

# 중간고사 보고서

제주도 렌터카 리뷰

20182793, 빅데이터 경영통계학과 김민식

## 00 목차

## 01 서론

## 02 데이터 수집

## 03 전처리

## 04 단어 빈도

## 05 감정 분석

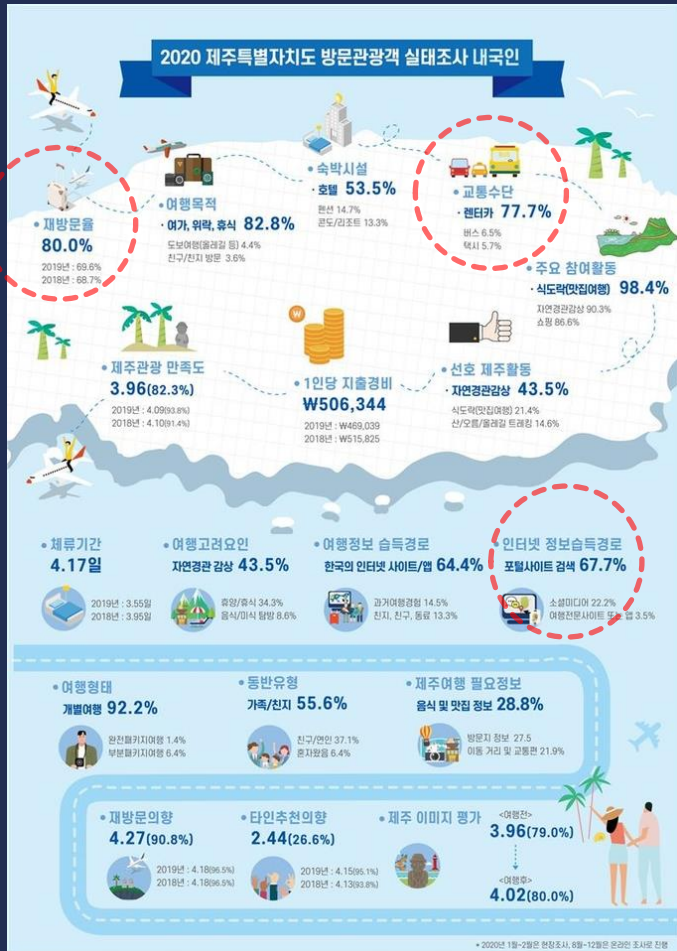
## 06 주제 분석

## 07 결론

## 08 자기 평가표



## 제주도 렌터카 리뷰 분석을 통한 발전 방향



### [현황]

코로나 19로 인해 '안전함과 품격'등의 이유로 해외 여행의 대안으로 **제주도 여행을 선택(관광객 증가)**

1. 제주도 여행 시 편리한 이동을 위해 상당수 (78%)가 렌터카 활용
2. 제주도 여행 시 여행 정보 습득 경로가 포털사이트 검색(67.7%)

∴ 제주도 렌터카 **리뷰 분석의 필요성 必**

(긍정 및 부정적 반응 분석을 통한 발전 및 개선사항)

[기대효과]: 소비자들의 반응을 통해 렌터카 회사 서비스의 발전방향을 제시, 이를 바탕으로 서비스의 질 향상 기대

# 이 요약 (진행 방식)

1

‘제주 패스 렌터카‘ 사이트에서 후기가 많은 상위 20개 회사의 리뷰데이터를 수집

2

데이터 EDA 작성 및 데이터 전처리를 실시(감성분석을 위한 Label 병행)

3

TDF, TF-IDF 단어 빈도 작성 및 시각화

4

단층 신경망 모델을 이용한 감성분석 모델 학습 및 금, 부정 단어 가중치 확인

5

LAD, LAS를 활용한 주제 분석

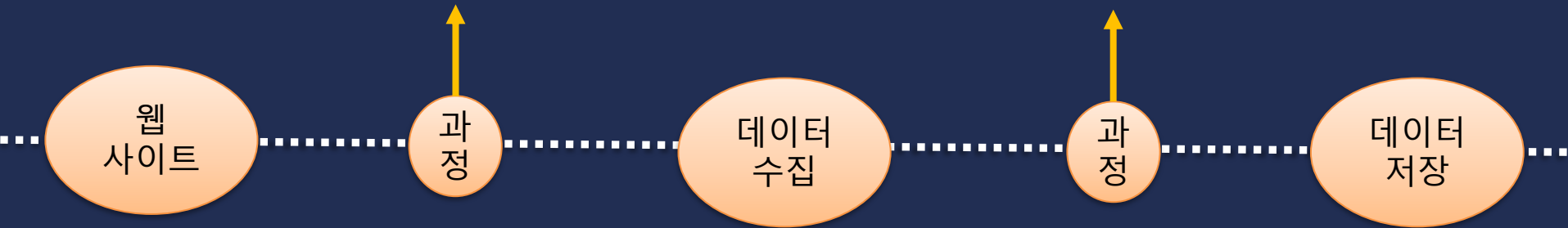
6

분석한 내용 바탕, 프로젝트 보고서 및 자기 평가 표 작성

## 02 데이터 수집

[2번 ~ 5번 과정 반복]

사이트 접속 → 후기 선택 → (자동화 위한) 페이지 정보 수집 → 회사별 리뷰 데이터 수집 → Excel 저장



데이터 수집 사이트



웹 스크롤 모듈: Selenium

수집 데이터 종류(약 3만건)

- 리뷰 점수(4개 분야), 리뷰 데이터
- 연도, 차량 종류, 회사명

추가사항

- 1) 오류 발생 시 해결 (Try, Except문)
- 2) 자동 및 함수화(For문)
- 3) 진행 상황 표시

데이터 저장: EXCEL

1. 회사별 데이터 수집  
(수집 완료 시, 자동 저장)

2. 각 회사 데이터 결합

# 03 전처리

## 맞춤법 검사

[리뷰데이터 맞춤법 정제]

웹 크롤링을 활용한 맞춤법 검사 자동화

(맞춤법 검사: 다음 “한글 맞춤법 검사기”)

## 정규 표현식 활용한 전처리

- 특수문자, 숫자, 영어 문자 제거
- 자음, 모음만 이루어진 문자 제거(예: ㅎㅎ)
- 리뷰 데이터의 길이가 10이하인 데이터 삭제
- 리뷰 데이터 中 한 글자 단어 삭제
- 양쪽 공백, 중복 공백 제거

## 형태소 분석

형태소 분석기: Kiwi

추출할 형태소: 체언(N), 용언(V)

특이점(성능 향상 진행 / But 의미가 없음)

1. 접두사(XRN), 일반 부사(MAG), 어근(XR)
2. 토큰 길이 2개 이상 추출

## 불용어 정의

TDM, 감정분석 中 단어빈도를 보고 불용어 정의

- 렌터카 회사명, 지시어(이것 등)
- 의미 없는 한 글자 단어

TDM, 감정분석 간 해당 단어를 불용어로 제거함

# 04 단어 빈도 (1) 시각화

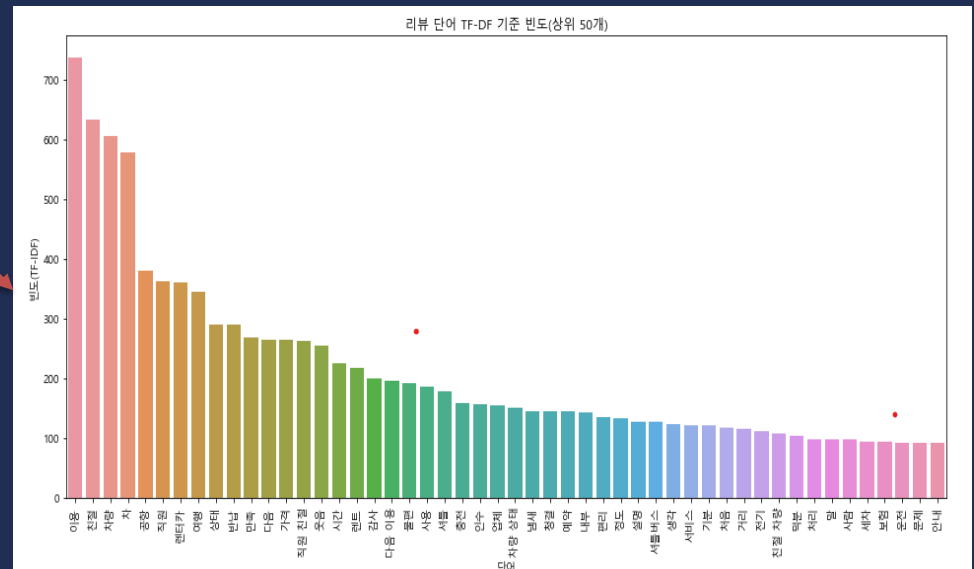
	단어	빈도(TDM)	빈도(TF-IDF)
618	이용	4660	737.567717
878	친절	4034	634.357039
796	차량	4863	606.719276
757	차	4282	579.514785
56	공항	2164	380.137905
740	직원	2390	364.152815
204	렌터카	2752	361.077893
533	여행	1872	345.224773
381	상태	1592	290.895412
284	반납	2161	290.858632
243	만족	1085	269.161536
152	다음	1316	266.340720
0	가격	1155	265.894746
747	직원 친절	1229	264.427649
589	웃음	975	254.804164
457	시간	1641	227.118676
225	렌트	1360	218.032673
20	감사	709	200.406467
158	다음 이용	719	196.764533
343	불편	872	192.290399

주요 단어 문서 정리(상위 20개)

[최소 문서: 5개 / n-gram 범위: 2단어]



시각화 (1): Word Cloud(차 모양)



시각화 (2): 상위 50 단어 막대 그래프

## 04 단어 빈도 (2)해석

### 1. 데이터: “렌터카 회사” 정보 사실 확인

- 단어: 차·차량·렌터카·렌트·반납·대여 등

[해석] 렌터카 회사의 리뷰 데이터이기 때문에 단어의 빈도가 많음 (렌터카 회사 리뷰 데이터 인식 가능)

### 2. 렌터카 대여 목적과 위치 정보

- 단어: 여행·공항·셔틀(버스)·거리·시간 등

[해석] 렌터카 활용하는 목적이 ‘여행’이며, 렌터카 업체 장소가 (제주)공항과는 거리가 있어, 셔틀버스를 타야 한다고 추측

(셔틀버스 시간과 서비스 자체의 리뷰 多)

→ 셔틀버스 서비스 관련 개선사항 필요 판단

### 3. 고객의 만족도 및 만족 대상 파악

- 단어: 상태·가격·응대·관리·보험·말·친절 등

[해석]

고객이 서비스를 받을 때, 만족·불만족을 느끼는 주요 대상들이다. (예: ‘차량 상태가 좋지 않다.’)

렌터카 회사는 해당 분야(가격, 관리 등)에 대한 서비스의 질 향상을 노력해야 한다.

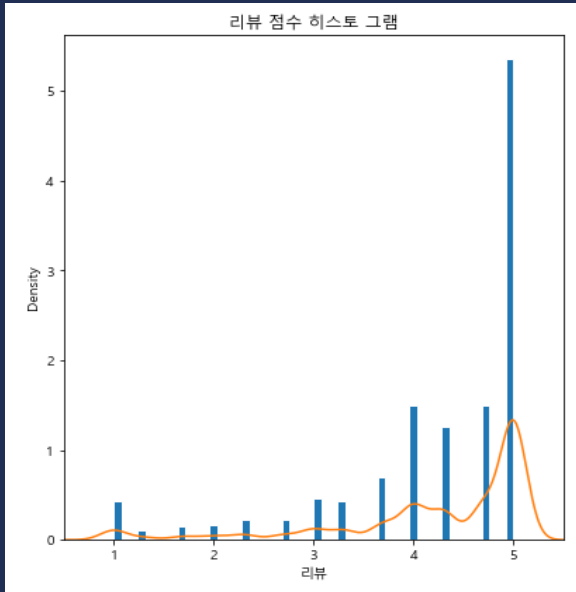
- 단어: 웃음·만족·편리·불편·최악 등등

[해석]

고객들이 서비스 간 느낀 감정. 서비스 대상과 연관해, 만족을 느끼는 서비스 발전, 불만족을 느끼는 서비스에 대해서는 개선



# 05 감성 분석 (1) Label



## [긍정 및 부정 Label 설정]

- 4점 이상 리뷰: 86.9% vs 2점 이하 리뷰: 10.4%
- 긍·부정 Label을 붙이기 위한 **세밀한 조정의 필요성 절감**

## 1. 리뷰 점수 3등분으로 나눔[긍정, 부정만 활용]

(부정: 2.7이하, 중도: 2.7 ~ 4.0 긍정: 4.3 이상)

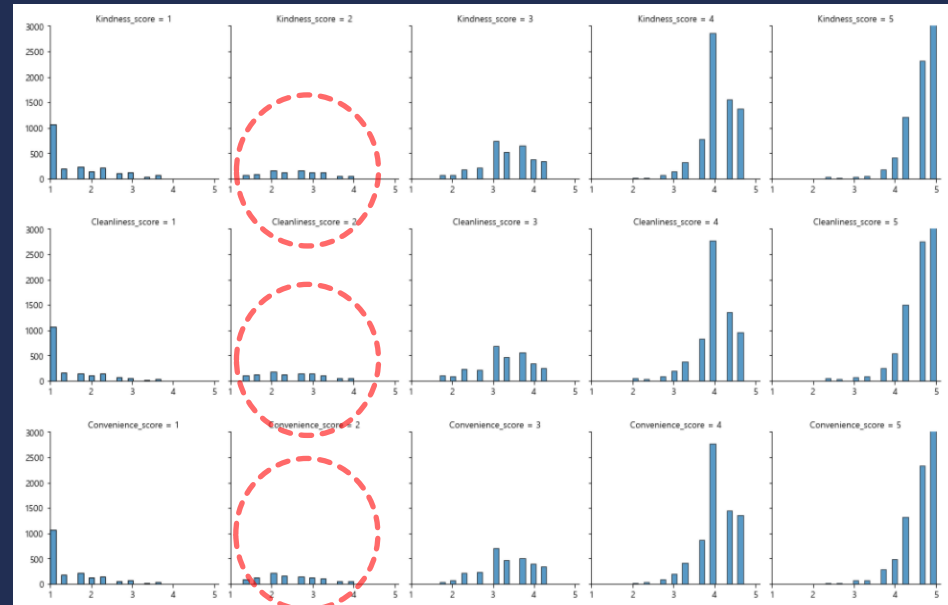
## 2. 세부 리뷰 점수 활용

- 청결도, 편의성, 친절도와 리뷰 점수 분포 비교
- 중도 혹은 긍정 中 세부 점수 1·2점 존재

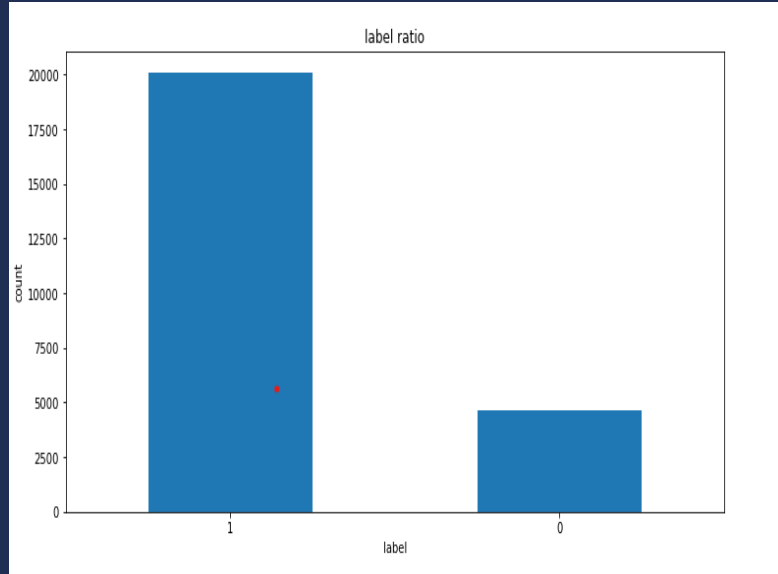
∴ 긍정: 리뷰 점수 4.3 이상

부정: 리뷰 점수 2.7 이하

또는 세부 리뷰 점수 2점 이하



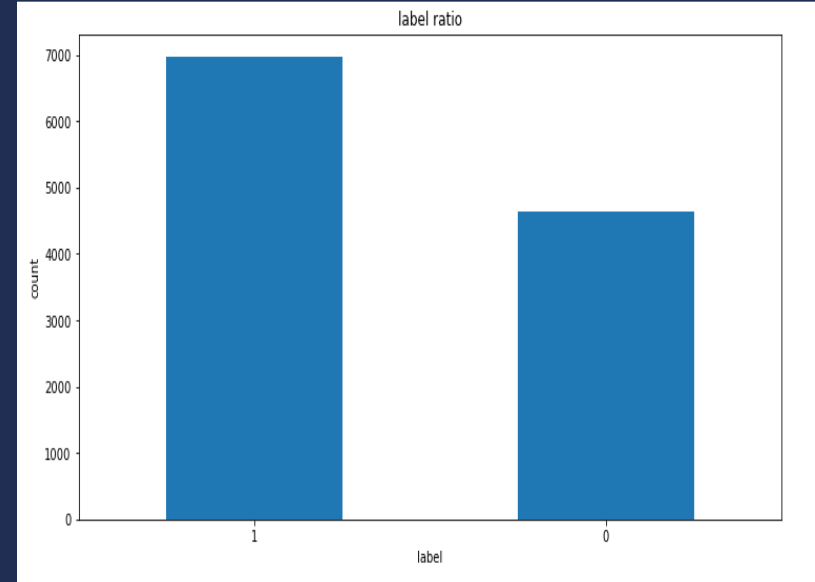
## 05 감성 분석 (2)데이터 불균형



### [데이터 불균형]

- 긍정 Label(81%) vs 부정 Label(19%)
- 검증 데이터 정확성(88%)
- Label 분포의 **편향**이라 판단

∴ 불균형 처리를 위한 **undersampling** 진행



### [Under Sampling]

- 전체 데이터 건수 1만 건 유지
- (긍정데이터 35% 랜덤 추출)

긍정 Label(60%) vs 부정 Label(40%)

- 검증 데이터 정확성(81%)
- 레이블 분포 불균형 문제 해소

## 05 감성 분석 (3) 모델 학습

데이터

$X$  = TF-IDF (TDF 비해 성능 좋음)

$Y$  = 긍정 · 부정 Label 값

Test 비율 = 20%

모델

- 단층 신경망(활성화 함수: 시그모이드 / 옵티마이저: Adam)
- Epochs = 100회 / batch\_size = 50회

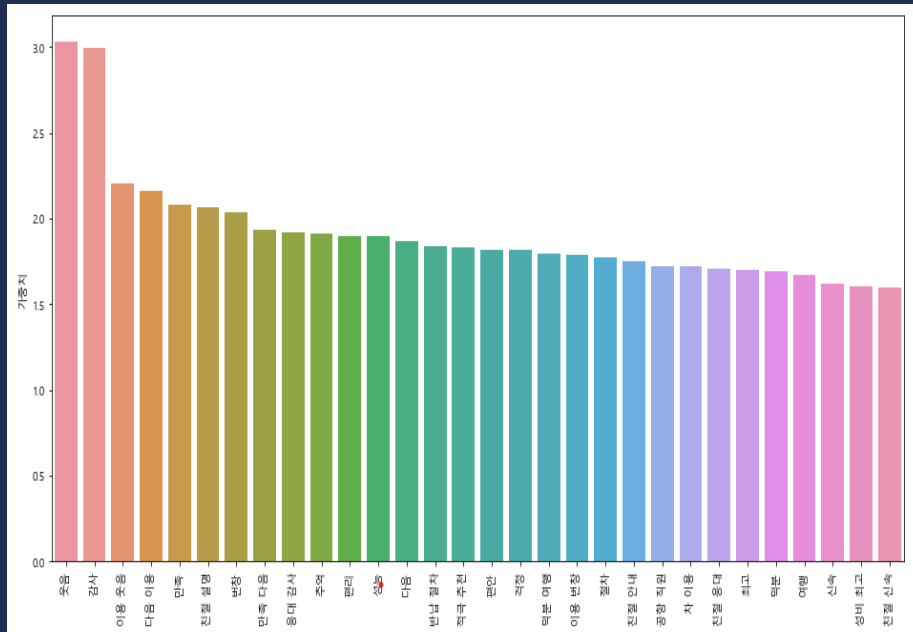
학습

- K-fold를 활용한 교차 검증( $n = 3$ )
- Early-Stopping: 검증 데이터 손실 값이 2회 이상 하락하지 않을 시 조기 종료
- Mode-Checkpoint: 검증 데이터 기준 정확도가 높은 모델을 따로 저장

성능

Test 결과: 약 0.821

## 05 감성 분석 (4) 긍정 단어와 해석



- 모델 가중치 中 양수: 긍정 단어
- 토큰(단어)를 확인해보면
- 긍정적인 단어(웃음, 감사, 번창 만족 등)들이 등장하는 것을 확인할 수 있음

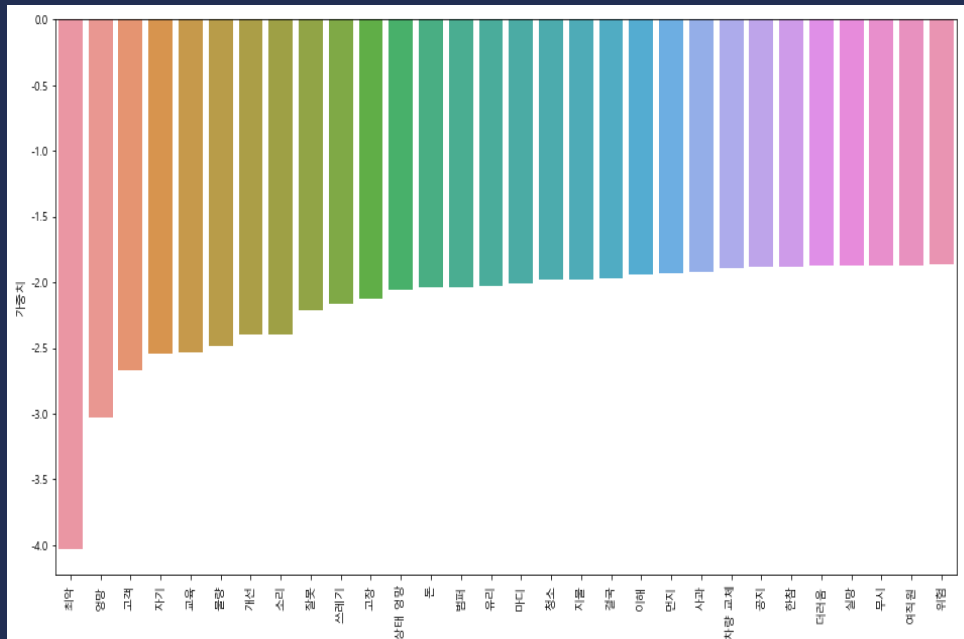
## [해석]

리뷰 데이터 中 긍정 단어 목록을 통해 고객들이 만족을 느끼는 부분을 파악할 수 있음

1. '성능'적인 측면: 고객들이 렌트 차량의 성능은 가성비에 만족함
2. '절차'적인 측면: 렌트 반납&예약 과정이 편리하며, 신속하게 처리됨
3. '서비스'적인 측면: 고객들을 '응대·설명·안내'할 때, 친절하고 일(기타)을 신속하게 처리함

→ 렌터카 서비스를 이용하는 것에 대해서 ‘웃음, 감사, 편리, 편안, 만족’의 감정을 느끼고, 이런 긍정적인 감정은 제주도 여행 간 **최고의 추억을 쌓는 것**의 역할을 한다. (재 이용 및 추천 향상)

## 05 감성 분석 (5) 부정 단어와 해석



- 모델 가중치 中 음수: 부정 단어

→ 토큰(단어)를 확인해보면

부정적인 단어(최악, 망망, 실망 등)들이 등장하는 것을 확인할 수 있음

[해석]

리뷰 데이터 中 부정 단어 목록을 통해 고객들이 불만을 느끼는 부분을 파악할 수 있음

1. '차량 상태' 측면: 고객들이 렌트 차량의 상태가 엉망이거나 불량일 경우

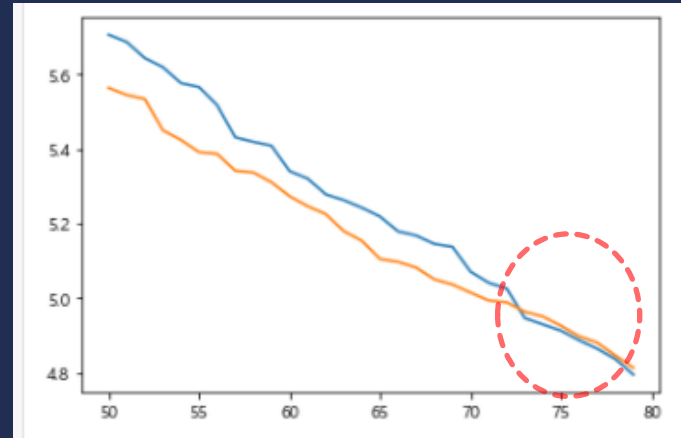
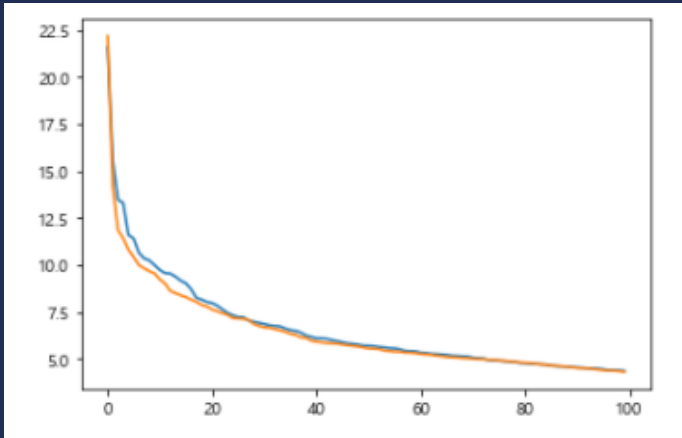
[1] 차량 내부의 청소가 되지 않아서 먼지 혹은 쓰레기 등이 있어서 더러운 상태

[2] 차량의 부품(범퍼 혹은 유리 등) 불량이거나 고장이 나서 소리가 남

2. '서비스'적인 측면: (일부) 고객을 무시하거나, 공지가 한참 뒤에 하는 경우

→ 렌터카 측에서는 차량 상태를 개선(차량 교체 등) 및 직원들의 서비스 교육을 통한 서비스 질 향상

# 06 주제 분석 (1) LSA 차원 결정 및 시각화



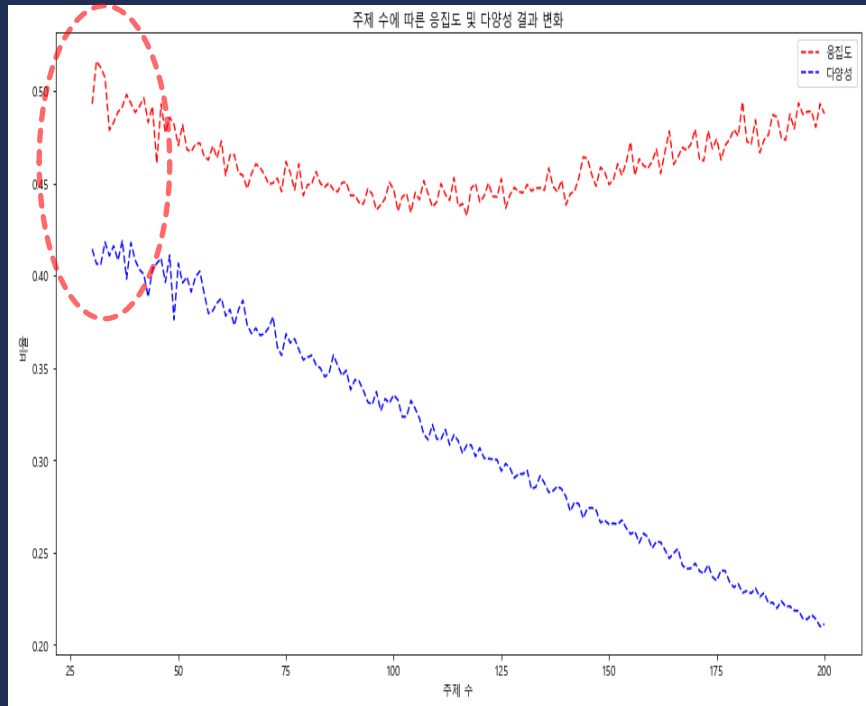
- 파란색 데이터: 기존 단어의 분포 데이터 vs 주황색 데이터: Random 뽑은 분포 데이터
- 주황 및 파랑 색 데이터가 70번 전후로 만남 (70차원 이후의 설명력의 큰 차이가 없다.)
- ∴ LSA와 NMF 분석 간 차원의 수: 72차원

이후

```
Topic 1: [(이름, 0.47997), (성명, 0.34963), (직업, 0.29819), (성, 0.32457), (직업, 0.21977)]
Topic 2: [(이름, 0.70023), (직업, 0.18472), (성명, 0.16008), (이름, 0.04732), (직업, 0.04348)]
Topic 3: [(직업, 0.32997), (성명, 0.32997), (직업, 0.32997), (성명, 0.32997), (직업, 0.32997)]
Topic 4: [(직업, 0.70700), (성명, 0.1134), (직업, 0.09133), (성명, 0.00919), (직업, 0.00920)]
Topic 5: [(직업, 0.74320), (성명, 0.26212), (직업, 0.22333), (성명, 0.22077), (직업, 0.17090)]
Topic 6: [(직업, 0.66400), (성명, 0.23644), (직업, 0.10771), (성명, 0.10541), (직업, 0.10541)]
Topic 7: [(직업, 0.7980), (성명, 0.37847), (직업, 0.37127), (성명, 0.34422), (직업, 0.1235)]
Topic 8: [(직업, 0.60100), (성명, 0.27163), (직업, 0.25720), (성명, 0.10879), (직업, 0.10441)]
Topic 9: [(직업, 0.30790), (성명, 0.30143), (직업, 0.21820), (성명, 0.15660), (직업, 0.17320)]
Topic 10: [(직업, 0.73540), (성명, 0.23807), (직업, 0.25720), (성명, 0.14104), (직업, 0.10821)]
Topic 11: [(직업, 0.45590), (성명, 0.34716), (직업, 0.22644), (성명, 0.19180), (직업, 0.11716)]
Topic 12: [(직업, 0.61610), (성명, 0.24982), (직업, 0.23848), (성명, 0.18120), (직업, 0.10277)]
Topic 13: [(직업, 0.43877), (성명, 0.34820), (직업, 0.25365), (성명, 0.15980), (직업, 0.25477)]
Topic 14: [(직업, 0.59154), (성명, 0.36910), (성명, 0.26201), (직업, 0.20241), (성명, 0.10380)]
Topic 15: [(성명, 0.30095), (직업, 0.38940), (직업, 0.17776), (성명, 0.16000), (직업, 0.14392)]
Topic 16: [(직업, 0.46120), (성명, 0.34540), (직업, 0.24111), (성명, 0.10260), (성명, 0.10260)]
Topic 17: [(직업, 0.70023), (성명, 0.18939), (성명, 0.17289), (성명, 0.13020), (성명, 0.12161)]
Topic 18: [(직업, 0.43820), (성명, 0.30440), (성명, 0.23807), (성명, 0.17310), (성명, 0.10040)]
Topic 19: [(직업, 0.49801), (성명, 0.40300), (성명, 0.21803), (성명, 0.2116), (성명, 0.20120)]
Topic 20: [(직업, 0.76130), (성명, 0.37130), (성명, 0.23807), (성명, 0.18120), (성명, 0.10277)]
Topic 21: [(직업, 0.54116), (성명, 0.29591), (성명, 0.23203), (성명, 0.17980), (성명, 0.10000)]
Topic 22: [(직업, 0.50190), (성명, 0.41390), (성명, 0.23807), (성명, 0.18120), (성명, 0.10277)]
Topic 23: [(직업, 0.40801), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 24: [(직업, 0.50190), (성명, 0.23807), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 25: [(직업, 0.75130), (성명, 0.41390), (성명, 0.23807), (성명, 0.18120), (성명, 0.10277)]
Topic 26: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 27: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 28: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 29: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 30: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 31: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 32: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 33: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 34: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 35: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 36: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 37: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 38: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 39: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 40: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 41: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 42: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 43: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 44: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 45: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 46: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 47: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 48: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
Topic 49: [(직업, 0.50190), (성명, 0.40251), (성명, 0.18120), (성명, 0.10277), (성명, 0.10277)]
```

1. LSA, NMF 시각화를 통한 주요 키워드 분석  
(함수를 통해 LSA 및 NMF 결과 한번에 출력)
2. Topic별 주요 단어 파악

## 06 주제 분석 (2) LDA 차원 결정 및 성능 향상



- 단어 : 최소 5개 문서 or 90% 미만 문서에서 등장
  - 불용어와 Kiwi를 활용해 토큰화 후 Corpus 형식 변환
  - Test 비율: 20%
  - 응집도·다양성 결과 변화를 바탕으로 주제 수 선정
- 주제 수: 31개(응집도: 0.516, 다양도: 0.406)

∴ LDA 분석 간 차원의 수: 31차원



위의 조건으로 모형 설정 후 Train-corpus 업데이트

최종 성능(혼란도): -7.8119 → -7.6256 (성능 향상)

## 06 주제 분석 (3) 해석

- (긍정) 주제 분석 주요 Key-word: 절차, 응대, 신속, 설명, 성능, 최고
- 선정 이유: 감정 및 긍정 표현(웃음, 감사 등) 제외하고 회사의 발전 방향을 제시할 만한 키워드

Key-word	관련 단어	Key-word	관련 단어
신속	대여·처리·반납·정확 (+)	설명	직원·차량·계약서·친절 (+)
절차	차량·인수·반납·과정 (+)	응대	직원·친절·태도·말투 (+)
최고	가성비·가격 대비·설명 (+)	성능	최고·(가)성비·연비·접근 (+)

### [해석]

렌터카 서비스를 받은 고객들의 리뷰의 단어들의 연관성(감성 분석 간 긍정 단어들 중심)

1. 고객들은 “예약 ~ 인수 및 반납” **절차 과정이 신속한 것을 선호**, 많은 반응(리뷰)를 보임
  2. 위 절차 속에서 필요한 것들(과정 설명, 계약서)에 대해 **직원들의 말투와 태도가 친절**할 경우호감을 보임(연관성이 높음)
  3. **가격 및 차량 성능(연비) 등 가성비 요소**에 장점에 대해 긍정적 표현과 연관성이 높음
- 렌터카 절차 과정을 manual화를 통한 신속한 대처, 직원들에 대한 지속적 교육(친절 교육)



## 06 주제 분석 (4) 해석

- (부정) 주제 분석 주요 Key-word: 불량, 돈, 여직원, 사과, 냄새, 엉망, 최악
- 선정 이유: 감정 및 긍정 표현(웃음, 감사 등) 제외하고 회사의 발전 반향을 제시할 만한 키워드

Key-word	관련 단어	Key-word	관련 단어
불량	세차·유리·실내·내부 청소(-)	여직원& 사과	말·전화 (-)
돈	기름(주유)·부족·주유 추가(-)	냄새	최악 엉망 단어 연관성 ↑
청소	먼지·유리창·얼룩·물때		담배 냄새· 차 냄새(-)

[해석]

렌터카 서비스를 받은 고객들의 리뷰의 단어들의 연관성(감성 분석 간 부정 단어들 중심)

1. 자동차실내가 먼지, 얼룩 등의 청소가 되지 않고 냄새(담배 등)가 날 때 부정적인 반응이 많음
2. 처음 이용하는 고객들의 경우 렌트 차량의 기름이 부족할 경우 대처 必
3. 부정적 단어 中 ‘여직원’ 단어가 말, 전화와 같은 단어 연관성이 높음을 통해서, 전화 응대 간의 불편함(말싸움 혹은 말 태도 등)을 느꼈을 것이라 추론됨

→ 자동차 렌트 과정에서 차량 점검(주유 여부·냄새·청소 상태) 매뉴얼 확립 및 주기적 점검

## 07 결론

### [제언 및 실행 방향성]

‘제주도 렌터카 리뷰’ 데이터의 감정 및 주제분석을 통한 발전방향:

- 온·오프라인 고객 응대 교육 및 프로그램 제정

(목적: 오프라인 유지 및 전화 서비스 질 향상)

- 차량 점검 매뉴얼 확립(인수 / 인계과정에서 최소 1회 점검 및 주기적 차량 점검

→ 주요 점검 항목: 차량 내부 상태(먼지, 얼룩) / 차량 상태(유리, 타이어 주유 상황) / 방향제 설치

- 렌터카 대여 과정의 매뉴얼 확립 및 주기적 점검(발전)

(목적: 렌터카 서비스에 대한 신속한 처리 / 예상치 못한 상황에 유동적 대처)

- (온·오프라인) 주기적인 점검 및 발전을 위한 고객 반응 정보 수집 및 분석 → 점검 및 발전방향 수립

### [결론 및 생각]

제주도 여행 시 가장 먼저 경험하는 서비스가 렌터카 사업이다. 그렇기에 제주도 관광산업 발전과의 지속적 선순환을 위해 고객 리뷰를 바탕으로 한 렌터카 사업의 지속적 발전방향 논의 및 서비스 질 향상은 렌터카 사업 뿐만 아니라, 제주도 관광산업의 핵심이라 생각한다.



감사합니다.

Thank you

20182793, 빅데이터 경영통계학과 김민식