

탐색적 자료 분석(EDA) 기법을 활용한 국내 11개 대표 온라인 쇼핑몰 BEST 100 비교

Comparison of Online Shopping Mall BEST 100 using Exploratory Data Analysis

강지천¹ · 강주영^{2*}

아주대학교 경영대학 경영학과¹
아주대학교 경영대학 e-비즈니스학과²

요 약

초기 온라인 쇼핑몰이 등장할 때부터 지금까지 BEST 100은 모든 쇼핑몰 웹사이트의 핵심 기능으로 제공되고 있다. BEST 100은 소비자들이 한눈에 인기 상품들을 확인할 수 있기 때문에 쇼핑몰의 매출 등에 미치는 영향이 높지만 온라인 쇼핑 선행 연구에서 BEST 100과 관련된 연구는 거의 이루어지지 않고 있다. 따라서 본 연구에서는 현 온라인 쇼핑몰 11곳을 대상으로 선정하여 쇼핑몰별 판매 특징을 분석하였다. 연구 방법으로 각 쇼핑몰 웹 사이트의 BEST 100의 구성요소인 판매문구, 가격, 무료배송의 유/무 확인을 크롤링 하여 탐색적 자료 분석 기법(EDA)을 활용하였다. 분석 결과 쇼핑몰 11곳의 종합 평균 가격은 72,891.41원으로 나타났으며 상품 가격이 저렴할수록 무료배송 비율이 낮음을 확인하였다. 가격 이외에 판매문구에서는 텍스트 마이닝을 통해 8개의 카테고리로 구분하였다. 가장 많은 카테고리는 fashion 부분이었으나 카테고리의 설정이 제품 속성이 아닌 마케팅 문구를 분석한 점에 의의가 있다. 본 연구는 EDA를 활용하여 현 온라인 시장 흐름을 파악하고 향후 방향을 제시하는데 시사점이 있다.

■ 중심어 : 온라인 쇼핑몰, 인기 순위, 웹 크롤링, 워드 클라우드, 텍스트 마이닝

Abstract

Since the beginning of the first online shopping mall, BEST 100 is being provided as the core of all shopping mall websites. BEST 100 is greatly important because consumers can identify popular products at a glance. However, there are only studies using sales outcome indicators, and prior studies using BEST 100 are insignificant. Therefore, this study selected 11 online shopping malls and compared their main characteristics. As a research method, exploratory data analysis technique (EDA) was used by crawling the BEST 100 components of each shopping mall website, such as product name, price, and free shipping check. As a result, the total average price of 11 shopping malls was 72,891.41 won. Sales texts were classified into 8 categories by text mining. The most common category was the fashion part, but it is significant that the setting of the category analyzed the marketing text, not the product attribute. This study has implications for understanding the current online market flow and suggesting future directions by using EDA.

■ Keyword : Online Store, Ranking, Web Crawl, Word Cloud, Text Mining

I. 서론

온라인 쇼핑의 등장으로 소비자들은 언제 어디서나 제품 및 서비스를 구매할 수 있게 되었다. 또한 소비자들은 온라인에서 풍부한 제품 및 서비스 정보를 받을 수 있게 되면서 오프라인보다 상대적으로 합리적 의사결정을 할 수 있게 되었다. 이와 함께, 편의성을 앞세운 모바일 시장의 발달로 소비자들의 소비패턴이 급격히 변화되었으며 쇼핑 플랫폼의 영역이 점차 세분되는 추세가 되었다. 실제로 12세 이상 인터넷이용자의 모바일 쇼핑 이용률(54.0%)이 컴퓨터(PC)를 통한 이용률(45.4%)보다 높다[21], 그 외에 다양한 결제방식, 빠른 유통체계 등의 발달로 지난 십 년간 온라인 쇼핑 시장은 급속한 성장을 하였으며, 2017년 온라인 쇼핑몰의 규모는 78조 8천억 원으로 2007년 15조 8천억 원 대비 약 5배 성장했다[17]. 또한 온라인 쇼핑의 수요가 많아진 만큼 인터넷 쇼핑몰 유통 채널도 늘어났다. 통계청 자료에서 2016년 전자상거래 사업체 수는 14,573개이며 2007년 2,825개보다 5.1배 증가하였다.

그러나 수많은 유통채널의 등장으로 극심한 경쟁은 불가피하게 되었고, 소비자들은 정보 과부하 현상을 겪고 있다. 한 예로, 같은 상품을 두고 채널별로 차별화된 마케팅 정책을 펼치지만 정작 소비자들은 어떤 채널에서 가장 합리적 구매를 할 수 있는지 고민하는 정보 스트레스를 받고 있다. 이에 대한 방안으로 기업들은 종합 상품 비교 검색, 인공지능 개인상품추천시스템을 개발하고 있으나, 상품 구매 빈도수가 적은 고객이거나 고객의 요구를 잘못 인식하여 정보를 제공할 수 있다는 한계점이 있다.

인터넷 초창기부터 포털사이트는 정보제공 방법의 하나로 이용자(이하 누리꾼)에게 인기 검색어 서비스를 제공해왔다. 포털사이트의 메인화면에 현재 이슈가 되고 있는 검색어를 표시하며 누리꾼들은 실시간으로 관심거리를 확인

할 수 있다. 더 나아가서 누리꾼들은 SNS를 통해 뉴스보다 빠르게 정보를 공유할 수 있게 되었다. 본 연구에서는 국내 종합 온라인 쇼핑몰에서 인기 검색어의 역할을 하는 BEST 100 항목을 분석하고 각 온라인 쇼핑몰 별 특징을 확인하고자 한다. 전통적으로 BEST 순위는 소비자들이 현 추세와 인기 상품을 확인하는 지표이기 때문에, 초기 쇼핑몰이 등장할 때부터 현재까지 온라인 쇼핑몰들은 BEST 순위 제공 서비스를 핵심으로 두며 정보를 제공하고 있다. 따라서 국내 11개의 온라인 쇼핑몰을 대상으로 웹 크롤링한 데이터를 통해 각 쇼핑 채널별 주력 특징을 파악하고 종합적으로 현 온라인 시장에 전략을 제시하고자 한다.

II. 이론적 배경 및 선행연구

2.1 온라인 쇼핑몰 구분

국내 온라인 쇼핑몰에 대한 구분을 살펴보면 박희영[9]은 통계청[17]과 한국온라인쇼핑협회[22]에서 정의한 온라인 쇼핑을 기반으로, “PC 및 모바일을 활용하여 온라인 구매가 이루어지는 온라인 쇼핑 부분, TV를 통해 구매가 이루어지는 TV 부분, 그리고 카탈로그 배포를 통해 전화상으로 구매가 이루어지는 카탈로그를 온라인 커머스 부분”으로 나누었다.

<표 1> 국내 온라인 쇼핑 시장 구분 및 정의¹⁾

구분	세분	정의	데이터 출처
온라인	온라인 쇼핑	인터넷 상에서 발생하는 상거래	통계청[17]
	소셜 커머스	SNS를 기반한 전자상거래	
	오픈마켓	판매자와 구매자에게 자유로운 인터넷 중개물	

1) <표 1>은 박희영[9]이 연구한 자료를 수정 보완한 것임.

구분	세분	정의	데이터 출처
TV	홈쇼핑	상품을 소개하는 방송을 통해 전화 또는 인터넷을 통해 상품을 판매	한국온라인쇼핑협회[17] 방송사업자 재산상황 공표집(2016) 각 홈쇼핑사 공시자료
	TV 커머스	디지털TV를 기반으로 리모컨 등을 활용하여 상품정보 검색, 구매, 결제 등의 상거래를 할 수 있는 서비스	방송사업자 재산상황 공표집(2016)
기타	카탈로그	오프라인 또는 우편으로 배포되는 인쇄물을 통해 상품을 구매	한국온라인쇼핑협회 [22]

2.2 BEST 100 현황

국내 온라인 쇼핑몰의 BEST 100 차트 선정 기준은 회사마다 다르지만, 기본적으로 시간대별 상품의 클릭, 판매량을 기준으로 선정된다. 세부적으로 A 쇼핑몰은 상품의 판매 실적과 상품가격대별 가중치를 주된 기준으로 삼으며, B 쇼핑몰은 최근 2시간 동안의 가격대별 가중치가 반영된 판매 점수와 판매 건수 일별증가점수 합으로 선정한다. BEST 100에 등재되어있는 구성요소로는 전 회사 공통으로 판매문구, 가격, 상품사진, 상품순위, 무료배송의 유/무 등으로 이루어져 있다. 그 외에 쿠폰 할인율, 사은품, 당일 배송의 확인란이 있으며, 쇼핑몰이 별도로 제공하는 서비스에 따라 세부적인 카테고리 인기 순위를 확인할 수 있다.

2.3 선행 연구

온라인 쇼핑몰의 등장 이후 학계에서는 온라인 쇼핑몰이 오늘날 대중화되기까지 다양한 연구를 지속해서 해왔으나 BEST 차트를 기반으로 한 연구는 전무하다. 온라인 쇼핑몰 등장 시에

는 주로 웹 접근성에 대한 연구, 소비자들의 구매 의사결정에 영향에 대한 중점을 두었다. 이정[12]은 초창기 전자상거래의 성공에 대해 시스템 성능, 접근성, 거래에 대한 신뢰성을 강조했다. 장은영[14]은 국내 인터넷 의류 쇼핑몰의 구성요소들을 분석하여 당시 옷은 입어보고 구매 한다는 문화에 맞서, 떠오르는 종합 쇼핑몰의 마케팅 전략 대비 국내 의류 시장의 인터넷 마케팅 전략을 제안하였다. 박민정[7]은 소셜 커머스 쇼핑몰이 생성된 초기 2013년 당시의 불만 사례와 문제점에 대한 과제를 분석하여 소셜 커머스 쇼핑몰의 콘텐츠 구성 및 향후 개선방안을 확인하였다. 온라인 쇼핑몰이 과도기와 들어서면서부터는 쇼핑몰 간의 과도한 경쟁으로 인한 소비자 피해에 대한 연구가 시작되었다. 김소라[4]는 쇼핑몰에서 제공하는 정보의 허위성과 잘못된 정보로부터 파생된 소비자 피해 가능성에 대해 연구를 했다. 당시 소비자보호 이슈와 함께 쇼핑몰의 문제점을 지적하고 개선방안을 제안했다.

그 외에 쇼핑몰별 비교하는 선행연구에서 박화진[8]은 국내 9개의 쇼핑몰의 시각요소를 분석하여 소비자의 긍정적 반응과 부정적 반응을 탐색했다. 손유석 외[10]는 국내 소셜커머스의 4P 전략을 비교 분석하여 온라인 쇼핑몰의 경쟁력 향상을 위한 소셜커머스 기업의 마케팅 전략을 비교 및 제시하였다. 가격 부분에서 임지원 외[13]는 온라인 쇼핑몰 브랜드 평가와 지각된 가격이 소비자의 유통업체 브랜드 태도에 양의 관계를 미치는 것을 연구했다. 또한 류권석 외[5]는 소비자의 구매 의사결정에 영향을 주는 온라인 쇼핑몰의 구성 요인을 분석하고, 제품 유형에 따라 각 요인의 기여도를 구분하여 컨조인트 분석을 통해 살폈다. 이 연구를 통해 소비자들은 제품 유형에 관계없이 가격에 민감한 결과를 얻었으며, 온라인 쇼핑몰에서 소비자들의 구매 요인에서 가격이 가장 중요한 결정 요인으로 꼽힌다는 것을 확인했다.

2.4 선행연구와의 차별성

본 연구에서는 설문조사, 동향, 판매지표를 주로 이용한 선행 연구들과 다르게 쇼핑몰별 BEST 100(인기 순위)을 비교·분석하여 연구하는데 시사점을 두고 있다. 소비자들은 온라인 쇼핑몰에서 상품을 구매할 때 기본적으로 상품사진, 가격 그리고 제품명을 통해 정보를 얻는다. 앞선 연구에서도 장은영[14]은 온라인 쇼핑몰의 정보 제공 방법으로 제품 관련 요소인 제품설명, 제품 가격, 제품 사진, 제품 구색을 분류하였는데, 이와 공통되게 현 온라인 쇼핑몰의 BEST 100에서도 판매문구(제품설명), 제품가격, 제품 사진 등으로 구성되어있다. 이러한 기존 문헌연구 사례를 바탕으로 BEST 100의 구성요소들인 가격, 판매문구, 카테고리 구분, 배송비 유/무 형식을 비교 분석하였다. 특히 카테고리는 수집한 자료가 아닌 전처리를 통해 추가한 자료이다. 수많은 제품 분류를 비교 분석하기에 자료의 한계성이 있기에 판매문구의 속성에 따라 크게 8가지의 카테고리를 분류하여 추가하였다.

본 연구는 기존의 판매 지표를 통한 분석과 달리 탐색적 자료 분석을 목적으로 웹 크롤링과 텍스트 마이닝을 통해 자료를 수집했다. 수집한 자료는 데이터 프레임 형식으로 정리하였다. 본 연구의 특징은 1. 가격대별 분포 비교, 2. 카테고리 분석, 3. 판매문구 텍스트 마이닝, 4. 유료배송과 무료배송의 요인, 5. 향후 온라인 쇼핑몰의

방향 제시를 하는 데 의미를 두고자 한다.

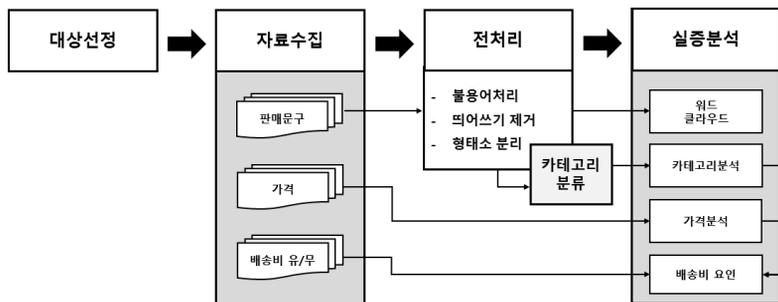
III. 연구 방법

본 연구는 국내 온라인 쇼핑몰에서 공통적으로 제공하는 BEST 100을 비교하기 위해 탐색적 자료 분석을 수행했다. 본 연구 절차는 <그림 1>과 같다. 연구 절차는 먼저 상기한 온라인 쇼핑몰 구분 및 정의를 기반으로 비교 분석할 대상을 선정한다. 두 번째로, 선정된 온라인 쇼핑몰들의 BEST 100 웹 페이지에서 자료를 수집한다. 마지막으로 수집한 자료 중 판매문구 데이터 열을 전처리 작업을 한 후 카테고리 분류를 한다.

실증분석에서는 완성한 데이터 프레임을 이용하여 상황에 따라 표, 그림과 같은 데이터 시각화 작업을 할 것이다. 완성된 시각화 자료를 탐색하여 판매문구 워드 클라우드 분석, 카테고리 분석, 가격 박스분포 분석 등을 수행할 것이다.

3.1 연구대상

본 연구는 상기한 <표 1> 국내 온라인 쇼핑 시장 구분 및 정의 표를 바탕으로 연구대상을 선정하였다. 연구 대상으로는 현재 활발히 운영하고 있는 11개의 국내 온라인 쇼핑몰을 무작위 대상으로 오픈마켓(AUCTION, 11st, Gmarket, SSG,²⁾ INTERPARK), 카탈로그 및 홈쇼핑(Hmall, CJmall,



<그림 1> 연구 절차

2) SSG닷컴의 경우 인기 상품 순위를 99위까지 보여주기에 때문에 100위 값을 NULL 값 처리.

GSSHOP), 소셜커머스(wemakeprice, TMON, G9)로 분류 선택하였다.

3.2 자료수집

11곳의 BEST 100 웹 사이트를 대상으로 자료를 수집하기 위해서 오픈소스인 R 언어를 통해 웹 크롤링(Web Crawling) 코드를 작성했다. 기본적으로 rvest 패키지를 활용하여 각 페이지의 가격, 판매문구, 배송비 유/무 표시가 있는 HTML nodes의 텍스트를 크롤링 하였다. 데이터를 수집하는 과정에서 웹 페이지 특성상 wemakeprice, CJmall, GSSHOP, AUCTION, TMON, 11st의 경우 기존 크롤링 방법이 적용되지 않았는데, 이런 경우 Selenium 패키지를 통해 웹 크롤링을 진행하였다. BEST 100은 등재된 리스트가 수시로 변하기 때문에 웹 크롤링 날짜를 2018년 5월 27일과 2018년 6월 3일로 고정하였으며, 결과적으로 2200개의 행과 3개의 열로 이루어진 데이터 프레임 생성했다.

3.3 데이터 전처리

BEST 100의 구성요소 중 판매문구는 마케팅의 일환으로 다양한 미사여구로 구성된다. 그러므로 전처리 없이 깊이 있는 비교분석에 어려움이 있다. 따라서 수집한 미가공 텍스트를 불용어 처리, 띄어쓰기 제거, 형태소 분리 등 데이터 정제 과정 수행하였다. 본 연구는 데이터 처리 과정에서 오픈소스 R 언어에서 제공하는 KoNLP tm, stringr 패키지를 사용했다. 특히 “?”, “%”, “[”, 숫자 등 특수 문자를 제거하고, 특정 쇼핑몰에서 자주 사용된 마케팅 문구를 제거하였다. 정제된 텍스트들이 포함된 각 행은 gsub함수를 사용하여 8개의 카테고리로 분류하였다. 카테고리 분류 기준은 2017년 통계청 온라인쇼핑 동향을 참조

하며 재구성하였다. 카테고리를 구분하면서 텍스트의 특징에 따라 농·축·수산물+음료+식품 등을 food, 의복+신발+패션 용품 및 액세서리+가방+스포츠 등을 fashion, 가구+사무+문구+애완용품+생활용품 등을 living, 레저+음반+비디오+악기+여행 및 예약서비스 등을 activity, 가전+전자+통신기기 등을 digital, 화장품+미용용품 등을 beauty, 아동+유아용품+아동서적 등을 baby, 나머지는 etc로 구분하였다. 새로운 카테고리 열을 분류하면서 총 수집한 자료는 2,200개의 행과 4개의 열을 가진 데이터 프레임으로 완성되었다.

IV. 실증 분석

4.1 가격 분석

전체 종합결과 11개의 온라인 쇼핑몰(이하 유통채널) BEST 100의 가격 분석 결과를 정리하면 <표 2>와 같다. 종합적으로 2,200개의 상품 데이터들은 표에서와 같이 평균 가격은 72,891.41원이며, NULL 값인 0원(SSG닷컴)을 제외한 최소가격은 100원, 최대 가격은 4,646,920원, 표준편차 243,646원으로 확인된다.

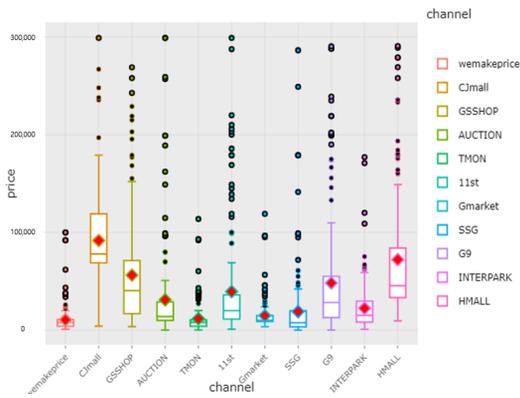
<표 2> 11개 유통채널 BEST 100 가격 분석표

channel	min	max	mean	sd
All	0	4646,920	72,891	243,646
wemakeprice	890	100,000	7,900	14,421
CJmall	4,100	1,340,000	78,900	119,392
GSSHOP	3,610	2,496,000	47,000	312,248
AUCTION	100	3,324,000	15,660	380,221
TMON	190	329,000	8,010	27,312
11st	890	1,728,000	20,250	240,634
Gmarket	3,500	119,000	9,950	14,924
SSG	0(NULL) ³⁾	3,515,484	8,900	295,014
G9	100	2,380,000	29,300	206,656
INTERPARK	800	379,000	15,175	34,912
HMALL	9,520	4,646,920	51,075	426,658

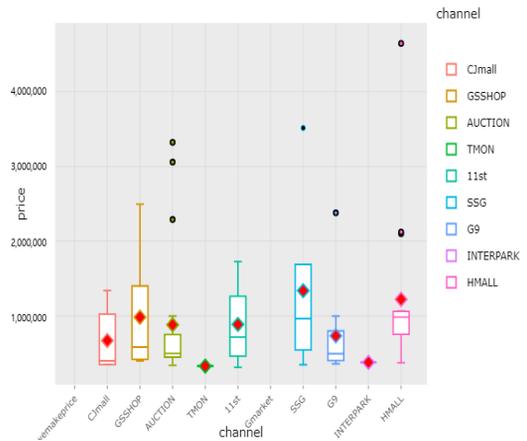
3) SSG닷컴은 NULL 값이 포함되기에 최소 가격이 0으로 처리된다. 두 번째 최소 가격 : 380원(삼다수 500ml).

유통채널별 가격 분석 결과를 박스플롯(<그림 2>, <그림 3>)으로 자세히 살펴보면 30만 원 이하 가격대에서 변동 편차가 심한 채널은 CJmall이며, 변동 편차가 좁은 채널은 Gmarket임을 쉽게 파악할 수 있다. 특히 CJmall은 91,693원의 평균치로 11개 소포물 중 상대적으로 높은 가격 군이 분포되어 있고, Gmarket은 판매 상품의 가격대가 평균 14,866원에 몰려있다.

공백이라 해도 무방하다. 고 가격대에서는 SSG닷컴이 분포가 높음을 알 수 있다. 가장 높은 가격의 제품은 한샘 부엌 리모델링 세트(EURO 9000 리치스페셜) 4,646,920원이다. 박스플롯은 높은 가격 범위의 도표 표현 한계점으로 인해 평균값과 표준편차를 합한 약 30만 원을 기준으로 두 개의 박스플롯으로 나누었다.



<그림 2> 30만 원 이하 가격 박스플롯



<그림 3> 30만 원 이상 가격 박스플롯

30만 원 이상 가격대에서는 wemakeprice과 Gmarket 유통채널이 공백값을 보였으며, INTERPARK와 TMON의 데이터값이 한 개씩 등재되어 있어서

카테고리로 구분하여 가격 평균을 유통채널별로 정리하면 <표 3>과 같다. activity 부분에서

<표 3> 카테고리별 가격평균 상태표

(단위 : 원)

channel	activity	baby	beauty	digital	fashion	food	living	etc
wemakeprice	30,937	14,689	8,130	7,627	6,512	9,938	9,830	15,756
CJmall	109,845	332,000	93,304	340,633	90,992	60,612	88,990	134,531
GSSHOP	65,721	482,56	77,587	418,326	85,623	67,578	146,079	84,481
AUCTION	93,388	26,464	111,484	413,957	35,415	21,798	98,087	243,450
TMON	31,217	17,698	6,875	9,425	8,050	10,018	11,810	11,979
11st	183,961	23,133	26,550	345,990	19,806	25,688	35,382	33,130
Gmarket	9,323	11,512	14,270	47,022	11,729	15,353	18,830	12,400
SSG	17,358	23,418	31,900	923,169	61,896	8,506	11,512	49,035
G9	54,366	74,776	20,223	276,236	32,359	26,641	42,500	40,287
INTERPARK	19,156	39,790	18,263	37,741	19,688	18,999	26,370	12,912
HMALL	210,838	151,732	52,405	415,290	54,954	88,446	365,392	924,220

Gmarket이 평균가격이 제일 낮으며 Hmall이 가장 높다. baby 부분에서 Gmarket이 평균가가 낮고 CJmall이 가장 높다. beauty 부분에서 TMON이 평균가가 가장 낮으며 AUCTION이 높다. digital 부분에서 wemakeprice의 평균가가 가장 낮으며 SSG가 높다. food 부분에서는 SSG닷컴이 가장 낮으며 Hmall이 높다. fashion 부분에서는 wemakeprice와 TMON이 비슷하게 평균가가 적으며 CJmall이 평균가가 높다. living 부분에서 역시 wemakeprice의 평균가가 가장 낮으며 Hmall의 가격이 높다.

순이다.



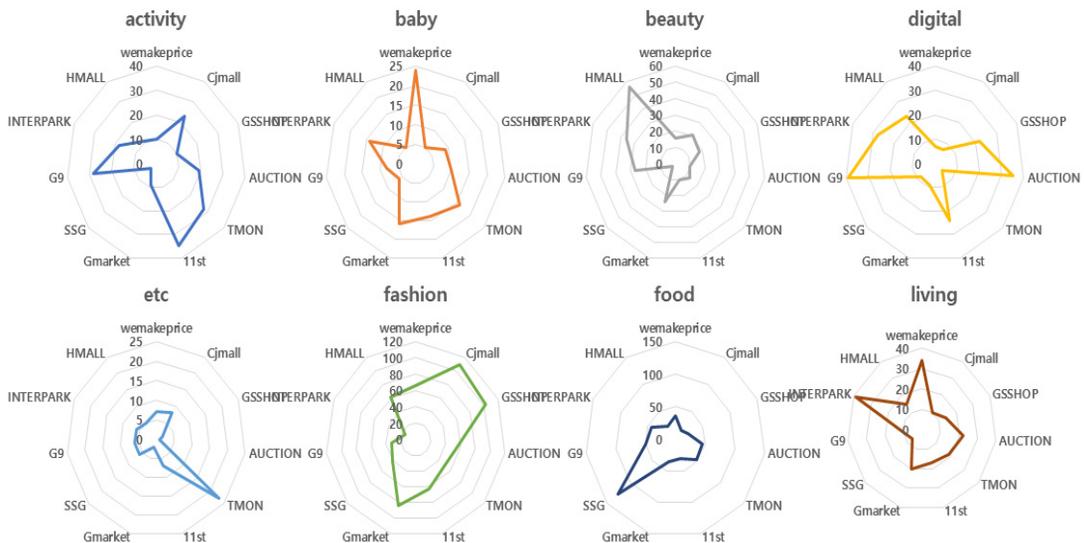
〈그림 4〉 카테고리 분포 막대그래프

4.2 카테고리 분석

11개 유통채널의 2,200개의 상품 데이터의 카테고리 분포는 <그림 4>와 같다. 전체 제품 중에서 높은 등재 비율을 보이는 카테고리는 fashion 이다. 이는 2017년 한국인터넷진흥원에서 실시한 인터넷이용실태조사에서 인터넷쇼핑 이용자의 주 구매 품목 1위인 의류, 신발, 스포츠용품, 액세서리(83.2%)와 상응한다. 그 뒤로 food, beauty, digital, living, activity, etc, baby 카테고리

유통채널별로 카테고리 분포를 구분하였을 때에는 <그림 5> 결과와 같다. activity 카테고리 판매량이 많은 쇼핑몰은 11번가이다. 또한 baby 카테고리는 wemakeprice, beauty 카테고리는 HMALL, digital 카테고리는 G9, fashion 카테고리는 CJmall, food 카테고리는 SSG가 압도적으로 높으며 living 카테고리는 INTERPARK가 높다.

가장 다양한 종류의 제품을 보유하고 있는 유통채널을 확인하기 위해 각 유통채널별 카테고리



〈그림 5〉 유통채널별 카테고리 분포 방사형 그래프

분포 표준편차를 구했다. 그 결과 INTERPARK가 13.3844로 가장 작은 표준편차임을 <표 4>에서 확인할 수 있는데, 이는 INTERPARK가 판매하고 있는 제품들의 종류가 상대적으로 고르게 분포되어있음을 의미한다. 반대로 SSG닷컴은 표준편차가 42.7250로 가장 크게 나타났다. 이는 food 카테고리의 보유량이 압도적으로 높기 때문에 높은 표준편차 형성에 영향을 주었다고 해석된다.

<표 4> 11개 유통채널 카테고리별 보유량 표준편차

INTER-PARK	G9	TMON	11st	wemake price
13.3844	15.3063	15.4550	17.6068	18.3458
AUCTION	HMALL	Gmarket	GSSHOP	CJmall
19.2576	21.4276	26.0494	30.7849	33.8652
SSG	-	-	-	-
42.7250	-	-	-	-

4.3 워드 클라우드 분석

워드 클라우드를 통한 결과는 <그림 6>, <그림 7>과 같다. 단어들을 나열할 때 평균값과 표준편차를 합한 약 30만 원을 기준으로 두 개의 워드 클라우드를 표시했다. 워드 클라우드 분석을 통해 fashion, beauty, food 카테고리 관련 키워드들이 30만 원 이하 가격대에 많이 분포되었다. 키워드는 주로 아디다스, 나이키 같은 스포츠 패션브랜드와 나스, 니베아 같은 화장품 브랜드 그리고 농심, 슈퍼, 칼라만시 등의 food 카테고리가 전체 유통채널 BEST 100에 자주 사용됨을 확인하였다. 30만 원 이상 워드 클라우드에서는 경험상 예상했던 것처럼 digital 카테고리 제품군 관련 키워드들이 대부분이었다. 대체로 고가격 대의 백색가전이 눈에 띄었으며 키워드로 다이슨, 휘센, 위니아 등이 있음을 확인하였다.



<그림 6> 30만 원 이하 워드 클라우드 <그림 7> 30만 원 이상 워드 클라우드

가격 구분 없이 전체적으로 워드 클라우드를 그려보면 쿠폰이 가장 많이 사용됨을 <그림 8>과 <표 5>를 통해 확인하였는데, 이는 마케팅 문구로 쿠폰이라는 키워드가 가장 소비자들에게 효과적이라고 해석된다. 그 외의 주로 사용된 마케팅 문구에서는 만 원의 행복, 국내산, 등급, 셀럽샵, 특가 등의 할인이나 신뢰를 의미하는 단어 선택이 있음을 워드 클라우드를 통해 확인할 수 있다. 브랜드 명에서 농심, 아디다스, 롯데, 정관장, 노브랜드, 입생로랑, 베라왕, 나스, 나이키 등이 주로 사용된다. 특히 영부인이 입은 저렴한 브랜드로 인기를 올리고 있는 베라왕, 최근 저렴한 가격에 고가의 청소기 브랜드 수준의 기능을 보이고 있는 차이슨이 눈에 띈다. 그 외 슈퍼 등 다양한 생활필수품이 가장 많이 판매되고 있음을 확인할 수 있다.



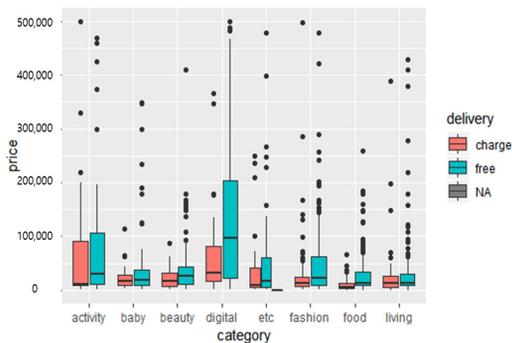
<그림 8> 전체 판매문구 워드 클라우드

〈표 5〉 전체 판매문구 키워드 순서표

순위	단어	빈도	카테고리
1	쿠폰	35	all
2	노브랜드	16	food, living
3	슈퍼	14	food, living
4	국내산	13	food
5	베라왕	13	fashion
6	아디다스	13	fashion
7	엣지	13	fashion
8	정관장	11	food
9	나스	10	beauty
10	농심	10	food
11	롯데	10	food
12	입생로랑	9	beauty
13	나이키	8	fashion

4.4 무료배송과 유료배송의 요인

〈그림 9〉에서 배송비와 제품의 종류에 따라 무료배송의 여부를 박스플롯으로 시각화하였다. 전체적으로 유료배송은 낮은 가격대에 분포되어 있으나 무료배송은 고루 분포되어 있다. 박스플롯을 통해 activity의 경우 가격이 낮아도 무료배송과 유료배송의 가격분포의 차이가 적음을 확인할 수 있다. 반대로 food, fashion, digital인 경우 저가일수록 배송비가 청구됨을 확인하였다.



〈그림 9〉 카테고리별 가격별 배송비 유무차이 박스플롯

V. 결론

5.1 분석 결과

본 연구에 사용된 웹 크롤링(Crawling)은 분산되어있는 대용량의 데이터를 수집하여 색인 작업을 하거나 검색엔진에 사용되는 기법이다. 자동으로 정보를 수집하기 때문에 주로 웹사이트의 텍스트, 그림, 멀티미디어 등의 정보를 얻는데 유용하게 사용된다. 웹 크롤링 기법을 통해 수집한 11곳의 유통채널(이하 온라인 쇼핑몰) 웹사이트의 텍스트를 통해, 일일이 홈페이지를 확인하여 텍스트를 타이핑하는 것보다 경제적으로 자료 수집을 할 수 있었다.

본 연구에서 30만 원 이하에서 wemakeprice가 평균가 10,524.55원 기준으로 가장 저렴한 제품들을 판매하고 있으며, 91,693.18원의 평균가를 보인 CJmall이 가장 높은 순위를 차지하였다. 특히 wemakeprice는 living, fashion, digital 부분에서 가장 평균가가 낮게 나오면서 저렴한 상품들이 타 유통채널에 비해 많이 판매되고 있음을 예측했다. 30만 원 이상에서는 digital 부분이 압도적으로 많았는데 워드 클라우드를 통해 D사 청소기, 여름대비 에어컨, 냉장고 등의 제품들을 가장 많이 찾고 있음을 확인할 수 있었다. 또한 가전제품의 평균가가 높은 곳은 923,169.5원의 SSG닷컴이 있는데 이는 식품을 주력으로 보이면서 digital 카테고리 품목 구성 부족이라 판단했다. 마지막으로 배송비 관련해서 상품의 판매 가격이 낮을수록 유료배송의 횟수가 높았음을 확인하였다.

본 연구를 통해 얻은 온라인 쇼핑몰별 인상적인 특징을 파악하자면 SSG닷컴은 food의 평균가 8,506원으로 가장 저렴하면서 카테고리 분포도 다른 쇼핑몰에 비해 food의 보유량이 가장 많았다. 또한 전체 평균가인 72,891.41원 이하 판매문구 워드 클라우드에서도 노브랜드, 이마트 등의 키워드가 나왔다. 이는 SSG닷컴이 이마트

PB상품을 통해 식품을 주력해서 판매하고 있음을 알 수 있다. 그 외 CJmall은 fashion 평균가가 타 쇼핑몰 대비 가장 높지만 fashion 보유량도 1위이며, wemakeprice는 living 평균가가 타 쇼핑몰 대비 최저이지만 living 보유량은 2위이다. 마지막으로 Gmarket은 baby 평균가가 가장 적는데 baby 보유량이 2위이다. 상기한 온라인 쇼핑몰과 카테고리의 관계는 보유량과 가격에 의해서로 강점이라 판단한 것이다.

또한 카테고리의 표준편차가 적은 INTERPARK, G9를 통해 다른 쇼핑몰에 비해 여러 가지 상품을 골고루 판매하고 있음을 볼 수 있었다. 카테고리 분포의 표준편차가 큰 상위 5개 HMALL, Gmarket, GSSHOP, CJmall, SSG의 경우 SSG를 제외하고 fashion 카테고리가 강세를 보이는데 이 중에서 절반 이상이 의류, 뷰티 홈쇼핑 채널에 기반임을 알 수 있었다. 반대로 보완해야 할 쇼핑몰을 파악하자면 wemakeprice는 digital의 평균가가 타 쇼핑몰보다 가장 낮으면서 digital 카테고리 점유율도 가장 낮다. AUCTION은 beauty의 평균가가 가장 높지만 beauty 점유율도 최하위이다. <표 6>은 본 연구를 통해 파악한 인상적인 특징이 나타나는 쇼핑몰을 골라서 SWOT을 판단한 표이다. 나머지 쇼핑몰은 카테고리별 표준편차가 작거나 보통의 경우이다.

<표 6> 쇼핑몰별 특징파악

온라인 쇼핑몰	카테고리	평균 가격	보유량	SWOT ⁴⁾
SSG닷컴	food	최저	최다	S
CJmall	fashion	최고	최다	S
wemakeprice	living	최저	2위	S
	digital	최저	최소	W
Gmarket	baby	최저	2위	S
INTERPARK	ALL	-	low deviation	-
G9				
AUCTION	beauty	최고	최하	W

4) SWOT : 강점(strength), 약점(weakness), 기회(opportunity), 위협(threat) 요인.

5) You Only Live One의 약자로 현재의 행복에 중심을 두고 사는 사람들을 지칭.

6) 소소하지만 확실한 행복의 약자.

5.2 제안

상기한 분석들을 통해 쇼핑몰 별 주력상품 · 유통체계가 다름을 확인하였고 이에 따라 유통채널별 경쟁력에 차이가 있음을 판단하였다. 이에 따라 본 연구 결과를 통해 몇 가지 제안을 제시하고자 한다.

SSG닷컴의 경우 기존 오프라인 유통채널의 장점을 살려 식품에 중점을 두었기에 가장 저렴한 가격에 다양한 food 카테고리 제품을 판매할 수 있음을 파악했다. 하지만 living 카테고리에서 현백화점 · 마트의 판매율 상승에 따른 인터넷 시장 확대의 잠재력이 있으나 타 쇼핑몰보다 최저의 보유량을 보인다. 워드 클라우드에서 자주 보이는 노브랜드 PB 상품을 지표 삼아 생활용품 품목의 마케팅 증진방안을 모색하는 것을 제안한다.

activity의 경우 보유량 1위인 11st의 평균 가격은 183,961원이며 2위의 보유량을 가진 g9의 평균가격은 54,366으로 약 3배 차이를 보인다. 지혜경[16]은 오픈마켓에서 충동구매에 대한 원인을 가격, 제품의 새로움 순으로 영향력 정도를 분석했다. 현재 여행, 율로족⁵⁾, 소확행⁶⁾등의 트렌드 추세로 activity의 판매 상승을 예상하며 가격에 민감한 온라인 쇼핑몰 소비자의 성향에 따라 11st에게 상대적으로 저렴한 문화, 체험 쿠폰 등의 판매 확대로 평균 가격을 낮추는 방안을 제안한다.

마지막으로 본 연구가 웹 크롤링을 통해 온라인 쇼핑몰 BEST 100을 연구하는 첫 번째 발자취로서 향후 온라인 쇼핑몰 경향 추세 연구에 도움이 되기를 희망한다. 추가적인 연구로는 주기적인 자료 수집을 통해 시계열 나열 비교 및 계절별 경향 파악 연구 가능성을 제시하고자 한다. 덧붙여서 본 연구의 탐색 대상이었던 온라인 쇼핑몰의 판매문구, 가격, 카테고리 분석을 통해

실질적인 탐색 연구를 할 수 있는 밑거름이 되어 추후 각 유통기업, 공공기관, 연구원 등에서 실질적인 최신 동향 리포트 및 통계 연구를 제안한다.

5.3 의의

본 연구는 각 제품의 카테고리 분류 과정에서 판매문구에 따라 카테고리 키워드 겹침 현상이 있었다. 예를 들어 “클리너”라는 문구는 바닥을 청소(living)를 의미하거나 화장(beauty)을 지우는 의미로 사용된다. 이에 따라 카테고리의 분류가 제품의 특성이 아닌 마케팅 문구에서 가져온 한계점이 있다. 또한 본 연구는 국내 온라인 쇼핑몰 11곳만을 대상으로 선정하여 정해진 기간 내에 분석했다는 점이다. 따라서 본 연구결과를 보편적으로 해석하기에 의의가 있다.

참 고 문 헌

- [1] 김경재, 김병국, “데이터 마이닝을 이용한 온라인 쇼핑몰 상품추천시스템”, 지능정보연구, 제11권, 제1호, pp.191-20, 2005.
- [2] 김경희, “온라인 쇼핑몰 이미지 포지셔닝 연구”, 한국콘텐츠학회논문지, 제8권, 제1호, pp.48-58, 2008.
- [3] 김도형, 강병호, 이승룡, “인기 검색어의 순위 변화 예측”, 정보과학회논문지, 제44권, 제8호, pp.782-790, 2017.
- [4] 김소라, “온라인 쇼핑몰 사이트에서의 정보성과 소비자 오도가능상”, 한국가정관리학회지, 제26권, 제1호, pp.15-29, 2008.
- [5] 류권석, 김성수, “온라인 쇼핑몰 구매의도에 영향을 주는 요인 : 평점분포를 고려한 컨조인트 분석을 중심으로”, 경영교육연구, 제31권, 제6호, pp.43-71, 2016.
- [6] 박경자, 장희영, 노희옥, “온라인 정보원천이 온라인 쇼핑몰 소비자만족에 미치는 영향”, 정보시스템연구, 제20권, 제1호, pp.21-40, 2011.
- [7] 박민정, “소셜커머스 웹사이트의 정보구성요소 : 상품, 가격, 서비스, 판매원 정보를 중심으로”, 한국콘텐츠학회논문지, 제13권, 제9호, pp.369-381, 2013.
- [8] 박화진, “국내 온라인 쇼핑몰 사이트의 색채분석”, 한국상품문화디자인학회논문집, 제39호, pp.71-85, 2014.
- [9] 박희영, “우리나라 온라인 커머스 시장 동향”, 정보통신방송정책, 제29권, 제15호, pp.26-45, 2017.
- [10] 손유석, 김창수, 안현숙, “국내 주요 소셜커머스 기업의 마케팅 전략에 관한 연구”, 한국정보기술학회논문지, 제12권, 제12호, pp.181-194, 2014.
- [11] 이영미, ““소확행(小確幸)’의 힘”, 새가정, 제65권, pp.68-69, 2018.
- [12] 이정, 이상설, “AHP를 이용한 온라인 쇼핑몰 선택에 대한 연구”, 한국산업경영시스템학회지, 제28권, 제1호, pp.16-23, 2005.
- [13] 임지원, 장기원, 이라경, 박진용, “소매업체 브랜드 자산이 온라인 쇼핑몰 유통업체 브랜드 태도에 미치는 영향”, 한국유통물류정책학회, 제4권, 제1호, pp.37-60, 2017.
- [14] 장은영, “국내 인터넷 의류 쇼핑몰의 유형별 사이트 구성요소 비교분석”, 복식, 제55권, 제6호, pp.76-92, 2005.
- [15] 장평, 황종택, 문희철, “중국 3대 C2C 오픈마켓의 온라인 쇼핑성과에 관한 비교연구”, e-비즈니스연구, 제15권, 제4호, pp.383-404, 2014.
- [16] 지혜경, “온라인 쇼핑몰 유형별 의류제품 충동 구매에 미치는 가격, 제품, 판촉의 영향”, 한국 의류산업학회지, 제15권, 제4호, pp.543-553, 2013.
- [17] 통계청, 온라인쇼핑몰 취급상품 범위별 거래액, 2017.
- [18] 통계청, 전자상거래 사업체수, 2017.
- [19] 통계청, 2017년 2월 온라인쇼핑 동향, 2017.

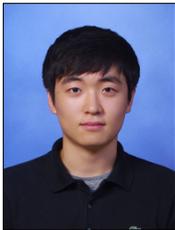
- [20] 통계청, 주요유통업체매출동향조사 : 대형마트·백화점 매출 동향(품목별), 2017.
- [21] 한국인터넷진흥원, 인터넷이용실태조사, 2017.
- [22] 한국온라인쇼핑협회, 2016 온라인쇼핑 시장에 대한 이해와 전망, 2017.
- [23] 한상일, “온라인 쇼핑몰 인터페이스 구성 항목에 관한 연구”, 사회과학논총, 제23집, pp.263-278, 2007.



강 주 영(Juyoung Kang)

- 1995년 : 포항공과대학교 컴퓨터공학과 (공학사)
- 1997년 : 서울대학교 컴퓨터공학 (석사)
- 2005년 : 한국과학기술원 경영공학 (박사)
- 2005년~현재 : 아주대학교 경영대학 e-비즈니스 학과 교수
- 관심분야 : 텍스트 마이닝, 빅데이터 분석, 클라우드 컴퓨팅, 지능형 전자상거래

저 자 소 개



강 지 천(Jicheon Kang)

- 2014년~현재 : 아주대학교 경영대학 경영학과 재학
- 관심분야 : 빅데이터, 텍스트 마이닝, ICT, 유통