



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위 논문

온라인 리뷰의 양, 방향성, 감성지수가
매출에 미치는 영향력 분석
브랜드 명성의 조절효과를 중심으로

The Impact of Online Review
Volume, Valence, and Sentiment
Score on sales
Focusing on Moderating Effect of Brand Strength

2019년 12월

승실대학교 대학원

IT유통물류학과

김 현 아

석사학위 논문

온라인 리뷰의 양, 방향성, 감성지수가
매출에 미치는 영향력 분석
브랜드 명성의 조절효과를 중심으로

The Impact of Online Review
Volume, Valence, and Sentiment
Score on sales
Focusing on Moderating Effect of Brand Strength

2019년 12월

승실대학교 대학원

IT유통물류학과

김 현 아

석사학위 논문

온라인 리뷰의 양, 방향성, 감성지수가
매출에 미치는 영향력 분석
브랜드 명성의 조절효과를 중심으로

지도교수 최 자 영

이 논문을 석사학위 논문으로 제출함

2019년 12월

숭실대학교 대학원

IT유통물류학과

김 현 아

김 현 아 의 석 사 학 위 논 문 을 인 준 함

심 사 위 원 장 안 승 호 인

심 사 위 원 최 자 영 인

심 사 위 원 전 홍 식 인

2019년 12월

승실대학교 대학원

목 차

| | |
|-----------------------------------|-----|
| 국문초록 | v |
| 영문초록 | vii |
| | |
| 제 1 장 서론 | 1 |
| 1.1 연구 배경 및 필요성 | 1 |
| 1.2 연구 방법 및 목적 | 3 |
| 1.3 연구 구성 | 4 |
| | |
| 제 2 장 이론적 배경 및 가설 설정 | 5 |
| 2.1 온라인 리뷰 | 5 |
| 2.1.1 리뷰의 양 | 6 |
| 2.1.2 리뷰의 방향성 | 7 |
| 2.1.3 리뷰의 유용성 | 9 |
| 2.1.4 텍스트 감성 분석 | 10 |
| 2.2 브랜드 명성 (Brand Strength) | 13 |
| | |
| 제 3 장 연구 방법론 | 15 |
| 3.1 분석 프로세스 | 15 |
| 3.1.1 데이터 크롤링 | 15 |
| 3.1.2 리뷰 데이터 정제 | 20 |
| 3.1.3 감성사전 선정 | 22 |
| 3.1.4 감성 분석 | 24 |

| | |
|--|----|
| 3.2 회귀분석 모델 | 26 |
| 3.2.1 매출 | 28 |
| 3.2.2 리뷰의 내용 및 제목의 감성지수(sentiment score) | 29 |
| 3.2.3 브랜드 명성(Brand Strength) | 29 |
| 제 4 장 결과 분석 | 30 |
| 4.1 기술통계 | 30 |
| 4.2 가설 검증 | 34 |
| 제 5 장 결론 | 41 |
| 5.1 연구결과 요약 및 시사점 | 41 |
| 5.2 한계점 및 향후 연구 방향 | 44 |
| 참고문헌 | 46 |
| 부 록 | 54 |

표 목 차

| | |
|--|----|
| [표 3-1] 아마존의 블루투스 스피커 베스트셀러 제품 정보 예시 | 19 |
| [표 3-2] 아마존 블루투스 스피커 베스트셀러 리뷰 정보 예시 | 19 |
| [표 3-3] 감성사전의 긍정적인 단어와 부정적인 단어의 예시 | 22 |
| [표 3-4] 수정 단계를 거친 단어 | 23 |
| [표 4-1] 기술 통계 분석 | 30 |
| [표 4-2] 회귀 분석 결과 | 35 |

그 립 목 차

| | |
|---|----|
| [그림 3-1] 분석 프로세스 | 15 |
| [그림 3-2] 크롤링 프로세스 | 15 |
| [그림 3-3] 아마존 리뷰 페이지와 수집한 데이터의 예시 | 18 |
| [그림 3-4] 리뷰의 helpfulness의 수가 나타난 아마존 리뷰 페이지 | 19 |
| [그림 3-5] 부정어가 포함된 단어의 예시 | 21 |
| [그림 3-6] 한 단어로 된 리뷰가 분할되지 않은 예와 이 단어들을 word2 에 할당하여 전처리한 예 | 22 |
| [그림 3-7] 아마존 블루투스 스피커 리뷰의 예시 | 23 |
| [그림 3-8] 분할된 단어와 감성분석 결과 | 25 |
| [그림 4-1] price와 rating의 분포 그래프 | 31 |
| [그림 4-2] 리뷰의 양과 평점의 산점도 | 32 |
| [그림 4-3] 제목의 긍정점수와 부정점수 분포도 | 33 |

국문초록

온라인 리뷰의 양, 방향성, 감성지수가
매출에 미치는 영향력 분석
브랜드 명성의 조절효과를 중심으로

김 현 아

IT유통물류학과

송실대학교 대학원

온라인 리뷰의 중요성은 온라인 전자 상거래가 활성화 된 시기부터 인식되기 시작하였다. 제품이나 서비스를 구매한 후 작성한 리뷰는 많은 정보를 담고 있기 때문에 소비자가 제품을 구매하기까지의 의사결정에 많은 영향을 끼친다. 이러한 이유에서 기업과 연구자들은 온라인 리뷰의 효과와 영향력에 대해 많은 관심을 가지고 연구를 진행해왔다.

과거에는 온라인 리뷰의 별점과 양 등 정량적 정보에 대한 연구가 진행되었는데, 보통 전자상거래 웹사이트에서 쉽게 얻을 수 있는 정보들인 리뷰의 개수, 별점 등을 이용하여 리뷰의 영향력을 파악하였다면 현재에는 텍스트 분석을 통한 리뷰의 감성을 인식하여 심도 있는 정성적 정보에 대한 분석이 진행되고 있다. 따라서 본 연구에서는 정량적 정보뿐만 아니라 온라인 리뷰 텍스트 데이터를 분석하여 감성 정도를 파악하고 매출에 미치는 영향에 대해 살펴보았다.

데이터 수집은 2019년 8월 13일부터 9월 7일까지 약 4주 동안 이루어

졌으며 통계 프로그램인 R을 이용하여 아마존 블루투스 스피커 베스트 셀러 제품들을 크롤링(crawling)을 수행하였다. 그 이후 리뷰 데이터를 수집하여 감성분석을 실시하였으며 이를 통해 리뷰의 긍정/부정 정도(감성 지수)가 매출에 미치는 영향에 대해 분석하였다.

먼저 정량적 정보에 해당하는 리뷰의 양, 별점(rating)과 유용성의 영향력을 조사하기 위하여 회귀 분석을 수행하였다. 그 결과 리뷰의 양, 별점 그리고 유용성(helpfulness)이 높을수록 매출에 긍정적인 영향을 미친다는 결과를 확인하였으며, 매출에 가장 큰 영향을 미치는 요인은 리뷰의 양인 것으로 나타났다.

그 다음 리뷰 데이터를 제목과 내용으로 분류하여 텍스트 분석을 통해 나온 결과를 바탕으로 회귀분석을 실시하였다. 그 결과, 리뷰의 내용보다 제목에서 감지된 감성지수의 영향력이 더 큰 것으로 나타났고, 리뷰의 제목에서 긍정(부정)적인 단어가 많이 쓰일수록 매출에 긍정(부정)적인 영향을 미친 반면 리뷰의 내용에서는 부정적인 단어만이 매출에 부정적 영향을 미친 것으로 드러났다.

마지막으로 브랜드 명성과 리뷰의 양, 리뷰의 감성지수의 상호작용 살펴보기 위해 조절회귀분석을 실시하였다. 그 결과 브랜드 명성이 낮은 경우, 높은 경우보다 리뷰의 양이 많을수록 매출에 미치는 영향력이 큰 것으로 나타났고, 리뷰의 제목과 내용에서 나타난 긍정(부정) 점수가 매출에 긍정(부정)적인 영향을 미치는 것으로 밝혀졌다. 이를 통해 브랜드 명성이 낮을수록 리뷰의 영향력이 더 중요하다는 것을 알 수 있었다.

ABSTRACT

The impact of Online review volume, valence and sentiment score on sales:

focusing on Moderating effect of brand strength

KIM HYEON A

Department of IT distribution & logistics

Graduate School of Soongsil University

The importance of online reviews has long been recognized because it influences consumers' decisions to purchase a product. This affects the interest of companies and researchers about the impact of online reviews.

In the past, research on quantitative information such as the rating and quantity of online reviews has been conducted. At present, analysis of in-depth qualitative information, such as identifying the sentiment of reviews through text analysis, is in progress.

Therefore, in this study, not only the quantitative information but also the online review text data was analyzed to identify the degree of sentiment and effect on sales.

Data collection took place for about four weeks, from August 13 to September 7, 2019, using a statistical program R to crawl Amazon Bluetooth speaker best-selling products. After that, review data was collected and sentiment analysis was conducted.

First, regression analysis was performed to investigate the influence of the quantity, rating, and usefulness of the review as quantitative information. As a result, it was confirmed that the higher the quantity, star rating, and helpfulness of the reviews had a positive effect on sales, and the factor that most influences the sales was the quantity of reviews.

Then, the review data was classified into titles and contents, and regression analysis was performed based on the results from text analysis. As a result, the sentiment scores detected in the title are more influential than the contents of the review, and the more positive(negative) words are used in the title of the review, the more positively(negatively) the sales result, while the contents of the review are only negative words turned out to have a negative impact on sales.

Finally, if the brand strength is weak, the larger the volume of reviews, the greater the impact on sales, and the positive (negative) scores shown in the title and content of the reviews have a positive effect on sales.

제 1 장 서 론

1.1 연구 배경 및 필요성

소비자는 더 이상 기업이 주는 광고 메시지에 의존하지 않고 온라인 리뷰와 같은 온라인 구전에 더 관심을 두고 있다. CMA(Competition & Markets Authority)가 2015년에 조사한 설문에 따르면, 소비자의 50% 이상이 온라인 리뷰를 참조한다는 결과가 나타났다. 대부분의 소비자는 기존 광고 형식보다 다른 사람들의 추천을 더 신뢰하는 경향이 있기 때문이다(Nielsen 2012).

온라인 리뷰는 소비자가 물건을 구매하거나, 영화를 볼 것인지를 결정하거나, 스포츠클럽과 같은 취미활동의 여부까지도 결정하도록 도와준다. 온라인 리뷰는 점점 마케팅의 주요 추진력이 되고 있으며(Cui, Lui and Guo 2012), 많은 웹사이트에서 사용하는 요소가 되었다. 온라인으로 제품을 구매하는 소비자들은 제품을 직접 시험해볼 기회가 없기 때문에 소비자가 리뷰를 신뢰하고 의존하는 경향이 크다. Willemsen, Neijens, Bronner and Ridder (2011)에 따르면, 온라인 리뷰와 같은 다른 소비자로부터 얻은 정보는 더 설득력 있고, 신뢰할 만 할 가치가 있고 온라인 환경에서 중요한 요소이다. 이와 같은 이유로 많은 연구자들과 기업이 온라인 리뷰가 가진 설득력과 파급력에 대해 관심을 가지고 꾸준히 연구하고 있다.

기존 연구들은 대체로 온라인 리뷰의 정량적인 정보에 초점을 두고 진행된 반면 정성적 정보에 대한 분석은 많이 이루어지지 않은 상태이다. 예를 들어, 리뷰의 양은 소비자의 매출과 수익에 긍정적인 상관관계가 있다고 알려져 있으며, 이에 대한 많은 선행 연구가 진행되었고(Bazzar

2015; Matfield 2011), 온라인 리뷰에서 높은 별점은 도서 판매량과 영화 박스오피스 매출에 긍정적인 영향을 미친다는 연구 결과가 있었다 (Chevalier and Mayzlin 2006; Chintagunta, Gopinath and Venkataraman 2010).

그러나 리뷰의 양과 별점에 관한 연구에 비해 리뷰 내용에 대한 선행 연구는 다소 적게 이루어진 상태이다. 온라인 리뷰의 직관적 정보 또한 매출에 중요한 요소로 작용하지만, 오늘날 다양해진 소비자의 니즈를 분석하기 위해 정량적인 정보보다 리뷰 텍스트와 같은 정성적 정보 분석의 필요성이 대두되었다. 리뷰 텍스트는 정성적 정보에 해당하는 비정형 데이터로, 숫자로 이루어진 정량적 데이터와 달리 분석이 어렵다는 특징을 가지고 있다. 그러나 최근 정성적 데이터의 중요성이 인식되면서 이에 따른 연구도 활발히 진행되고 있는데 대표적인 연구로는 텍스트 분석을 통해 리뷰의 내용에서 감지된 감성의 정도가 매출에 직접적인 영향을 준다는 Hu and Reddy (2014)의 연구가 있다. 또한 현재 국내에서도 텍스트 마이닝 연구가 활발히 진행되고 있으며, 대체로 단어빈도 분석을 통해 감성사전을 구축하고 텍스트 분석을 진행하는 방향으로 연구되고 있다(심영석·김홍범 2016; 김도경·김인신 2017; 임영희 2019).

따라서 본 논문에서는 리뷰 텍스트의 중요성을 인식하고 이에 초점을 맞추어 리뷰에서 나타난 감성의 정도가 매출에 미치는 영향을 살펴보았다. 또한 리뷰와 다른 요소들 간의 상호작용 또한 살펴보았다.

1.2 연구 방법 및 목적

본 연구는 온라인 리뷰의 중요성을 인식하고 매출에 미치는 영향력을 파악하기 위하여 텍스트 분석을 진행하였다. 기존의 연구에서는 제품 및 리뷰의 정량적 정보(가격, 별점 등)에 집중했다면, 본 논문에서는 정량적 정보뿐만 아니라 정성적 정보(리뷰 텍스트)가 매출에 미치는 영향에 초점을 맞추었다.

리뷰에서 나타난 감성의 정도를 파악하기 위하여 2019년 8월 14일부터 9월 7일 동안 총 8번의 자료 수집과정을 거쳤으며 아마존에서 판매되는 블루투스 스피커를 제품으로 분석하였다. 블루투스 스피커는 탐색재의 일종으로, 리뷰의 영향력을 상대적으로 많이 받는 경험재와 달리 정보 탐색으로 의사결정을 할 수 있는데, 본 논문에서는 탐색재를 통해 소비자가 다양한 리뷰의 정보 형태가 어떻게 영향을 받는지에 대해 연구를 진행하였다. 또한 종속변수인 매출을 추정하기 위해 아마존 판매 순위 1위부터 100위까지의 제품을 수집하여 연구를 진행하였다.

연구 진행방법은 다음과 같다. R을 이용하여 판매 순위와 제품 정보 및 리뷰를 크롤링한 뒤, 텍스트를 정제하는 과정을 거쳐 감성분석을 진행하였는데, 감성 분석에는 Liu가 제작한 ‘Opinion Lexicon’이 사용되었으며 제품 특성에 따라 단어를 수정하여 진행하였다. 감성 분석 프로세스를 통해 나온 감성지수와 제품의 정보들을 통합한 다음, SPSS를 이용해 조절회귀분석을 수행하였다.

1.3 연구 구성

본 연구는 다음과 같이 구성된다. 2장에서는 연구에 필요한 온라인 구전과 감성분석에 대한 이론적 배경을 제시하였다. 온라인 구전에는 리뷰의 양과 리뷰의 유용성에 대해 설명하고 감성분석에서는 리뷰의 방향성과 텍스트 분석에 대한 선행연구를 제시하며 향후 연구의 가설을 설정하였다.

3장에서는 연구 방법론에 대해 설명하였다. 첫 번째로 분석 프로세스에 대해 설명하는데, 데이터 수집 방법과 데이터 정제, 감성 사전 선정에 대해 설명하고 감성 분석의 과정에 대해 설명하였다. 그 다음 회귀 분석의 모델을 제시하며 변수를 정의하는 부분에서 매출과 감성지수의 산출 방법에 대해 설명하였다.

4장에서는 3장에서 설명한 연구 방법을 토대로 분석한 결과에 대해 설명하였다. 마지막으로 5장에서는 해당 연구의 시사점 및 한계점을 서술하고 향후 연구 방안을 제시하며 마친다.

제 2 장 이론적 배경 및 가설 설정

2.1 온라인 리뷰

소비자는 인지 능력에 제한이 있기 때문에 의사결정 시 많은 노력을 할 애하기보다 휴리스틱을 이용하는 경향이 있다(Maslowaska, Malthouse and Bernritter 2017). 그러나 소비자는 온라인에서 물건을 구매할 때 기업에서 제공하는 정보를 신뢰하지 않는 경향을 보인다. 기업이 제공하는 정보는 상업적인 의도를 가지고 있다고 간주하기 때문이다(Schiffman and Kanuk 2000). 반면 리뷰는 소비자가 익명으로 제품에 대한 사용 후기를 직접 작성한 것이기 때문에 신뢰성이 높다. 또한 리뷰는 제품의 품질과 소비자의 경험이 나타나있는 지표로 작용하여(Simonson 2014), 쉽게 접근 가능한 정보이다. 예를 들어 제품의 평균 별점이나 리뷰의 양은 소비자가 쉽게 얻을 수 있기 때문에 제품을 구매할 때 참고할 수 있는 정보가 된다. 이를 통해 제품의 리뷰는 소비자에게 제품 구매 의사결정시 많은 영향을 미친다(Forman, Ghose, and Wiesenfeld 2008)는 점을 알 수 있다.

또한 ‘Power review’ 설문조사(2010)에 따르면 64%의 온라인 소비자들이 리뷰를 읽는 데 10분 이상을 소요한다고 대답하였으며, 68%의 온라인 소비자들은 한 물건을 구매하기 전 최소 4개의 리뷰를 읽는다고 밝혔다. 이를 통해 리뷰는 단순한 개인의 의견이 아니라 다른 소비자들의 향후 구매 의사결정에도 막대한 영향을 미치게 된다.

2.1.1. 리뷰의 양

Liu (2006)는 구전의 양이 많을수록 소비자가 제품에 대해 정보를 많이 얻을 수 있고 이는 당연히 매출의 향상으로 이어진다고 주장했다. 온라인 리뷰(온라인 구전)는 기존의 구전과 같은 역할을 하는데, 온라인 리뷰와 같이 온라인을 통해 사람들에게 전달되는 정보들은 비동시적이고 짧은 시간에 수많은 사람에게 도달할 수 있다는 특징을 가진다. 다시 말해서 리뷰(온라인 구전)를 통해 소통하는 사람과 소비자 모두 기존의 구전 방식보다 의견을 전파하고 소비하는 데 훨씬 더 많은 선택권을 가지게 되며 정보에 대한 접근성을 높인다(Kiecker and Cowles 2002). 이러한 점들을 종합적으로 고려해봤을 때, 온라인 리뷰의 양이 많을수록 소비자가 구매 결정을 하는 데에 용이하다는 점을 알 수 있다.

선행연구에 의하면 Ambler and Bui (2011)는 리뷰가 있는 디지털 제품이 리뷰가 없는 제품보다 매출이 월등히 높은 결과를 나타냈다. 또한 Matfield (2011)는 리뷰의 방향성(긍정/부정)과 관계없이 리뷰의 양이 매출과 판매 수익 간에 양의 상관관계가 있다고 주장했다. Chiou and Cheng (2003)은 소비자에게 각각 6개와 12개의 휴대전화 포스팅 글을 보여줬을 때 12개의 포스팅 글을 읽은 소비자가 제품을 더 긍정적으로 평가했다는 연구 결과를 얻었다. 또한 Ghose and Ipeirotis (2010)은 리뷰의 양이 증가할수록 DVD와 디지털 카메라의 판매량에 긍정적인 영향을 준다는 점을 밝혔다. 이외에도 영화에 관련된 온라인 게시물의 수는 영화 박스오피스 매출에 큰 영향을 준다는 결과가 나타났으며(Duan, Gu, and Whinston 2008), 많은 연구들을 통해 리뷰의 양이 매출에 긍정적인 관계가 있음을 밝혔다. 또한 최근 Schoenmüller, Netzer and Stah (2019)의 연구에서는 대부분의 평가자가 별점을 매길 때 긍정적으로 평

가하는 경향이 있기 때문에 별점으로 판단하기 어려운 반면, 리뷰의 양이 많다는 것은 소비자가 해당 제품에 대해 만족한다는 것을 암시하므로 리뷰의 양이 더 유익한 정보라고 주장하며 리뷰의 별점보다 리뷰의 양이 매출에 훨씬 더 강력한 영향을 미친다는 결과가 나타났다. 따라서 본 논문에서는 여러 선행연구들로부터 리뷰의 양과 매출이 긍정적인 관계가 있다는 것을 토대로 이를 검증하기 위해 다음과 같은 가설을 수립하였다.

가설 1. 리뷰의 양이 많을수록 매출에 긍정적인 영향을 미칠 것이다.

2.1.2. 리뷰의 방향성

리뷰의 방향성은 크게 긍정과 부정으로 나눌 수 있다. 다수의 선행 연구에서 리뷰의 방향성은 별점에 따라 나누어 진행하였는데, 5점 만점의 평점 중 1-2점은 부정, 4-5점은 긍정으로 분류하였다. Yang, Sarathy and Walsh (2016)는 유사 실험설계와 온라인 설문조사를 통해 별점으로 분류한 리뷰의 방향성(4점과 5점은 긍정, 1점과 2점은 부정, 3점은 중립)과 리뷰의 양(많음 vs 적음)이 소비자의 구매의사 결정에 미치는 영향을 분석하였는데, 연구 결과 리뷰의 방향성이 리뷰의 양보다 소비자의 의사 결정에 더 큰 영향을 미치는 것을 발견하였다.

리뷰의 방향성이 매출에 미치는 영향에 관한 연구를 살펴보자면, Huang and Chen (2006)은 부정적인 내용의 리뷰는 제품 평가를 손상시키고 구매 가능성과 매출을 감소시킨다고 주장하였다. 조춘한·안승호·손영돈 (2012)은 소비자가 댓글의 방향성에 따른 구매의도의 차이를 연구하였는데, 그 결과 소비자가 긍정적인 댓글에 노출된 경우 양면(긍정

과 부정 모두 포함)보다 구매 의도가 높은 것으로 나타났다. 또한 Gopinath, Thomas and Krishnamurthi (2014)는 리뷰의 방향성이 휴대전화 판매량에 직접적인 영향을 미친다고 밝혔으며 Chevalier et al. (2006)는 아마존에서 별점 1개의 리뷰들이 판매 랭킹에 부정적인 영향을 미치며 별점 1개의 리뷰가 별점 5개의 리뷰보다 도서 판매량에 더 큰 영향력을 가진다고 밝혔다.

반면 리뷰의 방향성은 매출에 영향을 미치지 않는다는 연구결과도 존재하였다. Yahoo.com의 온라인 영화 리뷰를 조사한 결과, 리뷰의 방향성은 박스오피스의 매출에 영향을 미치지 않는 것으로 드러났다(Duan et al. 2008). 마찬가지로 Chen, Wu and Yoon (2004)은 온라인 리뷰가 아마존닷컴의 도서 판매랭킹에 영향을 미치지 않는다고 밝혔다. 마지막으로 Ambler et al. (2011)은 리뷰의 방향성이 디지털 소형가전의 구매를 예측할 수 없다는 결과를 밝혔다.

또한 리뷰의 방향성의 극단성에 대한 연구도 진행되었다. 극단적으로 긍정적이거나 부정적인 리뷰가 지나치게 많은 경우, 소비자가 리뷰에 대한 판단이 어려워지기 때문이다. 기존의 선행연구에서 긍정적인 리뷰의 수가 많을수록 매출에 긍정적인 영향을 준다는 결과가 밝혀졌지만, 일부 연구들에 따르면 오히려 지나치게 많은 긍정적 리뷰의 수는 소비자로 하여금 리뷰 신뢰성을 떨어뜨린다는 연구 결과가 있었다(Chevalier et al. 2006). 이와 비슷하게 Bosman, Boshoff and Rooyen (2013)의 연구에서는 리뷰의 방향성이 리뷰의 신뢰성에 큰 영향을 미치며 별점이 올라갈수록 리뷰 신뢰성이 평균 2.39% 하락하는 결과가 나타났다. 다시 말해서 소비자들은 5점 만점의 리뷰들을 의심스럽게 여기는 경향이 있고 (Dholakiya 2014), 지나치게 긍정적인 리뷰는 신뢰성에 부정적인 영향을 미칠 수 있다는 것이다. Mudambi and Schuff (2010)는 경험재를 구입할

때 적당한 내용(moderate)의 리뷰가 극단적인 내용(extreme)의 리뷰보다 더 신뢰성이 높다는 결과를 밝혔다.

여러 선행 연구들을 조사한 결과, 리뷰의 방향성이 매출에 영향을 미친다는 연구와 영향을 미치지 않는 연구결과가 양면적으로 나타났다. 따라서 본 논문은 리뷰의 방향성이 매출에 미치는 영향력을 검증하기 위하여 다음과 같은 가설을 세운다.

가설 2. 리뷰의 방향성이 매출에 영향을 미칠 것이다.

2.1.3 리뷰의 유용성

최근 들어 리뷰의 유용성이 매출과 상관관계가 높다는 연구결과가 나오면서 중요성이 커지고 있다. 리뷰의 유용성은 소비자가 생성한 리뷰를 읽은 다른 소비자들이 유용하다고 판단된다면 투표나 버튼을 눌러 유용성을 표시하는 것으로 대표적인 예로 아마존의 ‘helpful’ 버튼과 11번가의 ‘도움 돼요’ 버튼이 있다. 이러한 기능은 웹 사이트를 보다 사용자 친화적으로 만들어서 더 많은 소비자를 끌어당기는 효과가 있다. 많은 선행 연구들이 리뷰의 유용성에 영향을 미치는 요인에 대해 조사하였는데, 리뷰의 감성과 리뷰어의 등급, 리뷰의 길이, 리뷰 경과일, 리뷰의 사진 여부가 리뷰의 유용성에 영향을 미친다고 밝혔다(황침 · 권영진 · 이상용 2018). 또한 권진영 · 이문용 (2012)에 따르면, 리뷰의 유용성에 가장 영향을 미치는 요인은 리뷰어의 평판인 것으로 나타났으며, 다른 연구에서는 별점이 리뷰의 유용성에 가장 큰 영향을 미친다는 결과를 얻었다(Singh et. al. 2017). 이를 통해 리뷰의 유용성은 단지 소비자들이 투표

한 수를 집계해 놓은 것이 아니라, 다양한 요인에 의해 이루어지는 것이므로 높은 유용성을 보이는 리뷰는 리뷰의 질이 높다는 것을 알 수 있다(Jónsson 2016).

또한 높은 유용성을 보이는 리뷰가 실제 구매로 이어지는 경향을 보이며 매출에 영향을 미치는 요인으로 인식되고 있다. 선행 연구에 따르면, 리뷰의 ‘좋아요’ 수가 판매 랭킹에 긍정적인 영향을 미치는 것으로 밝혀졌다(황침 외 2018). 또한 아마존에서 helpfulness 기능을 통해 약 27억 달러의 추가 수익을 얻을 수 있다고 추정하였다(Spool, 2009). 박찬·유창조 (2006)는 온라인 리뷰만을 보여준 그룹과 온라인 리뷰와 해당 리뷰에 생성된 댓글을 보여준 그룹을 나누어 상표 평가의 차이를 연구하였는데, 그 결과 리뷰와 댓글을 함께 본 그룹이 상표 평가 태도가 더 크게 변화하였다. 이때, 긍정적인 리뷰에 동의하는 댓글이 많은 경우 상표 평가를 더 긍정적으로 하는 것으로 나타났으며 구매의사도 더 높게 나타났다. 지금까지 살펴본 선행연구들을 종합해보면 리뷰의 유용성이 매출에 긍정적인 영향을 미친다는 것을 예상할 수 있으며 논문에서는 앞서 진행된 선행연구들의 결과들을 검증하기 위해 다음과 같은 가설을 설정한다.

가설 3. 리뷰의 유용성이 높을수록 매출에 긍정적인 영향을 미칠 것이다.

2.1.4 텍스트 감성 분석

리뷰의 방향성이 매출에 영향을 미친다는 선행연구는 많이 수행되었지만, 대부분의 연구에서 리뷰의 방향성을 별점(1-5점, 5점 만점)으로 나누

어 긍정적(4-5점)과 부정적(1-3점)으로 나누어 진행하였다. 하지만 별점으로 방향성을 나누는 것만으로는 소비자의 의견을 정확히 구분할 수 없기 때문에 경험을 서술한 텍스트에서 감성을 분석하여 매출에 미치는 영향을 연구할 필요성이 제기되었다.

가격이나 평점같이 숫자로 된 정보들은 비교하기 쉬운 반면, 경험을 적은 리뷰는 주관적이기 때문에 평가하기가 어렵다는 특징을 가지고 있다 (Daft and Lengel 1983). 또한 가격이나 제품의 사양, 평점 등과 같이 숫자로 구성된 정보들은 소비자가 정보를 쉽게 얻을 수 있고 이해하기 위해 많은 시간이 소요되지 않는 반면, 소비자의 경험이나 감정과 같은 정성적 리뷰는 텍스트를 읽고 처리하기 위하여 많은 시간이 소요된다는 특징이 있다. Hu et al. (2014)은 별점이 판매에 간접적으로 영향을 미치는 반면 감성은 더 직접적으로 영향을 미치므로 텍스트에 나타난 감성의 중요성을 강조하였다. 이를 통해 리뷰에서 감지된 주관적인 감정들은 소비자에게 평점과 가격 같은 숫자 지표들보다 큰 영향을 미친다는 점을 알 수 있다. 본 논문에서는 감성분석의 중요성을 인지하고 소비자 리뷰에 나타난 감성을 추출하기 위해 감성분석을 활용하였다.

감성분석은 텍스트에서 긍정과 부정을 분류하기 위하여 수행하는데, 문제 인식에 따라 다른 방법이 사용된다(Liu 2010). Archak, Ghose and Ipeirotis (2011)는 두 개의 전자제품 카테고리 나누어 ‘제품의 특징 기반 감성분석(feature-based sentiment analysis)’을 수행하였다. 예를 들어 카메라의 사진 품질등과 같은 제품 별 특징을 구분하여 그에 따른 감성의 정도가 매출에 미치는 영향을 분석하였다. 그 결과, ‘적당한’, ‘괜찮은’ 같은 극단성이 미미한 단어들을 사용한 리뷰는 오히려 매출에 부정적인 영향을 미쳤으며, 제품의 세부 기능을 묘사하지 않은 ‘best camera’나 ‘perfect camera’ 같은 일반적인 리뷰들은 매출에 부정적인 영향을 미

친다는 결과가 나타났다. Ghose and Ipeirotis(2011)는 주관성 분석(subjectivity analysis)을 통해 작문 스타일(주관성의 정도, 가독성, 철자 오류의 정도)이 매출에 영향을 미친다는 것을 검증하였다. Hu et al. (2014)은 단어 빈도 분석(term frequency approach) 방법을 사용하여 감성을 분류하였는데, electronic inquirer dictionary에서 긍정/부정 단어를 지정한 다음 아마존 리뷰에 사용된 극성단어를 추출하여 감성을 매우 긍정, 긍정, 매우 부정, 부정으로 나눈 뒤, 리뷰의 텍스트에서 나타난 감성을 분석하였다. 이 때, 리뷰의 제목과 내용의 영향력 차이를 보기 위하여 리뷰의 내용(body)와 제목(title)을 분리하여 감성분석을 수행하였다. 그 결과 리뷰의 평균 별점이 매출에 미치는 영향은 유의하지 않고, 리뷰의 감성정도가 매출에 미치는 영향은 유의한 것으로 나타났다. 해당 연구에서는 평균 별점이 감성을 통해 매출에 간접적으로 영향을 미치는 반면, 감성은 매출에 직접적인 영향을 미친다고 밝혔다. 앞서 나온 선행연구들을 살펴보았을 때, 리뷰의 텍스트에서 감지된 감성이 매출에 영향을 미치는 것을 예상할 수 있었다. 따라서 본 논문에서는 Hu et al.. (2014)의 연구 내용을 바탕으로 리뷰의 제목과 내용에서 감지된 감성이 매출에 미치는 영향을 구체적으로 살펴보고자 가설 4를 다음과 같이 수립한다.

가설 4. 리뷰에서 나타난 감성의 정도가 매출에 영향을 미칠 것이다.

가설 4-1. 리뷰의 제목이 긍정(부정)적인 경우, 매출에 긍정(부정)적인 영향을 미칠 것이다.

가설 4-2. 리뷰의 내용이 긍정(부정)적인 경우, 매출에 긍정(부정)적인 영향을 미칠 것이다.

2.2 브랜드 명성(Brand Strength)

브랜드 명성(또는 가치)은 경쟁업체에 비해 지속가능하며 차등적인 우위를 구축하기 위해 설계된 장기 투자의 결과물이다(Doyle 1990). 브랜드 신호에 관한 선행 연구에 따르면, 제품에 대한 품질과 정보에 대한 불확실성은 리스크를 초래하게 된다(Erdem, Swait and Valenzuela 2006). 이러한 리스크에 대처하기 위해 소비자들은 구매 시 제품의 품질이나 정보를 나타내는 신호에 의존하는 경향이 있다(Shimp and Bearden 1982). 예를 들어 가격(Stiglitz 1989), 광고(Kirmani 1990) 등과 같은 마케팅 믹스 요소들이 믿을 만한 신호로 여겨지는데, 연구자들은 이 때 브랜드가 유독 강하고 효과적인 신호로 작용된다는 점을 밝혔다(Rao, Qu and Ruekert 1999). 이 때, 강한 브랜드는 약한 브랜드보다 기존의 브랜드 자산 손실(Erdem and Swait 1998)에 대한 위험성이 크기 때문에 강한 브랜드가 약한 브랜드보다 소비자들에게 더 많은 신뢰를 준다. 따라서 선행연구들을 살펴보았을 때, 소비자가 제품을 구매할 때 브랜드 또한 중요한 요소라는 점을 알 수 있다.

브랜드 명성과 온라인 리뷰 간의 상관관계에 대한 선행 연구들을 살펴보자면, 박찬 외 (2006)는 브랜드 명성이 높을 때보다 낮을 때 사용후기의 부정적인 내용의 영향력이 더 크다고 밝혔다. 브랜드 명성이 높을 때(유명 브랜드인 경우) 긍정적 사용 후기는 부정적 사용후기와 영향력에 별 차이가 없었지만, 무명 브랜드의 경우 긍정적 사용후기와 부정적 사용후기가 매출에 미치는 영향력이 유의하다는 결과가 나타났다. 또한 Ho-Dac et al. (2013)은 블루레이 플레이어와 DVD 플레이어의 리뷰와 브랜드의 상호작용효과를 살펴보았다. 그 결과 긍정(부정)적인 리뷰는 약한 브랜드의 판매량에 긍정(부정)적인 영향을 주지만 강한 브랜드에서

는 큰 영향을 미치지 않는다는 점을 밝혔다. 또한 이 때 부정적인 리뷰는 영향이 없는 반면 긍정적인 리뷰는 브랜드 명성을 상승시킨다는 점을 확인하였는데, 매출의 증진이 긍정적인 리뷰의 생성을 촉진시키며 이것이 약한 브랜드에서 매출 상승과 리뷰 양의 증가 간에 긍정적인 순환작용을 일으킨다고 주장했다. 결과적으로 긍정적인 리뷰는 약한 브랜드의 제품이 경쟁력을 갖는 동시에 브랜드 명성을 상승시키는 효과를 가진다. 그러나 이러한 리뷰와 매출의 순환작용은 리뷰의 영향을 덜 받는 강한 브랜드에서는 나타나지 않았다. 앞의 연구들을 통해서 브랜드 명성에 따라 리뷰의 영향력이 달라진다는 결과를 얻을 수 있었고, 선행 연구들의 결과에서 리뷰의 정량적 정보들이 그랬던 것처럼 본 논문에서는 리뷰의 정성적 정보 또한 유사한 패턴을 보일 것이라고 예상하였다. 따라서 다음과 같은 가설을 수립한다.

가설 5. 브랜드 명성에 따라 리뷰의 영향이 다를 것이다.

가설 5-1. 브랜드 명성이 낮은 경우, 높은 경우보다 리뷰의 양이 매출에 미치는 영향이 클 것이다.

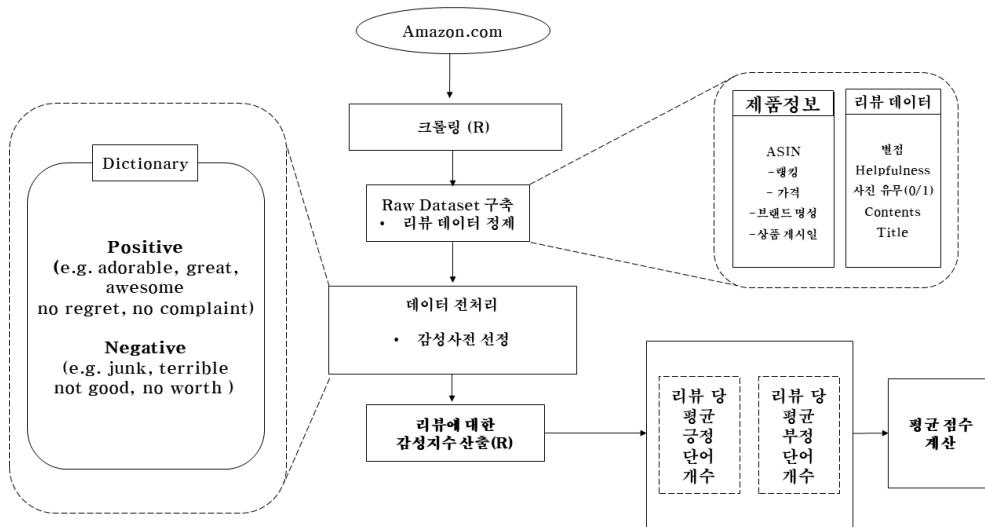
가설 5-2. 브랜드 명성이 낮은 경우, 높은 경우보다 리뷰의 내용에서 나타난 긍정(부정)점수가 매출에 미치는 영향이 클 것이다.

가설 5-3. 브랜드 명성이 낮은 경우, 높은 경우보다 리뷰의 제목에서 나타난 긍정(부정)점수가 매출에 미치는 영향이 클 것이다.

제 3 장 연구 방법론

3.1 분석 프로세스

[그림 3-1] 은 분석 프로세스를 나타낸 것이다.



[그림 3-1] 분석 프로세스

3.1.1. 데이터 크롤링

본 연구는 소비자의 리뷰의 별점, 양, 유용성, 방향성이 매출에 끼치는 영향을 분석하기 위해 아마존에서 판매되는 블루투스 스피커 데이터를 이용하여 분석을 진행하였다.

블루투스 스피커는 탐색재에 속하는 제품으로, 기존의 연구에서는 경험재가 탐색재에 비해 리뷰의 영향을 많이 받는 제품군이라는 결과가 있었다(Bei, Chen and Widdows 2004; 이호근·곽현 2013). 연구 결과에 따르면, 탐색재는 경험재와 달리 외재적 단서에 의해 평가되므로 충분한

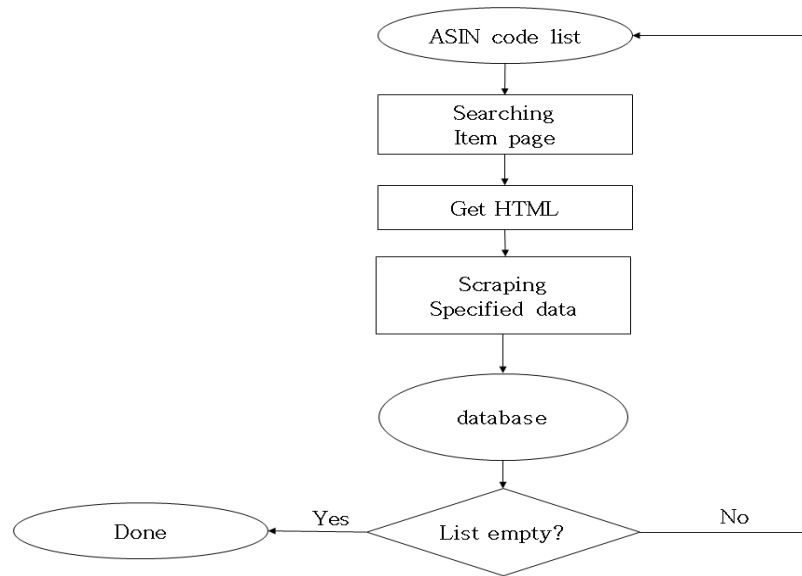
정보 탐색을 통해 구매 결정을 내릴 수 있기 때문에 소비자는 경험재를 구매할 때 리뷰의 영향을 많이 받는다. 그렇기 때문에 경험재에 대한 리뷰의 영향력은 많이 연구된 반면, 탐색재에 대한 연구는 많이 이루어지지 않았다. 그러므로 본 논문에서는 탐색재를 구매할 때 소비자가 받는 리뷰의 영향력에 대해 연구하게 되었다. 블루투스 스피커는 소비자에게 친숙한 전자기기인 동시에 상품의 기술적 정보를 근거로 하여 평가되는 탐색재의 특성을 가지므로(이진화 · 김지은 · 안현아 2016) 적절하다고 평가되어 선정하게 되었다. 또한 블루투스 스피커는 대중들에게 친근한 전자기기인 동시에 Mordor Intelligence의 조사에 따르면, 블루투스 스피커 시장은 대규모 소비자 집단을 찾기가 어려운 파편화한 시장(Fragmented)에 가까우며 우위를 점하는 브랜드의 수가 많지 않은 것으로 나타났다. 따라서 정보 탐색의 중요성이 더욱 크다고 판단하였다.

제품 데이터를 수집하는 크롤링에는 R을 이용하였다. 매출 데이터를 직접 구하기 어려우므로 매출을 가늠하기 위해 판매 순위 1위부터 100위까지의 제품을 수집하였다. 판매 순위는 기존의 많은 선행 연구에서 매출로 대체되어 사용되었으며 파레토 법칙에 의해 매출과 판매 순위 간에 선형관계를 보인다는 것을 입증하였다(Archak et al. 2011; Ho-Dac et al. 2013).

수집 기간은 2019년 8월 14일부터 9월 7일까지 이루어졌는데, 4주 동안 100개의 제품 정보를 8번의 크롤링 과정을 거쳐 총 800개의 제품데이터로 구성하였다. 아마존의 베스트셀러 랭킹은 한 시간 단위로 갱신되는데 랭킹의 변동을 살펴본 결과, 하루 단위로 바뀌는 순위의 변동은 약 5%, 리뷰의 증가량은 0.47% 정도로 눈에 띄는 차이를 보이지 않았다. 그러므로 1주일에 두 번, 주중인 수요일과 주말인 토요일로 나누어 데이터를 수집하였다. 1주일에 두 번씩 데이터를 수집하여 변화량을 살펴본 결과

랭킹은 약 8%, 리뷰 증가량은 약 1.4%의 변화를 보였고 적절한 변화량을 보였으므로 1주일에 두 번씩 데이터를 수집하는 방법을 택하였다.

데이터는 R의 ‘rvest’ 패키지를 이용하여 웹 크롤링을 통해 수집하였다. [그림 3-2]처럼 구현한 웹 크롤러가 웹 페이지를 탐색하면서 정의된 규칙에 따라 제품 정보를 수집한다.

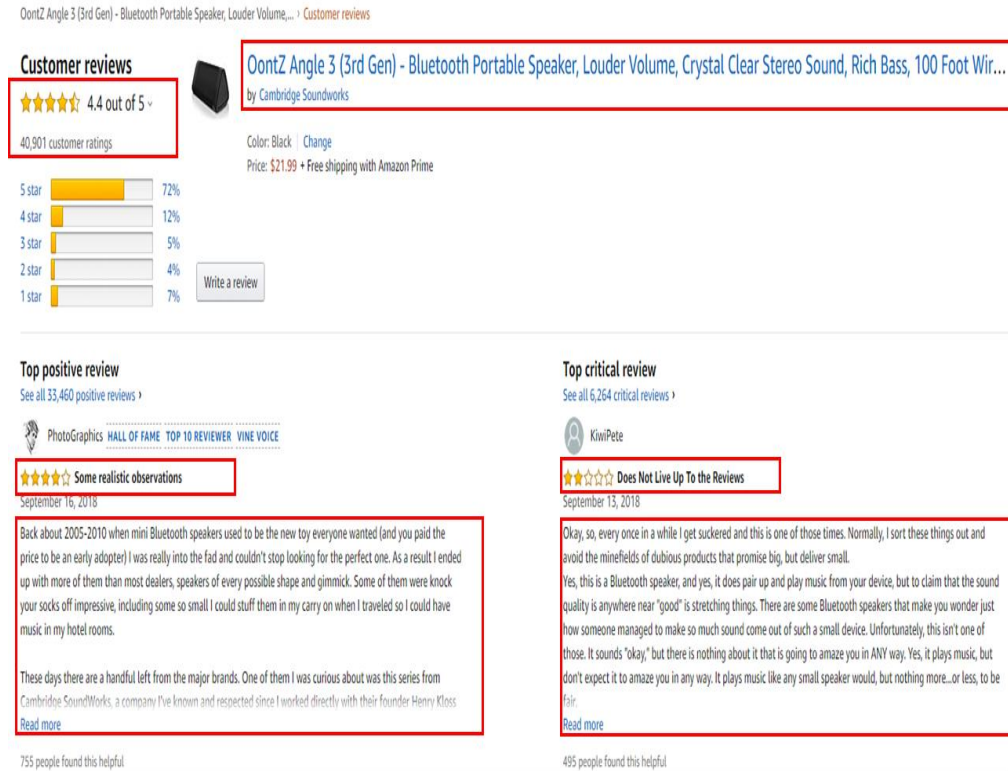


[그림 3-2] 크롤링 프로세스

제일 먼저 아마존 베스트셀러 페이지에서 상품 고유코드번호를(ASIN)을 100개 수집한 뒤, 고유 코드번호에 따른 URL이 새로 생성한다. 이 때 웹 크롤러는 생성된 URL에서 평점, 리뷰의 양, 가격, 브랜드, 리뷰 컨텐츠, 유용성(helpfulness)의 총 개수, 사진이 포함된 리뷰의 개수를 수집하여 데이터베이스에 저장한다. 마지막 100순위의 제품 정보 수집을 마치게 되면 알고리즘을 종료한다.


수집한 데이터는 블루투스의 랭킹, 상품명, 브랜드, ASIN코드, 가격, 평점, 리뷰의 양, 사진이 포함된 리뷰의 개수, Helpfulness의 총 개수, 상품 게시일, 온라인 리뷰로 이루어져있다. [그림 3-3]은 실제 아마존의 블루

투스 스피커 제품 리뷰 페이지를 나타낸 것이며, 수집한 데이터들의 예시를 나타낸다.



[그림 3-3] 아마존 리뷰 페이지와 수집한 데이터의 예시

Helpfulness는 다른 소비자들이 개인의 리뷰를 읽고 도움이 됐다고 생각하면 클릭하는 것으로, 아마존에서는 최근 순과 Helpfulness의 숫자를 기준으로 리뷰를 노출시키고 있다. [그림 3-4]는 소비자의 리뷰들 중 다른 소비자들이 투표한 helpful 버튼의 클릭 수를 집계한 helpfulness 수를 나타낸다.


KiwiPete

☆☆☆☆☆
Does Not Live Up To the Reviews

September 13, 2018

Color: Black
Verified Purchase

Okay, so, every once in a while I get suckered and this is one of those times. Normally, I sort these things out and avoid the minefields of dubious products that promise big, but deliver small.

Yes, this is a Bluetooth speaker, and yes, it does pair up and play music from your device, but to claim that the sound quality is anywhere near "good" is stretching things. There are some Bluetooth speakers that make you wonder just how someone managed to make so much sound come out of such a small device. Unfortunately, this isn't one of those. It sounds "okay," but there is nothing about it that is going to amaze you in ANY way. Yes, it plays music, but don't expect it to amaze you in any way. It plays music like any small speaker would, but nothing more...or less, to be fair.

Not sure if I'll keep this or not.

508 people found this helpful

Helpful
▼ 9 comments
Report abuse

[그림 3-4] 리뷰의 helpfulness의 수가 나타난 아마존 리뷰 페이지

8번의 수집 기간 동안 모은 데이터의 모델 수는 총 152개이며 <표3-1>는 t1 시점에서 크롤링을 완료한 제품 데이터의 예시이다.

<표 3-1> 아마존의 블루투스 스피커 베스트셀러 제품 데이터 정보 예시

| Rank | Product | Brand | ASIN | Price | Rating | Review quantity | Total picture | Total helpfulness | Days on amazon |
|------|---|----------------------|------------|-------|--------|-----------------|---------------|-------------------|----------------|
| 1 | OontZ Angle 3 (3rd Gen) | Cambridge Soundworks | B010OYASRG | 25.99 | 4.4 | 36151 | 74 | 526 | 1545 |
| 2 | OontZ Angle Solo | Cambridge Soundworks | B07DPVF8DQ | 20.99 | 4.6 | 879 | 21 | 247 | 350 |
| 3 | OontZ Angle 3 PLUS | Cambridge Soundworks | B01LZV6TYQ | 22.94 | 4.4 | 5607 | 75 | 2372 | 372 |
| 4 | Bluetooth Speakers, Anker Soundcore Bluetooth Speaker | Anker | B016XTADG2 | 27.95 | 4.5 | 13725 | 92 | 960 | 1433 |
| 5 | DOSS Touch Wireless Bluetooth V4.0 Portable Speaker | DOSS | B01CQOV3YO | 39.99 | 4.4 | 9391 | 86 | 666 | 1293 |

리뷰 데이터는 날짜, 개별 리뷰의 평점, 리뷰의 제목, 리뷰의 내용으로 구성되었다. 리뷰 데이터는 각 상품 당 크롤링 할 수 있는 리뷰의 최대 개수가 5000개로 제한되어 있었다. 따라서 5000개가 넘는 리뷰를 가진 제품들은 최근 순으로 5000개의 리뷰를 수집하였다. <표 3-2>는 수집한 리뷰 데이터의 예시이다.

<표 3-2> 아마존 블루투스 스피커 베스트셀러 리뷰 데이터 정보 예시

| Rating | Title | Contents |
|--------|----------------------|---|
| 5 | Love It! | Works perfect! Love it! |
| 1 | Breaks easily | Case broke the first time I used it |
| 5 | Great Little Speaker | I use this when my friends and I go to the beach and it's a great speaker for the price |
| 3 | Perfect fit! | Awesome fitting case that really seems to cushion the speaker. Started to come apart after a month of use. |
| 5 | Fits to a ... | Good case to hold my Angle 3. Keeps it protected, and I can attach it if needed. |
| 5 | Excellent product | Economical way to get great sound! |

3.1.2. 리뷰 데이터 정제

리뷰의 감성분석(텍스트 분석) 또한 크롤링과 마찬가지로 R을 이용하여 진행되었다. 텍스트 분석에 널리 통용되는 ‘tidyr’ 과 ‘tidytext’ 패키지를 이용하여 분석을 수행하였다(Silge and David 2016; Wickham and Henry 2016). 이 때 리뷰의 제목과 리뷰의 내용의 영향력 차이를 구분하기 위하여 데이터를 둘로 나누어 분석을 수행하였다.

Raw data 수집을 마친 뒤 데이터 정제 및 전처리 작업을 수행하였다. 작업 과정은 다음과 같다. 첫째로, 리뷰의 제목(Title)과 내용(Contents)을 분류하였다. 그 다음 분석 과정에 사용되는 데이터를 적합하게 가공하는 과정을 거쳤다. 이 과정에서는 결측값과 특수문자를 제거하고 텍스트

트 분석을 원활하게 진행하기 위해 모든 텍스트를 소문자로 변형하였다.

해당 과정을 거친 후 텍스트를 두 단어로 분할하는 bigram 프로세스를 수행하였는데, bigram은 바이너리, 텍스트 등 전체 문자열을 N값으로 나눈 통계방법인 N-gram의 일종으로 N의 값에 따라 Unigram, Bigram, Trigram 등으로 나누어진다고(Michel et al. 2011). 이 때 bigram이란 연속된 두 개의 단어를 하나로 묶어 분할하는 방법이며 추출된 단어가 바로 전 단계에 추출된 단어에 의존하는 경우를 분석할 때 주로 사용된다. bigram 프로세스를 통해 두 개씩 묶인 단어들은 ‘separate’ 함수를 사용하여 하나의 단어로 분할하는 과정을 거친다. 이 단어들은 각각 word1과 word2로 분류된다. 해당 기법을 사용하면 부정어가 포함된 단어를 추출할 수 있다. 부정어(no, not, never 등)가 포함된 단어는 본래의 단어와 반대의 의미를 가지므로 하나의 단어로 분할하는 unigram 방식으로 는 골라낼 수 없다. 이러한 이유로 보다 정확한 분석결과를 얻기 위해 bigram방식을 사용하였다. [그림3-5]는 word1에 부정어가 포함된 단어의 일부이다.

| title | word1 | word2 |
|--|-------|-----------|
| does not work after 1 year | not | work |
| good speaker firmware update does not work and can ... | not | work |
| i would not recommend buying this speaker | not | recommend |
| it's okay but not worth the higher price | not | worth |

[그림 3-5] 부정어가 포함된 단어의 예시

‘separate’함수를 실행하면 단어를 2개씩 묶어서 나누기 때문에 한 단어로만 이루어진 리뷰들이 분할되지 않고 결측값(NA)으로 나오게 된다. 본 논문에서는 이를 방지하기 위해 해당 단어를 word2에 할당하여 word2에서 감정 단어가 분류되도록 하였다. [그림3-6]은 한 단어로만 이

루어진 리뷰의 제목이 포함되지 않고 결측값으로 나온 경우이며 이 단어들을 word2에 할당하여 해결한 예시이다.

| contents | word1 | word2 |
|----------|-------|-------|
| Good | NA | NA |
| Greatttt | NA | NA |
| Great | NA | NA |
| Good | NA | NA |
| Perfect | NA | NA |
| Great | NA | NA |
| Great | NA | NA |
| Great | NA | NA |
| Flawless | NA | NA |

| contents | word1 | word2 |
|----------|-------|----------|
| Good | na | Good |
| Greatttt | na | Greatttt |
| Great | na | Great |
| Good | na | Good |
| Perfect | na | Perfect |
| Great | na | Great |
| Great | na | Great |
| Great | na | Great |
| Flawless | na | Flawless |

[그림 3-6] 한 단어로 된 리뷰가 분할되지 않은 예(좌) 와 이 단어들을 word2에 할당하여 전처리한 예(우)

3.1.3. 감정사전 선정

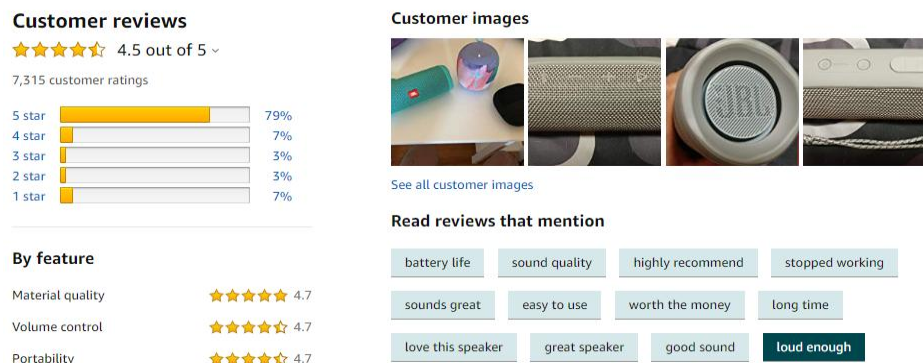
감성 사전은 Liu가 제작한 Opinion Lexicon (2015)을 이용하였다. 해당 감성사전은 단일단어(Unigram)에 대한 감성을 Positive, Negative 두 분류로 나누었다. tidytext에는 ‘Afinn’, ‘Bing’, ‘NRC’ 3개의 단어사전을 제공하고 있는데 각각의 어휘집은 다른 개발자 팀에 의해 만들어졌으며 모든 텍스트의 감정 분석을 목표로 하고 있다. Afinn은 단어 개수가 적고, NRC는 단어의 긍,부정 정확도가 떨어져 Opinion Lexicon을 채택하게 되었다. Opinion Lexicon은 총 6786개의 단어로 구성되어 있고 2005개의 긍정단어, 4781개의 부정단어로 이루어져 있다. 긍정적인 단어에는 good, wonderful, amazing 등이 있고 부정적인 단어에는 bad, poor, terrible 등

이 있다. <표3-3>는 감성사전의 긍정적인 단어와 부정적인 단어의 예시이다.

<표 3-3> 감성사전의 긍정적인 단어와 부정적인 단어의 예시

| word | sentiment | word | sentiment |
|------------|-----------|------------|-----------|
| 2-faces | negative | abounds | positive |
| abnormal | negative | abundance | positive |
| abolish | negative | abundant | positive |
| abominable | negative | accessable | positive |
| abominably | negative | accessible | positive |
| abominate | negative | acclaim | positive |

그러나 문맥에 따라 단어의 감정이 바뀌는 경우가 있기 때문에 단어 사전의 감정(sentiment)을 수정하거나 단어를 삭제하는 과정을 거치게 되었다. 예를 들어 [그림 3-7]과 같이 음량의 크기가 크다는 의미의 ‘loud’는 기존의 사전에서 부정적인 감정으로 분류되었다. 그러나 블루투스 스피커의 리뷰의 내용을 살펴보면 ‘loud enough’라는 제품의 특징이 있다.



[그림 3-7] 아마존 블루투스 스피커 리뷰의 예시

이를 통해 ‘loud’는 소리가 잘 들린다는 뜻으로 긍정적인 의미라는 것을 알 수 있다. 반면 ‘stopped working’이라는 특징은 갑자기 제품이 작동되지 않거나 멈추었을 때 쓰이므로 부정적인 의미로 사용되었다. 그러므로 변형된 형태를 포함한 ‘stop’은 부정적인 단어로 분류한다. 수정 단계를

거친 단어들은 <표 3-4>과 같다.

<표 3-4> 수정 단계를 거친 단어

| 수정 단어 | 수정 후 | 수정 이유 |
|------------|----------|------------------------------|
| Loud | 부정 -> 긍정 | 소리가 크다는 긍정적인 의미 |
| Quiet | 삭제 | 강조의 의미와 소리의 의미 중복 |
| Work | 삭제 | ‘일한다’의 긍정적 의미와 ‘작동한다’의 의미 중복 |
| Disconnect | 추가(부정) | ‘소리나 연결이 끊긴다’는 부정적 의미로 사용 |
| Stop | 추가(부정) | ‘소리나 연결이 멈췄다’는 부정적 의미로 사용 |
| insane | 삭제 | 긍정과 부정의 의미 중복 |

3.1.4 감성 분석

정제된 텍스트를 두 단어로 분할하는 과정을 거친 후 감성 분석을 수행한다. 감성 분석은 word2의 단어들과 감성사전의 단어들을 매치시켜서 해당되는 단어들을 추출해낸다. 이 때 추출된 단어의 sentiment(감정)이 긍정적인 단어에 포함되는 경우 sentiment 열에 positive로 분류되고 부정적인 경우 negative로 분류된다. 이렇게 분류된 감성들을 긍정단어 개수와 부정 단어 개수를 계산한다. 이 과정을 통해 한 리뷰에 어떤 긍정적인 단어와 부정적인 단어가 쓰였는지 파악할 수 있다.

Hu et al.,(2014)의 연구에서는 산출된 긍정 점수에 부정점수를 차감하는 계산식을 사용하였지만, 해당 계산식을 본 논문에 사용하면 긍정 점수와 부정 점수가 같은 경우 상쇄되어 0의 값을 가지게 된다. 이는 감성 단어가 추출되지 않은 리뷰의 감성지수와 같아지게 되므로 차별을 두기 위해 본 논문에서는 긍정점수와 부정점수를 따로 산출하였다. 다시 말해

서, 개별 리뷰의 긍/부정 단어를 모두 합쳐 전체 리뷰의 개수로 나누어 제품의 평균 긍/부정 점수를 계산하는 방법을 택하였다. 감성분석을 통해 도출된 감성지수의 계산식은 다음과 같다.

$$Sentiment_i = \frac{RS_1 + RS_2 + RS_3 + \dots + RS_n}{n}$$

RS_n : 제품 i 의 n 번째 리뷰의 감성 점수
 n : 리뷰의 총 개수

이 과정은 리뷰 데이터의 제목(Title)과 내용(Contents)에 똑같이 적용된다. [그림 3-8]에서는 감성 분석의 과정의 일부를 나타내고 있다.

| title | contents | word1 | word2 | sentiment |
|--|-----------------------------------|--------|---------------|-----------|
| Great quality | good | na | good | positive |
| Great Product | this blue tooth speaker is a big | not | great | positive |
| Great Product | this blue tooth speaker is a big | is | decent | positive |
| Started out great... | was nice. however eight months | really | liked | positive |
| Started out great... | was nice. however eight months | was | good | positive |
| Started out great... | was nice. however eight months | was | nice | positive |
| Started out great... | was nice. however eight months | very | disappointing | negative |
| Good when it workse off. had a hard time getting p | | jack | broke | negative |
| Good when it workse off. had a hard time getting p | | a | hard | negative |
| Good when it workse off. had a hard time getting p | | same | problem | negative |
| Great product! | d - connects easily - phone calls | in | clear | positive |

[그림 3-8] 분할된 단어와 감성분석 결과

3.2 회귀분석 모델

본 논문에서는 블루투스 스피커의 매출을 예측하기 위하여 종속변수는 매출(sales)로 지정하였다. 또한 독립변수로는 브랜드 명성(Brand Strength), 리뷰의 양(Review_qty), 별점(Rating), 유용성(Sum of helpfulness), 리뷰 내용의 긍정점수(Contents positive score), 리뷰 내용의 부정점수(Contents negative score), 리뷰 제목의 긍정점수(Title positive score), 리뷰 제목의 부정점수(Title negative score)가 있다.

통제변수로는 가격(Price), 사진을 포함한 리뷰의 개수(The Number of reviews including Picture), 상품 게시일(Days on amazon), 리뷰 제목의 단어 개수(Title word count), 리뷰 내용의 단어 개수(Contents word count)로 이루어졌고, 추가로 수집기간을 t1부터 t8까지 구분하여 지정하였다.

독립변수 중 리뷰의 양(Review_qty)은 한 제품에 대해 소비자들이 평가한 리뷰의 총 개수이다. 이 때, 한 제품의 옵션이 여러 개인 경우, 베스트셀러 모델의 리뷰만 수집하였다. 별점(Rating)은 소비자가 평가한 제품의 평균 별점으로 1점에서부터 5점까지 있으며 5점을 만점으로 한다. 유용성(Sum of helpfulness)은 소비자가 제품의 리뷰가 도움이 된 경우 누르는 helpful의 수를 합한 것이다. 리뷰 내용의 긍정점수(Contents positive score)와 부정점수(Contents negative score)는 내용에서 추출된 긍정과 부정 단어의 개수이다. 마찬가지로 리뷰 제목의 긍정점수(Title positive score), 리뷰 제목의 부정점수(Title negative score) 또한 제목에서 추출된 긍정과 부정 단어의 개수를 뜻한다.

통제변수 중 가격(Price)은 상품의 가격을 의미한다. 상품 게시일(Days

on amazon)은 상품이 아마존에 게시된 날로부터 지난 시간을 나타낸다. 숫자가 클수록 오래전에 게시된 상품이다. 사진을 포함한 리뷰의 개수(The Number of reviews including Picture)는 사진을 포함한 리뷰의 개수로 소비자가 리뷰에 사진을 첨부한 수를 계산하였다. 리뷰 제목의 단어 개수(Title word count)와 리뷰 내용의 단어 개수(Contents word count)는 각각 제목과 내용에 쓰인 단어의 개수를 의미한다. 또한 Ho-Dac et al. (2013)의 연구에서는 시간차를 두고 랭킹의 변화를 측정하였는데, 본 논문 위의 연구내용을 기반으로 t8을 기준으로 오래된 순서부터 t1부터 t7까지 수집 기간을 더미변수를 표현하였다. 이를 통해 지난 시점의 랭킹이 최근 시점의 랭킹에 주는 영향을 알아보고자하였다. 연구에 사용된 회귀 모델은 아래와 같다.

$$\begin{aligned} \ln(SR_{jt}) = & \beta_0 + \beta_1 \ln(NR_{jt}) + \beta_2 \ln(AR_{jt}) + \beta_3(SH_j) + \beta_4(CPO_j) + \beta_5(CNE_j) \\ & + \beta_6(TPO_j) + \beta_7(TNE_j) + \beta_8(BS_j) + \beta_9 \ln(NR_{jt}) \times BS_j + \beta_{10}(TPO_j) \times BS_j \\ & + \beta_{11}(TNE_j) \times BS_j + \beta_{12}(CPO_j) \times BS_j + \beta_{13}(CNE_j) \times BS_j + \beta_{14}(P_{jt}) + \beta_{15}(DA_j) \\ & + \beta_{16}(TWC_j) + \beta_{17}(CWC_j) + \beta_{18}(RP_j) + \beta_{19}(t_1) + \beta_{20}(t_2) + \beta_{21}(t_3) + \beta_{22}(t_4) \\ & + \beta_{23}(t_5) + \beta_{24}(t_6) + \beta_{25}(t_7) \end{aligned}$$

SR_{jt} : t기간에 제품j의 판매 랭킹

NR_{jt} : t기간에 제품 j의 리뷰 총 개수

AR_{jt} : t기간에 제품 j의 평균 별점

SH_j : 제품 j의 리뷰의 총 helpfulness 개수

CPO_j : 제품 j의 리뷰의 내용에서 계산된 긍정점수

CNE_j : 제품 j의 리뷰의 내용에서 계산된 부정점수

TPO_j : 제품 j의 리뷰의 제목에서 계산된 긍정점수

TNE_j : 제품 j의 리뷰 제목에서 계산된 부정점수

BS_j : 제품 j의 브랜드 명성 여부

P_t : t기간에 제품 j의 가격

DA_j : 제품 j의 아마존 게시일(일수)

TWC_j : 제품 j의 리뷰 제목에 쓰인 단어의 개수

CWC_j : 제품 j의 리뷰 내용에 쓰인 단어의 개수

RP_j : 제품 j의 리뷰에 사진이 포함된 개수

t_n : n번째 수집기간

3.2.1 매출

종속변수는 매출(Sales)로 지정하였으나 아마존에서는 매출 데이터를 제공하지 않아서 정확한 매출을 확인할 수 없었다. 그러나 아마존은 베스트셀러 카테고리에 판매 순위를 1위부터 100위까지 제공하므로 이를 이용하여 매출을 추론할 수 있다. 선행 연구들에 의하면 상품들의 매출과 판매 순위 간의 관계가 파레토 법칙을 통해 설명된다. 파레토 법칙에 의하면 매출(Sales)과 판매 순위(Sales rank)는 선형을 이루고 있는데, 해당 식은 다음과 같다.

$$\ln(\text{sales}) \approx a + b \times \ln(\text{sales rank})$$

이러한 선형 관계는 도서, 소프트웨어, 여성 의류, 전자 기기 등 다양한 카테고리에서 나타났다(Brynjolfsson, Hu, and Simester 2011; Brynjolfsson, Hu, and Smith 2003; Ghose and Sundararajan 2005). Ho-Dac et al. (2013)에 의하면, 파레토 법칙이 유지된다고 가정했을 때 $\ln(\text{sales})$ 을 사용하는 대신 $\ln(1/\text{sales rank})$ 을 사용하면 추정 계수와 표준 오차가 상수로 스케일링되고 추정된 절편이 다른 상수만큼 이동된다.

이것은 계수의 의미나 부호를 바꾸지 않기 때문에 두 모델 간에 큰 차이를 가지지 않는다. 계수의 가독성을 높일 수 있다는 장점이 있다. 본 논문에서는 이와 같은 이유로 판매순위의 역수를 로그로 변환하여 $\ln(1/\text{rank})$ 과 같은 형식으로 매출로 정의하였다.

3.2.2 리뷰의 내용 및 제목의 감성지수(sentiment score)

리뷰의 감성지수는 리뷰의 제목과 내용의 긍정 점수(Contents positive score)와 부정 점수(Contents negative score)로 나누어지는데, 이는 감성 분석을 통해 계산된 긍정(부정) 점수이다. 개별 리뷰에서 긍정(부정)단어의 수를 추출한 뒤 전체 리뷰의 수로 나누어 평균을 산출하였다. 이 계산을 통해 한 제품에 평균 긍정(부정)단어가 몇 개가 쓰였는지를 파악할 수 있다. positive score가 높을수록 긍정적인 단어가 많이 쓰였다는 의미이며 마찬가지로 negative score가 높을수록 부정적인 단어가 많이 쓰였다는 것을 의미한다.

3.2.3 브랜드 명성(Brand Strength)

브랜드 명성(Brand strength)은 강한 브랜드와 약한 브랜드를 나눈 것으로 강한 브랜드를 1, 약한 브랜드를 0으로 코딩하여 나타냈다. 해당 브랜드 명성은 리서치 회사인 Arizton의 Bluetooth Speaker Market in US - Industry Outlook and Forecast 2018-2023에서 발표한 Major Brand를 통해 선정하였다. 선정된 강한 브랜드(높은 브랜드 명성)로는 Bose, JBL, Sony, Philips, Harman Kardon, Logitech로 총 53개의 브랜

드 중 6개의 브랜드로 이루어졌다.

제 4 장 결과 분석

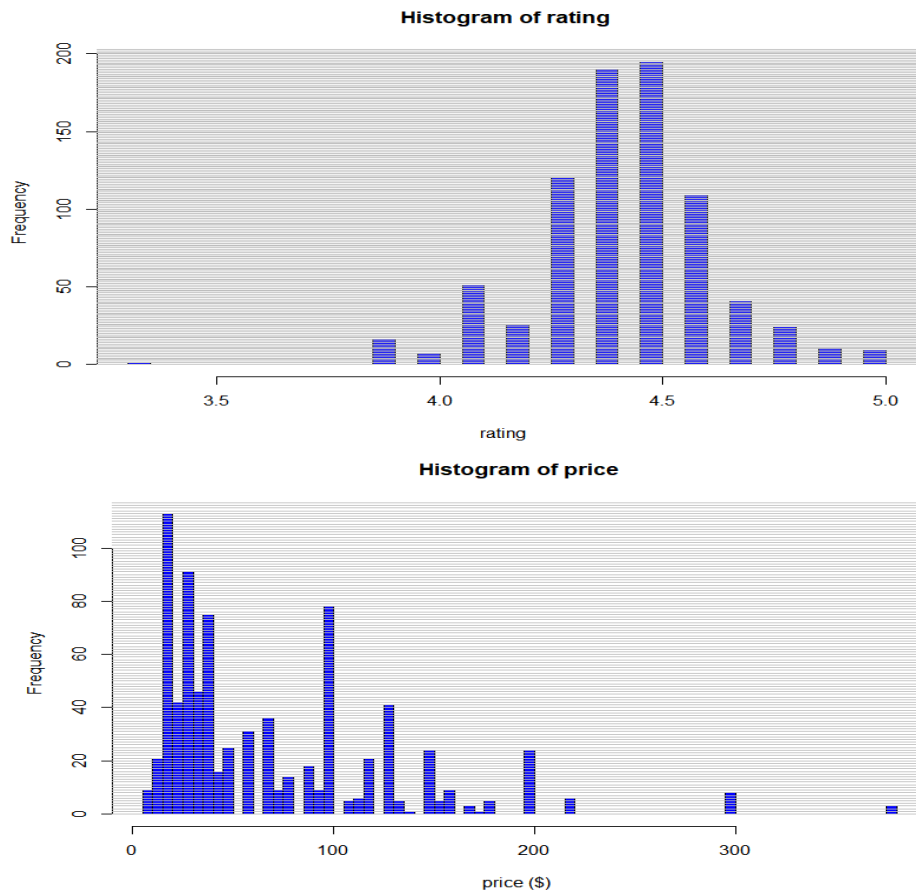
4.1 기술 통계

변수들의 기술통계는 <표 4-1>과 같다.

<표 4-1> 기술 통계 분석

| | Min | Max | Mean | SD |
|----------------------------|------|--------|---------|---------|
| 랭킹(Rank) | 1 | 100 | 50.50 | 28.89 |
| 가격 | 7.98 | 379.95 | 69.33 | 58.50 |
| 별점(Rating) | 3.3 | 5.0 | 4.44 | .20 |
| 리뷰의 양 | 3 | 36722 | 1618.86 | 3977.97 |
| 사진을 포함한 리뷰의 수 | 0 | 1219 | 24.44 | 72.17 |
| Helpfulness의 총 개수 | 0 | 5069 | 423.78 | 971.49 |
| 상품 게시일 (Days on amazon) | 65 | 3714 | 787.46 | 620.86 |
| 리뷰 내용의 긍정점수 | 0 | 5.42 | 2.50 | .72 |
| 리뷰 내용의 부정점수 | 0 | 1.53 | .61 | .24 |
| 리뷰 제목의 긍정점수 | 0.2 | 4.87 | 1.17 | 1.15 |
| 리뷰 제목의 부정점수 | 0 | 1.83 | .26 | .29 |
| 제목의 단어 개수 | 0 | 5.40 | 4.14 | .43 |
| 내용의 단어 개수 | 0 | 66.67 | 35.84 | 10.04 |

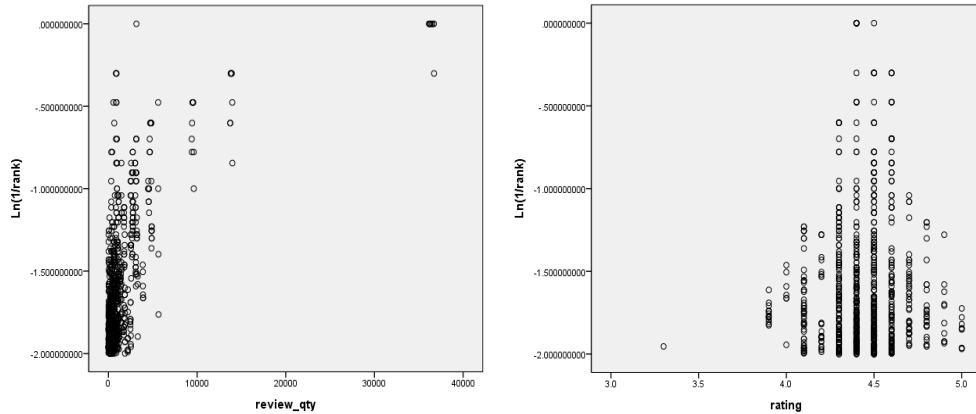
판매 랭킹(Rank)은 아마존 베스트셀러 랭킹은 1위부터 100위까지를 의미한다. 가격(price)은 베스트셀러 제품들의 가격을 의미하는 것으로 이때 가격의 평균은 약 69.33달러이며 최저 가격은 7.98달러, 최고 가격은 379.95달러이다. 별점(rating)은 소비자가 평가한 제품의 평균 별점이다. 베스트셀러 제품들의 별점은 평균 약 4.44점을 기록하고 있는데 5점을 만점으로 보았을 때 전반적으로 상품들이 높은 평점을 가지고 있음을 알 수 있다. [그림 4-1]은 가격과 별점의 분포 그래프이다.



[그림 4-1] 별점(rating)(상)과 가격(price)(하)의 분포 그래프

리뷰의 양(review_qty)은 소비자들이 평가한 리뷰의 개수로 평균적으로 각 품목당 약 1618개의 리뷰를 가지고 있다. 그러나 표준편차가 3977.97임을 보았을 때 제품 간의 리뷰의 수 격차가 상당히 크다는 점을 알 수 있다.

[그림 4-2]는 별점과 리뷰의 양의 종속변수에 대한 산점도를 나타내고 있는데, 산점도의 분포를 살펴보았을 때 별점과 리뷰의 양이 한쪽으로 치우쳐 있는 것을 알 수 있다. 이 같은 분포를 비교하기 위해 별점과 리뷰의 양의 값을 각각 로그로 변환하여 사용하였다.

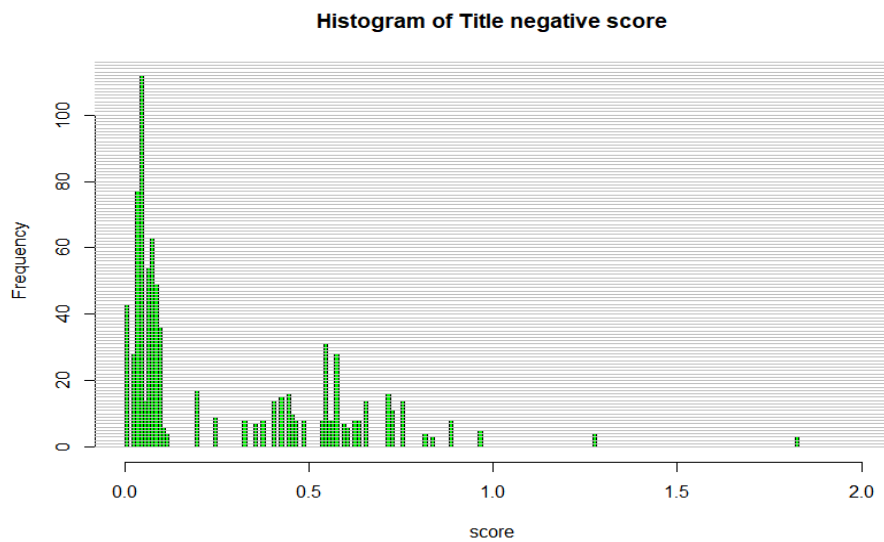
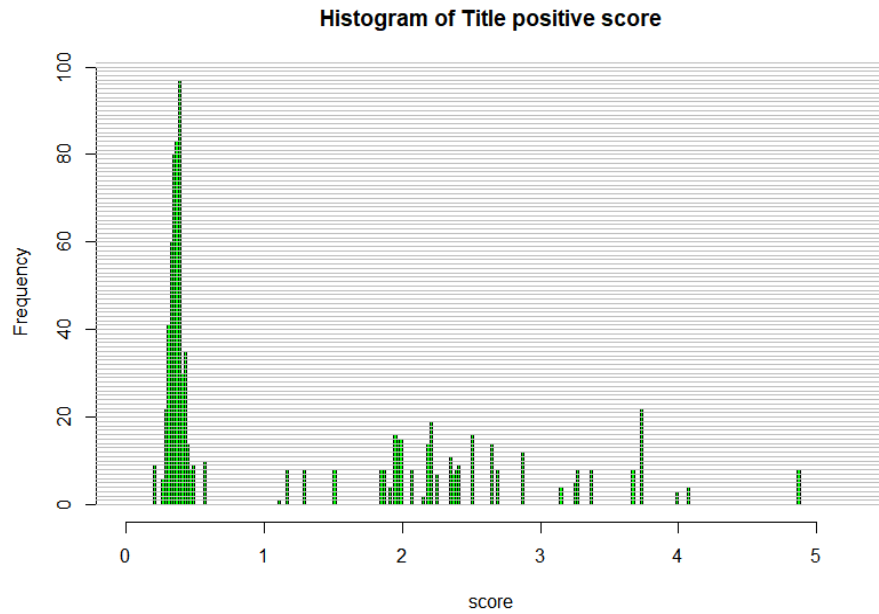


[그림 4-2] 리뷰의 양과 평점의 산점도

사진을 포함한 리뷰의 수는 사진을 포함한 리뷰의 개수들로, 제품 당 평균 24.44개의 사진이 포함되었다.

상품 게시일(Days on amazon)은 상품이 아마존에 게시된 날로부터 분석날짜까지의 일수를 나타낸다. 숫자가 클수록 오래전에 게시된 상품이며 가장 최근에 게시된 일수는 65일(약 9주)전 이었고 가장 오래된 일수는 3714일(약 530주) 전이었다. 그러나 Dataset에서 8일차 데이터를 모두 병합하였을 때 152개의 모델 중 51개 모델의 게시일이 나와 있지 않아 총 800개의 데이터 중 712개를 구할 수 있었다.

내용의 긍정점수(Contents positive score)의 평균은 약 2.5개로 이는 제품의 리뷰 당 평균 2.5개의 긍정적인 단어를 쓴 것을 의미한다. 리뷰 내용의 부정점수(Contents negative score)의 평균은 약 0.61로, 소비자가 리뷰 내용에 평균적으로 부정적인 단어를 0.61개 사용했다는 것을 알 수 있다. 제목의 긍정점수(Title positive score)는 제목에 쓰인 긍정적인 단어의 개수로 평균 1.17개 정도의 긍정적 단어를 사용한다는 것으로 나타났다. 마찬가지로 제목의 부정점수(Title negative scores)는 제목에 쓰인



[그림 4-3] 제목의 긍정점수(상)와 부정점수(하)의 분포도

부정적인 단어의 개수로 평균 0.26개의 부정적 단어가 쓰였음을 알 수 있다. [그림 4-3]은 제목의 긍정점수와 부정점수의 분포도이다.

제목의 단어 개수(Title word count)는 제목에 쓰인 단어의 개수를 의미

하며 평균적으로 4개의 단어가 사용되었다. 내용의 단어 개수(Contents word count) 또한 내용을 이루고 있는 단어의 개수로 평균적으로 약 36개의 단어가 쓰인 것으로 나타났다.

4.2 가설 검증

본 연구의 가설 검증을 위해 조절회귀분석(regression)을 실시하였다. 종속변수로는 매출(rank)로 지정했고, 독립변수로는 브랜드 명성, 리뷰의 양, 별점, 유용성, 리뷰 내용의 긍정점수, 리뷰 내용의 부정점수, 리뷰 제목의 긍정점수, 리뷰 제목의 부정점수로 지정하였다. 통제변수로는 가격, 사진을 포함한 리뷰의 개수, 상품 게시일, 리뷰 제목의 단어 개수, 리뷰 내용의 단어 개수로 이루어졌다. 또한 상호작용에 대한 변수로는 브랜드력*리뷰의 양, 브랜드력*리뷰 내용의 긍정(부정)점수, 브랜드력*리뷰 제목의 긍정(부정)점수가 있었다.

독립변수와 종속변수 간에 상관관계는 $R = .687$ 로 나타났고 수정된 R^2 은 .452로 45.2%의 설명력을 가졌다. 또한 F 값이 $p < .00$ 에서 24.480로 회귀모형이 적합한 것으로 드러났다. <표 4-2>는 회귀분석 결과표이다.

<표 4-2> 회귀 분석 결과

| 모형 | 비표준 계수 | 표준 계수 | t | 유의수준 |
|-------------------|----------|-------|--------|------|
| | B | 베타 | | |
| (상수) | -3.700 | | -8.474 | .000 |
| 가격 | .001 | .071 | 2.018 | .044 |
| log(리뷰의 양) (A) | .487 | .774 | 18.626 | .000 |
| 리뷰 내용의 긍정 점수 (B) | -.044 | -.074 | -1.621 | .105 |
| 리뷰 내용의 부정 점수 (C) | -.143 | -.085 | -2.049 | .041 |
| 리뷰 제목의 긍정 점수 (D) | .094 | .268 | 3.616 | .000 |
| 리뷰 제목의 부정 점수 (E) | -.333 | -.240 | -3.030 | .003 |
| log(별점) | 2.123 | .103 | 3.365 | .001 |
| Helpfulness의 총 개수 | 4.783E-5 | .118 | 4.079 | .000 |
| 브랜드 명성 | -.133 | -.154 | -4.571 | .000 |
| 브랜드 명성 * (A) | -.190 | -.183 | -4.876 | .000 |
| 브랜드 명성 * (B) | -.132 | -.145 | -3.298 | .001 |
| 브랜드 명성 * (C) | .380 | .118 | 3.170 | .002 |
| 브랜드 명성 * (D) | -.122 | -.198 | -2.542 | .011 |
| 브랜드 명성 * (E) | .417 | .189 | 2.300 | .022 |
| 상품 게시일 | .000 | -.188 | -5.232 | .000 |
| 리뷰 제목의 단어 개수 | -.108 | -.111 | -2.927 | .004 |
| 리뷰 내용의 단어 개수 | .002 | .046 | 1.191 | .234 |
| 사진을 포함한 리뷰의 개수 | .000 | .090 | 3.065 | .002 |
| t1 | .031 | .024 | .663 | .508 |
| t2 | .025 | .020 | .537 | .591 |
| t3 | .036 | .028 | .764 | .445 |
| t4 | .039 | .031 | .847 | .397 |
| t5 | .008 | .006 | .168 | .867 |
| t6 | .020 | .016 | .441 | .660 |
| t7 | .009 | .008 | .212 | .832 |

가설 1에서는 온라인 리뷰의 양이 많을수록 매출에 긍정적인 영향을 미칠 것이라고 예상하였다. 회귀 분석 결과, 리뷰의 양이 많을수록 매출

이 증가하는 것으로 나타났고($\beta=.774$, $p<.000$), 변수들 중에 리뷰의 양이 매출에 가장 큰 영향력을 가지는 것으로 나타났다. 그러므로 가설 1을 채택하였다.

가설 2에서는 별점이 높을수록 매출이 높을 것이라고 예상하였다. 회귀 분석 결과, $\log(\text{별점})$ 이 매출에 미치는 영향은 유의한 것으로 나타났으며($\beta=.103$, $p<.001$), β 가 양의 값을 가지므로 별점이 높을수록 매출에 긍정적인 영향을 미친다는 것이 나타났다. 따라서 가설 2를 채택하였다.

가설 3에서는 리뷰의 유용성이 높을수록 매출이 높다고 예상하였는데, 분석 결과 Helpfulness의 총 개수가 종속변수에 미치는 영향이 유의한 것으로 나타났다($\beta=.118$, $p<.000$). 이를 통해 helpfulness의 수가 높을수록 매출에 긍정적인 영향을 미친다는 결과를 얻을 수 있다. 따라서 가설 3을 채택하였다.

가설 4에서는 리뷰에서 나타난 감성의 정도(감성지수)가 매출에 미치는 영향력을 살펴보고자 수립하였다. 회귀 분석 결과 리뷰의 제목에서 나타난 감성정도가 리뷰의 내용에서 나타난 감성정도보다 매출에 미치는 영향이 더 큰 것으로 나타났다. 이는 소비자가 전체적으로 리뷰의 내용보다 리뷰의 제목에서 나타난 감성정도에 더 많은 영향을 받는다는 것을 알 수 있다.

가설 4-1은 리뷰 제목의 긍정(부정) 점수가 매출에 영향을 미칠 것이라고 예상하였는데 회귀 분석 결과, 긍정 점수($\beta=.268$, $p<.000$)와 부정 점수($\beta=-.240$, $p<.003$)의 영향이 유의한 것으로 나타났다. 긍정 점수의 경우 β 값이 양의 값을 가지므로 리뷰의 제목에 긍정적인 단어가 많이 쓰일수록 매출에 긍정적인 영향을 미친다는 것을 알 수 있다. 반면에 부정 점수의 β 값은 음수를 가지므로 부정 단어가 많이 쓰일수록 매출에 부정적인 영향을 미친다는 것을 알 수 있다. 따라서 가설 4-1을 부분적으로

채택하였다.

가설 4-2에서는 리뷰의 내용에서 쓰인 긍정(부정) 점수가 매출에 영향을 미칠 것이라고 수립하였는데 긍정 점수의 경우, 매출에 유의하지 않은 것으로 나타난 반면, 리뷰의 내용에서 쓰인 부정 점수는 매출에 영향을 미치는 것으로 나타났다($\beta = -.085, p < .05$). 이 때 β 값이 음의 값을 가졌는데, 이는 리뷰의 부정단어가 많을수록 매출에 부정적인 영향을 미친다는 것을 나타낸다.

마지막으로 가설 5를 검증하기 위해 브랜드 명성과 리뷰의 양, 감성지수의 상호작용을 살펴보았다. 먼저 브랜드 명성이 매출에 미치는 영향을 살펴보았는데, 회귀분석 결과 브랜드 명성이 매출에 미치는 영향은 유의한 것으로 나타났다($\beta = -.154, p < .000$). 또한 β 값이 음수의 값을 가지므로 강한 브랜드보다 약한 브랜드의 경우가 매출이 높은 것을 알 수 있다.

다음은 브랜드 명성과 리뷰의 상호작용에 대해 분석하였는데 그 결과, 브랜드 명성과 리뷰의 양, 감성 지수의 상호작용 모두 영향을 미치는 것으로 나타났다. 브랜드 명성과 리뷰의 양의 β 값은 음수로 나타났으며($\beta = -.183, p < .000$) 이를 통해 약한 브랜드의 경우 리뷰의 양이 매출에 미치는 영향은 강한 브랜드보다 더 크다는 의미를 도출할 수 있다. 따라서 가설 5-1을 채택하였다.

브랜드의 명성과 리뷰의 내용에서 나타난 감성지수의 상호작용 또한 유의한 것으로 나타났다. 특히, 가설 4-3에서 리뷰의 내용에서 나타난 긍정 점수는 매출에 미치는 영향에 대해서 유의하지 않은 결과가 나왔지만 브랜드와 내용의 긍정 점수의 상호작용은 유의한 것으로 나타났다($\beta = -.145, p < .001$). 즉 브랜드의 명성이 약할 때, 강한 브랜드의 제품보다 리뷰에서 나타난 긍정적인 단어가 매출에 미치는 영향이 높다고 해석할 수 있다. 브랜드 명성과 리뷰의 내용에서 나타난 부정 점수의 상호작용

또한 위와 같이 유의하였는데($\beta=.118$, $p<.002$), 브랜드가 약한 경우, 강한 경우보다 리뷰의 내용에서 부정 점수가 매출에 미치는 영향이 컸으며, 이는 약한 브랜드인 경우 강한 브랜드보다 리뷰에서 사용된 부정적 단어가 적을수록 매출이 높은 것으로 나타났다. 따라서 가설 5-2를 채택하였다.

브랜드 명성과 리뷰의 제목에서 나타난 감성 지수의 상호작용 또한 유의한 것으로 나타났다. 리뷰의 제목에서 나타난 긍정 점수의 경우, β 값이 음수로 나타났는데($\beta=-.198$, $p<.011$), 이는 약한 브랜드인 경우(-) 강한 브랜드보다 리뷰의 제목에서 나타난 긍정 점수가 높을수록(+) 매출에 긍정적인 영향을 준다는 것을 의미한다. 반면 브랜드 명성과 리뷰의 제목에서 나타난 부정 점수의 상호작용의 β 값이 양수로 나타났다($\beta=.189$, $p<.022$). 이를 통해 브랜드가 약한 경우에 강한 경우보다 리뷰의 제목에서 부정적인 단어가 적게 쓰일수록 매출이 높다는 결과를 얻을 수 있다. 따라서 가설 5-3 또한 채택한다.

이외에도 통제변수가 종속변수에 미치는 영향을 살펴보면, 상품의 가격이 매출에 미치는 영향이 유의한 것으로 나타났다($\beta=.071$, $p<.044$). 이때 β 값이 양수를 가지므로 가격이 높은 상품일수록 매출이 높은 것으로 드러났다. 상품 게시일 또한 종속변수에 미치는 영향력이 유의했으며($\beta=-.188$, $p<.000$), β 값이 음수를 가지는 것으로 나타났다. 이는 아마존에 상품 게시일이 오래되지 않을수록, 다시 말해서 상품이 최신이거나 아마존에 게시된 시간의 경과가 짧을수록 매출이 높은 것을 의미한다.

리뷰 제목의 단어 개수가 매출에 미치는 영향 또한 유의하게 나타났다($\beta=-.111$, $p<.004$). β 값이 음수를 가지고 있으므로, 단어의 개수가 짧을수록 매출이 높은 것을 뜻하는데, 이는 소비자가 리뷰의 제목이 긴 것보다 단순하고 정확한 것을 선호한다고 해석할 수 있다. 반면 리뷰 내용의 단

어 개수는 매출에 영향을 미치지 않는 것으로 나타났다.

사진을 포함하고 있는 리뷰의 개수가 매출에 미치는 영향은 유의한 것으로 나타났다($\beta=.090$, $p<.01$). 이는 사진이 포함된 리뷰의 개수가 많을수록 매출이 높은 제품이라고 해석할 수 있다.

또한 마지막으로 가장 최근인 t8을 기준으로 수집 기간이 랭킹에 미치는 영향을 살펴보았다. 그 결과, t1부터 t7까지 모두 유의하지 않은 것으로 나타났다. 이는 특정 기간의 랭킹이 다른 기간의 랭킹에 영향을 미치지 않는다는 것을 뜻하며, 기간 별 랭킹의 차이 또한 없다는 것을 의미한다.

제 5 장 결론

5.1 연구 결과 요약 및 시사점

본 논문에서는 온라인 구전의 중요성을 인식하고 온라인 구전의 대표적 형태인 브랜드 명성에 따른 리뷰의 영향력에 대해 연구하였다. 리뷰의 양과 별점, 유용성, 리뷰에서 나타난 감성의 정도가 매출에 영향을 미치는지를 분석하였는데, 특히 정량적 정보(별점, 가격 등)보다 정성적 정보(텍스트에서 감지된 감성)가 매출에 더 큰 영향을 미칠 것이라고 판단하여 연구를 진행하였다. 또한 브랜드 명성에 따라 리뷰가 매출에 미치는 영향이 다를 것이라고 예상하고 브랜드 명성과 리뷰의 상호작용을 분석하였다.

연구에 사용된 데이터는 아마존의 블루투스 스피커의 베스트셀러 제품으로 2019년 8월 14일부터 9월 7일까지 주중/ 주말로 나누어 약 8번의 데이터 수집과정을 거쳐 데이터를 완성했다. R을 이용하여 제품 데이터를 크롤링을 하였고, 정제 및 병합 과정을 거쳐 감성분석을 수행하였다. 감성분석에는 ‘Opinion lexicon’을 바탕으로 단어 수정 및 변형을 거쳐 사용하였고 그 이후 SPSS를 이용하여 회귀분석을 진행하였다.

연구 결과, 온라인 리뷰의 별점과 유용성과 같은 정량적 정보보다 리뷰에서 나타난 감성(긍정/부정)의 정도가 매출에 더 큰 영향을 미치는 것으로 나타났다. 결론을 요약하자면 다음과 같다.

첫째, 리뷰의 양과 별점이 매출에 미치는 영향이 유의한 것으로 나타났다. 특히 매출에 영향을 미치는 요인은 리뷰의 양인 것으로 나타났으며, 많은 연구에서 밝혀졌듯이, 본 연구에서도 정량적 정보에 해당하는 별점과 리뷰의 양이 매출에 영향을 미치는 것으로 드러났다.

둘째, 감성의 정도가 매출에 영향을 미친다는 점을 확인하였다. 연구 결과, 리뷰의 제목에서 나타난 긍정점수와 부정점수 모두 매출에 긍정적, 부정적인 영향을 끼쳤던 반면 리뷰의 내용에서는 부정점수만이 매출에 부정적인 영향을 주었고 긍정점수는 영향을 미치지 않았다. 또한 소비자는 리뷰의 제목과 내용 중 제목의 감성지수의 영향을 더 많이 받는 것으로 나타났다. 이는 소비자가 휴리스틱에 의해 비교적 짧은 길이의 제목을 보고 구매 의사결정을 하는 경향이 크다고 해석할 수 있다.

셋째, 브랜드 명성에 따른 리뷰의 양과 감성정도의 상호작용 또한 매출에 영향을 미치는 것으로 나타났다. 브랜드 명성이 낮은 경우 높은 경우보다 리뷰의 양이 많을수록 매출이 높은 결과가 나타났고, 리뷰의 제목과 내용의 긍정(부정)점수가 매출에 긍정(부정)적인 영향을 미친다는 결과가 나타났다. 이는 제품의 브랜드의 인지도가 낮은 경우, 강한 브랜드에 비해 신뢰도가 낮으므로 소비자가 리뷰를 통해 더 많은 정보를 파악하고 구매 의사결정에 적극적으로 활용한다는 결과로 해석할 수 있다.

본 연구의 학문적 시사점은 다음과 같다.

첫째, 빅데이터를 통해 리뷰의 영향력을 분석하였다. 세계적인 전자상거래 시장인 아마존의 블루투스 데이터를 이용하여 매출에 영향을 미치는 요인을 다양하게 확장하였다. 기존의 선행 연구에서 매출에 대한 독립변수를 리뷰의 양과 별점, 가격 등과 같은 정량적인 요소들에 집중하여 분석한 반면 본 논문에서는 이외에도 텍스트의 긍,부정의 정도, 유용성, 텍스트의 단어 개수, 브랜드를 포함하여 심도 있는 분석을 통합적으로 진행하였다. 또한 브랜드 명성에 집중하여 브랜드 명성과 리뷰의 여러 요소 간에 상호작용을 분석하여 매출에 영향을 미치는 다양한 요인을 발견하고 유의한 결과를 나타냈다.

둘째, 감성 분석의 새로운 방식을 제시하였다. 세 가지 감성사전을 비교

분석하여 Bing Liu가 제작한 Opinion Lexicon을 이용하여 감성분석을 시행하였다. 또한 bigram 방식을 통해 부정어의 의미를 잡아내었고 특징별 쓰이는 단어의 감성을 수정하여 소비자의 감성분석의 정확도를 향상시켰다. 기존의 연구에서는 감성의 정도를 파악하기 위하여 긍정 점수에서 부정 점수를 차감하여 도출한 계산식을 사용하였는데, 본 연구에서는 긍정 점수와 부정 점수를 따로 계산하여 긍정적인 감정과 부정적인 감정이 매출에 각각 어떤 영향을 미치는 지를 파악하였다.

본 연구의 실무적 시사점은 다음과 같다.

첫째, 기업들에게 리뷰의 중요성을 강조하였다. 연구 결과에서 리뷰의 양과 별점뿐만 아니라 리뷰의 내용 또한 소비자의 구매 의사결정에 영향을 미치는 것으로 나타났는데, 이는 기업들에게 리뷰의 중요성을 인식시키고 리뷰 관리에 더욱 더 집중할 수 있도록 시사점을 제안하였다. 특히 긍정적인 내용의 리뷰보다 부정적인 리뷰의 영향력이 더 크므로, 부정적인 리뷰의 처리방안이나 대응책에 대해 강구할 필요성을 제시한다.

둘째, 약한 브랜드의 마케팅 전략에 도움을 줄 수 있다. 본 논문에서는 브랜드 명성이 높을 때와 낮은 경우로 구분하여 리뷰의 영향력을 분석하였는데, 브랜드 명성이 낮은 경우 높은 경우보다 리뷰의 감성지수가 매출에 미치는 영향이 크다는 결과를 얻었다. 내용과 제목에서 나타난 긍정/부정 점수와 브랜드 명성의 상호작용효과가 매출에 미치는 영향이 유의하게 나타났고, 특히 주 효과였던 내용의 긍정 점수는 매출에 미치는 영향이 유의하지 않은 반면, 상호작용을 분석하였을 때 유의한 것으로 나타났다. 이러한 결과를 통해 브랜드 명성이 낮은 경우 긍정적인 내용의 리뷰 생성에 초점을 맞출 필요성을 제시한다. 긍정적인 리뷰는 매출에 직접적인 이익을 주는 것으로 드러났기 때문이다. 긍정적인 리뷰의 양이 증가하면 전체 리뷰의 양이 증가하게 되고, 이것이 매출에 영향을

주게 되어 긍정적인 순환작용이 일어나게 된다. 이는 낮은 명성의 브랜드가 시장에서 주도권을 잡을 수 있도록 도움을 줄 수 있음을 명시한다. 결론적으로 해당 논문에서 얻은 결과는 기업이 소비자의 의견과 입장을 분명하게 이해하고 향후 매출을 증진시키기 위한 전략을 제시하는 데에 도움이 될 것이다.

5.2 연구의 한계점 및 향후 연구 방향

본 연구의 한계점은 다음과 같다. 첫째, 감성분석의 개선이 필요하다고 판단하였다. 본 연구에서는 전반적인 감성에 대해 분석을 시행하였지만 향후에 보다 정교한 감성 연구의 진행이 필요하다. 연구에서 강도 측정을 수행하지 않았는데, 정확한 극성단어를 구분해내어 텍스트 분석을 수행한다면 매출에 미치는 영향을 더욱 심도 있게 이해할 것이다. 더불어 단어들의 상관관계 분석의 필요성이 대두된다. 제품의 특징을 나누어 특징별 감성을 분석하면 보다 세밀한 소비자 의견을 얻을 수 있을 것이라 예상한다.

또한 감성사전의 개선이 필요하다고 판단되었다. 제품별 특징에 따라 단어의 감성이 바뀌는 경우가 많아서 제품 특징 별 단어의 유동적인 변환이 필요하다. 또한 한 단어가 이중적 의미를 가지는 경우가 있기 때문에 감성의 구분을 지어야 보다 정확한 감성분석을 수행할 수 있다. 마지막으로 신조어나 인터넷 채팅용어가 반영되어야 한다. 사회 현상이나 흐름에 따라 많은 단어가 사라지고 생성되므로 이를 반영할 감성단어사전이 꾸준히 개선되고 갱신되어야 할 필요성을 가지고 있다.

둘째, 제품의 특징에 따른 다른 제품(경험재)의 분석이 필요하다.

본 논문에서는 탐색재에 대한 리뷰의 영향력 분석에 대해 연구하였으므

로 경험재와 비교 분석 연구가 필요하다. 특히 기존 연구에서는 경험재가 리뷰의 영향력을 더 많이 받는다는 결과가 있었는데, 이는 정량적 특성(리뷰의 양, 별점 등)에 따른 영향력을 분석한 결과가 대다수였다. 따라서 텍스트 분석을 통해 정성적 정보의 측면에서 리뷰의 영향력을 재비교할 필요성이 있다.

참고문헌

- 권진영 · 이문용 (2012), “온라인 리뷰 유용성의 결정 요인에 대한 연구”,
한국지능정보시스템학회 학술대회논문집, 205-211.
- 김도경 · 김인신 (2017), “텍스트 마이닝을 이용한 온라인 리뷰의 호텔
선택 속성 분석”, 관광학연구, 41(9), 109-127.
- 박찬 · 유창조 (2006), “온라인에서의 구전커뮤니케이션이 상표평가에
미치는 영향에 관한 연구 : 상표 사용 후기와 답글을 중심으로”,
소비학연구, 제17권, 제1호, pp.73-93.
- 심영석 · 김홍범 (2016), “텍스트 마이닝을 이용한 관광지 이미지 구성
요인 및 측정에 관한 연구”, 관광학연구, 40(7), 221-245.
- 이진화 · 김지은 · 안현아 (2016), 소비자의 가격민감도에 따른 상품
특성별 멀티채널 선택에 관한 연구-경험재로서의 의류상품과
탐색재로서의 디지털 가전제품을 중심으로-.한국의류학회지, 40(6),
967-978.
- 이호근 · 광현 (2013), 온라인 소비자 리뷰의 효과에 영향을 미치는
요인에 대한 고찰. 정보화정책저널, 20(3), 0-0
- 임영희 (2019), “온라인 리뷰 빅데이터를 활용한 외래 관광객의 국내호텔
이용 만족도에 관한 연구”, 세종대학교 대학원
- 조영 · 김세범 (2014), “온라인 구전의 방향성과 유형이 구매의도에
미치는 영향”, 한국경영학회 통합학술발표논문집, 280-304.
- 조준한 · 안승호 · 손영돈 (2012), “넷글의 방향성과 웹사이트의 신뢰성이
서비스 브랜드 태도와 구매의도에 미치는 영향”, 한국 IT 서비스
학회지, 11(2): 69-91
- 황침 · 권영진 · 이상용 (2018), “온라인 리뷰 유용성과 상품매출에 영

- 향을 주는 요인”, *Journal of Information Technology Applications & Management*, 25(2), 53-72.
- Amblee, Naveen, and Tung Bui (2011), "Harnessing the influence of social proof in online shopping: The effect of electronic word of mouth on sales of digital microproducts." *International journal of electronic commerce* 16(2), 91-114.
- Archak, Nikolay, Anindya Ghose, and Panagiotis G. Ipeirotis (2011), "Deriving the pricing power of product features by mining consumer reviews." *Management science* 57(8), 1485-1509.
- Arizton (2018), Bluetooth Speaker Market in US - Industry Outlook and Forecast 2018-2023, <https://www.arizton.com/market-report/us-bluetooth-speaker-market-analysis-size-share-growth/snapshots>
- Bei, L. T., Chen, E. Y. and Widdows, R.(2004), "Consumers' online information search behavior and the phenomenon of search vs. experience products." *Journal of Family and Economic Issues*, 25 (4), 449-467.
- Bosman, Dirk J., Christo Boshoff, and Gert-Jan Van Rooyen (2013), "The review credibility of electronic word-of-mouth communication on e-commerce platforms." *Management Dynamics: Journal of the Southern African Institute for Management Scientists*, 22 (3), 29-44.
- Brynjolfsson, Erik, Yu Hu, and Duncan Simester. (2011), "Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales." *Management Science* 57.8, 1373

- Brynjolfsson, Erik, Yu Hu, and Michael D. Smith.(2003) "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers." *Management Science* 49 (1), 1580-1596.
- Campbell, Colin, Leyland. F. Pitt, Michael Parent, and Pierre. R. Berton (2011), "Understanding consumer conversations around ads in a Web 2.0 world," *Journal of Advertising*, 40(1), 87-102.
- Chen, Pei-Yu, Shin-yi Wu, and Jungsun Yoon (2004), "The impact of online recommendations and consumer feedback on sales." *II S 2004 Proceedings*: 58.
- Chevalier, Judith A., and Dina Mayzlin (2006), "The effect of word of mouth on sales: Online book reviews." *Journal of marketing research*, 43(3), 345-354.
- Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets." *Marketing Science*, 29(5), 944-957.
- Chiou, J.S. and Cheng, C. (2003), "Should a company have message boards on its web sites", *Journal of Interactive Marketing*, Vol. 17 No. 3, pp. 50-61.
- Cui, Geng, Hon-Kwong Lui, and Xiaoning Guo (2012), "The effect of online consumer reviews on new product sales." *International Journal of Electronic Commerce*, 17(1), 39-58.
- Daft, Richard L., and Robert H. Lengel (1983), *Information richness. A*

- new approach to managerial behavior and organization design.*
No. TR-ONR-DG-02. Texas A and M Univ College Station Col
l of Business Administration.
- Dholakiya, P. (2014), Are excessively positive customer reviews hurting
g your conversion rates? [Study].[http://unbounce.com/conversion
-rate-optimization/customer-reviews-conversion-rates](http://unbounce.com/conversion-rate-optimization/customer-reviews-conversion-rates).
- Doyle, P., 1990. "Building successful brands: The strategic options." *J
ournal of Consumer Marketing* 7 (Spring), S-20.
- Erdem, Tülin, and Joffre Swait (1998), "Brand equity as a signaling p
henomenon." *Journal of consumer Psychology*, 7(2), 131-157.
- Erdem, Tülin, Joffre Swait, and Ana Valenzuela. "Brands as signals:
a cross-country validation study." *Journal of marketing* 70(1), 3
4-49.
- Forman, Chris, Anindya Ghose, and Batia Wiesenfeld (2008), "Examini
ng the relationship between reviews and sales: The role of revi
ewer identity disclosure in electronic markets." *Information syste
ms research*, 19(3), 291-313.
- Ghose, Anindya, and Panagiotis G. Ipeirotis (2011), "Estimating the he
lpfulness and economic impact of product reviews: Mining text
and reviewer characteristics." *IEEE Transactions on Knowledge
and Data Engineering*, 23(10), 1498-1512.
- Ghose, Anindya, and Arun Sundararajan (2005) "Evaluating Pricing an
d Product Line Strategy Using e-Commerce Data: Evidence and
Estimation Challenges." *Statistical Science*, 27 (6), 995-1011.
- Gopinath, S., Thomas, J.S., & Krishnamurthi, L. (2014), "Investigating

- the Relationship Between the Content of Online Word of Mouth, Advertising, and Brand Performance." *Marketing Science*, 33, 241-258.
- Ho-Dac, Nga N., Stephen J. Carson, and William L. Moore (2013), "The effects of positive and negative online customer reviews: do brand strength and category maturity matter?." *Journal of Marketing*, 77(6), 37-53.
- Huang, Jen Hung, and Yi Fen Chen (2006), "Herding in online product choice." *Psychology & Marketing* 23(5), 413-428.
- Hu, Nan, Noi Sian Koh, and Srinivas K. Reddy (2014), "Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales." *Decision support systems*, 57, 42-53.
- Rao, Akshay R., Lu Qu, and Robert W. Ruekert. (1999), "Signaling unobservable product quality through a brand ally." *Journal of Marketing Research* 36(2), 258-268.
- Shimp, Terence A., and William O. Bearden (1982), "Warranty and other extrinsic cue effects on consumers' risk perceptions." *Journal of Consumer research* 9(1), 38-46.
- Schiffman, Leon G., Leslie Lazar Kanuk, and Joseph Wisenblit. "Consumer behavior." *Upper Saddle River, NJ* (2007).
- Schoenmüller, Verena, Oded Netzer, and Florian Stahl (2018), "The extreme distribution of online reviews: Prevalence, drivers and implications." *Columbia Business School Research Paper* 18-10.
- Singh, J. P., Seda Irani, Nripendra P. Rana, Yogesh K. Dwivedi, Sunil

- Saumya,, and Pradeep Kumar Roy (2017), "Predicting the "helpfulness" of online consumer reviews.", *Journal of Business Research*, 70, 346-355.
- Simonson, Itamar (2014), "Mission Accomplished: What's Next for Consumer BDT-JDM Researchers?."
- Silge, Julia, and David Robinson. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. J. Open Source Software, 1(3), 37.
- Spool, J. M. (2009), The Magic Behind Amazon's 2.7 Billion Dollar Question, from www.uie.com/articles/magicbehindamazon/2009
- Stiglitz, Joseph E. (1989), "Imperfect information in the product market." *Handbook of industrial organization* 1, 769-847.
- Jónsson, Kristján, and Devin Platt (2016), "CSE 255 Assignment 1: Helpfulness in Amazon Reviews."
- Kiecker, Pamela, and Deborah Cowles (2002), "Interpersonal communication and personal influence on the Internet: A framework for examining online word-of-mouth." *Journal of Euromarketing*, 11 (2), 71-88.
- Kirmani, Amna. (1990), "The effect of perceived advertising costs on brand perceptions." *Journal of consumer research* 17(2), 160-171.
- Khare, Adwait, Lauren I. Labrecque, and Anthony K. Asare. (2011), "The assimilative and contrastive effects of word-of-mouth volume: An experimental examination of online consumer ratings.", *Journal of Retailing* 87(1), 111-126.
- Liu, Bing.(2010), "Sentiment analysis and subjectivity." *Handbook of n*

atural language processing 2.2010, 627–666.

Liu, Yong (2006), "Word of mouth for movies: Its dynamics and impact on box office revenue." *Journal of marketing* 70(3), 74–89.

Maslowska, Ewa, Edward C. Malthouse, and Stefan F. Bernritter (2017), "Too good to be true: the role of online reviews' features in probability to buy." *International Journal of Advertising* 36.1, 142–163.

Matfield, K (2011), "The Impact of Customer Reviews and Ratings on Conversion Rates."

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, ... & Steven Pinker (2011). "Quantitative analysis of culture using millions of digitized books." *science*, 331(6014), 176–182.

Mordor Intelligence (2018), LUETOOTH SPEAKER MARKET – GROWTH, TRENDS, AND FORECAST (2019 – 2024), <https://www.mordorintelligence.com/industry-reports/global-bluetooth-speaker-industry>

Mudambi, Susan M., and David Schuff (2010), "What makes a helpful review? A study of customer reviews on Amazon. com." *MIS quarterly*, 34(1), 185–200.

Nielsen (2012), "Consumer trust in online, social and mobile advertising grows."

Power Review (2012), <http://socialcommercetoday.com/2010-social-shopping-study-top-line-results/>

Wernerfelt, Birger (1988), "Umbrella branding as a signal of new prod

- uct quality: An example of signalling by posting a bond." *The RAND Journal of Economics*, 458–466.
- Wickham, Hadley, and Henry Lionel (2017), Tidy: Easily Tidy Data with ‘spread ()’and ‘gather ()’functions. *R package version 0.6.1*.
- Willemsen, Lotte M., Peter C. Neijens, Fred Bronner, Jan A. de Ridder (2011), ""Highly recommended!" The content characteristics and perceived usefulness of online consumer reviews." *Journal of Computer-Mediated Communication*, 17.1, 19–38.
- Yang, Jing, Rathindra Sarathy, and Stephen M. Walsh (2016), "Do review valence and review volume impact consumers' purchase decisions as assumed?." *Nankai Business Review International* 7 (2), 231–257.
- Zhao, Jun, Kang Liu, and Liheng Xu. (2016), "Sentiment analysis: mining opinions, sentiments, and emotions.", 595–598.

부 록

[부록1] 상관관계 표

| | Ln(1/rank) | Price | Brand strength | log (Reivew_quantity) | log(Rating) | Sum of helpfulness | Days on amazon |
|-------------------------------------|------------|-------|----------------|-----------------------|-------------|--------------------|----------------|
| Pearson 상관계수 | Ln(1/rank) | | | | | | |
| Price | 1.000 | | | | | | |
| Brand strength | -.191 | 1.000 | | | | | |
| log(Reivew_quantity) | -.176 | .470 | 1.000 | | | | |
| log(Rating) | .561 | -.181 | -.181 | 1.000 | | | |
| Sum of helpfulness | .055 | -.092 | .161 | -.148 | 1.000 | | |
| Days on amazon | .232 | -.073 | -.019 | .131 | .075 | 1.000 | |
| Contents positive score | .122 | -.039 | -.101 | .512 | -.298 | .006 | 1.000 |
| Contents negative score | -.145 | .191 | -.035 | .011 | .056 | .020 | -.071 |
| Title positive score | -.026 | .006 | -.019 | .126 | -.006 | -.024 | .193 |
| Title negative score | .103 | -.042 | -.002 | .045 | .082 | .032 | .029 |
| Title wordcount | .062 | -.033 | .021 | .040 | .063 | .030 | .048 |
| Contents wordcount | -.046 | -.058 | -.072 | .026 | -.035 | .012 | .039 |
| Number of reviews including Picture | -.094 | -.099 | -.043 | -.170 | .060 | -.033 | -.086 |
| Brand * Review quantity | .125 | -.060 | .008 | .048 | -.016 | .146 | .024 |
| Brand * Contents positive score | .219 | -.033 | -.200 | .616 | -.185 | .069 | .339 |
| Brand * Contents negative score | -.093 | .270 | -.017 | .109 | .037 | -.053 | .004 |
| Brand * Title positive score | .044 | .058 | .005 | .103 | .001 | -.002 | .045 |
| Brand * Title negative score | -.004 | -.098 | .013 | -.082 | .096 | .049 | -.096 |
| t1 | -.018 | -.058 | .038 | -.088 | .073 | .052 | -.076 |
| t2 | .001 | -.008 | .048 | .000 | .003 | .015 | .051 |
| t3 | -.006 | .003 | .007 | -.001 | .029 | -.001 | .065 |
| t4 | .016 | .010 | -.006 | .014 | .000 | .013 | .079 |
| t5 | .014 | .011 | .010 | .010 | .018 | .016 | .038 |
| t6 | -.004 | -.006 | -.017 | .007 | -.016 | .008 | -.005 |
| t7 | -.004 | -.013 | -.028 | -.006 | .000 | -.023 | -.024 |
| | -.015 | -.002 | -.007 | -.010 | -.023 | -.019 | -.088 |

| | Contents positive score | Contents negative score | Title positive score | Title negative score | Title wordcount | Contents wordcount | Number of reviews including Picture |
|-------------------------------------|-------------------------|-------------------------|----------------------|----------------------|-----------------|--------------------|-------------------------------------|
| Pearson 상관계수 | | | | | | | |
| Ln(1/rank) | -.145 | -.026 | .103 | .062 | -.046 | -.094 | .125 |
| Price | .191 | .006 | -.042 | -.033 | -.058 | -.099 | -.060 |
| Brand strength | -.035 | -.019 | -.002 | .021 | -.072 | -.043 | .008 |
| log(Reivew_quantity) | .011 | .126 | .045 | .040 | .026 | -.170 | .048 |
| log(Rating) | .056 | -.006 | .082 | .063 | -.035 | .060 | -.016 |
| Sum of helpfulness | .020 | -.024 | .032 | .030 | .012 | -.033 | .146 |
| Days on amazon | -.071 | .193 | .029 | .048 | .039 | -.086 | .024 |
| Contents positive score | 1.000 | .453 | .015 | -.005 | .014 | -.018 | -.026 |
| Contents negative score | .453 | 1.000 | -.010 | -.074 | .053 | -.031 | .074 |
| Title positive score | .015 | -.010 | 1.000 | .888 | -.088 | .004 | -.016 |
| Title negative score | -.005 | -.074 | .888 | 1.000 | -.113 | .005 | -.026 |
| Title wordcount | .014 | .053 | -.088 | -.113 | 1.000 | .643 | .038 |
| Contents wordcount | -.018 | -.031 | .004 | .005 | .643 | 1.000 | .005 |
| Number of reviews including Picture | -.026 | .074 | -.016 | -.026 | .038 | .005 | 1.000 |
| Brand * Review quantity | .122 | .094 | -.074 | -.089 | -.008 | -.124 | -.025 |
| Brand * Contents positive score | .655 | .190 | -.041 | -.010 | -.097 | -.095 | -.037 |
| Brand * Contents negative score | .237 | .524 | -.036 | -.042 | .004 | -.012 | -.017 |
| Brand * Title positive score | -.048 | -.033 | .567 | .565 | -.075 | .015 | .035 |
| Brand * Title negative score | -.011 | -.036 | .510 | .628 | -.075 | .048 | .015 |
| t1 | -.046 | -.022 | .047 | .038 | -.032 | -.032 | -.040 |
| t2 | -.002 | .003 | .031 | .038 | -.025 | -.001 | -.059 |
| t3 | -.022 | .017 | -.002 | -.010 | .010 | .001 | -.051 |
| t4 | .005 | .010 | -.005 | -.006 | .023 | .021 | -.034 |
| t5 | -.022 | -.022 | -.007 | .002 | .011 | -.018 | -.046 |
| t6 | -.014 | -.011 | -.018 | -.007 | .020 | .006 | -.048 |
| t7 | .058 | .019 | -.025 | -.031 | .017 | .018 | .050 |

| | Brand * Review quantity | Brand * Contents positive score | Brand * Contents negative score | Brand * Title positive score | Brand * Title negative score | t1 | t2 | t3 | t4 | t5 | t6 | t7 |
|-------------------------------------|-------------------------------|---------------------------------------|---------------------------------------|---------------------------------|---------------------------------|-------|-------|-------|-------|-------|-------|-------|
| Pearson 상관계수 Ln(1/rank) | .219 | -.093 | .044 | -.004 | -.016 | .001 | -.006 | .016 | .014 | -.004 | -.004 | -.015 |
| Price | -.033 | .270 | .058 | -.098 | -.058 | -.008 | .003 | .010 | .011 | -.006 | -.013 | -.002 |
| Brand strength | -.200 | -.017 | .005 | .013 | .038 | .048 | .007 | -.006 | .010 | -.017 | -.028 | -.007 |
| log(Review_quantity) | .616 | .109 | .103 | -.082 | -.088 | .000 | -.001 | .014 | .010 | .007 | -.006 | -.010 |
| log(Rating) | -.185 | .037 | .001 | .096 | .073 | .003 | .029 | .000 | .018 | -.016 | .000 | -.023 |
| Sum of helpfulness | .069 | -.053 | -.002 | .049 | .052 | .015 | -.001 | .013 | .016 | .008 | -.023 | -.019 |
| Days on amazon | .339 | .004 | .045 | -.096 | -.076 | .051 | .065 | .079 | .038 | -.005 | -.024 | -.088 |
| Contents positive score | .122 | .655 | .237 | -.048 | -.011 | -.046 | -.002 | -.022 | .005 | -.022 | -.014 | .058 |
| Contents negative score | .094 | .190 | .524 | -.033 | -.036 | -.022 | .003 | .017 | .010 | -.022 | -.011 | .019 |
| Title positive score | -.074 | -.041 | -.036 | .567 | .510 | .047 | .031 | -.002 | -.005 | -.007 | -.018 | -.025 |
| Title negative score | -.089 | -.010 | -.042 | .565 | .628 | .038 | .038 | -.010 | -.006 | .002 | -.007 | -.031 |
| Title wordcount | -.008 | -.097 | .004 | -.075 | -.075 | -.032 | -.025 | .010 | .023 | .011 | .020 | .017 |
| Contents wordcount | -.124 | -.095 | -.012 | .015 | .048 | -.032 | -.001 | .001 | .021 | -.018 | .006 | .018 |
| Number of reviews including Picture | -.025 | -.037 | -.017 | .035 | .015 | -.040 | -.059 | -.051 | -.034 | -.046 | -.048 | .050 |
| Brand * Review quantity | 1.000 | .179 | .171 | -.134 | -.142 | .006 | -.019 | .001 | .009 | -.020 | -.012 | .012 |
| Brand * Contents positive score | .179 | 1.000 | .362 | -.073 | -.016 | -.050 | -.008 | -.034 | -.007 | -.011 | -.007 | .065 |
| Brand * Contents negative score | .171 | .362 | 1.000 | -.063 | -.067 | -.052 | .008 | .019 | .003 | -.028 | -.017 | .036 |
| Brand * Title positive score | -.134 | -.073 | -.063 | 1.000 | .900 | .024 | .009 | -.028 | -.003 | .000 | .000 | -.003 |
| Brand * Title negative score | -.142 | -.016 | -.067 | .900 | 1.000 | .021 | .002 | -.035 | -.008 | .010 | .014 | -.008 |
| t1 | .006 | -.050 | -.052 | .024 | .021 | 1.000 | -.128 | -.127 | -.133 | -.135 | -.139 | -.143 |
| t2 | -.019 | -.008 | .008 | .009 | .002 | -.128 | 1.000 | -.130 | -.136 | -.139 | -.143 | -.147 |
| t3 | .001 | -.034 | .019 | -.028 | -.035 | -.127 | -.130 | 1.000 | -.135 | -.137 | -.141 | -.145 |
| t4 | .009 | -.007 | .003 | -.003 | -.008 | -.133 | -.136 | -.135 | 1.000 | -.144 | -.147 | -.152 |
| t5 | -.020 | -.011 | -.028 | .000 | .010 | -.135 | -.139 | -.137 | -.144 | 1.000 | -.150 | -.155 |
| t6 | -.012 | -.007 | -.017 | .000 | .014 | -.139 | -.143 | -.141 | -.147 | -.150 | 1.000 | -.159 |
| t7 | .012 | .065 | .036 | -.003 | -.008 | -.143 | -.147 | -.145 | -.152 | -.155 | -.159 | 1.000 |