

# 온라인 쇼핑물 구매리뷰 데이터를 활용한 리뷰평점 영향인자 도출



팀명: 타이거JK와 아이들  
팀원: 권민수  
김이영  
김종인  
문지현

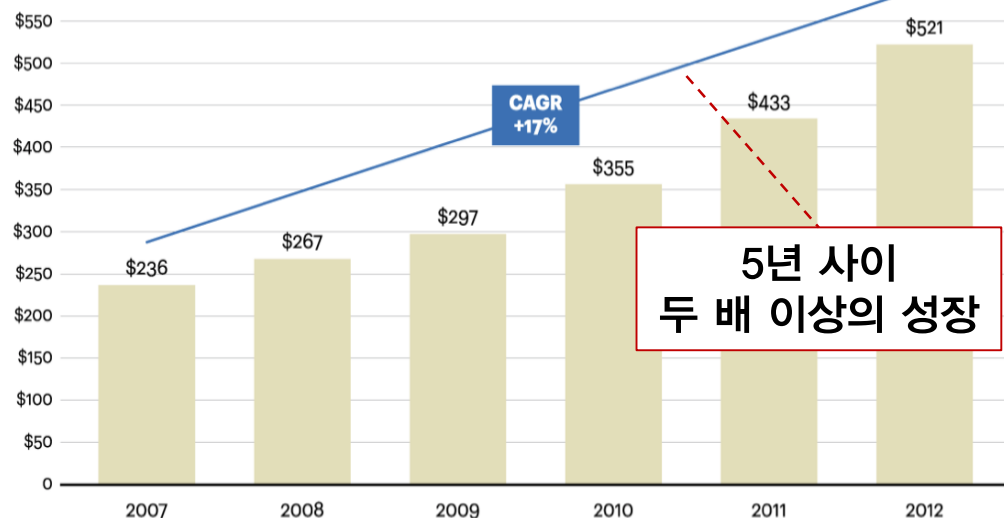


## 온라인 쇼핑몰의 규모

Global online retail sales have increased 17 percent yearly since 2007

Global online retail sales

(US\$ billion)



Note: Online retail sales exclude sales tax and are at constant 2012 exchange rates.

Source: Euromonitor

폭발적으로 성장하는  
온라인 쇼핑몰의 시장규모



## 리뷰

= 온라인 쇼핑몰의 중요한 평가기준



88%

의 고객은 온라인 리뷰를  
지인 추천만큼 신뢰한다

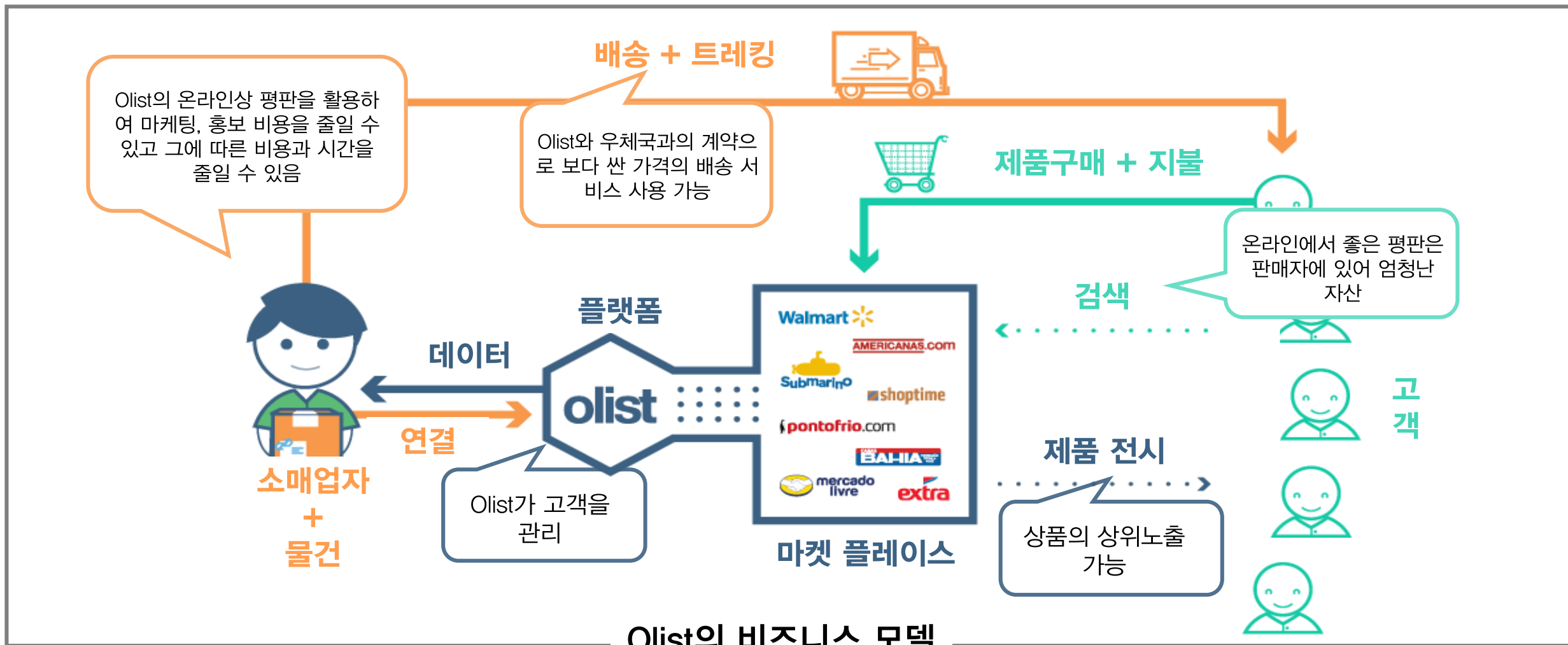
80%

의 고객은 부정적인 리뷰를  
읽고 구매심리가 바뀐다리뷰평점 1점이 올라갈 때  
이익이 5~9% 상승한다내용출처: Northwestern Digital and database research center,  
Harvard Business School working journal

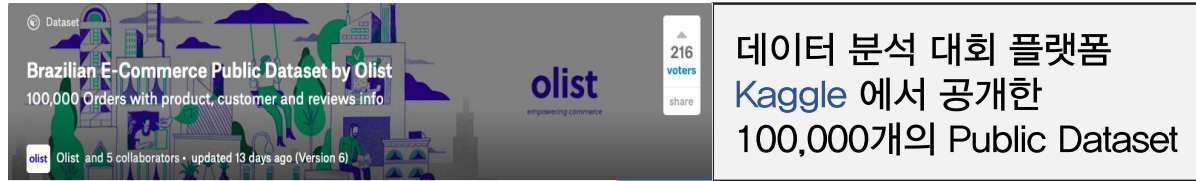
쇼핑몰의 리뷰평점 관리가  
매출에 큰 영향력을 끼친다

**olist**

소상공인들과 브라질 시장에 진출하기 위한 기업들의 원활한 브라질 온라인 시장 진출을 도와 주기 위한 플랫폼



## 02 데이터 수집과 가공 우리의 데이터의 출처는?



데이터 분석 대회 플랫폼  
Kaggle 에서 공개한  
100,000개의 Public Dataset

### 두 종류의 데이터

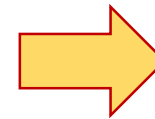
#### 정형 데이터

데이터 변수 그룹	테이블 설명	컬럼 속성
orders	주문과 관련된 정보	주문 아이디, 배송상태, 총 제품 가격, 배송료, 주문 내 제품수, 주문 내 판매자 수, 구매 시간, 구매 승인 시간, 배송예정 날짜, 고객 물품수령 날짜
customers	고객정보를 관리한다	고객 아이디, 고객 시 주소, 고객 주, 고객 zip코드, 고객 unique id,
payments	결제와 관련된 정보	할부 개월, 지불 방법
products	제품과 관련된 정보	제품 아이디, 제품 품목 이름, 제품명 길이, 제품 설명 길이, 제품 사진 개수, 제품 무게, 제품 길이, 제품 높이, 제품 넓이
sellers	판매자와 관련된 정보	판매자 아이디, 판매자 zip코드, 판매자 도시, 판매자 주
geolocation	(zipcode기준)지리적 정보를 관리한다	우편번호 앞 3자리, 우편번호의 도시, 우편번호의 주, 우편번호의 위도, 우편번호의 경도
review	고객의 리뷰와 관련된 정보	리뷰 아이디, 리뷰 점수, 리뷰 제목, 리뷰 내용, 리뷰 요청 날짜, 리뷰 작성 날짜

#### 비정형 데이터



포르투갈어로  
작성 되어있는  
40,000개의  
리뷰 코멘트



구글 API를  
사용하여  
영어로 번역

## 변수선정

로지스틱 회귀모델

[종속변수]

Positive / Negative (리뷰 평점의 긍정 부정 여부)

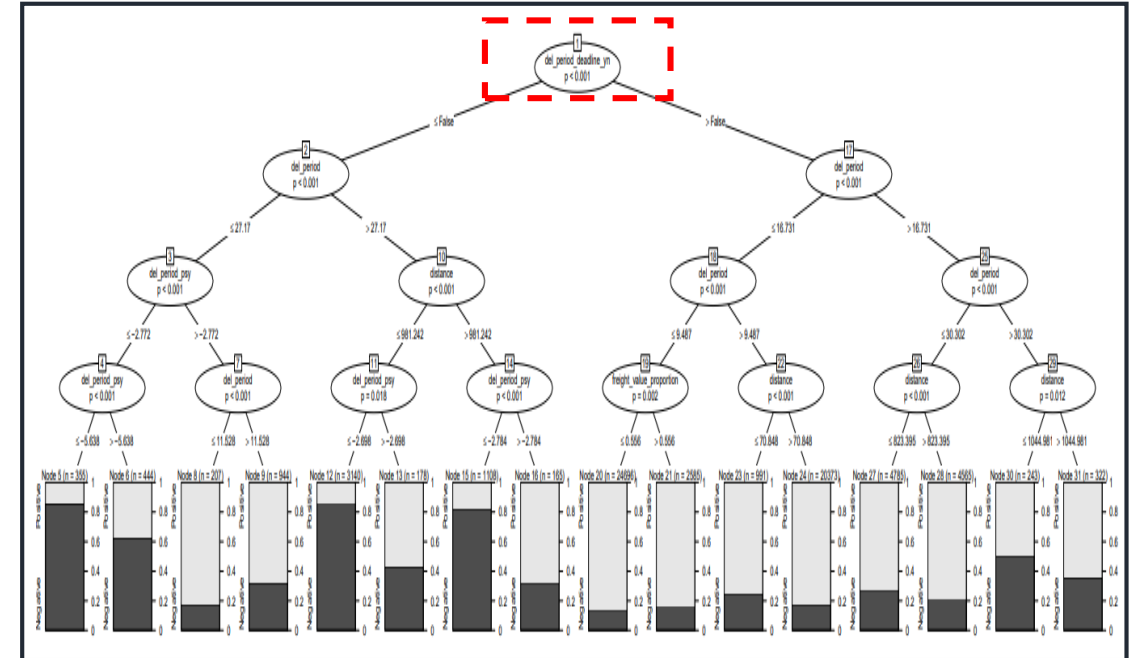


[예측변수들의 중요도]

변수명	한글 변수명	Estimate	Pr(> z )
del_period_deadline_ynTrue	배송 기한 준수 여부 (약속 이행)	1.303	0.00
del_period	절대적 배송기간	-0.048	0.00
distance	판매자 - 고객 직선 거리(km)	0.000	0.00
order_product_value	제품 가격	0.000	0.00
freight_value_proportion	배송료 비중	-0.150	0.00
del_period_psy	심리적 배송 기간	0.004	0.02
order_day.L	주문 요일	-0.059	0.03
installments_ynTrue	할부 여부 (True)	-0.060	0.04

## 모델생성

의사결정 나무



배송기한 준수 여부	심리적 배송기간
절대적 배송기간	판매자-고객 직선 거리

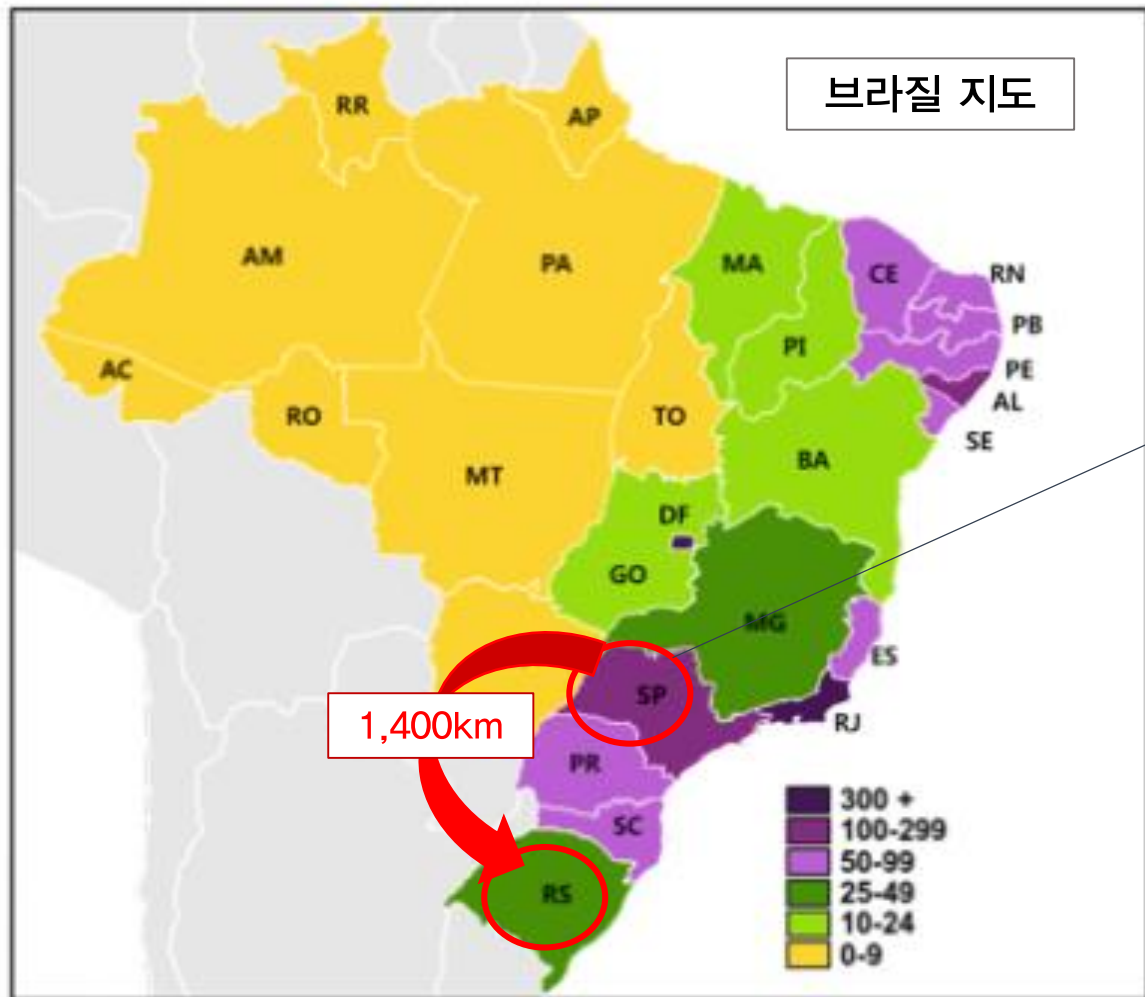
리뷰 평점을 결정짓는 요인은 바로 **배송**



# 03 정형 데이터 분석 결과에 따른 활용방안



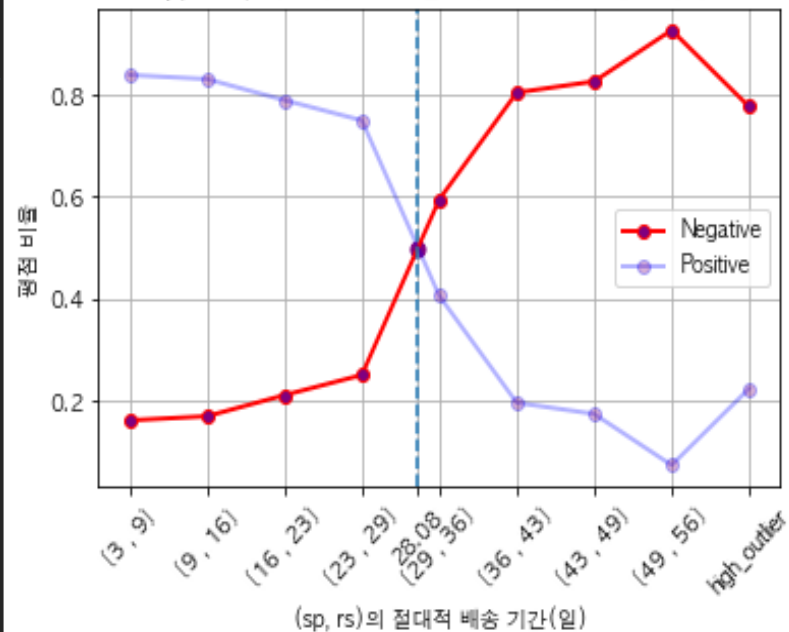
배송을 어떻게 효율적으로 할 수 있을까?



거래 빈도수가 높은 (판매자 State - 고객 State) 별 허용 가능한 배송 기간 임계치 탐색

순위	조합	빈도 (비율)	임계 배송기간 (일)
1	SP - SP	28990 (31%)	18.7 일
2	SP - RJ	8423 (9%)	22.3 일
3	SP - MG	7349 (8%)	22.8 일
4	SP - RS	3463 (4%)	28.1 일
5	SP - PR	2972 (3%)	26.9 일

(sp, rs)의 절대적 배송 기간(일)에 따른 평점 상관관계



거리

$\propto$

배송기간  
임계치

# 04 비정형 데이터

## 딕셔너리 생성

[새롭게 재정의한 13개의 카테고리]

주제	카테고리
만족	만족
	배송 지킴
제품 관련	제품 불량
	제품 훼손
	과대 광고
	포장 불량
Seller Service	서비스
	잘못된 제품
배송 관련	배송 지연
	부분 배송
	기타 배송
기타	기타
	리턴

## 딕셔너리 예시

Phrase Count기반

Hello.  
I just received two pendants,  
but I bought three and I sent an  
e-mail complaining to notify that it is  
missing an item for you so far.  
I **have not received an answer.**  
I do not know about the fault of the  
mail or the shop request.

	배송지연	부분배송	서비스
투표수	4	2	5
표현	So far, have not received, not received, so far i have not received	missing an item, just received	sent an e-mail complaining, have not received an answer, complaining to notify, not received an answer, sent an e-mail

## 딕셔너리 정확도

	카테고리	정확도	정밀도
만족	만족	80.4%	70%
	배송지킴	84.1%	66%
불만족	제품불량	96.8%	62%
	제품훼손	98.9%	63%
	과대광고	98.3%	57%
	포장불량	99.4%	65%
	서비스	95.4%	62%
	잘못된 제품	97.5%	70%
	배송지연	90.4%	60%
	부분배송	95.7%	80%
	기타 배송	96.2%	60%
	기타	91.4%	41%
	리턴	97.9%	70%



## 04 딕셔너리 활용방안

정형 데이터로만 알 수 있는 부분

컨셉	카테고리	비중
배송 기간	배송 지연	33%

비정형 데이터를 통해 추가적으로  
영향인자를 파악할 수 있음

딕셔너리를 통해 알 수 있는 부분

컨셉	카테고리	비중
제품	제품 불량	18%
	제품 훼손	
	포장 불량	
	과대 광고	

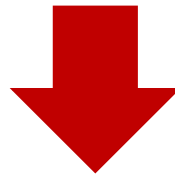
컨셉	카테고리	비중
기타 배송	기타 배송	22%
	부분 배송	

컨셉	카테고리	비중
기타	기타	17%
	리턴	

컨셉	카테고리	비중
Seller Service	잘못된 제품	11%
	서비스	

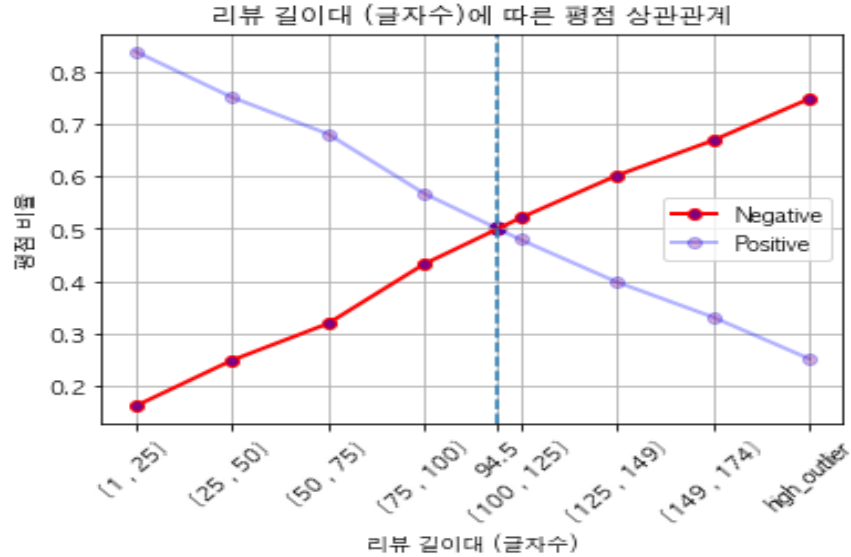
## 05 정형 & 비정형 결론

구분	데이터	분석	결론	리뷰 평점 개선 방안 시사점
정형 데이터	** 고객 정보 부재 배송 정보 주문 정보	긍정/ 부정 리뷰 평점에 영향 주는 요인 발견	배송 이외 정보 확인 어려움	배송 기간 전략 차별화
비정형 데이터	리뷰 정보	주제별 딕셔너리 형성	배송 이외 고객 이슈 확인	고객의 다양한 불만 해결 방안 마련 가능



온라인 쇼핑 리뷰 평점 영향 인자 도출

## 긍정적 리뷰 부족



긍정적 리뷰의 글자수는 대부분 100자 이상이  
넘지 않는 짧은 글



**양질의 긍정적 후기를 작성하도록 유도**

1. 베스트 후기를 남겨준 고객에게 리워드 지급
2. 리뷰 작성 길이에 최소한의 제한(90자)을 주어 리뷰의 질을 높이기

## 리뷰 카테고리 데이터 수집

〈기존의 리뷰 작성 항목〉

리뷰 제목  
리뷰 메시지

〈추가할 리뷰 작성 항목〉

리뷰 카테고리



### 리뷰 카테고리 예시

제품 불량

☐

제품훼손

☐

포장 관련 이슈

☒

과대 광고

☐

기타 배송

☐

부분 배송



리턴

☐

기타

☐

만족

☐

잘못된 배송

☐

서비스

☐