

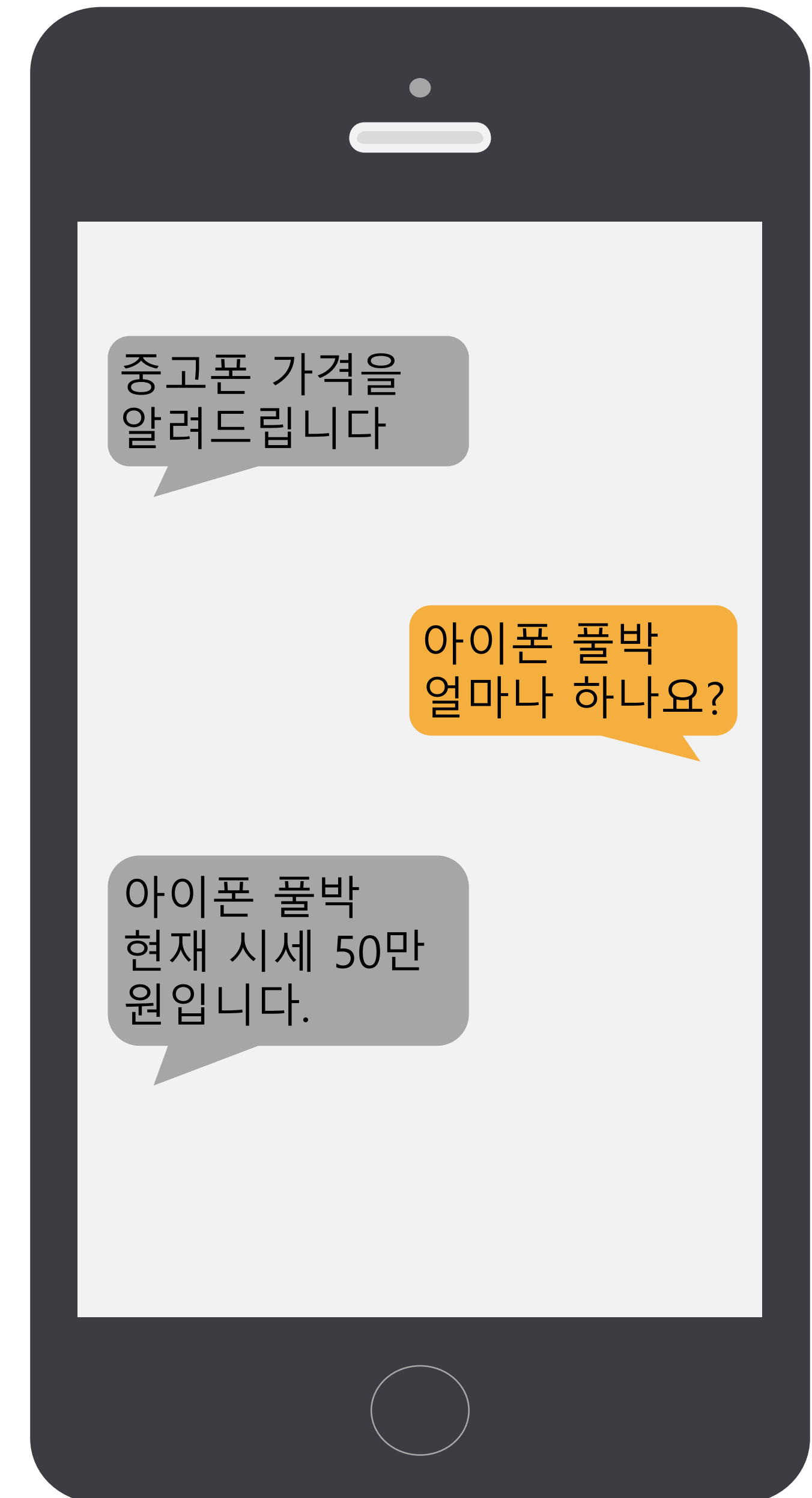
빅데이터를 활용한 중고폰 적정가격 예측 서비스

Team Scoop

Scoop은 '뜻밖의 행운'이라는 뜻으로, 빅데이터 분석을 통하여 뜻밖의 행운을 선사해주고자 이름을 선정하였습니다.

Project "AlmhaGo"

'얼마고'는 중고폰 적정가격을 예측 해주는 챗봇 명입니다.



CONTENTS

01. 상황분석

02. 문제점 도출

03. 데이터 분석

04. 챗봇 소개

05. 프로젝트 기대효과

공통

- 주제 선정
- 기획 및 문제 추출

팀원 수행 역할

허재훈

데이터 전처리 중 모델명 사전 제작 및 2차 가공 데이터에 대한 피쳐 추가 알고리즘을 담당 하였고, 텍스트 마이닝을 통한 분류 알고리즘 구현을 중심으로 전반적인 데이터 분석 프로세스를 수행하였습니다. 전처리에는 R, 모델링은 python 언어를 사용하였습니다

김은석

웹 크롤링을 통해 스마트폰 표준정보 데이터를 구축하여 데이터 전처리에 참여하였습니다. 또한, 최종 모델인 랜덤 포레스트 회귀 모델링 및 적용을 통해 전반적인 데이터 분석 프로세스를 수행하였습니다. 전처리에는 R, 모델링은 python 언어를 사용하였습니다

전흥진

아마존 AWS를 이용하여 서버를 구축하였고 Python언어와 Django Framework를 사용하여 카카오톡 가격예측시스템인 '얼마고'를 개발하였습니다

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

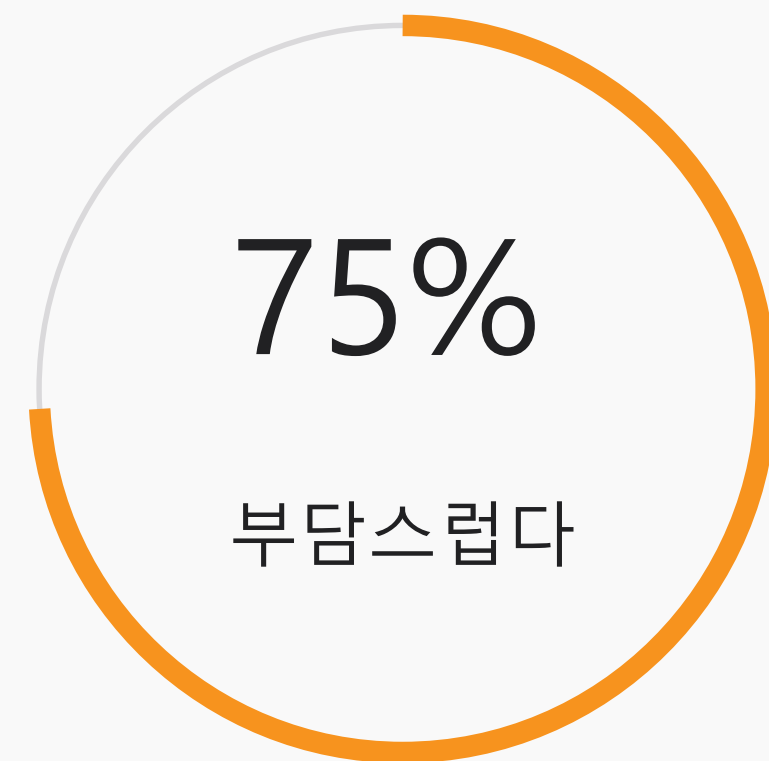
01

상황분석

상황분석

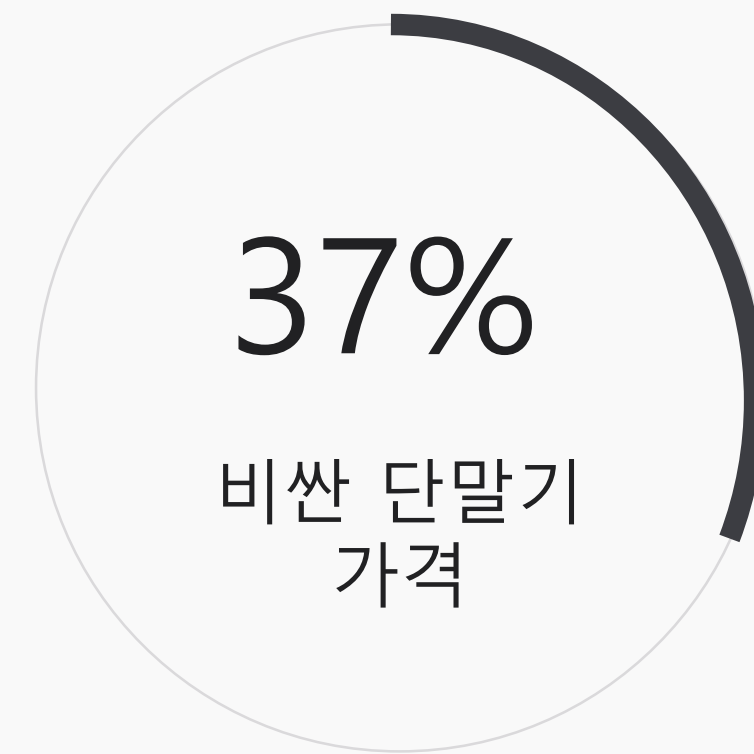
한 해 유통 중고폰 1000만대 추정

2017년 녹색 소비자 연대, n=1000



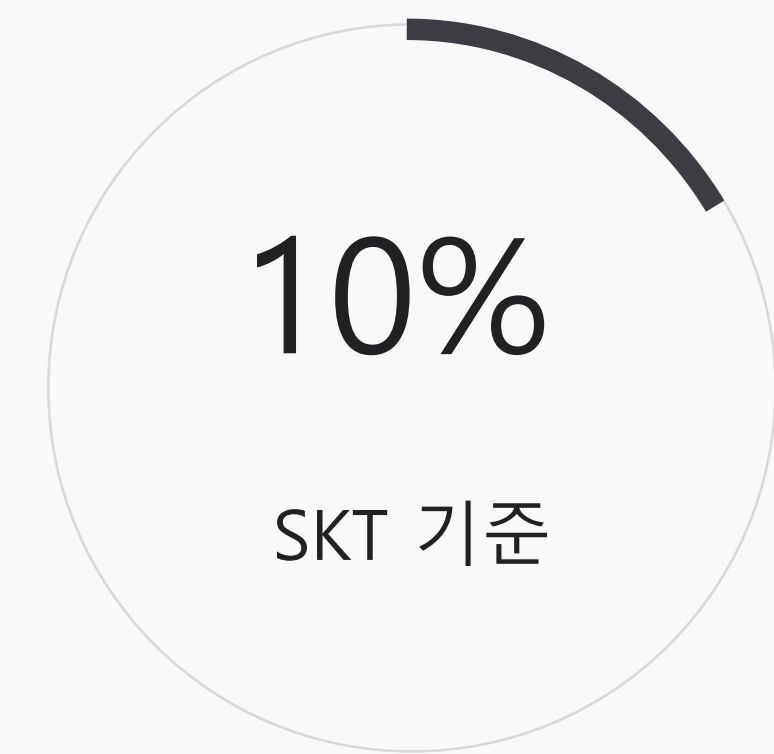
가계 통신비
부담스럽나

2014년 한국소비자원, n=1000



가계 통신비가
부담스러운 이유

2016년 전자신문



2016년
중고폰 개통
비중

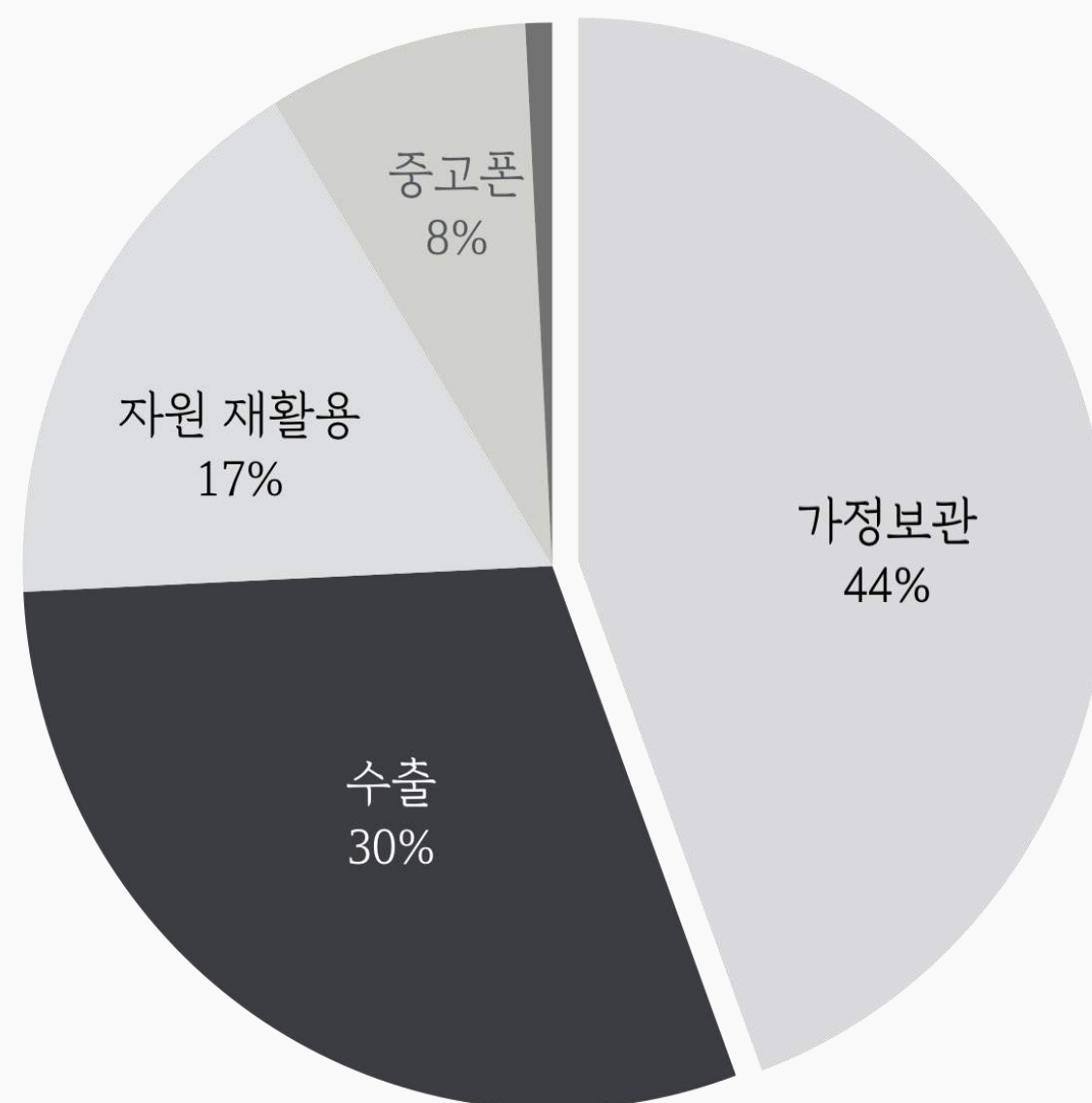
- 우리나라에서 한해 유통 중고폰 1000만대, 거래규모는 약 4조원에 육박하고 있다고 합니다.
- 또한 많은 사람들이 단통 법 이후로 핸드폰 단말기 가격을 부담스러워 하는 것을 다양한 조사를 통해 알 수 있었습니다.

상황분석

내 품 안의 중고폰. 48%에 육박

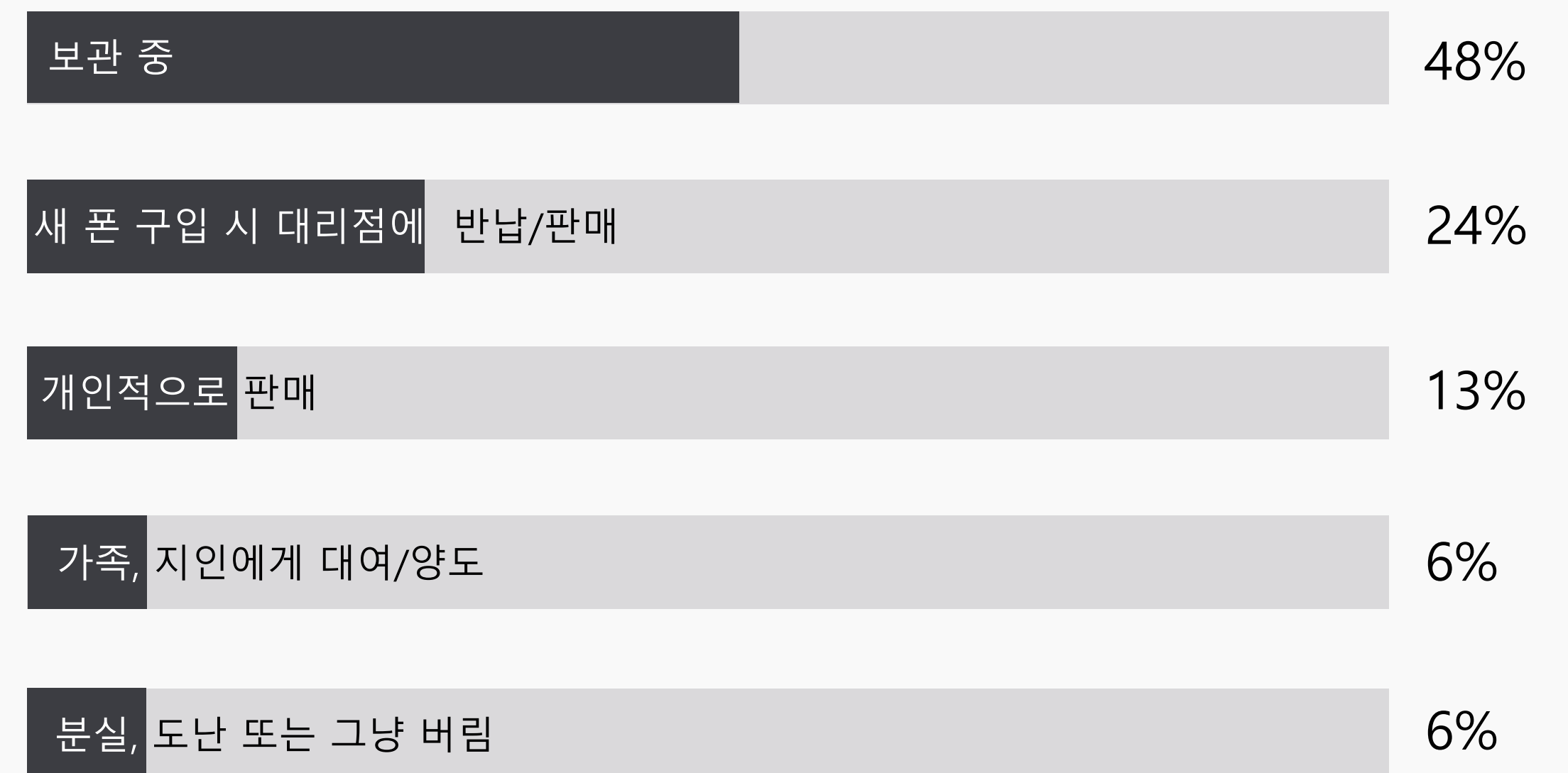
폐휴대폰 회수 현황

2010년 환경부, n=18440000



이전 휴대전화 처분 형태

2014년 한국갤럽, n=1000



- 하지만, 전체 중고폰 중 내 품 안의 중고폰이 더 많은 상황입니다.
환경부에 따르면, 1844만대의 폐휴대폰 중, 818만대의 휴대폰은 그냥 가정에서 방치되고 있다고 합니다.
또한, 2014년 한국갤럽 조사에 따르면, 48%에 가까운 사람들이 직전에 사용한 휴대폰을 가정에 보관 중이라고 답하였습니다.

상황분석

왜 이전 휴대폰을 보관하고 있을까요?

중고폰 미처분 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	기기에 남아있는 개인정보 유출 우려	130명
처분하기 번거롭거나 바빠서	140명	나중에 쓸 데가 있을 것 같아서	105명

- 이러한 중고폰을 미처분하는 가장 큰 이유는 중고폰 구매 사이트나 대리점에 팔아도 받을 수 있는 금액이 적기 때문으로 파악되었습니다.
- 그래서 저희 스쿱은 받을 수 있는 가격이 낮은 원인에 대해서 조사를 진행하였습니다.

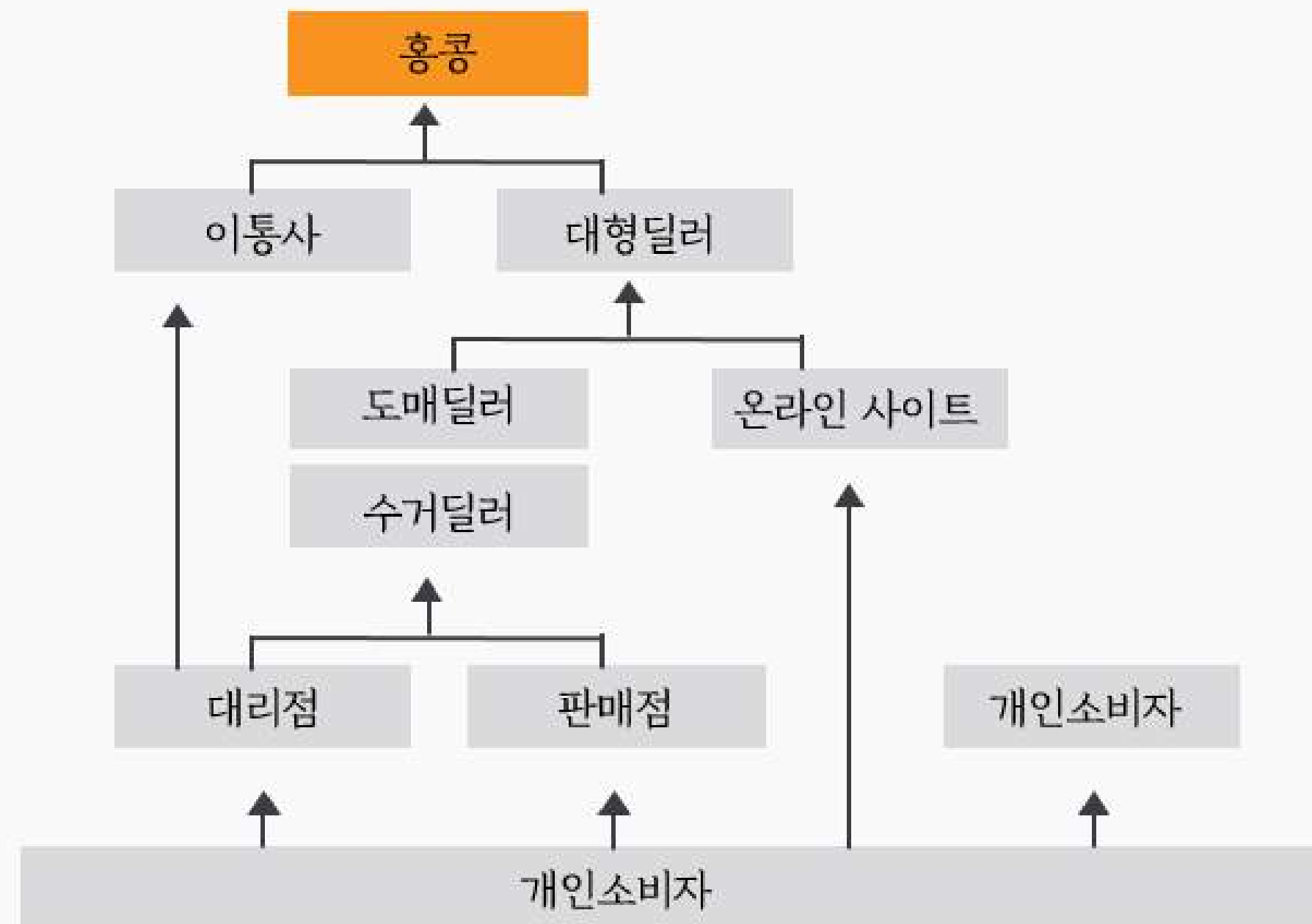
시장의 전반적인 상황을 토대로
문제가 되는 상황을 도출하고자
하였습니다.

02

문제도출

중고폰 유통 경로

원인은 중고폰 유통 경로에 있었습니다.



2014년 매경 이코노미 제 1781호

- 중고폰 구매 사이트나 도매상에 팔았을 때, 낮은 가격이 책정되는 이유는 **중고폰 유통 경로**가 원인이었습니다.

- 왼쪽 프로세스를 통해 알 수 있듯이, 홍콩 도매상이 90%의 중고폰을 가져가는 독점 현상이 이어지고 있습니다.

- 홍콩 도매상들은 많은 이익을 가져가기 위해 헐 값에 중고폰을 사들이고 있습니다. 결국 이러한 원인으로 고객들의 중고폰 판매 가격이 낮아졌습니다.

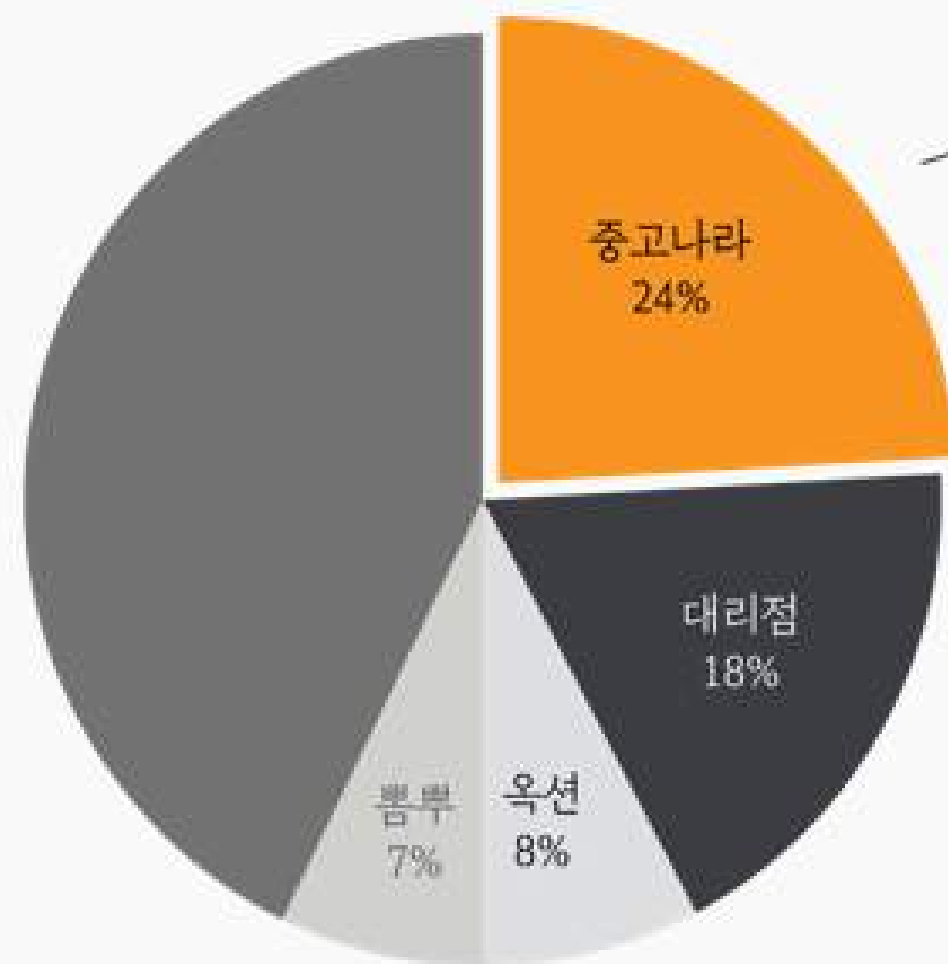
- 또한 이러한 도매상들의 중고폰 흡결 기준이 개개인마다 다르고, 명확하지 않기 때문에 소비자들은 중고폰 가격에 대한 큰 불만을 가지고 있었습니다.

문제점 도출에 따른 해결방안 제시

이용률이 높을 것이라 생각했던 중고나라, 문제점 존재

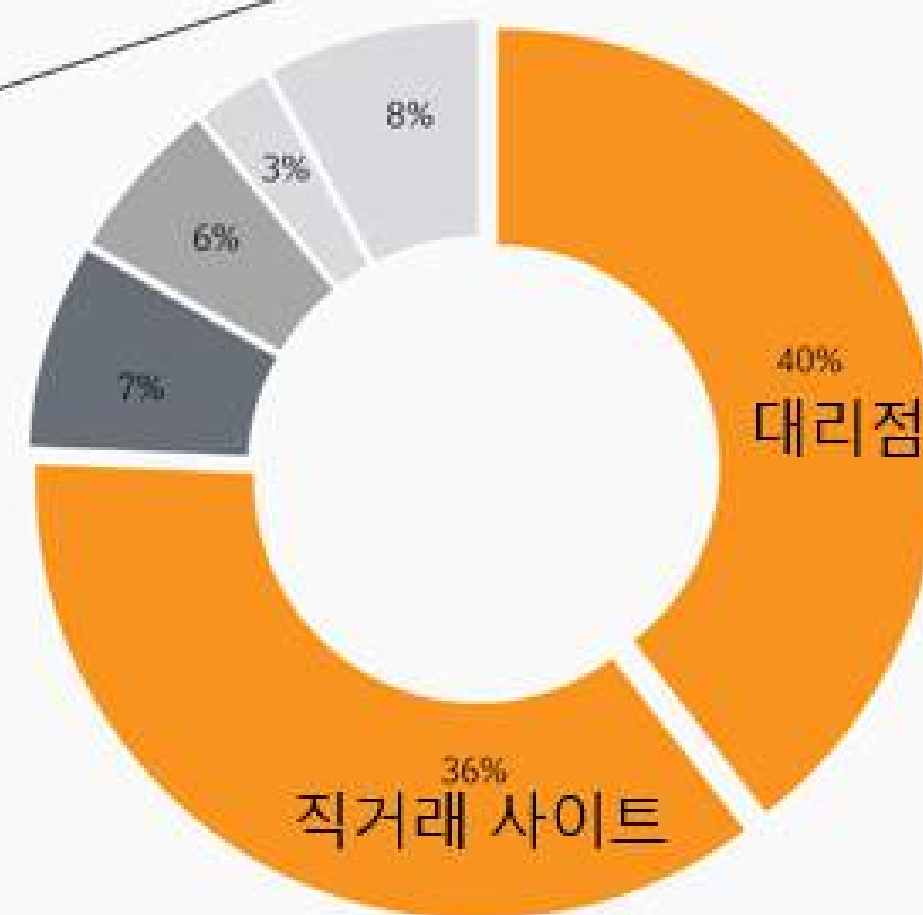
중고폰 판매 경로 인지도

2014년 한국갤럽, n=1000



중고폰 판매시 이용 경로

2014년 한국갤럽, n=482



- 이러한 사실들을 소비자들도 알고 있기 때문에, 많은 사람들이 직거래 중고나라 채널 판매 경로를 인지하고 있었습니다. 결과적으로 많은 사람들이 직거래 사이트를 이용할 것이라고 추측하였습니다.
- 하지만 중고폰 판매 시 이용 경로는 대리점과 직거래 사이트가 비슷하게 나왔음을 확인하였습니다
- 이는 직거래 사이트 사용시, 중고폰 거래 적정 가격 기준이 없어서 사용자들이 번거로워 하는 이유 때문이었습니다.
- 따라서 저희 팀 스콥은 **앞선 문제점들을 해결할 수 있는 중고폰의 적정 가격을 제시해줄 수 있는 서비스**를 구축하고자 하였습니다.
- 이때, 적정 가격을 제시해 줄 수 있는 예측 모델을 현재 중고나라 현황 데이터를 분석하여 구축하기로 결정하였습니다.

Problem

저조한 폐휴대폰 재활용률로 인한 자원낭비 등

Reason

비합리적인 중고폰 판매가격 & 알 수 없는 중고폰 적정 가격

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

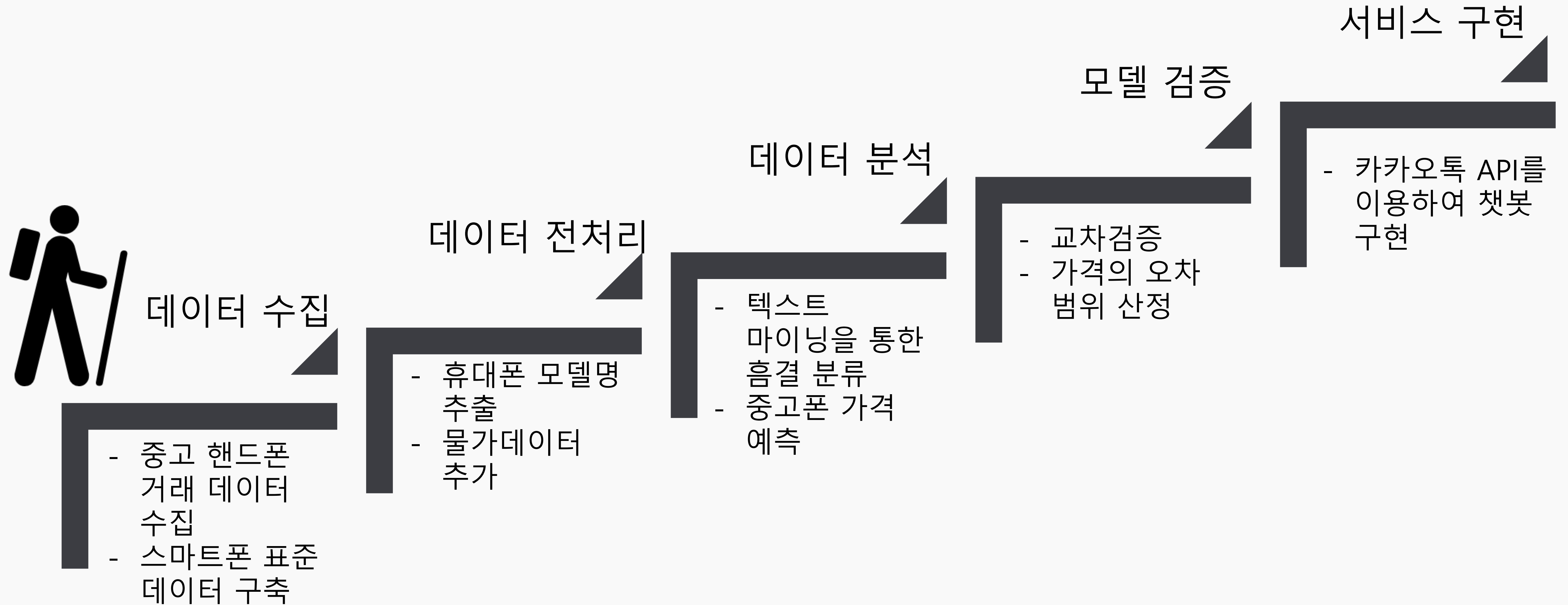
최종
문제도출

도출된 문제점을 해결을 위해
데이터 분석을 통해 솔루션을
제공하고자 합니다.

03

데이터 전처리/분석

프로젝트 전체 아키텍처



데이터 전처리 프로세스

Raw 데이터 → 정제된 데이터

1. 중고품목 유통현황 데이터 확보

(출처 : 한국 데이터베이스 진흥원의 데이터 스토어)

2015년부터 2017년까지, 네이버 카페 중고나라에 올라온
약 45만 건의 판매 글 데이터를 확보하였습니다.

2. 1차 데이터 정제

45만 건의 데이터에서, 휴대폰 카테고리에 해당하는
약 5만 건의 데이터를 추출하였습니다.

추출한 데이터에서
광고 글, 비 휴대폰 게시 글, 가격 미 기재 글, 구매 희망 글 등
불필요한 데이터를 제거하였습니다.

4. 기종에 따른 모델명 사전 제작

모델명을 애매하게 입력하는 경우를 고려하여
모델명 사전을 별도로 제작하였습니다.

ex) 갤럭시 노트 3 모델 : “갤럭시3”, “노트3”, “노트3” ... 등
판매자가 입력할 수 있는 경우의 수를 사전에 저장해두었습니다.

3. 스마트 폰 표준정보 데이터 구축

(출처 : 통신 3사 공식 홈페이지)

통신3사 홈페이지 크롤링을 통해, 150개 스마트 폰 기종의
(모델명 / 용량 / 시기별 출고 가격 / 제조사)을 조사하였습니다.
→ 스마트 폰 표준정보 데이터를 구축하였습니다.

5. 2차 데이터 정제

모델명 사전과 스마트 폰 표준정보 데이터를 활용하여,
개발한 알고리즘을 통해 1차 정제 데이터에,
(모델명 / 용량 / 출고가 평균 / 제조사) 변수를 추가하였습니다.

통신 물가지수 변수 추가

2015년의 휴대폰 기기값을 기준, 백분율로 표현한 데이터입니다.

ex) 2016년 통신 물가지수가 103이라면
2015년 대비 3% 증가했다는 의미입니다.

데이터 전처리 완료, 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	츨뎡 갤럭시s6엣지제품 단품상태 사진참고용량 32기가선택약정머...	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 발령나서 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 as가능하며 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 불...	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 유플라스이면 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 갔지만 사용하시는데 지장없어요불편하시...	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 민트색이에요원제는 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 퀵 신청하기 판 매 양 식아이...	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 퀵 신청하기 판 매 양 식아이...	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+ 피쳐폰 폴더 슬라이드 2g폰 3g폰 정리합니다 lgu+ 피쳐폰...	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 가급...	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 피쳐폰 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 데이터 변수

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

데이터 전처리 완료,
데이터 분석 시작!

5000개의 데이터, **150개**의 휴대폰 기종

데이터 분석을 실시하면서 위와 같은 가설들을 검정하였습니다.

가설 1. 판매 게시 글의 내용이 상품 상태에 유의미한 영향을 미칠 것이다.

가설 2. 물품의 상태가 중고가에 영향을 미칠 것이다.

-저희 팀은 판매 게시 글의 내용이 중고가에 영향을 미칠 것이라 판단하였습니다.

-따라서 각 모델 별 가격을 A/B/C로 삼등분 하여 y-label로 설정하였습니다.

이후 판매 게시 글의 텍스트 마이닝을 통해 분류 알고리즘을 적용하여 분류 정확도를 측정하였습니다.

가설 3. 출고가는 중고가 형성과 연관이 있을 것이다.

-휴대폰 출고가와 판매가격이 연관이 있을 것이라 판단하여,
연관성 검정을 실시하였습니다.

데이터 분석 프로세스 요약

2. 텍스트마이닝을 통한 흠결등급 분류 모델링

텍스트마이닝을 통한 흠결등급 분류 모델링을 통해,
가격을 기준으로 등급을 매긴 것이 정확도를 판별

텍스트마이닝 과정

- 1) 텍스트 전처리 : 특수문자, 전화번호 등 숫자 제거
- 2) 중고 폰 관련 불용어 사전 제작
- 3) 사전 기반 필터링 후, 한글 형태소를 Word Embedding

1. 탐색적 데이터 분석

목표변수인 판매 금액을 예측하기 위한 **탐색적 데이터 분석**을 진행하였습니다.

- 1) 각 휴대폰 모델의 가격에 대한 **정규성 검정을 실시**하였습니다. → 정규성을 보임
→ 흠결등급 변수를 추가, 각 휴대폰 모델마다 가격을 기준으로 3분류 (A/B/C)
ex) 갤럭시 노트 3 13만원 → A등급, 15만원 → B등급, 20만원 → C등급

- 2) 출고가와 판매가격 사이의 **연관성을 검정을 실시**하였습니다. → 연관성이 없음

“글로벌롬이라 아무 설정없이 sktkt 유심만 꼽으면
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 같습니다
감사합니다”

=> **A/B/C 등급**

4. 교차검증을 통한 모델 성능 평가

오버피팅을 위한 교차검증을 진행 → 86% ~ 94% 사이의 검증 값
검증 결과, **오버피팅의 가능성이 낮은 것**으로 확인하였습니다.

3. 가격 예측 모델링

최종 변수 (홈결등급 / 년-월 데이터 / 통신 물가지수 / 중고폰 모델명 / 출고가)를
통한 **랜덤 포레스트 회귀**를 실시하였습니다.

5. 가격예측의 오차 범위를 포함한 가격 산정

회귀 모델의 가격 예측 결과 : $\pm 10\%$ 의 가격 예측 범위

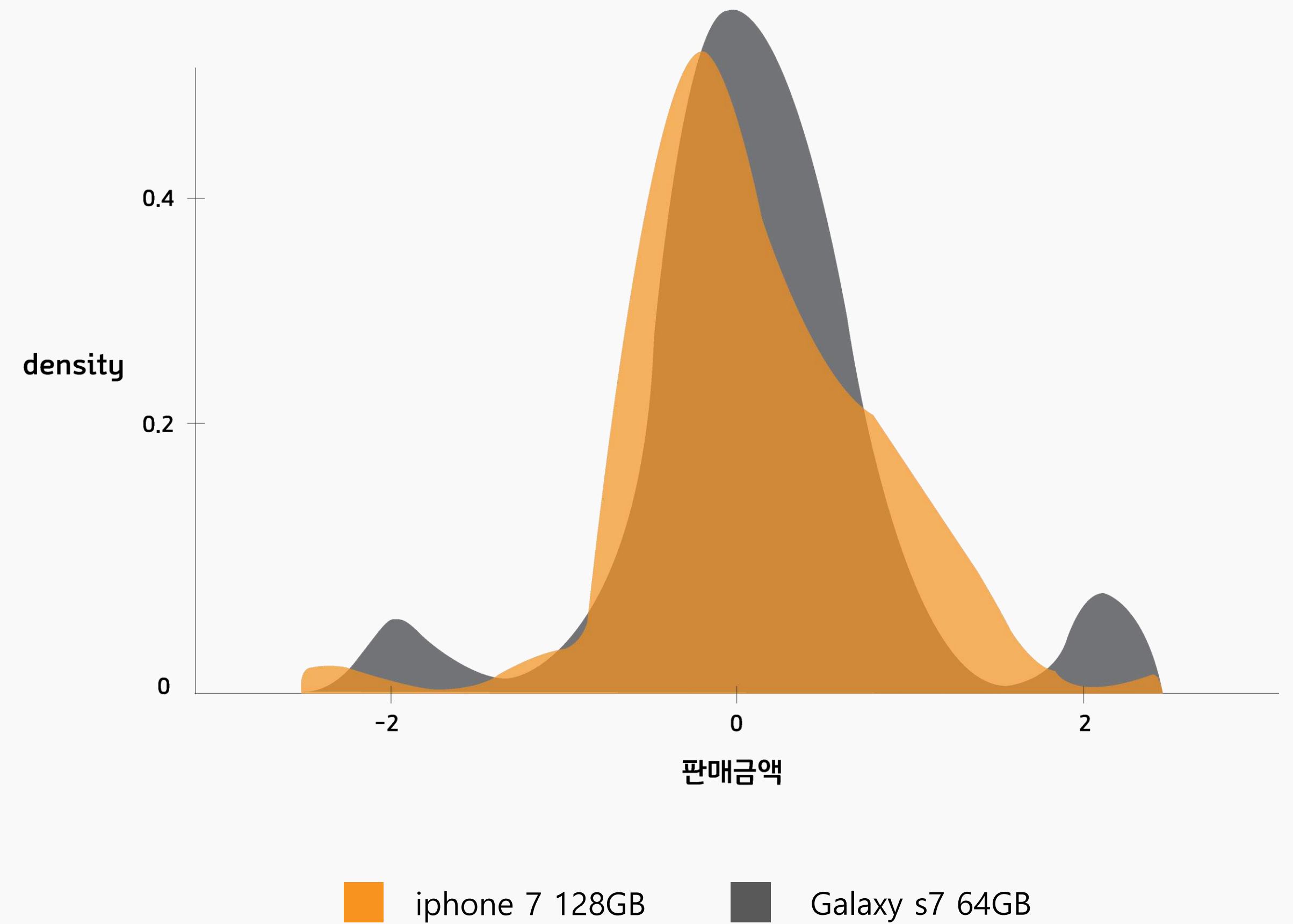
데이터 분석
프로세스 상세 내용

탐색적 데이터 분석 과정

기종 별 '판매금액'을 기준으로, 판매제품마다 등급을 부여하였습니다.

기종 별 판매금액 분포 검정

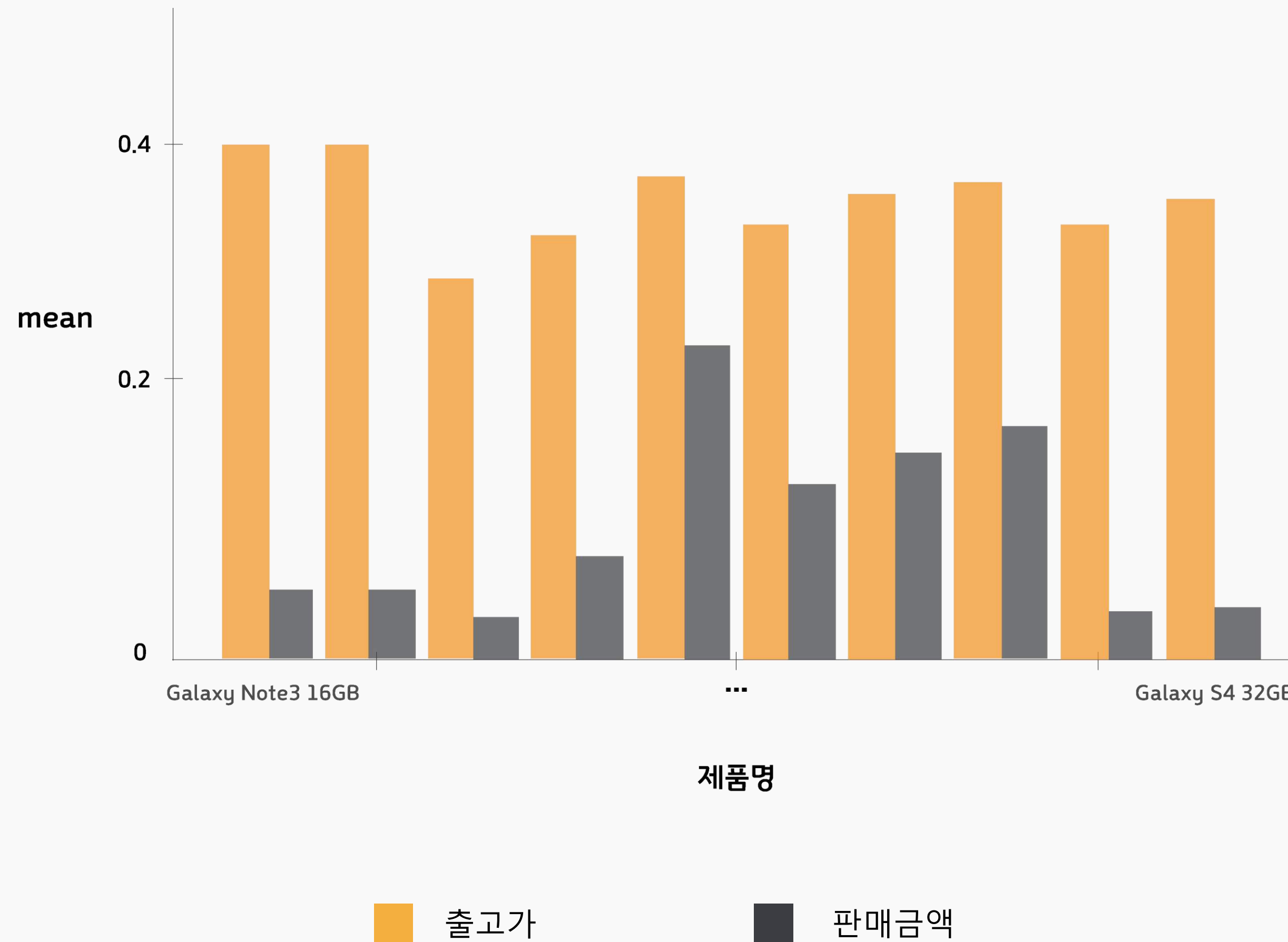
1. 기종 별 '판매금액' 변수가 정규분포인지 확인하고자, 두 기종을 선정하여 판매금액 변수를 **표준화**한 후, **확률 밀도함수 그래프**로 표현하였습니다.
2. 각 기종에 대해 **Shapiro-wilk Test**를 시행하여 정규성을 띠는지 검정
→ W: 0.99 , p-value : e-10
→ "선형적 성질이 정규분포적 성질을 띠지만, 정규분포라고는 볼 수 없다."
→ 하지만, **구간별로 데이터를 나누기에는 무리가 없다**"고 알려져 있습니다.
3. 기종간 분산분석 실시 → p-value가 모두 1에 가까운 값
→ 모든 기종 별 가격이 정규분포적 성질을 띠는 것을 파악하였습니다.



따라서 표준화된 판매금액 데이터의 z를 기준으로, 판매되는 제품의 등급을 **A / B / C** 분할하였습니다.

탐색적 데이터 분석 과정

판매금액에 영향을 미치지 않는 불필요한 변수를 판단하였습니다.



출고가와 판매금액간의 차이검정

출고가와 판매금액 사이의 연관성 검증을 실시하였습니다.

→ 핸드폰 기종 별 출고가와 평균 판매가격간의 상관관계를 히스토그램으로 나타내었습니다.

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

가설 3. 출고가는 중고가 형성과 연관이 있을 것이다.

→ 기각

텍스트 마이닝 과정

텍스트 마이닝을 통한 객관적인 중고 폰 등급 판정 기준을 수립하였습니다.

Step 01



텍스트 전처리

1. 데이터 내에서 분석에 필요한 모든 텍스트 데이터를 취합
→ Contents 칼럼 생성
2. 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

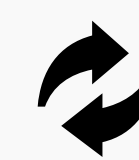
Step 03



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 텍스트 마이닝을 통한 분류 알고리즘 적용

Z값이 0.5를 기준으로 데이터의 A/B/C 등급을 나누었을 대의 모델 정확도와 Z값을 0.7을 기준으로 데이터를 나누었을 때의 모델 정확도에 대한 결과입니다.

Word Embedding 벡터의 차원의 10^n 개에 달했기 때문에, PCA, LDA로 차원축소 후, 모델학습을 진행하였지만 오히려 정확도가 평균적으로 20% 이상 감소하였습니다.

Z값을 올릴수록 정확도는 증가하였지만, 데이터를 균등하게 나누지는 못하였습니다. 따라서 데이터를 균등하게 나누면서 정확도가 비교적 정확한 0.7을 분할 기준으로 선정하였습니다.

- 1. 테스트 데이터에 대해 70%의 정확도
- 2. 한국어 문서 분류에 대한 최신 연구 동향에 따르면 70%는 높은 정확도

위의 두 가지 이유로

가설 2. 물품의 상태가 중고가에 영향을 미칠 것이다. → 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>

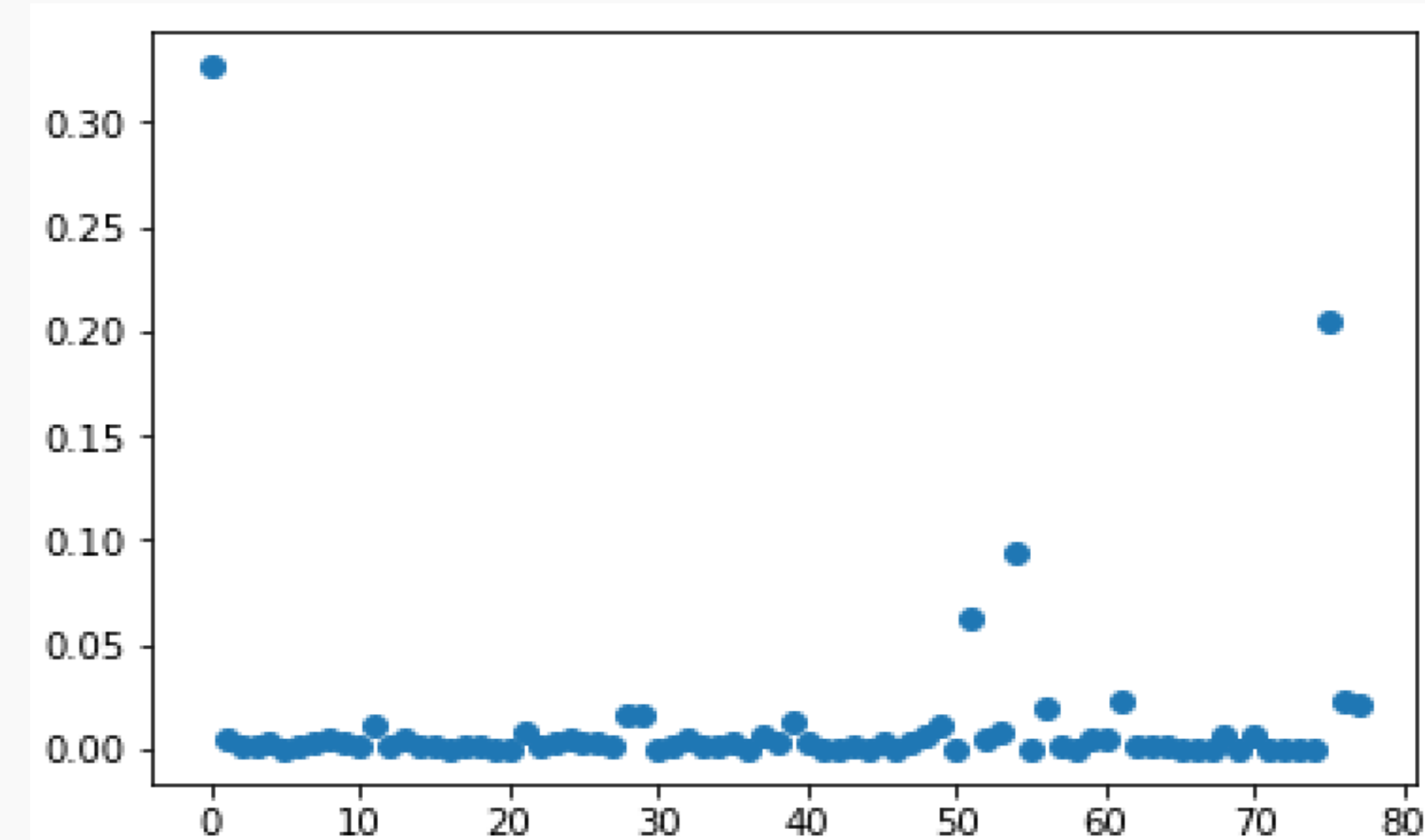
변수 중요도 평가

변수 중요도 추출을 통해 가격 예측에 필요한 최종 변수 선택

1. 변수 중요도 플롯

더미변수 추출 과정을 통해 생성된 80여 개의 변수 중,
중요도가 높은 변수는 다음과 같음

- 1.제품명 : 33%
- 2.홈결등급 (판매금액_z_rank) : 20%
- 3.상위 더미 변수 : 10%, 7%



가설 1. 물품의 상태가 중고가에 영향을 미칠 것이다 → 채택

```
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
> |
```

2. 단계적 회귀방법 사용(Stepwise)

AIC 통계량을 검출하여,
출고가 피처를 제거해야 한다는 것을 알 수 있음.

가설 3. 출고가는 중고가 형성과 연관이 있을 것이다. → 기각임을 재확인

가격 예측 모델링

앞의 결과를 토대로 최종 모델 학습에 들어가는 변수 선택
(최종 변수 3개 선정)

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

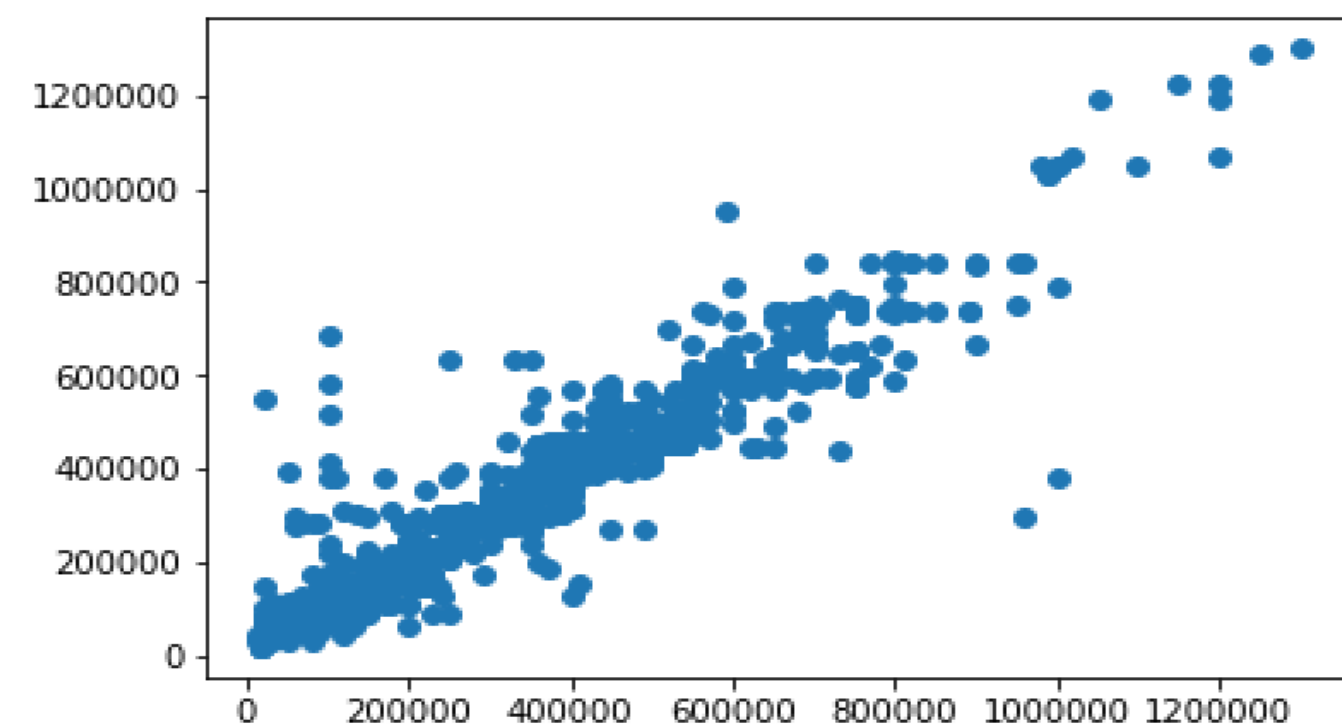
랜덤 포레스트 회귀 모델 적용 및 평가

수치형 벡터 혹은 더미변수로 변환된 변수

(**홈결등급 / 년-월 데이터 / 통신 물가지수 / 중고폰 모델명 / 출고가**)를
최종 변수로 랜덤 포레스트 회귀 모델을 학습한 결과입니다.

$R^2(\text{Training Score}) : 0.95$

$R^2(\text{Test Score}) : 0.89$



<모델에서 예측된 y값과 실제 y값을 각각의 축으로 한 플롯>

최종 모델의 검증

1. 가격 예측의 오차범위

테스트 데이터의 가격레이블과
테스트 데이터를 모델에 적용했을 때의 예측 가격 간의
차이를 오차로 하여 오차 범위를 계산하였습니다.
→ +/- 10%의 오차 범위 (아웃 라이어를 제거했을 때)

2. K-fold 교차검증

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

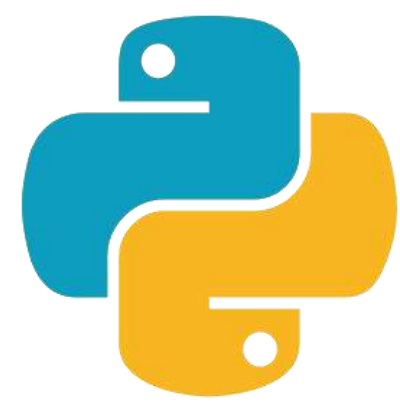
데이터 분석 과정을 거쳐 최종적으로
완성된 모델을 이용하여 챗봇의 형태로
예측 서비스를 구현하였습니다

04

챗봇구현

Application Server

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



모델 임베딩

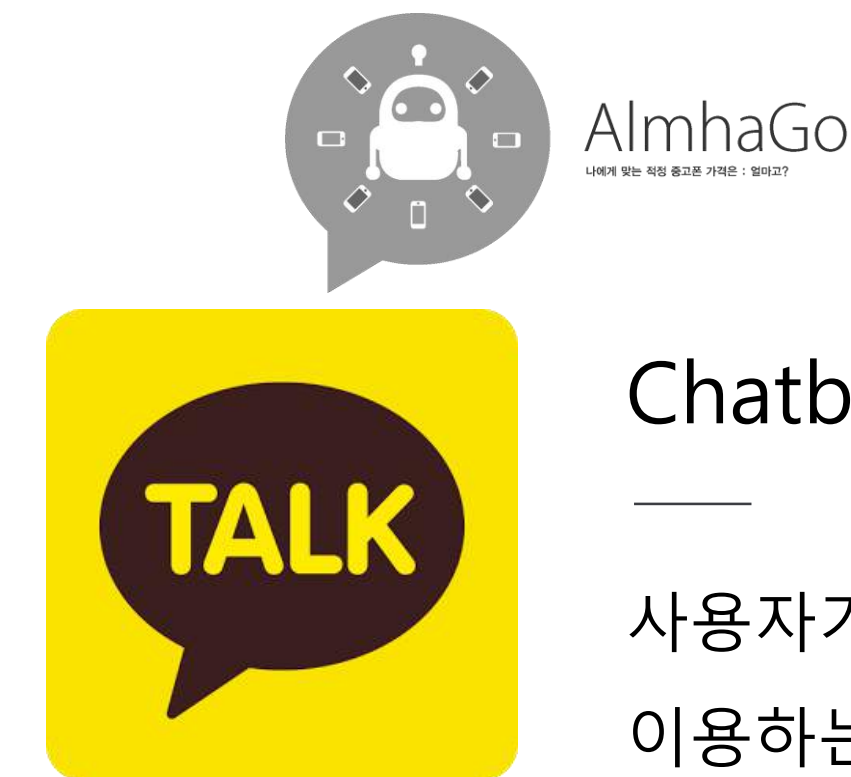


Machine Learning Model

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

챗봇 시스템 구조도

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



Chatbot Client

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

챗봇 개발 과정

서버생성부터 최종 구현까지

1. 아마존 **AWS** 아이디 생성 및 **Ubuntu** 가상서버 생성

2. **Python**언어와 **Django Framwork**를
사용하여 앱 개발 환경 구축

3. 로컬에서 생성한 앱개발 환경을 github에 업로드
Ubuntu서버에서 다운로드 환경을 설정

4. 카카오톡 플러스 친구 관리자 계정 생성 API TEST를 통하여 앱과 카카오톡 서비스 연동

6. 사용자가 버튼을 입력하여 핸드폰을 검색하고 적정 가격이 출력되는 흐름을 구현

시작하기 버튼 - 모델 버튼 - 핸드폰 기종 버튼 - 핸드폰 용량버튼 - 구매자, 판매자
버튼의 흐름으로 구현하였습니다. 구매자 판매자 버튼에서 판매할 시 모의로 판매글을
카카오톡 채팅창에 글을 올려보면 입력된 글 만으로 핸드폰등급을 정해주는 서비스를
구현 하고자 하였습니다.

8. SQLite의 데이터베이스에 메이커, 모델, 용량, 입력 및 사진 업로드

5. 카카오톡 챗봇 API 설명서를 참고하여 기본함수 생성

카카오톡 챗봇을 실행하는데에 필요한 def keyboard, def
message로 구성된 두개의 함수를 view.py에 선언하였습니다.

7. SQLite를 이용하여 데이터베이스 구축

model.py에 핸드폰의 메이커, 메이커를 상속받는 모델, 그리고 모델을 상속받는
용량으로 구성하였고 모델에는 핸드폰 사진을 업로드 할 수 있도록
설정하였습니다.

9. 이미 만들어진 불용어 사전을 기반으로 핸드폰상태에 따라
'A', 'B', 'C' 등급으로 나누는 **감성분석 알고리즘 임베딩**

Pickle 바이너리파일로 불용어 사전을 서버에 업로드하였고
pridiction_rank.py 모듈을 사용하여 감성분석을 시행하는 함수를 임베드
하였습니다.

10. 랜덤포레스트 함수를 이용하여 **가격예측함수 알고리즘 임베딩**

Pickle 바이너리 파일로 랜덤포레스트 알고리즘을 서버에 업로드 하였고
pridiction_price.py 모듈을 사용하여 랜덤포레스트 함수를 임베드
하였습니다.

11. 알파고 로고를 벤치마킹하여 **얼마고 로고 디자인 제작**

챗봇 최종구현

최종 프로젝트의 기대 효과

05

프로젝트 기대효과

기대효과

서비스 SWOT분석

Strength

1. C2C 중고폰 거래 시의 적정 가격 제안
2. 중고폰 구매시 모델별 적정가격 안내로 사기 매물 판단 기준 제시
3. 중고폰 판매시 시세 파악의 번거로움 최소화
4. 도매상에서 반영하지 않은 변수를 예측에 반영하여 예측 알고리즘 정교화

Opportunities

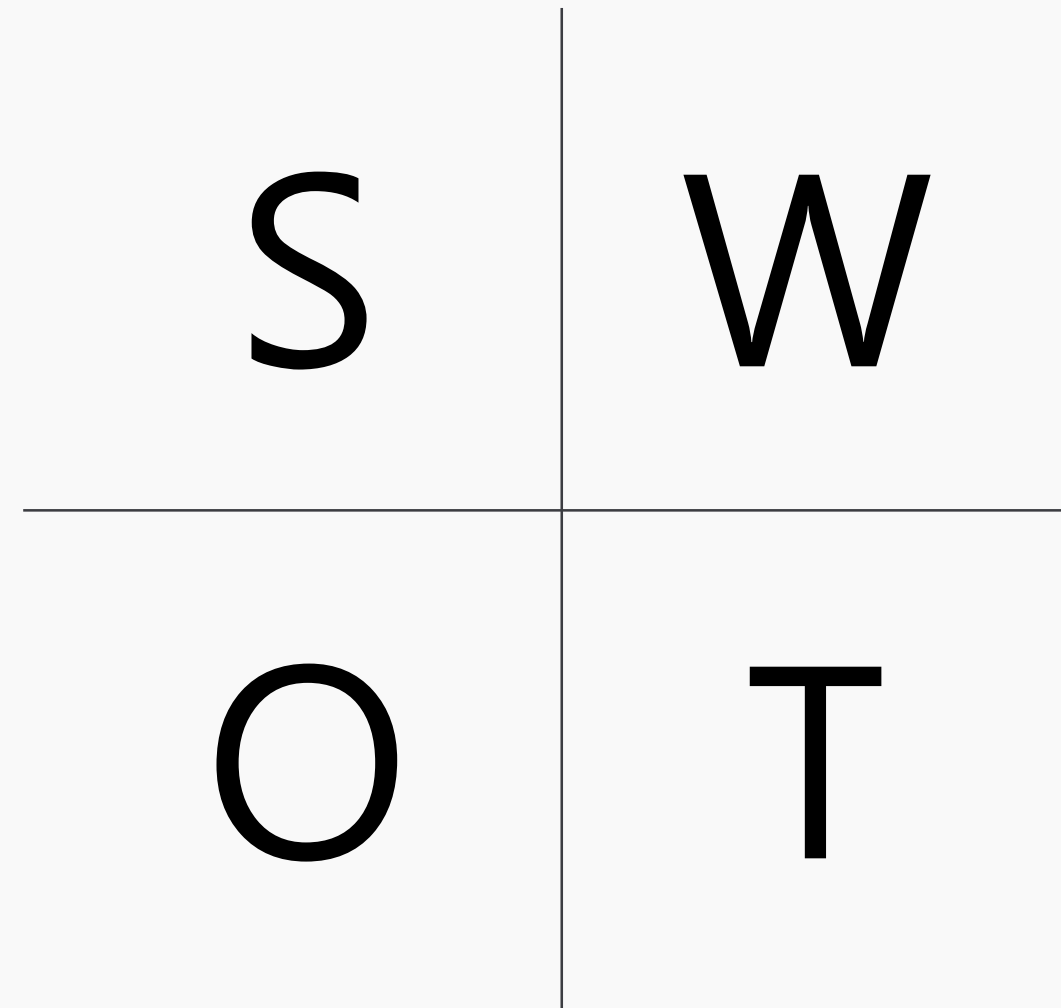
1. 휴대폰 기기 값 부담으로 인한 중고폰 관심 향상
2. 연내 중고폰 요금 할인 상승 예정이에 따른 중고폰 수요 증가
3. 중고나라 챗봇 서비스 도입으로 안전거래 실시 예정

Weaknesses

1. 판매자의 중고폰 흠결 오기입(또는 허위 기재)로 인한 예측정확도 하락 가능
2. 2015년 이전 출시된 핸드폰 정보 미반영

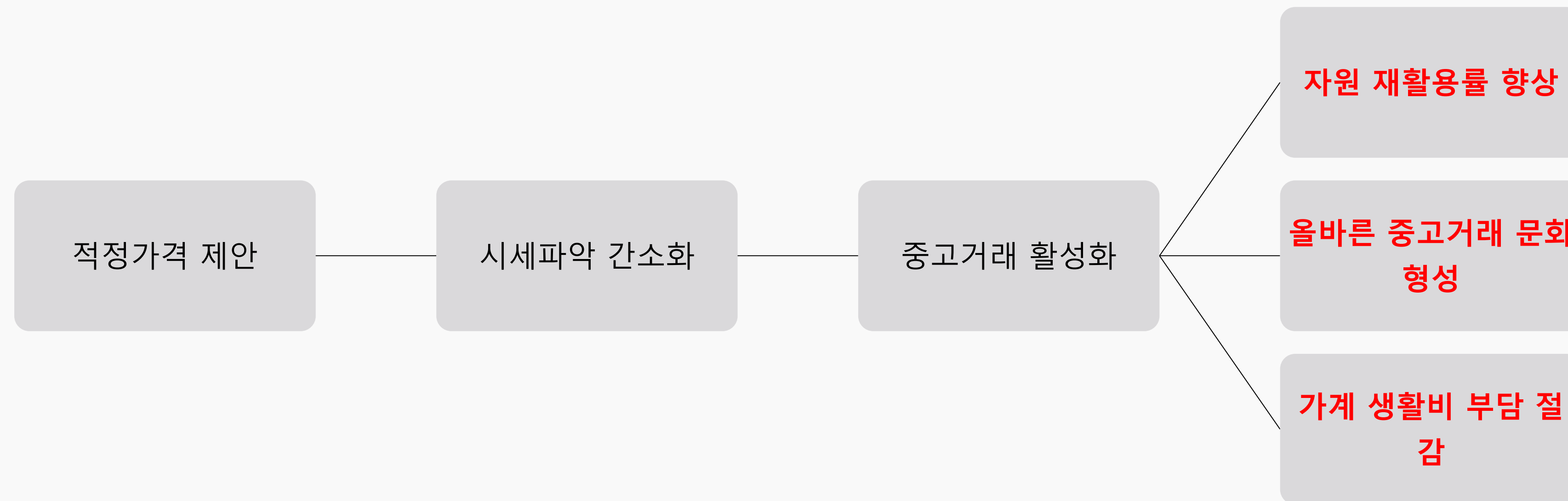
Threats

1. 첨단 기술을 반영한 신형 스마트폰의 지속적인 발매
2. 중저가 모델 출시



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/>