pandemic data predict the stock market. *Comput Syst Sci Eng* 36(3):451–460
- Alonso AD, Kok SK, Bressan A, O'Shea M, Sakellarios N, Koresis A, Solis MAB, Santoni LJ (2020) COVID-19, aftermath, impacts, and hospitality firms: An international perspective. *Int J Hosp Manag* 91:102654
- Birtus M, Lázároiu G (2021) The neurobehavioral economics of the covid-19 pandemic: consumer cognition, perception, sentiment, choice, and decision-making. *Anal Metaphys* 20:89–101
- Carneiro HA, Mylonakis E (2009) Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 49(10):1557–1564
- Carrière-Swallow Y, Labbé F (2013) Nowcasting with Google Trends in an emerging market. *J Forecast* 32(4):289–298
- Chen Y-C, Lee Y-H, Wu H-C, Sung Y-C, Chen, H-Y (2017) Online apparel shopping behavior: Effects of consumer information search on purchase decision making in the digital age. 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)
- Choi H, Varian H (2012) Predicting the present with Google Trends. *Econ Record* 88:2–9
- Dahiya S, Rokanas LN, Singh S, Yang M, Peha JM (2021) Lessons from internet use and performance during COVID-19. *J Inform Policy* 11:202–221
- Darougeh S (2021) Dispersed consumption versus compressed output: Assessing the sectoral effects of a pandemic. *J Macroecon* 68:103302
- Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 20(5):533–534
- Dunphy C, Miller GF, Rice K, Vo L, Sunshine G, McCord R, Howard-Williams M, Coronado F (2022) The impact of COVID-19 state closure orders on consumer spending, employment, and business revenue. *J Public Health Manag Pract* 28(1):43
- Dryhurst S, Schneider CR, Kerr J, Freeman AL, Recchia G, Van Der Bles AM, Spiegelhalter D, Van Der Linden S (2020) Risk perceptions of COVID-19 around the world. *J Risk Res* 23(7-8):994–1006
- Gizzi FT, Kam J, Porrini D (2020) Time windows of opportunities to fight earthquake under-insurance: evidence from Google Trends. *Humanit Soc Sci Commun* 7:61
- Grashuis J, Skevas T, Segovia MS (2020) Grocery shopping preferences during the COVID-19 pandemic. *Sustainability* 12(13):5369
- Gu S, Ślusarczyk B, Hajizada S, Kovalyova I, Sakhibieva A (2021) Impact of the covid-19 pandemic on online consumer purchasing behavior. *J Theor Appl Electron Commer Res* 16(6):2263–2281
- Guthrie C, Fosso-Wamba S, Arnaud JB (2021) Online consumer resilience during a pandemic: An exploratory study of e-commerce behavior before, during and after a COVID-19 lockdown. *J Retail Consum Serv* 61:102570
- Haley KJ, Drazen JM (1998) Inflammation and airway function in asthma: what you see is not necessarily what you get. *Am J Respir Crit Care Med* 157(1):1–3
- Harris TF, Yelowitz A, Courtemanche C (2021) Did COVID-19 change life insurance offerings? *J Risk Insur* 88(4):831–861
- Heng Y, Gao Z, Jiang Y, Chen X (2018) Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach. *J Retail Consum Serv* 42:161–168
- Horvath A, Kay B, Wix C (2023) The Covid-19 shock and consumer credit: Evidence from credit card data. *J Bank Financ* 152:106854
- Husain T, Sani A, Ardhiansyah M, Wiliani N (2020) Online Shop as an interactive media information society based on search engine optimization (SEO). *Int J Comput Trend Technol* 68(3):53–57
- Kam J, Park J, Shao W, Song J, Kim J, Gizzi FT, Porrini D, Suh Y-J (2021) Data-driven modeling reveals the Western dominance of global public interest in earthquakes. *Humanit Soc Sci Commun* 8:242
- Kam J, Stowers K, Kim S (2019) Monitoring of drought awareness from google trends: a case study of the 2011–17 California drought. *Weather Clim Soc* 11(2):419–429
- Kim S, Shao W, Kam J (2019) Spatiotemporal Patterns of US Drought Awareness. *Palgrave Commun* 5:107
- Kirk CP, Rifkin LS (2020) I'll trade you diamonds for toilet paper: Consumer reacting, coping and adapting behaviors in the COVID-19 pandemic. *J Bus Res* 117:124–131
- Lamos V, Majumder MS, Yom-Tov E, Edelstein M, Moura S, Hamada Y, Rangaka MX, McKendry RA, Cox IJ (2021) Tracking COVID-19 using online search. *NPJ Digit Med* 4(1):1–11
- Lee S, Peng T-Q, Lapinski MK, Turner MM, Jang Y, Schaaf A (2021) Too stringent or too lenient: antecedents and consequences of perceived stringency of COVID-19 policies in the United States. *Health Policy Open* 2:100047
- Li J, Hallsworth AG, Coca-Stefaniak JA (2020) Changing grocery shopping behaviours among Chinese consumers at the outset of the COVID-19 outbreak. *Tijdschrift voor economische en sociale geografie* 111(3):574–583. <https://doi.org/10.1111/tesg.12420>
- Linden G, Smith B, York J (2003) Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet Comput* 7(1):76–80
- Liu X, He M, Gao F, Xie P (2008) An empirical study of online shopping customer satisfaction in China: a holistic perspective. *Int J Retail Distrib Manag* 36(11):919–940
- Liu A, Kam J, Kwon S, Shao W (2023) Monitoring the impact of climate extremes and COVID-19 on statewide sentiment alterations in water pollution complaints. *npj Clean Water* 6:29. <https://doi.org/10.1038/s41545-023-00244-y>
- Lupton RC, Allwood JM (2017) Hybrid Sankey diagrams: Visual analysis of multidimensional data for understanding resource use. *Resour Conserv Recycl* 124:141–151

- Makki F, Sedas PS, Kontar J, Saleh N, Krpan D (2020) Compliance and stringency measures in response to COVID-19: a regional study. *J Behav Econ Policy* 4(S2):15–24
- Moghnieh R, Abdallah D, Bizri AR (2022) COVID-19: second wave or multiple peaks, natural herd immunity or vaccine—we should be prepared. *Disaster Med Public Health Preparedness* 16(2):718–725
- Mouratidis K, Papagiannakis A (2021) COVID-19, internet, and mobility: The rise of telework, telehealth, e-learning, and e-shopping. *Sustain Cities Soc* 74:103182
- Nasser N, Karim L, El Oudrhiri A, Ali A, Khan N (2021) n-Gram based language processing using Twitter dataset to identify COVID-19 patients. *Sustain Cities and Soc* 72:103048
- Nist N (2006) SEMATECH e-handbook of statistical methods. US Department of Commerce
- Park C-K, Kam J, Byun H-R, Kim D-W (2022) A Self-Calibrating Effective Drought Index (scEDI): Evaluation against Social Drought Impact Records over the Korean Peninsula (1777–2020). *J Hydrol* 613:128357
- Peiffer-Smadja N, Lucet J-C, Bendjelloul G, Bouadma L, Gerard S, Choquet C, Jacques S, Khalil A, Maisani P, Casalino E (2020) Challenges and issues about organizing a hospital to respond to the COVID-19 outbreak: experience from a French reference centre. *Clin Microbiol Infect* 26(6):669–672
- Pham VK, Do Thi TH, Ha Le TH (2020) A study on the COVID-19 awareness affecting the consumer perceived benefits of online shopping in Vietnam. *Cogent Bus Manag* 7(1):1846882
- Pratminingsih SA, Lipuringtyas C, Rimenta T (2013) Factors influencing customer loyalty toward online shopping. *Int J Trade Econ Financ* 4(3):104–110
- Rita P, Oliveira T, Farisa A (2019) The impact of e-service quality and customer satisfaction on customer behavior in online shopping. *Heliyon* 5(10):e02690. <https://doi.org/10.1016/j.heliyon.2019.e02690>
- Sheth J (2020) Impact of Covid-19 on consumer behavior: Will the old habits return or die? *J Bus Res* 117:280–283
- Smith A, Machova V (2021) Consumer tastes, sentiments, attitudes, and behaviors related to COVID-19. *Anal Metaphys* 20:145–158
- Spearman C (1904) The proof and measurement of correlation between two things. *Am J Psychol* 15:72–101
- Stiglitz JE (2021) The proper role of government in the market economy: The case of the post-COVID recovery. *J Gov Econ* 1:100004
- Stigler SM (2005) Correlation and causation: A comment. *Perspect Biol Med* 48(1):88–S94
- Tan D, Caponecchia C (2021) COVID-19 and the public perception of travel insurance. *Ann Tour Res* 90:103106
- Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, An X, Feng D, Tong Y (2017) Dynamic forecasting of Zika epidemics using Google Trends. *PloS One* 12(1):e0165085
- Tintori A, Cerbara L, Ciancimino G, Crescimbeni M, La Longa F, Versari A (2020) Adaptive behavioural coping strategies as reaction to COVID-19 social distancing in Italy. *Eur Rev Med Pharmacol Sci* 24:10860–10866. [https://doi.org/10.26355/eurrev\\_202010\\_23449](https://doi.org/10.26355/eurrev_202010_23449)
- Vătămănescu EM, Dabija DC, Gazzola P, Cegarro-Navarro JG, Buzzi T (2021) Before and after the outbreak of covid-19: Linking fashion companies' corporate social responsibility approach to consumers' demand for sustainable products. *J Clean Prod* 321:128945
- Vosen S, Schmidt T (2011) Forecasting private consumption: survey-based indicators vs. Google trends. *J Forecast* 30(6):565–578
- Watanabe C, Akhtar W, Tou Y, Neittaanmäki P (2021) Amazon's new supram-channel: realizing growing seamless switching for apparel during COVID-19. *Technol Soc* 66:101645
- Wilcoxon J, Follett L, Severe S (2020) Forecasting foreign exchange markets using Google Trends: Prediction performance of competing models. *J Behav Financ* 21(4):412–422
- Wilks DS (2011) Principal component (EOF) analysis. In *Int Geophys* 100:519–562. <https://doi.org/10.1016/B978-0-12-385022-5.00012-9>
- Woo J, Owen AL (2019) Forecasting private consumption with Google Trends data. *J Forecast* 38(2):81–91

## Acknowledgements

We thank the NAVER DataLab for making available the NAVER DataLab Shopping Insight (NDLSI) data. This study was supported by a grant from the National Research Foundation of Korea (NRF-2021R1A2C1093866).

## Author contributions

JS: conceptualization, data curation, investigation, formal analysis, methodology, software, validation, visualization, writing—original draft, writing—review and editing. JK: conceptualization, methodology, funding acquisition, project administration, resources, supervision, writing—review and editing. KJ: analysis, validation, writing—review and editing.

## Competing interests

The authors declare no competing interests.

## Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

## Informed consent

This article does not contain any studies with human participants performed by any of the authors.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-023-02183-y>.

**Correspondence** and requests for materials should be addressed to Jonghun Kam.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2023