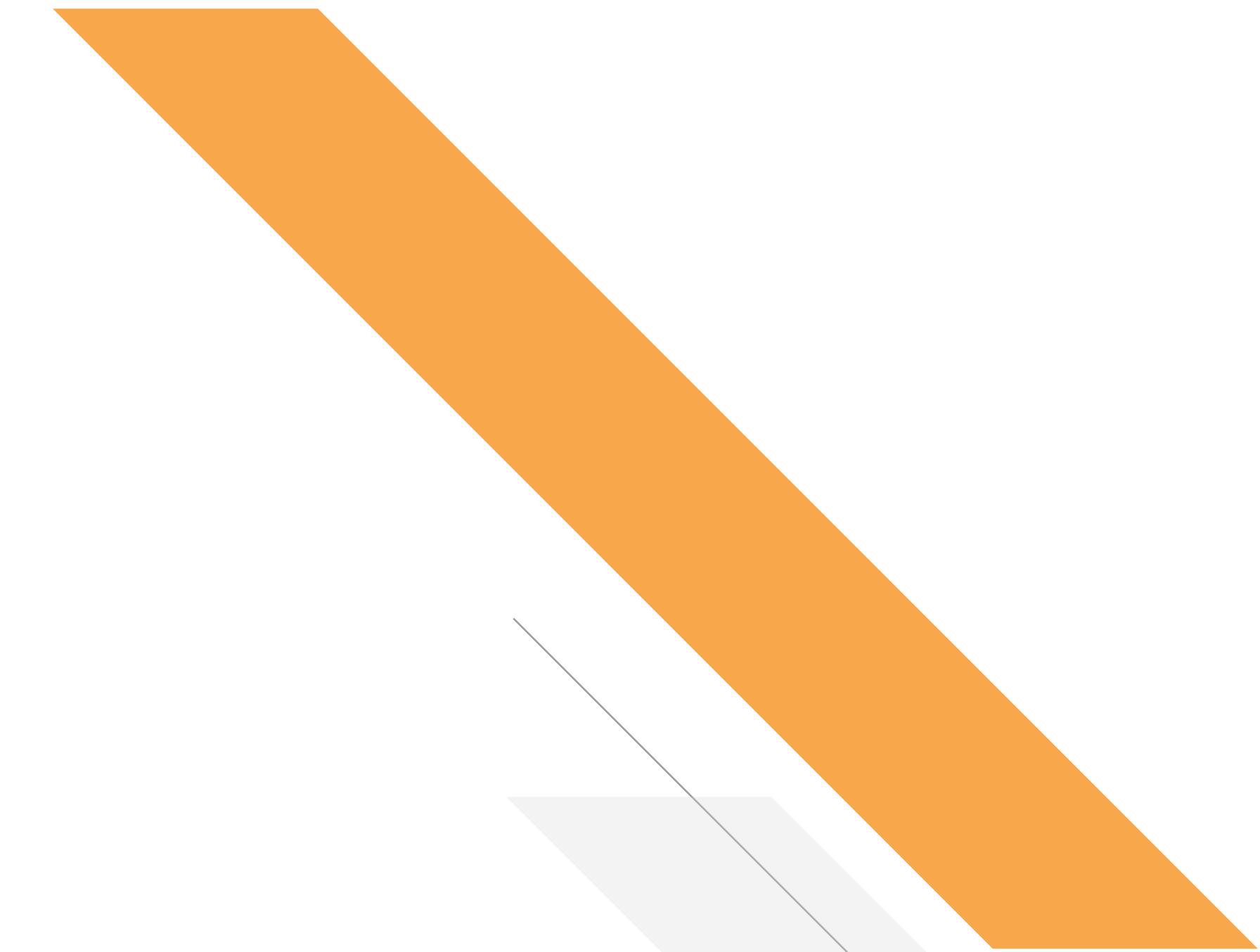




빅데이터를 활용한 중고폰 적정가격 예측 서비스



Team Scoop
Project "AlmhaGo"

Scoop 팀원소개



데이터 전처리 :
허재훈



기획 & 리서치 :
이고은



프로젝트 매니지먼트 :
윤기태



데이터 분석 :
김은석



챗봇 개발 :
전홍진



데이터 시각화 :
안지훈

CONTENTS

01. 상황분석

02. 문제점 도출

03. 데이터 분석

04. 챗봇 소개 및 시연

05. 프로젝트 기대효과

데이터 분석 프로젝트를 진행하기에 앞서, 중고폰
시장의 전반적인 상황을 살펴보며 문제점을
분석하고 인사이트를 도출하고자 하였다.

01

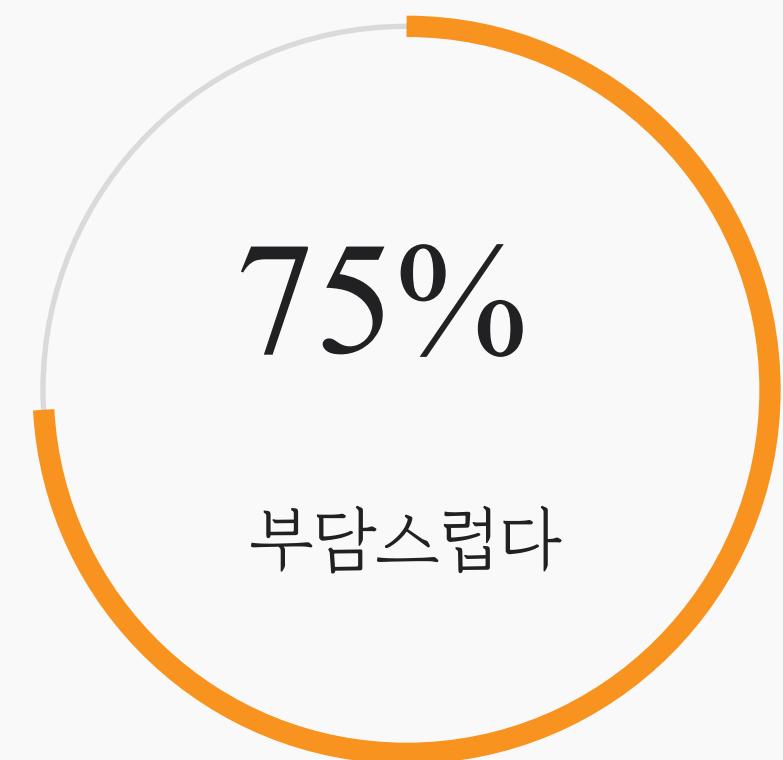
상황분석



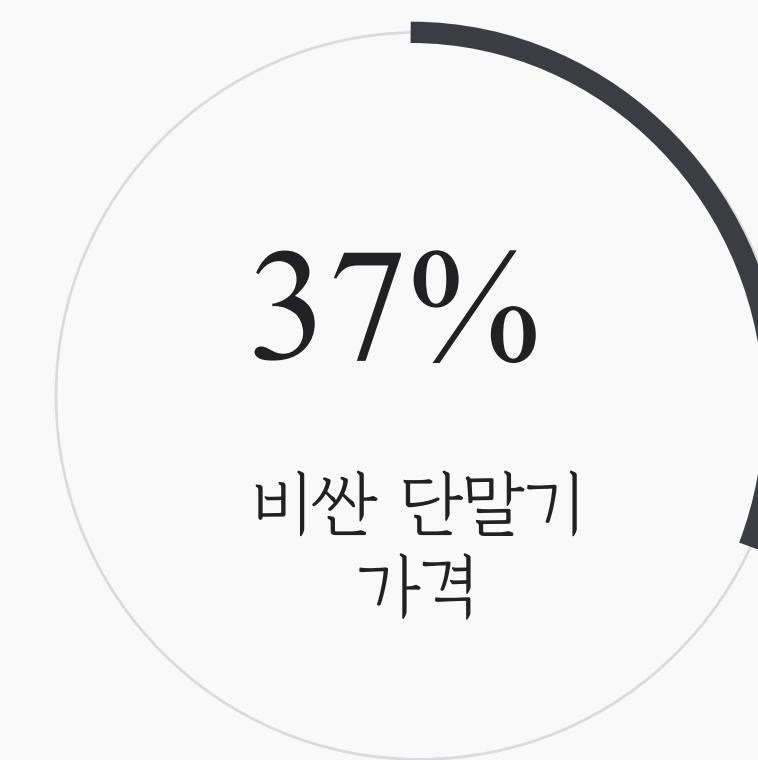
상황분석

연간 1000만대의 국내 중고폰 거래 시장규모

2017년 녹색 소비자 연대, n=1000

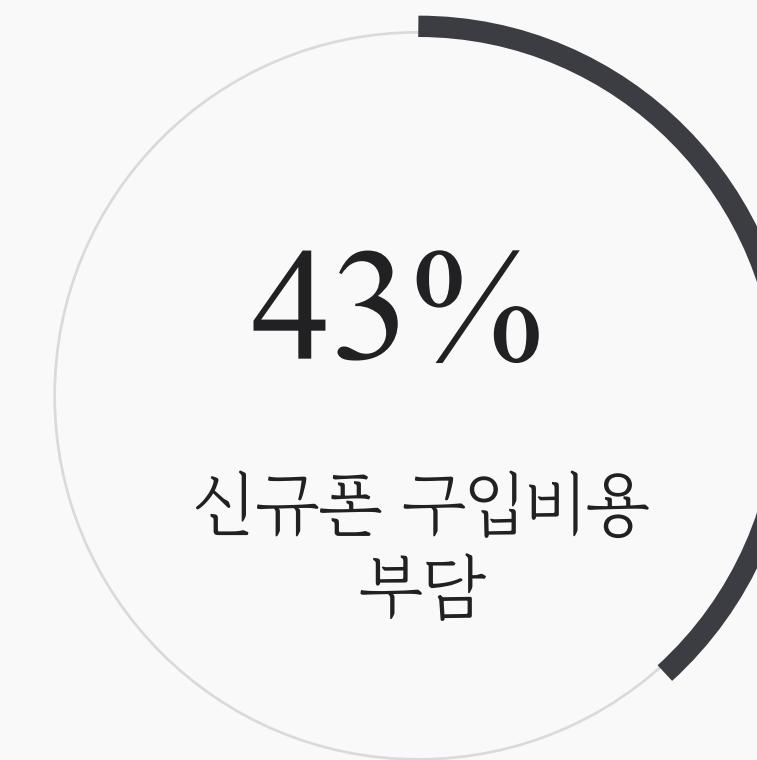


가계 통신비
부담스럽나



가계 통신비가
부담스러운 이유

2014년 한국소비자원, n=1000



중고 휴대전화
구입이유

2016년 전자신문



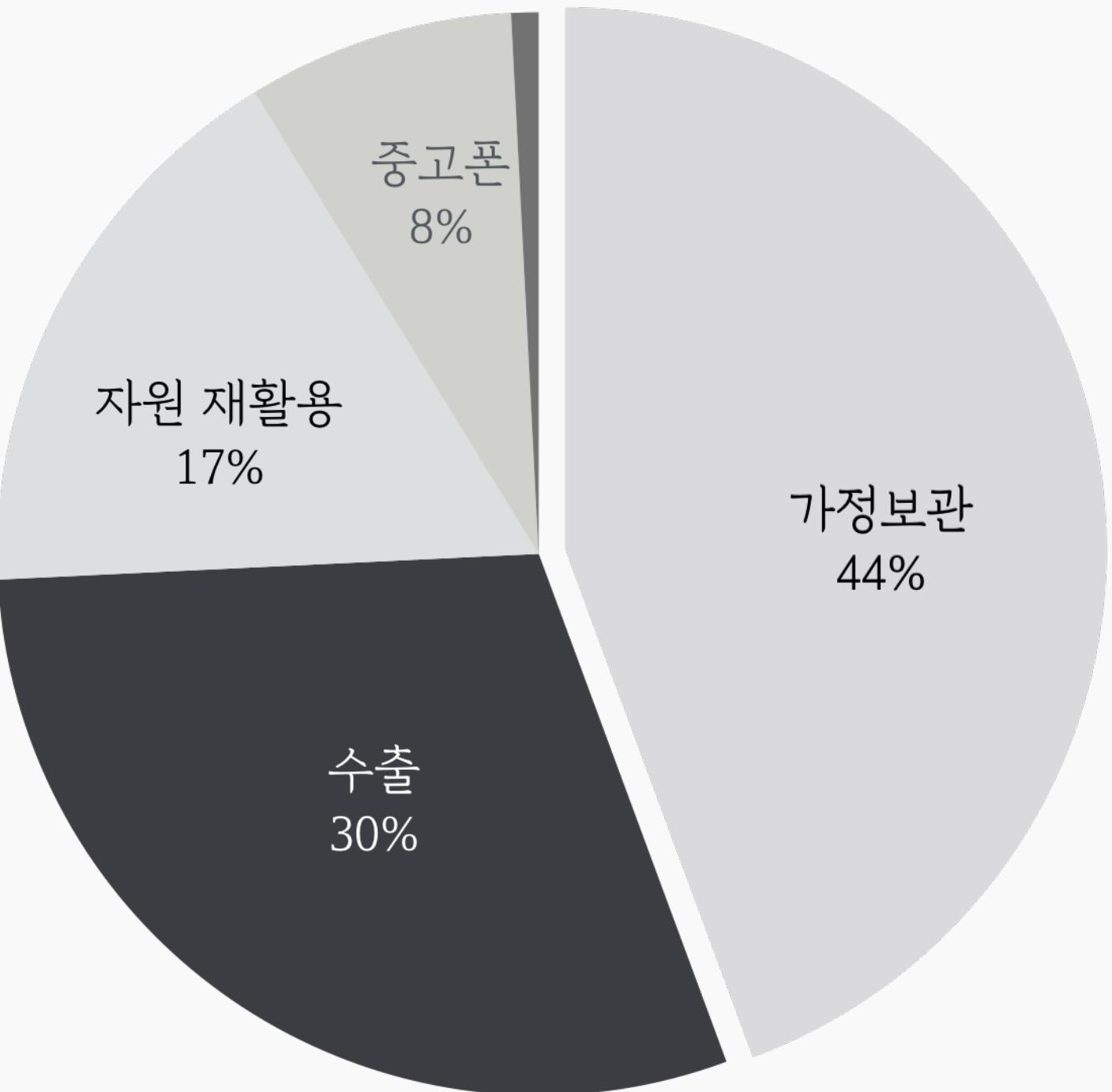
2016년
중고폰 개통
비중

상황분석

하지만, 내 품 안의 중고폰이 더 많은 상황

폐휴대폰 회수 현황

▶ 2010년 환경부, n=18440000



이전 휴대전화 처분 형태

▶ 2014년 한국갤럽, n=1000



상황분석

왜 이전 휴대폰을 보관하고 있을까?

중고폰 비처분 이유

▶ 2014년 한국갤럽, n=482, 복수응답

팔아도 받을 수 있는 금액이 적어서

165명

저장되어 있는 사진, 음악, 연락처 때문

132명

어떻게 팔아야 할지 몰라서

159명

기기애에 남아있는 개인정보 유출 우려

130명

처분하기 번거롭거나 바빠서

140명

나중에 쓸 데가 있을 것 같아서

105명

02

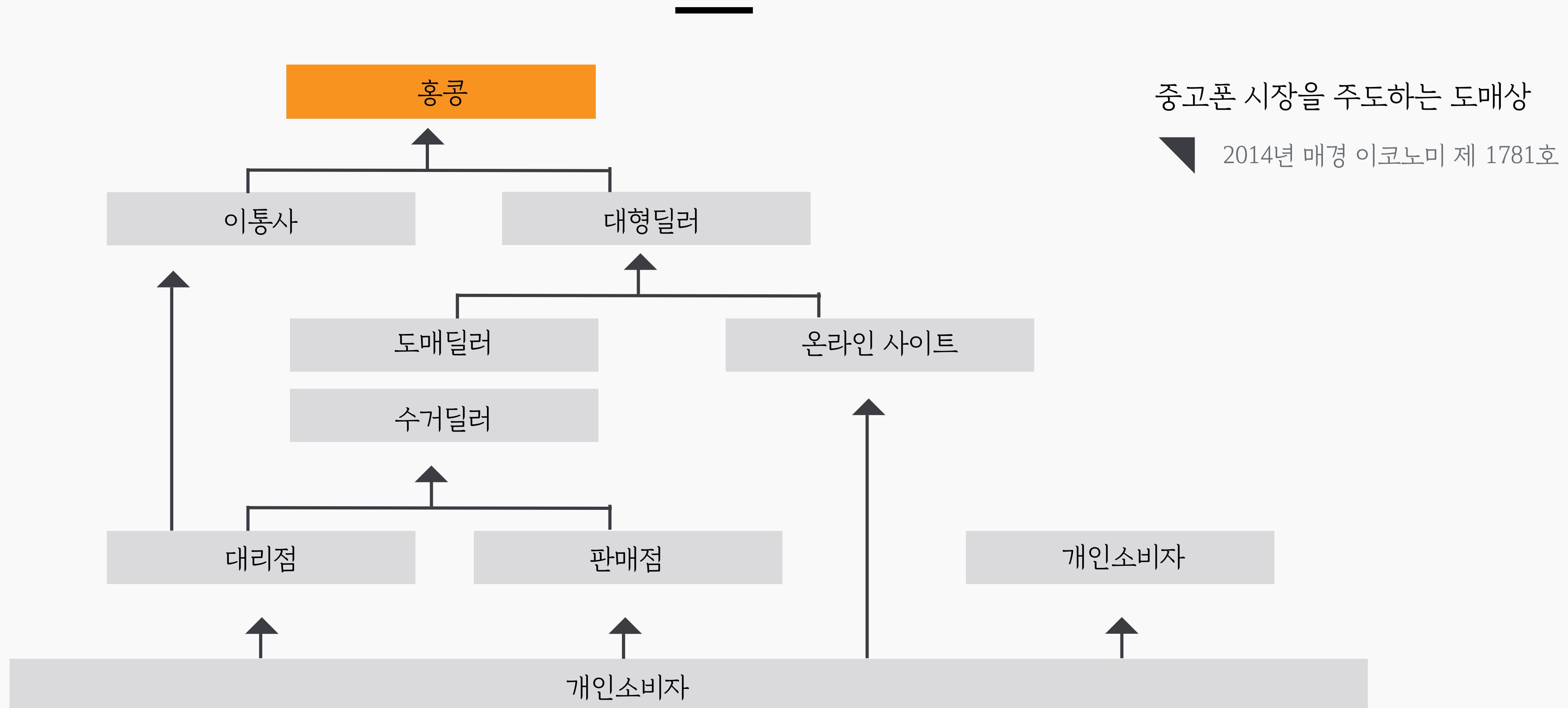
문제도출

시장의 전반적인 상황을 토대로
문제가 되는 상황을 도출하고자 하였다.



중고폰 유통 경로

중고폰 가격은 어떻게 정해지는가



도매상과 직거래간의 가격 차이

359,000원

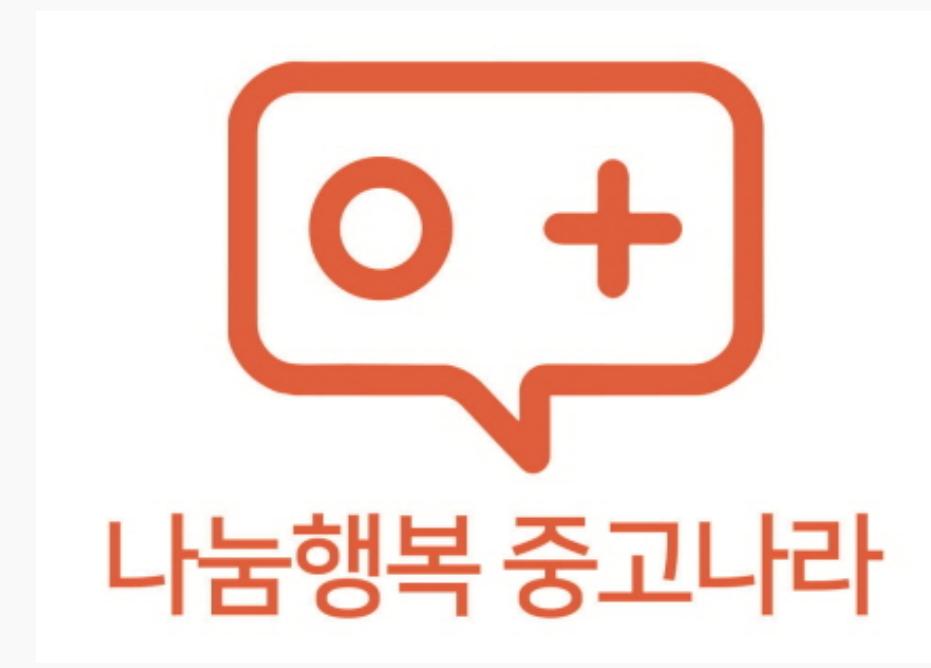
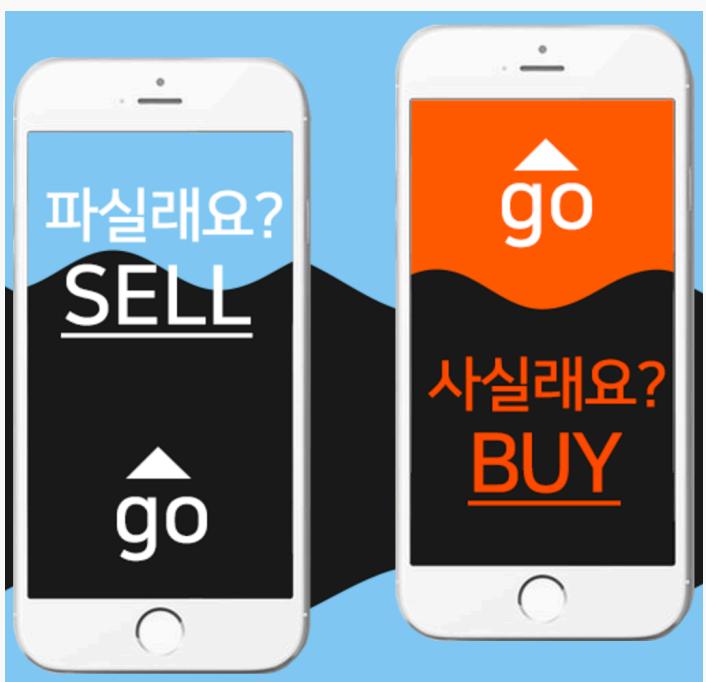
돈다오

430,000원

폰사닷컴

550,000원

중고나라

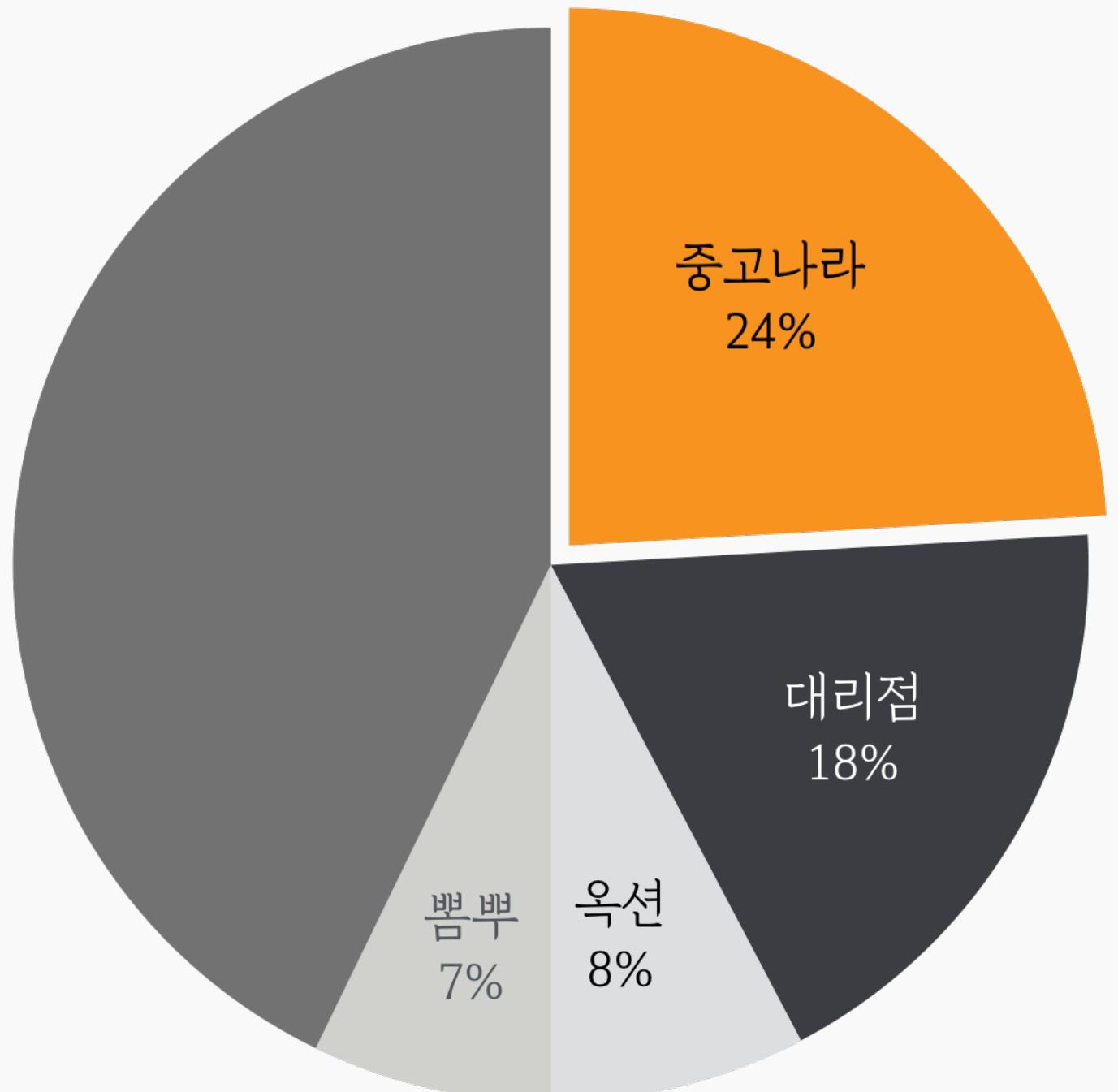


중고폰 판매 경로 인지도

중고폰 판매 채널로 높은 인지도를 보유한 중고나라

비교상기도

▶ 2014년 한국갤럽, n=1000



보조상기도

▶ 2014년 한국갤럽, n=1000

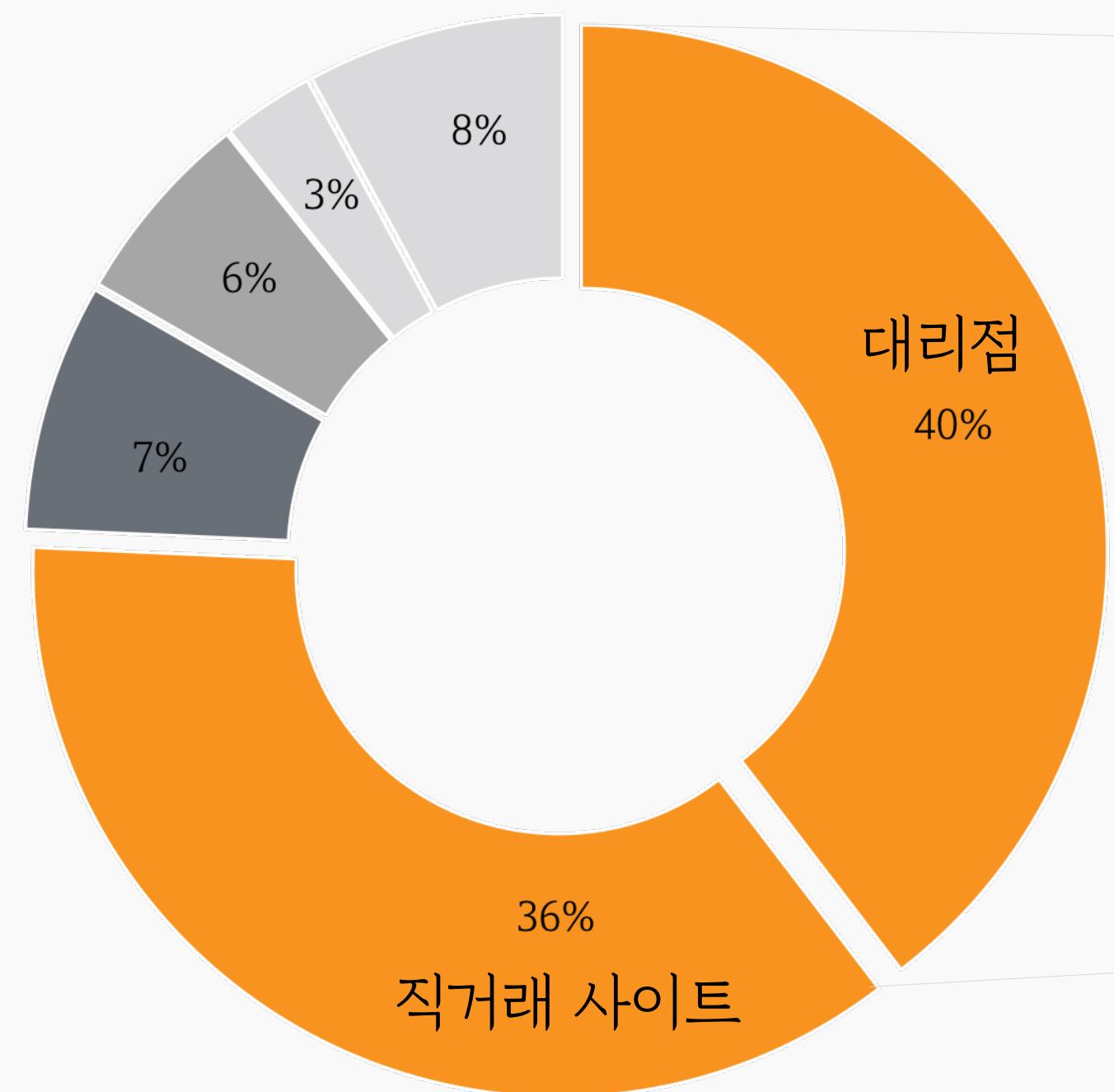


직거래 사이트를 이용하는 소비자

더 높은 중고폰 보상 금액을 위해

중고폰 판매시 이용 경로

▶ 2014년 한국갤럽, n=482



해당 경로 선택 이유

▶ 2014년 한국갤럽, n=482

판매 절차가 간편해서

49.5%

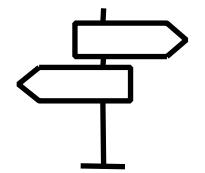
보상금액이 가장 높아서

46.9%



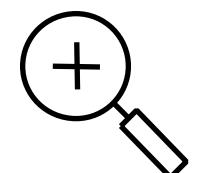
Problem

저조한 폐휴대폰 재활용률로 인한 자원낭비



Reason

비합리적인 중고폰 판매가격



Needs

판매자 : 적정 가격으로 중고폰 처분을 희망

구매자 : 사기 없는 안전한 중고폰 거래 희망



Solution

중고나라 휴대폰 카테고리 판매글 분석을 바탕으로 한 적정 시세 제안

최종
문제도출

03

데이터 분석

도출된 문제점을 해결을 위해
데이터 분석을 통해 솔루션을 제공하고자 한다.



데이터 전처리 프로세스

Raw 데이터를 정제된 데이터로



중고품목 유통 현황 데이터 확보

출처 : 한국 데이터베이스 진흥원의 데이터 스토어

2015년부터 2017년까지 네이버 카페 중고나라에 올라온 45만건의 판매글을 수집한 데이터



ID를 기준으로 중고폰 카테고리 추출



추출한 데이터에서 특수문자, 개인정보, 광고글, 비 휴대폰 게시글
가격이 없는 판매글, 구매 희망을 원하는 게시글 등 불필요한 데이터 제거



스마트폰 표준 정보 데이터 구축

출처 : 통신 3사 공식 홈페이지

웹 크롤링을 통하여, 150개 스마트폰 기종의
(모델명 / 제조사 / 저장 용량 / 시기별 출고 가격) 조사



기종에 따른 검색어 사전 제작

모델명을 오기입하는 경우를 고려하여
모델명에 따른 검색어 사전을 별도로 제작
(Ex) 갤럭시 노트 3 모델 : “갤노3”, “노트3” ...



```
# 25 Galaxy Tab S3 LTE
product_name23 = "galaxy tab s3 lte"
maker_name23 = "samsung"
dic23 = c("galaxy tab s3 lte", "galaxytab s3 lte", "galaxy tab s3lte", "galaxytab s3lte",
"갤럭시 tab s3 lte", "갤럭시tab s3 lte", "갤럭시 tab s3lte", "갤럭시tab s3lte",
"갤럭시 tab s3 lte", "갤럭시tab s3 lte", "갤럭시 tab s3lte", "갤럭시tab s3lte",
"galaxyTab s3 lte", "galaxyTab s3 lte", "galaxyTab s3 lte", "galaxyTab s3 lte",
"갤럭시Tab s3 lte", "갤럭시Tab s3 lte", "갤럭시Tab s3 lte", "갤럭시Tab s3 lte",
"갤럭시Tab s3 lte", "갤럭시Tab s3 lte", "갤럭시Tab s3 lte", "갤럭시Tab s3 lte",
"삼성 galaxy tab s3 lte", "삼성 galaxy tab s3 lte", "삼성 galaxy tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"galaxy tab s3 lte", "galaxytab s3 lte", "galaxytab s3lte", "galaxytab s3lte",
"갤럭시 tab s3 lte", "갤럭시tab s3 lte", "갤럭시 tab s3lte", "갤럭시tab s3lte",
"갤럭시 tab s3 lte", "갤럭시tab s3 lte", "갤럭시 tab s3lte", "갤럭시tab s3lte",
"galaxyTab s3 lte", "galaxyTab s3 lte", "galaxyTab s3 lte", "galaxyTab s3 lte",
"갤럭시Tab s3 lte", "갤럭시Tab s3 lte", "갤럭시Tab s3 lte", "갤럭시Tab s3 lte",
"갤럭시Tab s3 lte", "갤럭시Tab s3 lte", "갤럭시Tab s3 lte", "갤럭시Tab s3 lte",
"삼성 galaxy tab s3 lte", "삼성 galaxy tab s3 lte", "삼성 galaxy tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte", "삼성 갤럭시 tab s3 lte",
"sm-t825n0")
```



사전 기반 모델명 추출 알고리즘 개발

알고리즘을 활용하여 1차 정리 데이터의 모델명을 추출.
(모델명 / 용량 / 출고가 평균 / 제조사) 피쳐 추가

통신 물가지수 데이터 추가



2015년의 휴대폰 기기값을 기준, 백분율로 표현한 데이터
2016년 통신 물가지수가 103이라면 2015년 대비 3% 증가했다는 의미



탐색적 데이터 분석 시작

X	등록날짜	판매금액	상세설명	수집날짜	제품명	용량	출고가	제조사	date	rate
1	2017-03-19 4 35 00 PM	550000	아이폰6플러스 블랙+애플라이트 64기가 팝니다 아이폰6플러스...	2017-03-25 2 19 171 PM	iphone 6 64gb	64gb	924000	apple	Mar-17	95.96
2	2016-10-26 12 08 00 PM	380000	풀명 갤럭시s6엣지제품 단품상태 사진참고용량 32기가선택여정여...	2017-03-25 1 46 7760 PM	galaxy s6 edge 32gb	32gb	979000	samsung	Oct-16	103.05
3	2016-10-25 12 52 00 PM	300000	구매한지 1개월만에 해외 별령나서 갔다가 이번에 돌아와서 쓰지도...	2017-03-25 2 19 541 PM	galaxy s6 32gb	32gb	854000	samsung	Oct-16	103.05
4	2017-03-23 11 14 00 PM	290000	sk g5 티탄 폰 단품판매합니다 올 5월17일까지 as 가능하며 폰단...	2017-03-25 2 08 2397 PM	lg g5 32gb	32gb	836000	lg	Mar-17	95.96
5	2016-04-11 7 35 00 PM	280000	소니 엑스페리아 c5 울트라sony xperia c5 ultra e5506 16g...	2017-03-25 2 08 847 PM	lg u 32gb	32gb	396000	lg	Apr-16	102.59
6	2017-03-22 1 03 00 AM	280000	갤럭시 a7 2016년형 sm-a710s 기기 외관상태 및 상세설명 블...	2017-03-25 1 33 750 PM	galaxy a7 2016 16gb	16gb	599500	samsung	Mar-17	95.96
7	2017-03-23 8 12 00 AM	270000	갤럭시s6 블랙 32gb 팝니다통신사 유플러스이면 3사유심 모두사...	2017-03-25 1 46 9165 PM	galaxy s6 32gb	32gb	854000	samsung	Mar-17	95.96
8	2016-10-07 10 28 00 AM	40000	홍대 직거래액정에 금이 깃지만 사용하시는데 지장없어불편하시...	2017-03-25 2 08 1681 PM	galaxy note 2 32gb	32gb	1089000	samsung	Oct-16	103.05
9	2017-03-25 2 18 00 AM	150000	g4 배터리 커버 스카이블루 색상 팝니다 민트색이예요현재는 인천...	2017-03-25 2 19 654 PM	g4 32gb	32gb	825000	lg	Mar-17	95.96
10	2016-03-23 9 58 00 AM	15000	중고나라 공식 앱 다운받기 편한 택배 쿠 신청하기 판 매 양식이...	2017-03-25 1 33 2032 PM	galaxy note 2 32gb	32gb	1089000	samsung	Mar-16	94.90
11	2016-03-22 10 41 00 AM	15000	중고나라 공식 앱 다운받기 편한 택배 쿠 신청하기 판 매 양식이...	2017-03-25 1 46 10045 PM	galaxy note 2 32gb	32gb	1089000	samsung	Mar-16	94.90
12	2016-10-11 11 23 00 PM	15000	lgu+ 피처폰 폴더 슬라이드 2g폰 3g폰 정리합니다 lg+ 피처폰...	2016-10-29 2 58 08 AM	wine smart jazz 4gb	4gb	242000	lg	Oct-16	103.05
13	2016-10-11 9 19 00 PM	15000	skt2g폰 정리합니다 폴더 슬라이드 01x 011용▶소액이니 가금...	2016-10-29 2 58 04 AM	wine smart jazz 4gb	4gb	242000	lg	Oct-16	103.05
14	2016-10-26 2 03 00 PM	15000	skt kt 3g 유심방식 피처폰 정리합니다 폴더폰 슬라이드폰연락처 ...	2016-10-29 6 05 06 AM	lg u 32gb	32gb	396000	lg	Oct-16	103.05
15	2016-10-21 12 03 00 PM	15000		2017-03-25 2 19 835 PM	galaxy s4 32gb	32gb	954800	samsung	Oct-16	103.05
16	2016-10-17 9 16 00 AM	18000	정상 해지된 인터넷전화기 입니다구성품 폰 본체 , 거치대 , 거치대...	2017-03-25 1 46 7505 PM	lg u 32gb	32gb	396000	lg	Oct-16	103.05
17	2016-10-18 9 13 00 AM	18000	정상 해지된 인터넷전화기 입니다구성품 폰 본체 , 거치대 , 거치대...	2017-03-25 1 33 1544 PM	lg u 32gb	32gb	396000	lg	Oct-16	103.05
18	2016-10-20 10 06 00 AM	18000	정상 해지된 인터넷전화기 입니다구성품 폰 본체 , 거치대 , 거치대...	2017-03-25 1 46 7855 PM	lg u 32gb	32gb	396000	lg	Oct-16	103.05
19	2016-10-24 8 59 00 AM	18000	정상 해지된 인터넷전화기 입니다구성품 폰 본체 , 거치대 , 거치대...	2017-03-25 2 08 3365 PM	lg u 32gb	32gb	396000	lg	Oct-16	103.05
20	2016-10-26 9 29 00 AM	18000	정상 해지된 인터넷전화기 입니다구성품 폰 본체 , 거치대 , 거치대...	2017-03-25 1 33 1771 PM	lg u 32gb	32gb	396000	lg	Oct-16	103.05

최종 정리된 feature

- 등록날짜
- 용량
- 판매금액
- 출고가
- 상세설명
- 제조사
- 수집날짜
- 년-월
- 제품명
- 물가지수



중고가에 영향을
미치는 것은?



5000개의 데이터, 150개의 휴대폰 기종

기종 별 데이터 수가 많아서 가격도 다양할 것

가설 1. 물품의 상태가 중고가에 영향을 미칠 것이다

가설 2. 판매 게시글의 내용이 상품 상태에 유의미한 영향을 미칠 것이다

가설 3. 출고가, 물가 데이터는 중고가 형성과 연관이 있을 것이다.

→ 판매되는 제품의 등급을 A, B, C로 나누어서 향후 분석에 활용하고자 함

→ 중고가에 영향을 미치지 않는 불필요한 변수들 제거

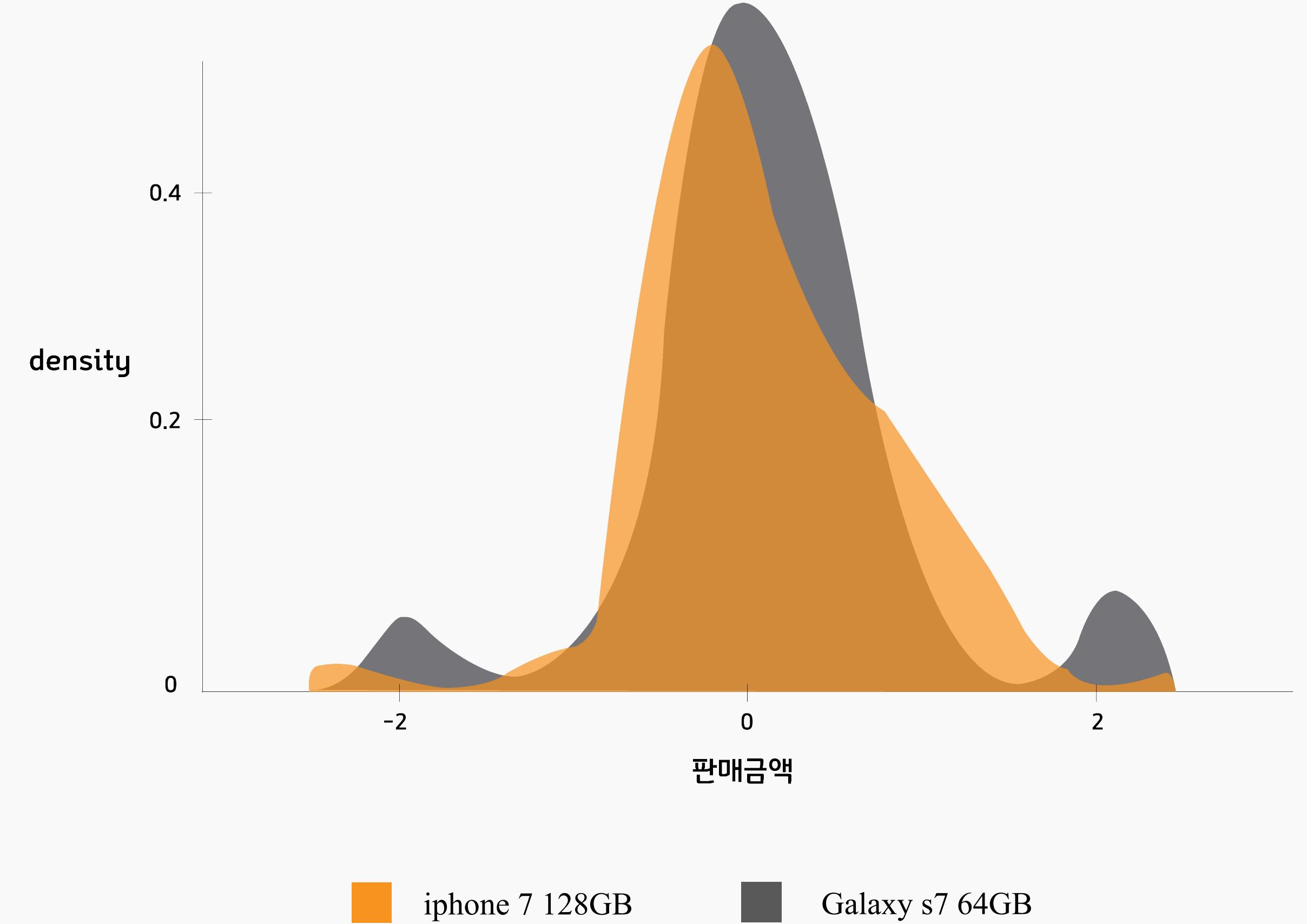
탐색적 데이터 분석 과정

중고물품 등록된 제품의 등급 산정하기

기종별 판매금액 분포 검정

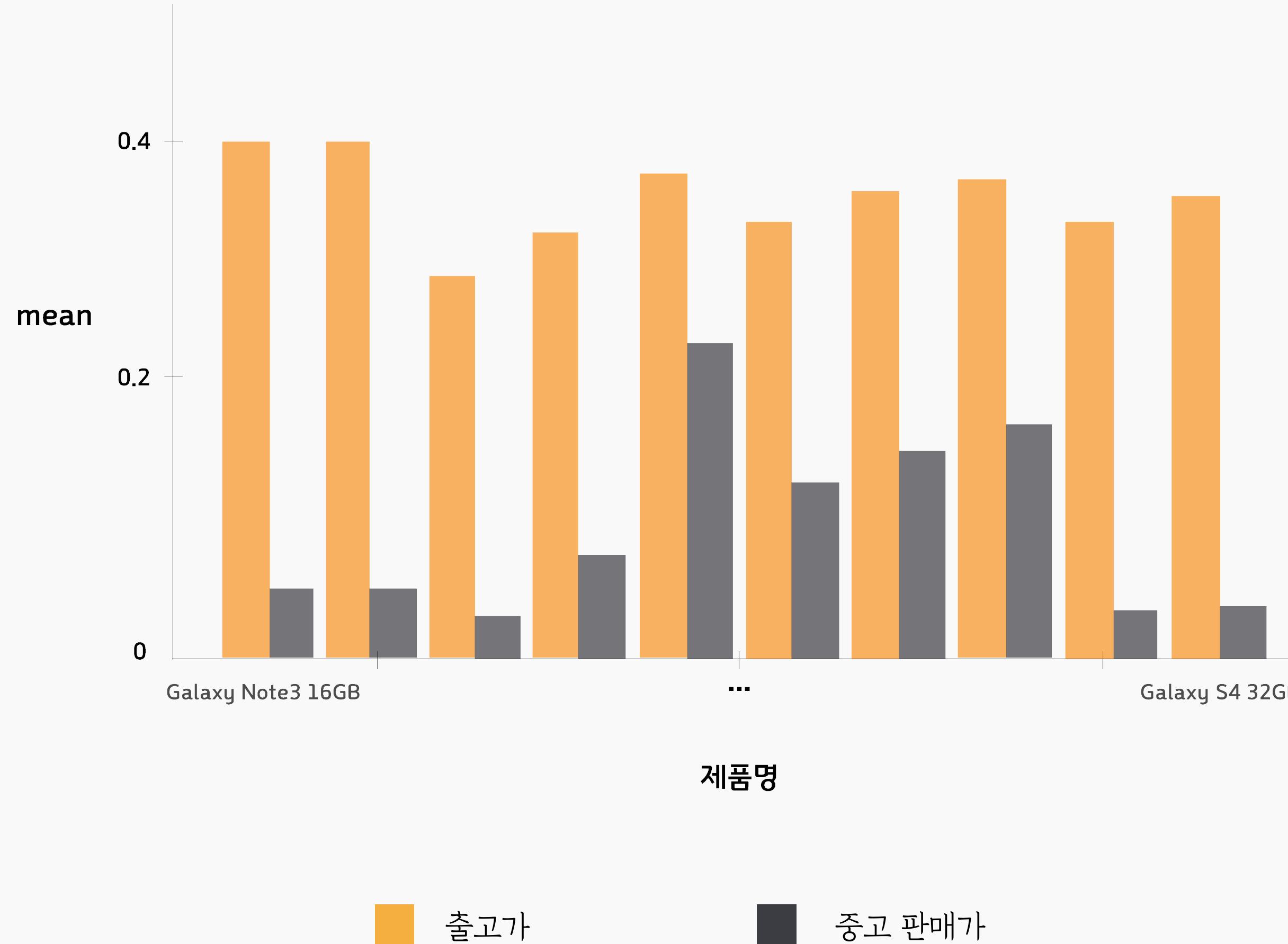
1. 기종별 가격 데이터가 정규분포인지 확인하고자, 두 기종을 선정하여 표준화 후, 확률 밀도함수 그래프로 표현
2. 각각 기종에 대해서 Shapiro-wilk Test를 시행하여 정규성을 띠는지 검정 → W: 0.99 , p-value : e-10 → 선형적 성질이 정규분포적 성질을 띠지만, 정규분포라고는 볼 수 없다. 하지만, 구간별로 데이터를 나누기에는 무리가 없다고 알려져 있음.
3. 기종간 분산분석 실시 → p-value가 모두 1에 가까운 값이 나옴
= 모든 기종 별 가격이 정규분포적 성질을 띤다는 것을 파악

→ 판매되는 제품의 등급을 표준화 된 데이터의 z를 기준으로
A / B / C 분할



탐색적 데이터 분석 과정

중고가에 영향을 미치지 않는 불필요한 변수



출고가와 판매금액간의
차이검정



출고가, 중고판매가 두 변수의 연관성 검증

→ 핸드폰 기종별 출고가와 평균 판매가격간의 상관관계를
히스토그램으로 나타냄

T-test 결과 p-value 값이 0.05보다 현저하게 작았기 때문에
두 변수간의 뚜렷한 상관관계를 발견하기 어렵다고 판단.

가설 3. 출고가, 물가 데이터는 중고가 형성과 연관이 있을 것이다.
→ 가설 3을 기각.

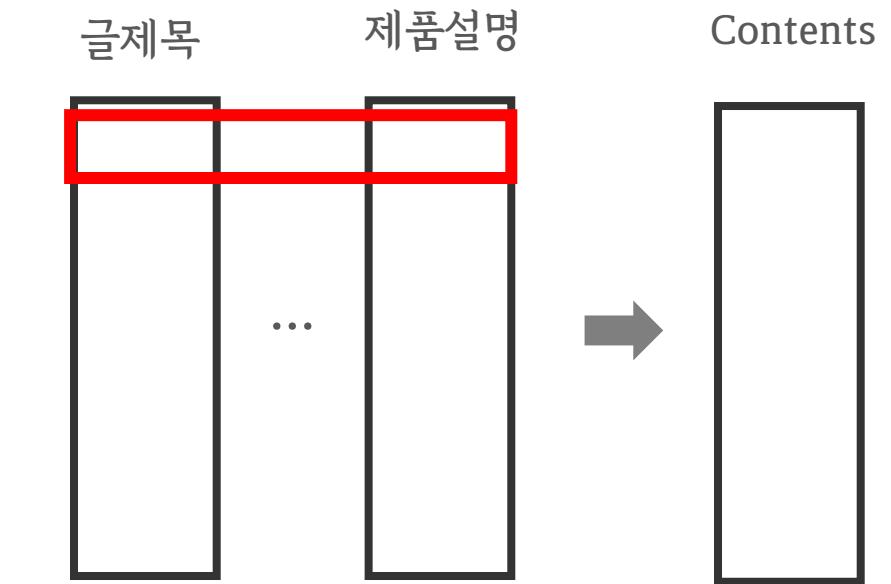


감성분석을 위한 텍스트 전처리



1. Contents Column 생성

데이터 내에서 분석에 필요한 모든 텍스트 데이터를 취합하여 새로운 Contents라는 칼럼을 생성함



2. 텍스트 전처리

특수문자, 전화번호 등 무의미한 숫자 제거

skt2g폰 정리합니다 폴더 슬라이드 01x 011용▶
▶▶ 소액이니 가급적...

3. 불용어 사전 제작

텍스트를 형태소 단위로 처리할 때, 불필요한 단어를 제거하기 위해
불용어 사전을 제작함.

감성분석 과정

감성분석을 통한 객관적인 중고폰 등급 판정 기준 수립

Step 01



텍스트 전처리

Contents 피쳐 생성
특수문자 / 숫자 제거

Step 02



불용어 사전 제작

1. 텍스트를 형태소 단위로 분할
2. 빈도수 상위 2000개의 단어 추출
3. 무의미한 중복 단어 사전 등록

Step 03



불용어 필터링

불용어 사전으로 필터링하여
Contents 데이터에서 형태소 추출

Step 04



Word Embedding

텍스트 데이터를 수치형 벡터화

1. occurrence count matrix
2. tf-idf
3. doc2vec

판매 글 감성분석을 위한 분류 알고리즘 적용

Z값이 0.5를 기준으로 데이터의 A / B / C 등급을 나누었을 때의 모델 accuracy와 Z값을 0.7을 기준으로 데이터를 나누었을 때의 모델 accuracy에 대한 결과. →

워드 임베딩 벡터의 차원이 10^n 개에 달했기 때문에 PCA, LDA로 차원축소 후 모델학습을 진행하였지만 오히려 accuracy가 평균적으로 20% 이상 감소하였음.

Z값을 올릴수록 정확도는 증가하였지만, 데이터를 균등하게 나누지 못함.

따라서 데이터를 균등하게 나누면서 accuracy가 비교적 정확한 0.7을 분할 기준으로 선정.

1. 테스트 데이터에 대해 비교적 정확한 0.7의 accuracy
2. 한국어 문서 분류에 대한 최신 연구 동향

위의 두가지 이유로

가설 2. 판매 게시글의 내용이 상품 상태에 유의미한 영향을 미칠 것이다

→ 가설 2를 채택

	word-embedding	model	accuracy(%)
z = +0.5	단순 빈도수	나이브베이즈	60
	단순 빈도수	선형 SVM	62
	tf-idf	로지스틱 분류	65
	tf-idf	SVM	67
	doc2vec	선형 SVM	50
	doc2vec	비선형 SVM	44
	doc2vec	나이브베이즈	44
	doc2vec	로지스틱 분류	50
z = +0.7	단순 빈도수	나이브베이즈	65
	단순 빈도수	선형 SVM	70
	tf-idf	로지스틱 분류	67
	tf-idf	SVM	70
	doc2vec	선형 SVM	60
	doc2vec	비선형 SVM	57
	doc2vec	나이브베이즈	57
	doc2vec	로지스틱 분류	60

<모델 학습 및 테스트 결과 표>

워드 클라우드

분류 된 결과 : 중고폰 등급별 특징 단어 시각화



단순 빈도수를 바탕으로 제작하였기 때문에
등급간 큰 차이를 발견하기 어려움



A등급



각 등급 별 작은 단어의 차이가 tf-idf 알고리즘에서는 가중치가 높아짐.



B등급

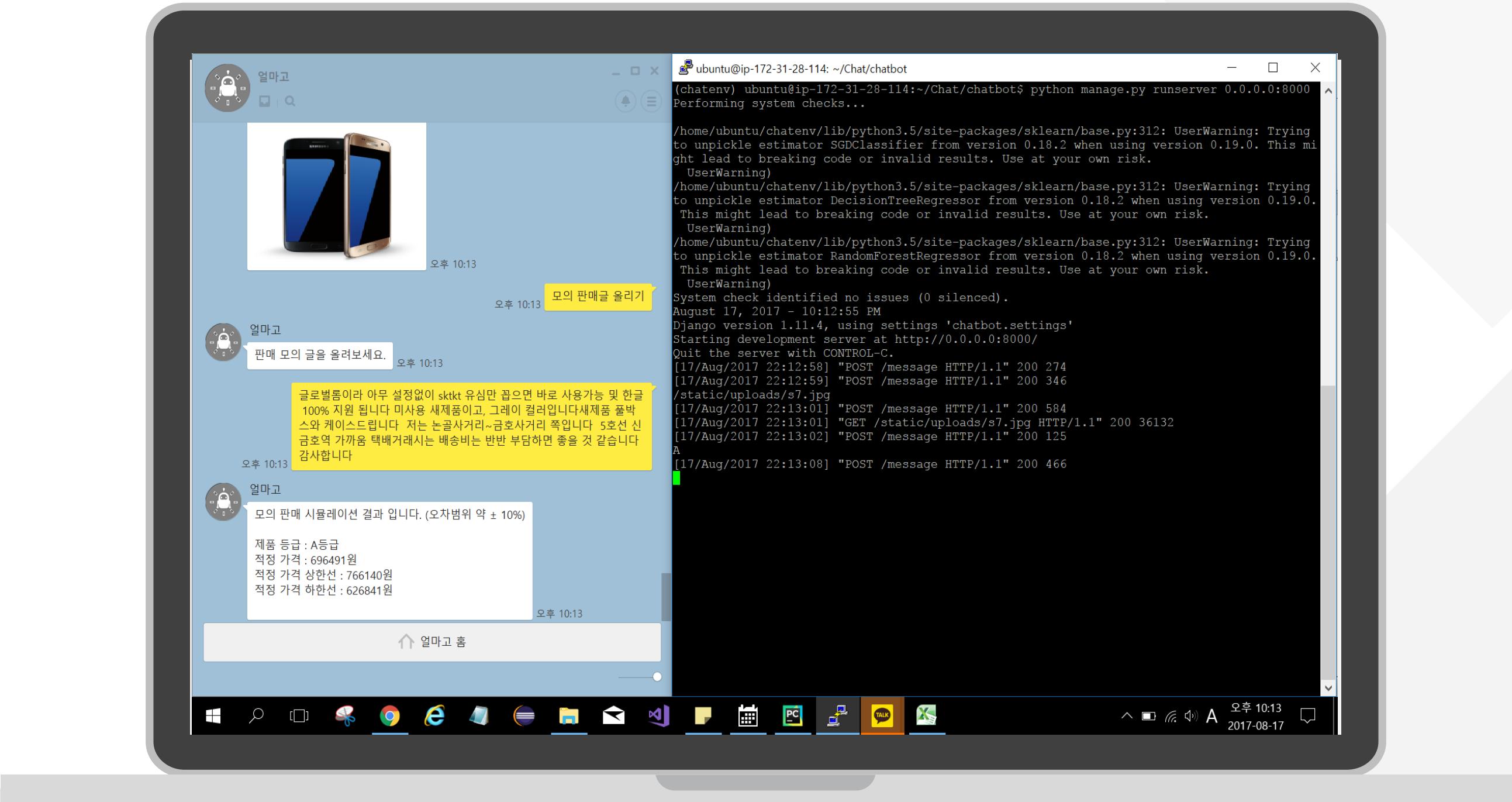


A/C등급에서 조금이라도 빈도수가 높은
특정 단어가 출현 할 경우 모델에 들어가는
가중치가 높아짐



C등급

감성 분석 테스트



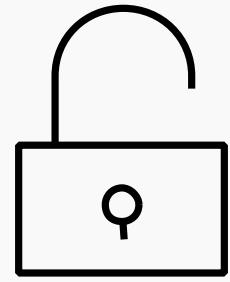
챗봇과 서버를 활용하여, 실제로 텍스트 데이터를 특정 등급으로
분류하는지 실제 거래 데이터로 테스트해 보았음

“글로벌롬이라 아무 설정없이 skkt 유심만 꽂으면 바로 사용가능 및
한글 100% 지원 됩니다 미사용 새제품이고, 그레이 컬러입니다새제품
풀박스와 케이스드립니다 저는 논골사거리~금호사거리 쪽입니다 5호선
신금호역 가까움 택배거래시는 배송비는 반반 부담하면 좋을 것
같습니다감사합니다”

→ ‘A 등급’ 으로 분류

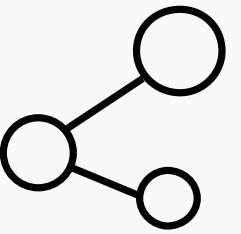
가격 예측 모델링

예측에 필요한 변수 선별



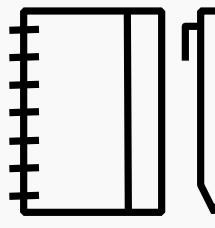
카테고리명 / 제품설명

결측 칼럼이므로 제거



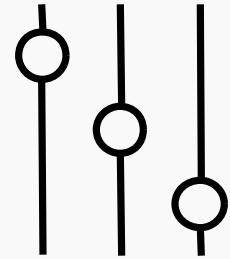
거래 구분 / 지불 방법

무의미한 데이터밖에 없으므로 제거



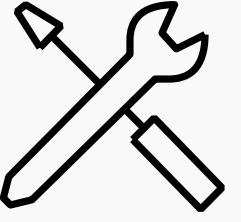
흡결등급

변수 중요도 20% 이상으로 중요한 변수



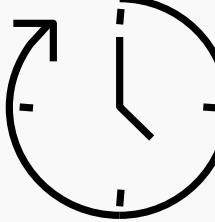
년-월 데이터 / 통신 물가지수

특정 조합에 변수 중요도 상승



중고폰 모델명

변수 중요도 30% 이상으로 가장 중요



출고가

변수 중요도 추출 결과 중요도 매우 낮음

가격 예측 모델링

예측에 필요한 변수 선별



카테고리명 / 제품설명

결측 칼럼이므로 제거



거래 구분 / 지불 방법

무의미한 데이터밖에 없으므로 제거



هم결등급

변수 중요도 20% 이상으로 중요한 변수



최종 모델링에 들어갈 피처 선택 -> 불필요한 피처 제거

년-월 데이터 / 통신 물가지수

특정 조합에 변수 중요도 상승

중고폰 모델명

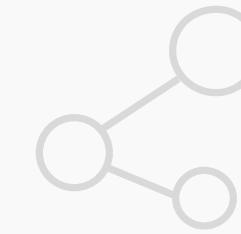
변수 중요도 30% 이상으로 가장 중요

출고가

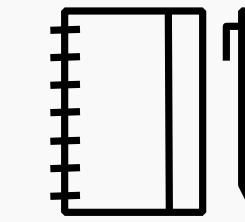
변수 중요도 추출 결과 중요도 매우 낮음

가격 예측 모델링

예측에 필요한 변수 선별

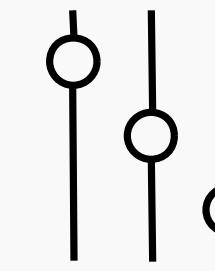


카테고리명 / 제품설명
범주형 데이터의 경우 수치형 벡터로 전환 (더미변수)
결측 칼럼이므로 제거
→ 연속 데이터 예측을 위해 **랜덤포레스트 회귀 모델** 시도
거래 그룹 / 판별 반복
무의미한 데이터밖에 없으므로 제거



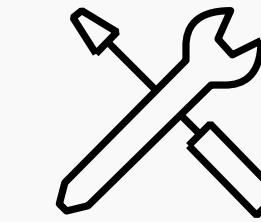
흡결등급

변수 중요도 20% 이상으로 중요한 변수



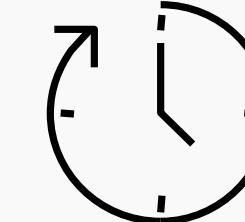
년-월 데이터 / 통신 물가지수

특정 조합에 변수 중요도 상승



중고폰 모델명

변수 중요도 30% 이상으로 가장 중요



출고가

변수 중요도 추출 결과 중요도 매우 낮음

R^2 (Training Score) : 0.95

R^2 (Test Score) : 0.89

랜덤 포레스트 회귀 모델

수치형 벡터 혹은 더미변수로 변환된 피처

(흡결등급 / 년-월 데이터 / 통신 물가지수 / 중고폰 모델명 / 출고가)

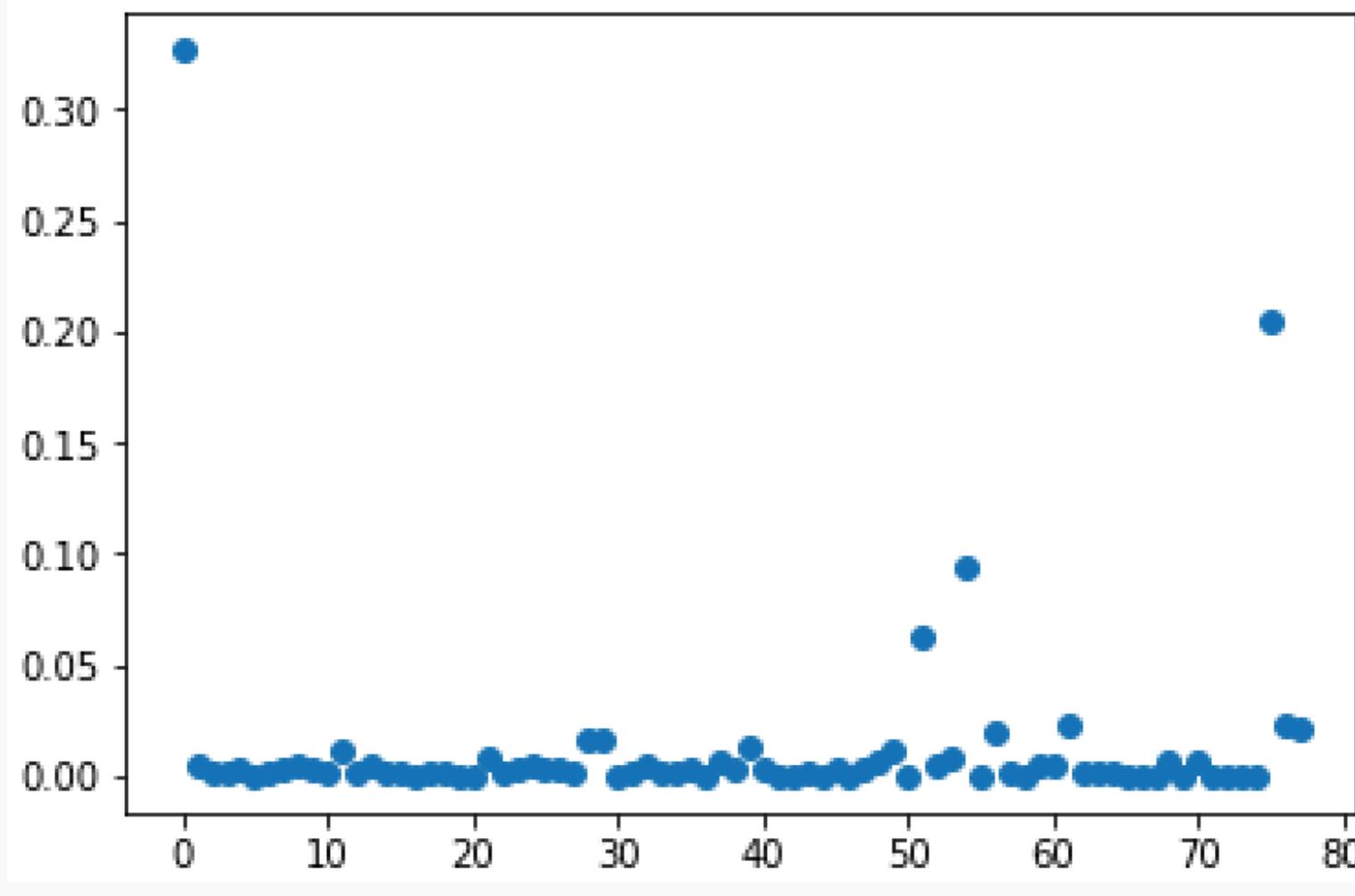
를 변수로 랜덤 포레스트 회귀 모델을 학습 결과

약 0.95의 수정 결정 계수 (Training score)와

0.89의 수정 결정 계수 (Test score)의 결과가 도출됨.

변수 중요도 평가

변수 중요도 추출을 통하여 최종 피처 선택



Step: AIC=-15653.41
판매금액 ~ X + 제품명 + date + rate + 판매금액_z_rank

	Df	Sum of Sq	RSS	AIC
<none>		203.36	-15653	
+ 출고가	1	0.03	203.33	-15652
- rate	1	0.29	203.65	-15648
- date	8	3.48	206.84	-15585
- 판매금액_z_rank	2	21.86	225.22	-15152
- 제품명	63	251.99	455.35	-11788
- X	1	394.38	597.74	-10317
>				

변수 중요도 플롯

더미변수 추출 과정을 통해 생성된 80여개의 피처 중,
중요도가 높은 변수는 다음과 같음

- 모델명 : 33%
- 판매 등급 : 20%
- 상위 더미 피처 : 10%, 7%

단계적 회귀 방법을 통해 AIC 통계량을 검출하여
출고가 피처를 제거해야 한다는 것을 알 수 있음.

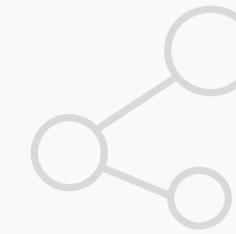
가설 1. 물품의 상태가 중고가에 영향을 미칠 것이다 → 채택

가격 예측 모델링

최종 피처 선택

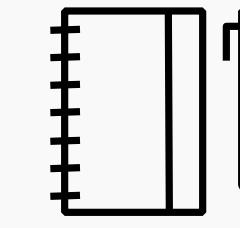


카테고리명 / 제품설명



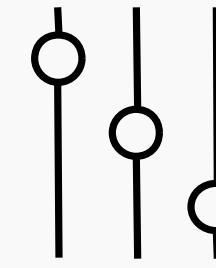
거래 구분 / 지불 방법

결측 **앞의 결과를 토대로 최종 모델 학습에 들어가는 변수 선택**



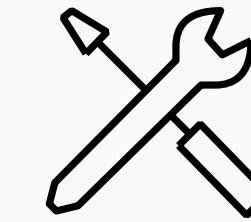
흡결등급

변수 중요도 20% 이상으로 중요한 변수



년-월 데이터 / 통신 물가지수

특정 조합에 변수 중요도 상승



중고폰 모델명

변수 중요도 30% 이상으로 가장 중요



출고가

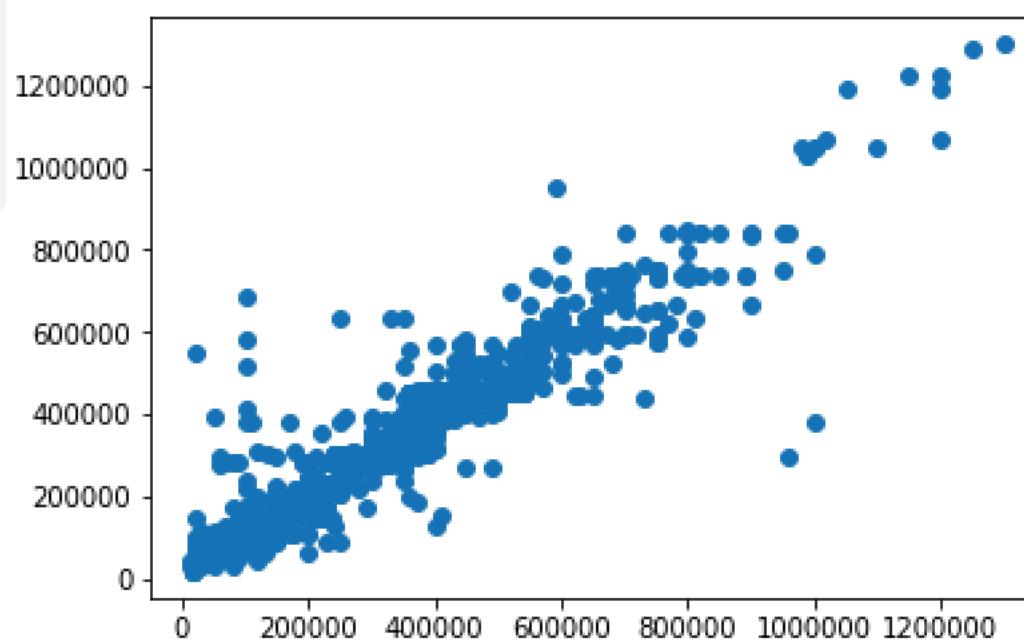
변수 중요도 추출 결과 중요도 매우 낮음

피처 선택 후 모델 학습 및 평가

R^2 (Training Score) : 0.95

R^2 (Test Score) : 0.89

최종적으로 선택한 피처로 모델링을 다시 진행한 결과,
결과는 동일했음.

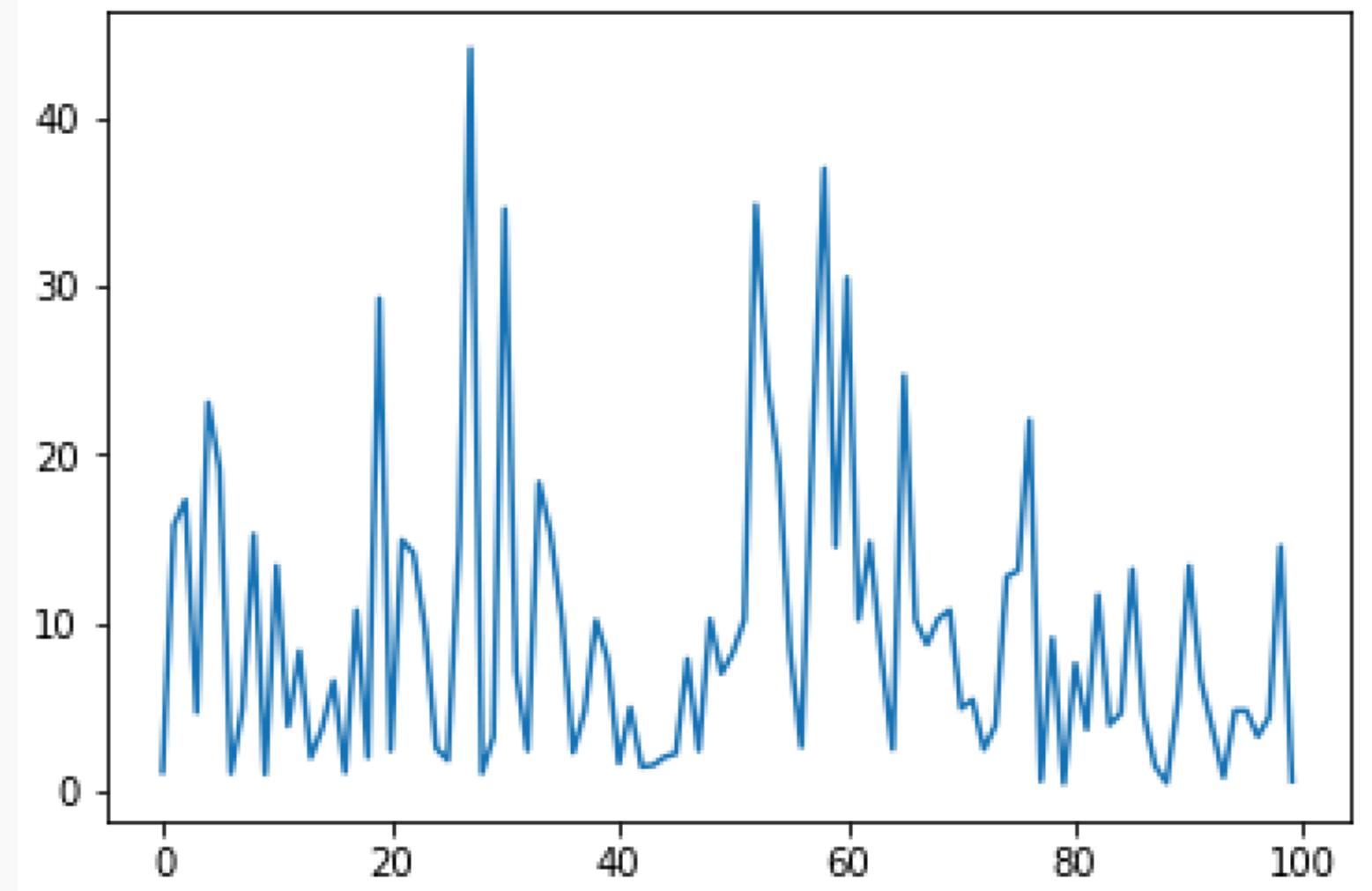


<모델에서 예측된 y 값과 실제 y 값을 각각의 축으로 한 플롯>

학습 모델의 검증

1. 가격 예측의 오차범위

테스트 데이터의 가격 레이블과 테스트 데이터를 넣었을 때의 예측 가격을 오차로 하여 오차 범위를 계산
→ ± 10%의 오차 범위 (아웃라이어를 제거했을 때)



2. K-fold 교차검증

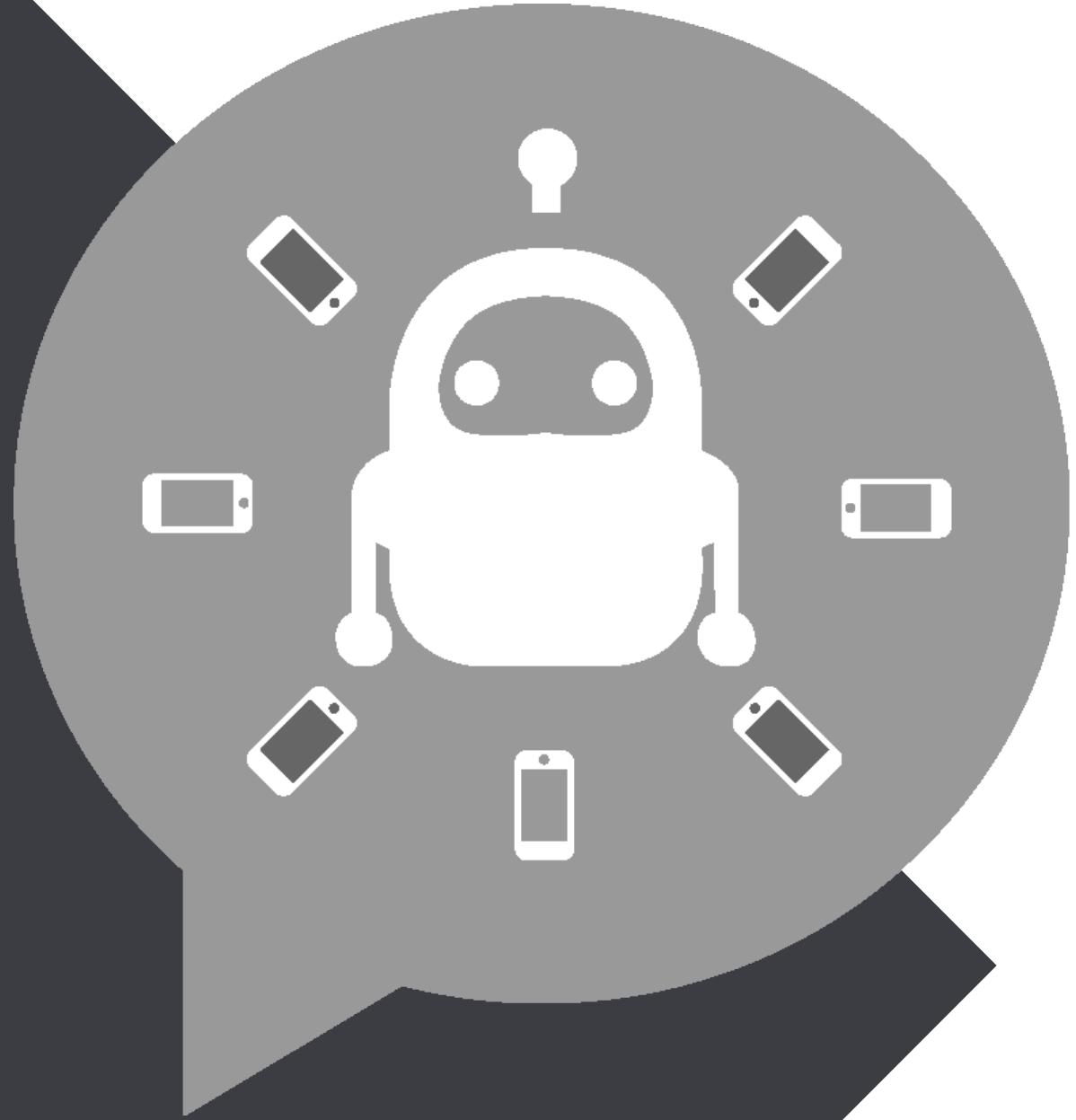
오버피팅 검증을 위해 교차검증을 진행. 층화 K-fold 방식으로 ($k=10$) 검증한 결과 평균 90%(최소 86% ~ 최대 94%)의 결과가 나와 오버피팅의 가능성은 낮은 것으로 판단.

04

챗봇 소개 및 시연

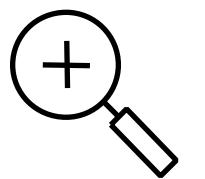
데이터 분석 과정을 거쳐 최종적으로 완성된
모델을 이용하여 챗봇의 형태로 예측 서비스 제공





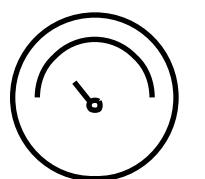
AlmhaGo

나에게 맞는 적정 중고폰 가격은 : 얼마고?



챗봇 소개

챗봇 얼마고는 카카오톡 채널을 통해 중고폰 적정가격 예측 서비스를 제공합니다.



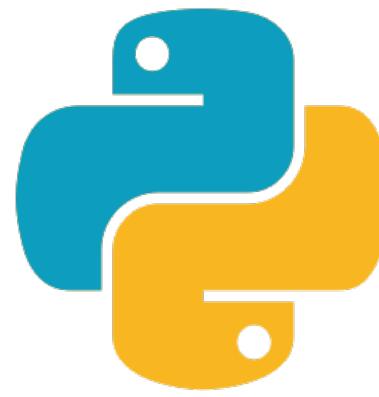
서비스 제공 범위

구매 희망자 : 최저 하한가, 최고 상한가, 적정 가격 예측 서비스 제공

판매 희망자 : 모의 판매 예측 서비스 제공

Application Server

Django Framework로 파이썬 모델을
임베딩하는 서버 구축. 사용자의 요청을
처리하고 피처를 입력받아 **예측 서비스 제공**



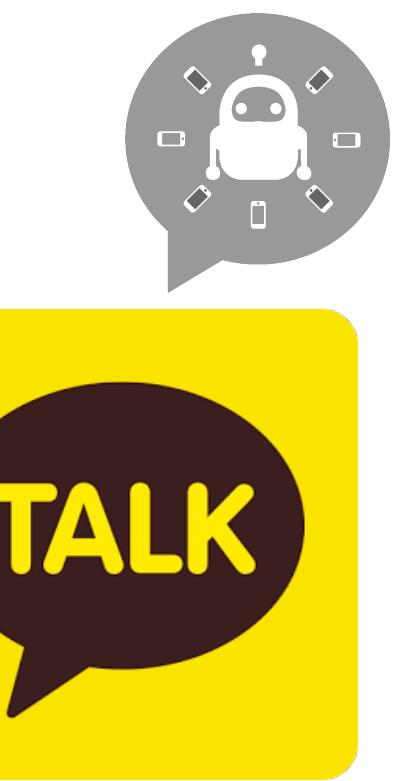
Machine Learning Model

파이썬으로 학습한 머신 러닝
모델을 서버에 직렬화된 pkl
확장자 파일로 저장하여 챗봇
application에 **임베딩**함.

모델 임베딩



1. Response,
 2. 예측 서비스 제공
-
1. Request,
 2. 예측에 필요한 피쳐



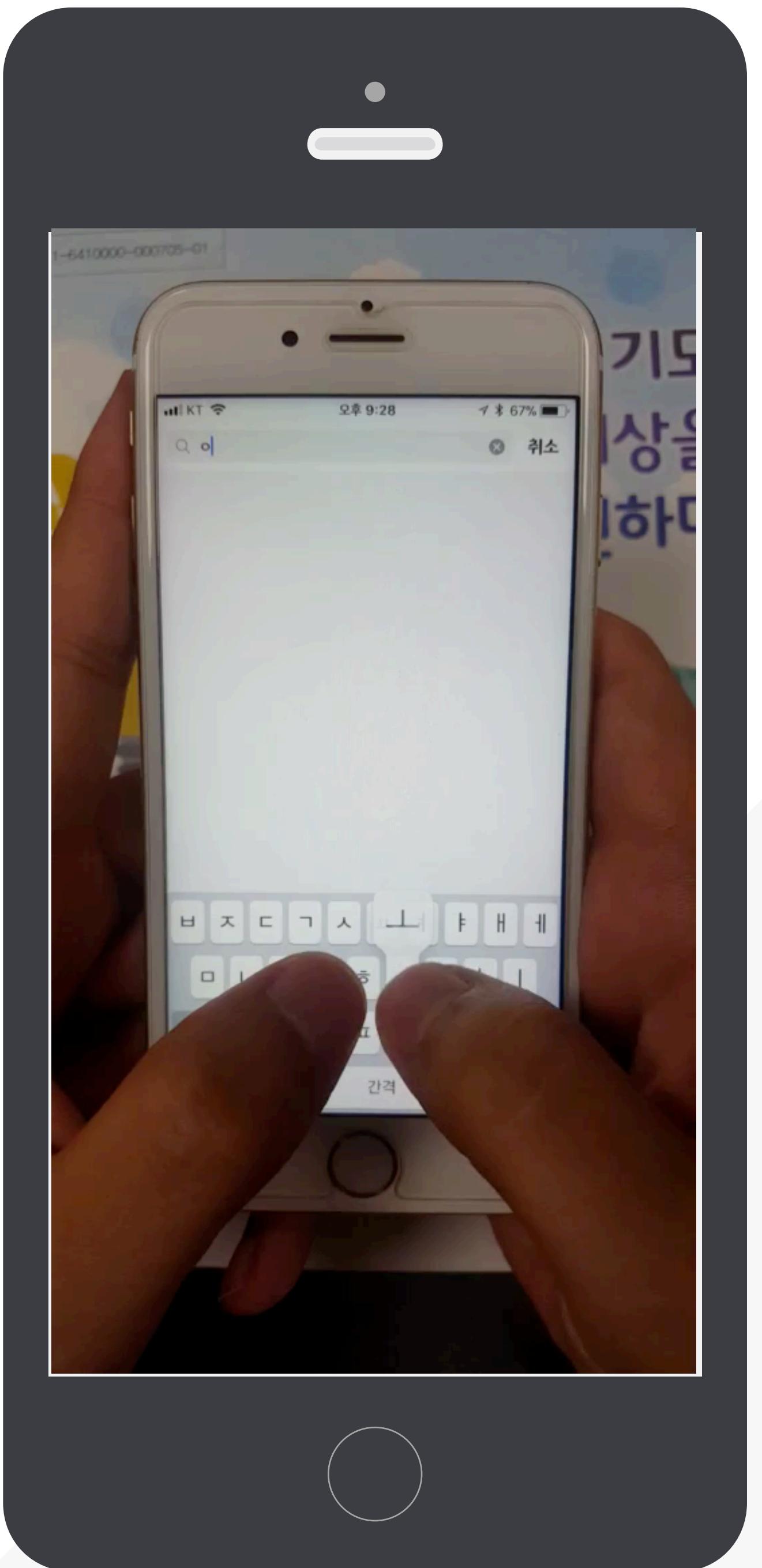
챗봇 시스템 구조도

AlmhaGo

나에게 있는 직장 중고문 가격은 : 얼마인가?

Chatbot Client

사용자가 직접적으로 이용하는
챗봇의 형태.
간단한 입력으로 머신 러닝
서비스를 제공받음.



챗봇 시연



중고나라 카페에서 예측 적중 확인

학습 데이터가 부족했던 모델의 경우, A등급을 부정확하게 예측하는 경향을 보이기도 하였으나 전반적으로 오차 범위 내에서 **꽤 정확한 예측**을 보여주었음.

실험 모델명 :

(갤럭시 S6, 갤럭시 노트3, 아이폰6, 베가 아이언2)

*SK/KT 갤럭시S6(G920S, G920K) 화이트 블랙 골드 특SS급(갤6, 갤6, 갤럭시6, galaxyS6)
*기스나 흠집 찍힘 없는 아주 깨끗한 제품입니다.(분실 도난 침수 폰은 절대 취급하지 않습니다)
*갤럭시S6S는 통신사 3사 호환이 모두되는 제품이라 SK KT LG 모두 사용 가능합니다.
*직거래는 서울 구로구 1호선 오류동역에서 하고 다른곳은 택배거래합니다.
*택배거래시 택배비 부담해주셔야 합니다. 3000원 같이 입금해주시면 선불로 배송 됩니다.
*유심만 끼우시면 바로 사용 가능하십니다.
*박스 디자인과 뒷커버는 sk입니다
*구성품은 골라서 구매 가능하십니다.
*풀박스는 선물용으로 구매 많이 하시고 단말기+충전기+이어폰+USB케이블+ 사용설명서 입니다.
*2개 이상 구매도 가능하니 연락주심 감사하겠습니다.
*단말기 = 27만
*풀박스 모든 구성품=28만
*케이스와 보호필름 사은품으로 드립니다
*가격 네고는 힘든점 양해 부탁드립니다.
*댓글 잘 안보니 전화나 문자 부탁드립니다.
*문의주실때 모델명 통신사 색상 구성 같이 말씀해주시면 좀 더 빠르고 정확하게 상담 가능합니다.

영주 직거래가능

67085 갤럭시S6 블루색상 32기가 구성품 포함 판매합니다 / 청주 직거래가능 [판매]

0257115 [중고폰#] 중고폰 최다 기종 보유매장!! 친절 상담을 약속 드리겠습니다 [판매]

70249569 갤럭시S6 골드 갤럭시S7엣지 핑크 팝니다 [판매]

370216302 갤럭시 s6 엣지팝니다..부천,광명 직거래 [판매]

369961877 [중고폰#] 중고폰 최다 기종 보유매장!! 친절 상담을 약속 드리겠습니다 [판매]

369954715 [중고폰#] 중고폰 최다 기종 보유매장!! 친절 상담을 약속 드리겠습니다 [판매]

.중고폰#] 중고폰 최다 기종 보유매장!! 친절 상담을 약속 드리겠습니다. [판매]

skt 갤럭시 노트 엣지 / note edge / n915s a+급 (천안) [판매]

노트3 정품배터리 거치대포함 팝니다. [판매] [1]

인천 노트3 레드 정장해지공기게 팝니다 내용 [판매] [1]

노트3 화이트골드/블랙골드(서울/동대문구) [판매]

노트3 블랙골드

[제품정보]

통신사 : SKT

모델/색상 : N900S 노트3 블랙골드

제품구성 : 단품

선택옵정 : 20%요금할인가능

확인기변 : 가능 / 정상해지 공기기

판매가격 : 140,000원

판매지역 : 서울 동대문구 회기역

배송방법 : 택배 / 빽빽이or신문지로 감싼후 안전하게 포장/

노트 네오 있습니다

010-9494-8219

010-구사구사-팔이일구

아이폰6 64기가 대구에서 판매합니다. 판매완료 [판매]

예약됐다가 평크나서 19.7에 바로팝니다.

오늘까지만 팔고 안팔리면 중고상에 넘기려구요

사용기간 1년3개월입니다.

정상작동합니다.

홈버튼 지문인식 버튼조절 모두 정상입니다.

다만, 터치가 먹는경우나 유령터치될때가 가끔 있습니다.

액정은 상단부분에 금간곳이있구요

테두리는 사진 확인해보시면 될 것 같습니다.

택포 19.7만입니다.

직거래는 남춘천역에서 가능하며 아니면 택배거래 가능합니다.

54823 액정깨진 아이폰6 64기가 골드팝니다 [판매]

247441 SK 아이폰6 16G 골드 팝니다 내고 그그 [판매]

0242675 아이폰7메트블랙256g 아이폰6s 아이폰6 [판매]

370217966 전주- 아이폰6+, 6S+(골드, 블랙, 화이트) A급 [판매]

370217769 전주- 아이폰6, 6S A급 직거래 [판매]

370216626 아이폰6 16GB 골드 대구지역 직거래 합니다. [판매]

010-2797-2328로 연락주세요 액정에 흠집 및 기스가 있습니다. 나머지 문제가 될 점은 없네요

구성품은 본체밖에 없습니다. 가격은 10입니다.

ps. 싱가폴판이라서 국내 as는 안 되는 것으로 알고있습니다.

방수캡은 헐거워져서 최근에 소니서비스센터에서 새로 구매한 녀석으로 바꾼이후 5번도 안 뺐습니다.

방수캡은 그냥 새것이라고 생각하면 됩니다.

소니 엑스페리아 Z3 부품용 구매합니다 [판매]

05

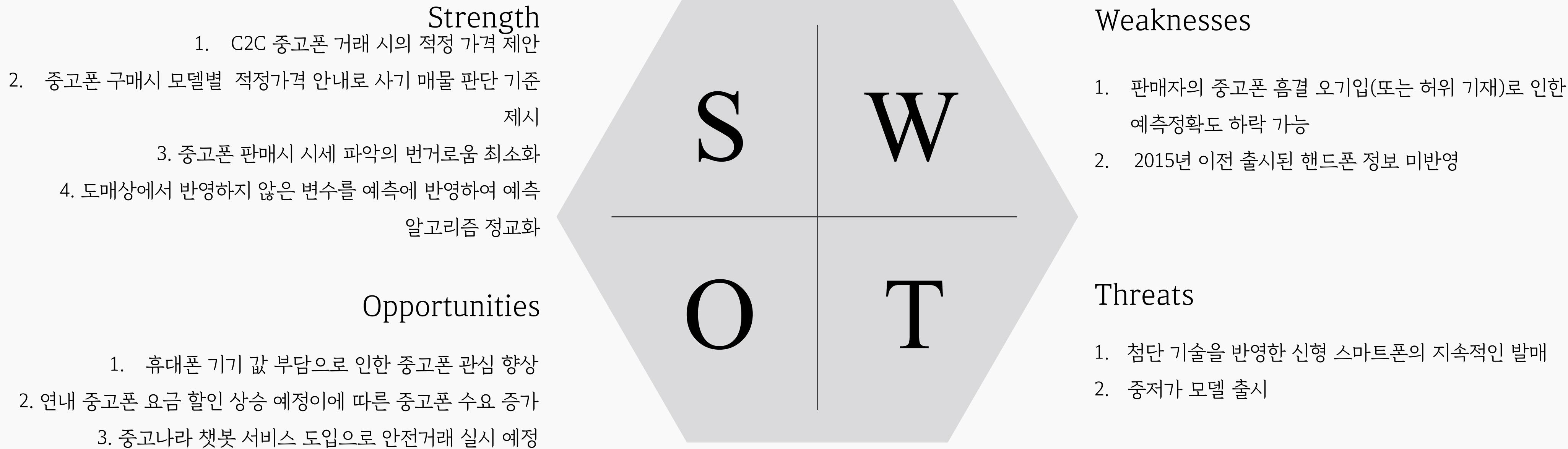
프로젝트 기대효과

최종 프로젝트의 기대 효과



기대효과

서비스 SWOT분석

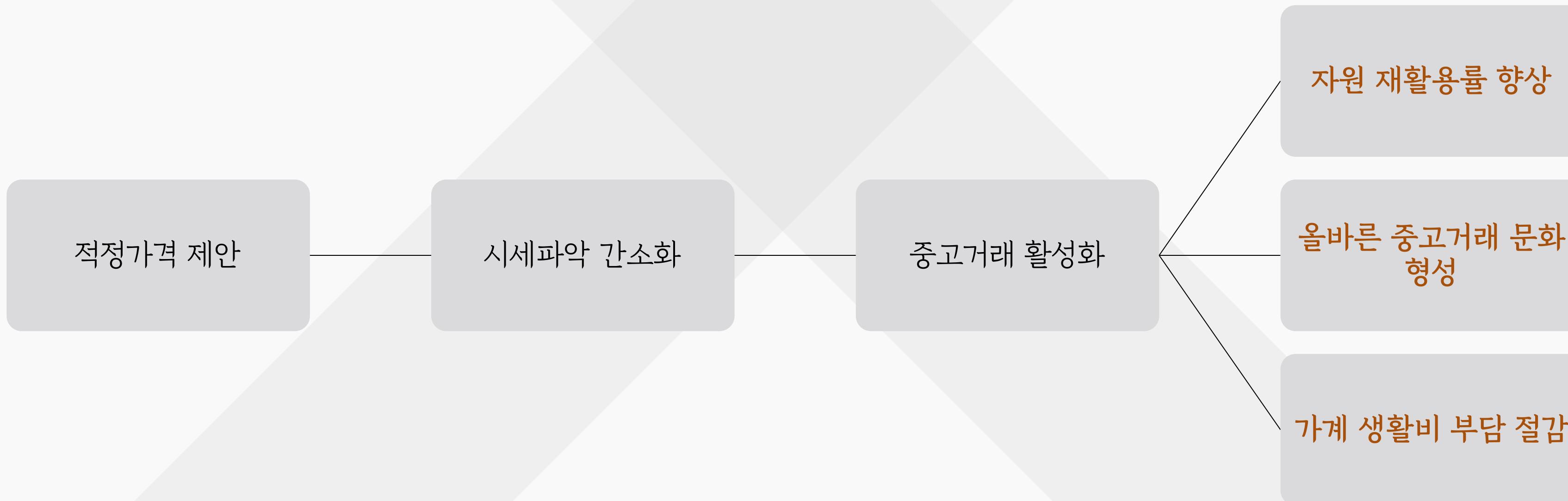


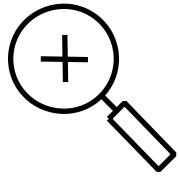
향후 발전 방향



중고나라 챗봇기반 상품 검색 및 안전거래 서비스 고도화 중 :

향후 적정가격 제안 서비스를 타 카테고리까지 확장 적용함으로써 중고나라 이용 편의성 향상





분석 측면

1. 데이터 양 부족
2. 카테고리 데이터를 직접 구축하는 것의 한계점(전체 휴대폰 모델을 반영하지 못한 아쉬움)
3. 텍스트 감성분석에 딥러닝(신경망 알고리즘)을 적용해보지 못한 아쉬움
4. 가격의 시계열 예측을 하지 못한 아쉬움



서비스 측면의 개선사항

1. 다양한 기능(실시간 시각화 등)을 삽입하여 사용자 편의성 향상
2. 도매업자들이 제시하는 매입가격과의 비교 부재

한계점 및
개선사항

참고 문헌

1. 중고 휴대전화의 소비자문제 및 개선방안, 한국소비자원 시장조사국 서비스조사팀, 2015.07
2. '1조 지하경제' 중고폰 유통 6. 세계, 동아닷컴, 2015.07.21, <http://news.donga.com/3/all/20150721/72615882/1>
3. 중고폰 확산일로 'SKT 개통 고객 열에 하나는 중고폰', 전자신문, 2016.06.07, <http://www.etnews.com/20160607000323>
4. 중고폰 시장 소비자 조사 현황, SK C&C 블로그, 2014.07.29, <http://skccblog.tistory.com/1878>
5. SK텔레콤 공식 온라인 스토어, "SK 텔레콤 공시지원금 안내", <http://shop.tworld.co.kr/handler/Dantong-SKT>, 2017/08/04
6. 올레샵, "단말기 지원금 안내", <https://shop.olleh.com/m/smart/supportAmtList.do>, 2017/08/04
7. LG U+, "상품 소개", https://www.uplus.co.kr/css/note/item/RetrieveItemDstrDisc.hpi?WEB_BNNR_ID=Intro_MO_Map_C01, 2017/08/04
8. 위키백과, <https://ko.wikipedia.org/wiki>, 2017/08/04
9. Yoon Kim, Convolutional Neural Networks for Sentence Classification, 2014
10. TF-IDF Calculate, <https://gist.github.com/clemsos/7692685>
11. PyCon Korea 2015, Korean NLTK, Gensim, <https://www.slideshare.net/lucypark/nltk-gensim>
12. KoNLP Document, <http://konlpy.org/en/v0.4.4/>
13. Python Scikit-learn Document, <http://scikit-learn.org/>
14. Wes mckinney, Python for data Analysis, 2013
15. Wes Mckinney & PyData Development team, pandas: powerful Python data analysis toolkit, 2016
16. Free Software Foundation, <http://ggplot2.tidyverse.org/index.html>
17. 카카오톡 API, https://github.com/plusfriend/auto_reply
18. THE R GRAPH GALLERY, <http://www.r-graph-gallery.com/about/>



THANKS!