

~~DOLORE~~  
*Lorem*



# KEYNOTE THEME

*Mirum est notare quam littera gothica*

---



## Google Store

리뷰, 게시날짜, 별점을 중심  
으로 정보수집

약 1800개의 데이터 수집



## INSTAGRAM

어플과 관련된 개인적 경험정  
보를 수집

약 1200개의 데이터 수집



## 따릉이 시민의견수렴

구글스토어의 적을 리뷰수를  
보안 하기 위해 진행함

약 8000천개의 데이터 수집



## 텍스트 마이닝 과정

크롤링

**Selenium,  
Beutifulsoup4**

REQUEST 응답 여부에 따라 Selenium과 Beutifulsoup4 라이브러리를 선택적으로 사용하여 웹페이지 정보를 수집함.

텍스트 분석

**Customized Konlpy**

형태소 분석이 되지 않는 단어장에 추가, 의미 없는 형태소 및 단어 조각 제거 후 품사별 빈도수를 추출함.

시각화

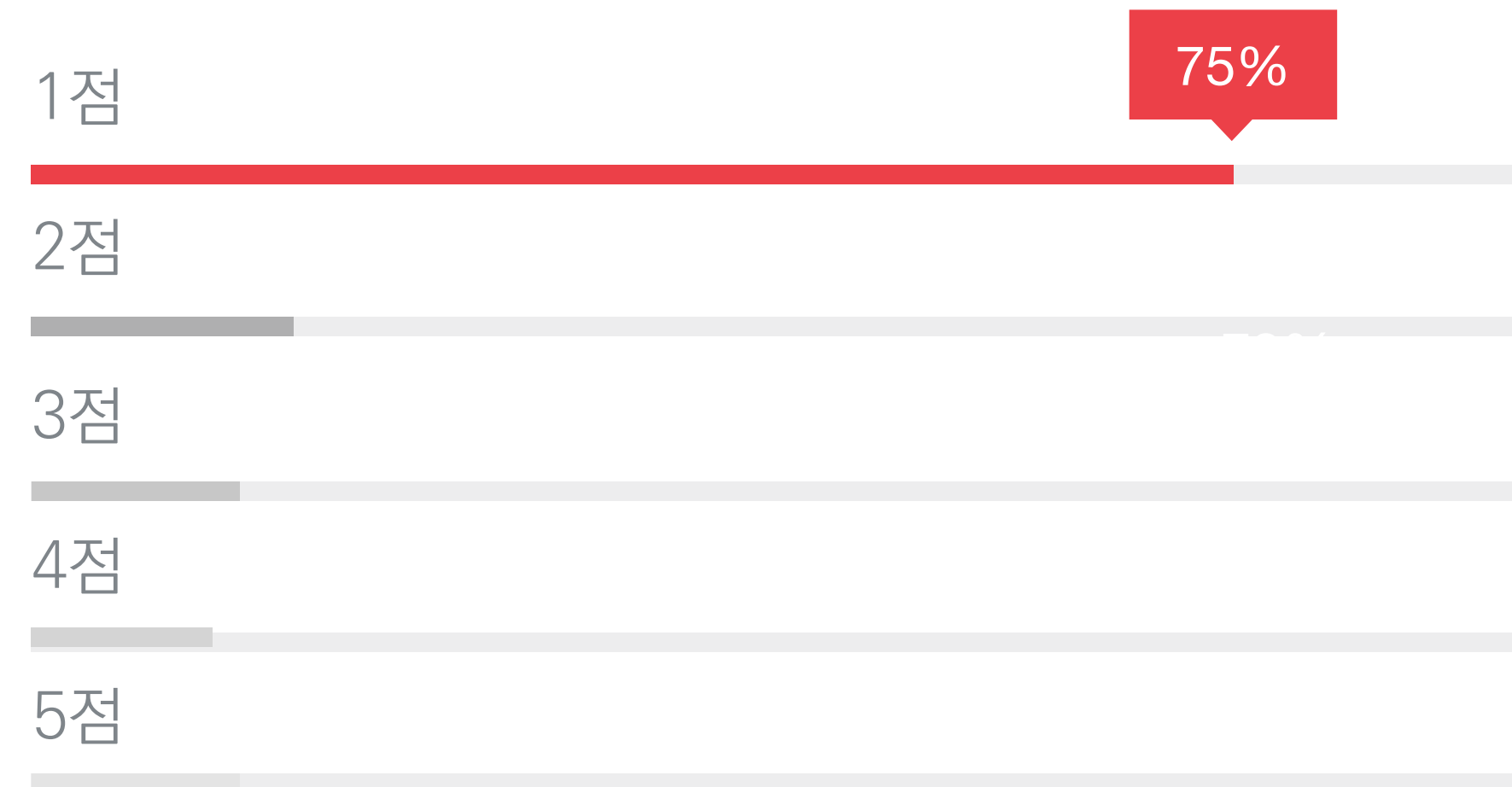
**DataFrame, Exel,  
WordCloud**

수집된 데이터를 시각화함



## 데이터 마이닝

Python Library(Selenium, BeautifulSoup4)



2015년 9월 부터 2019년 7월 까지 'Google Playstore'에 등록된 '서울 자전거 따릉이' 리뷰를 파이썬을 이용하여 크롤링함. 관리자 답글을 제외한 1,847개의 리뷰, 평점, 이용자명, url 항목을 수집함.

데이터 수가 시민수렴게시판에 비해 비교적 적었으나, 수치화된 어플의 평가와 어플에 관련된 사용자 경험을 얻을 수 있었음.

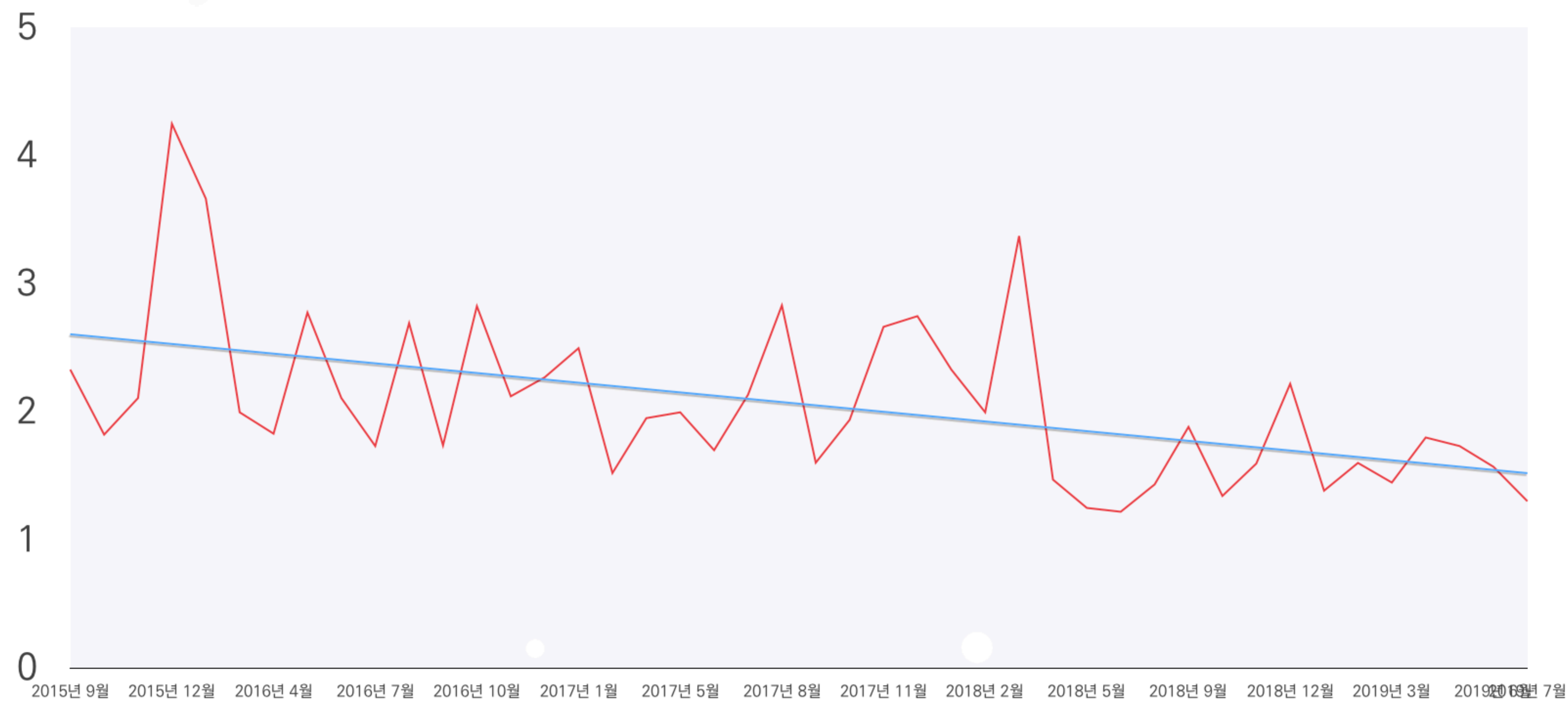


Google Playstore

총 리뷰수: 2,389개

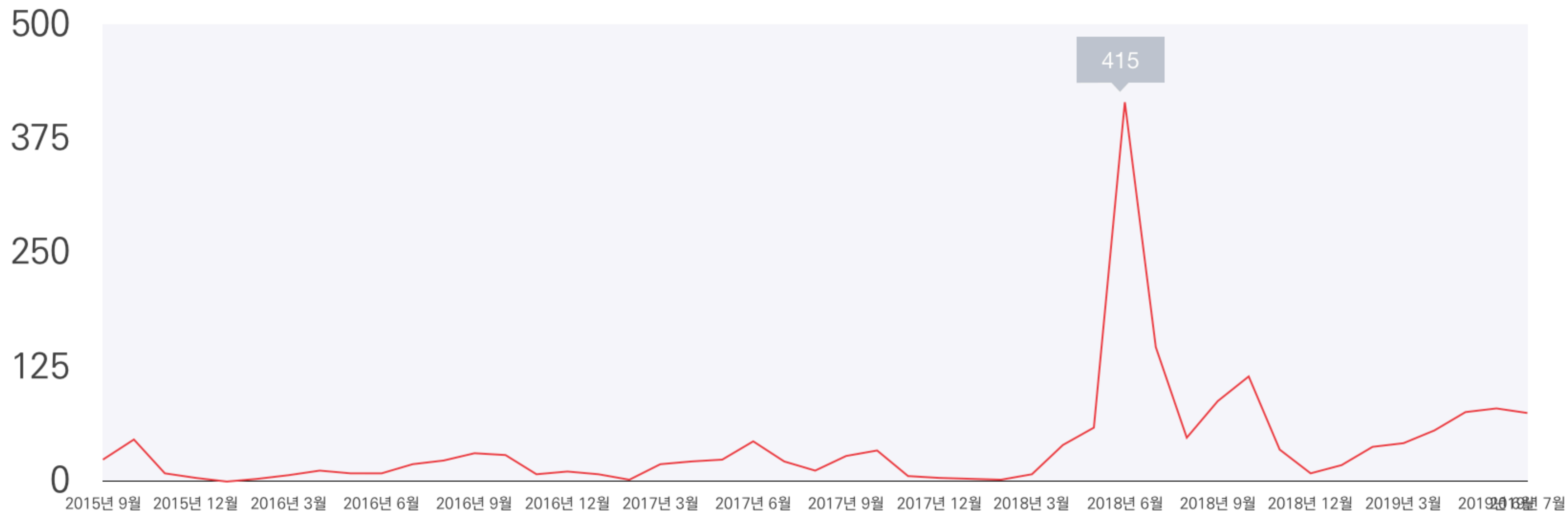
평균 별점: 1.6점

# GOOGLE STORE REVIEW SCORE



# GOOGLE PLAYSTORE REVIEW ANALYSIS

월별 리뷰수



Highly increasing the number of reviews

2015

2016

2017

2018

2019

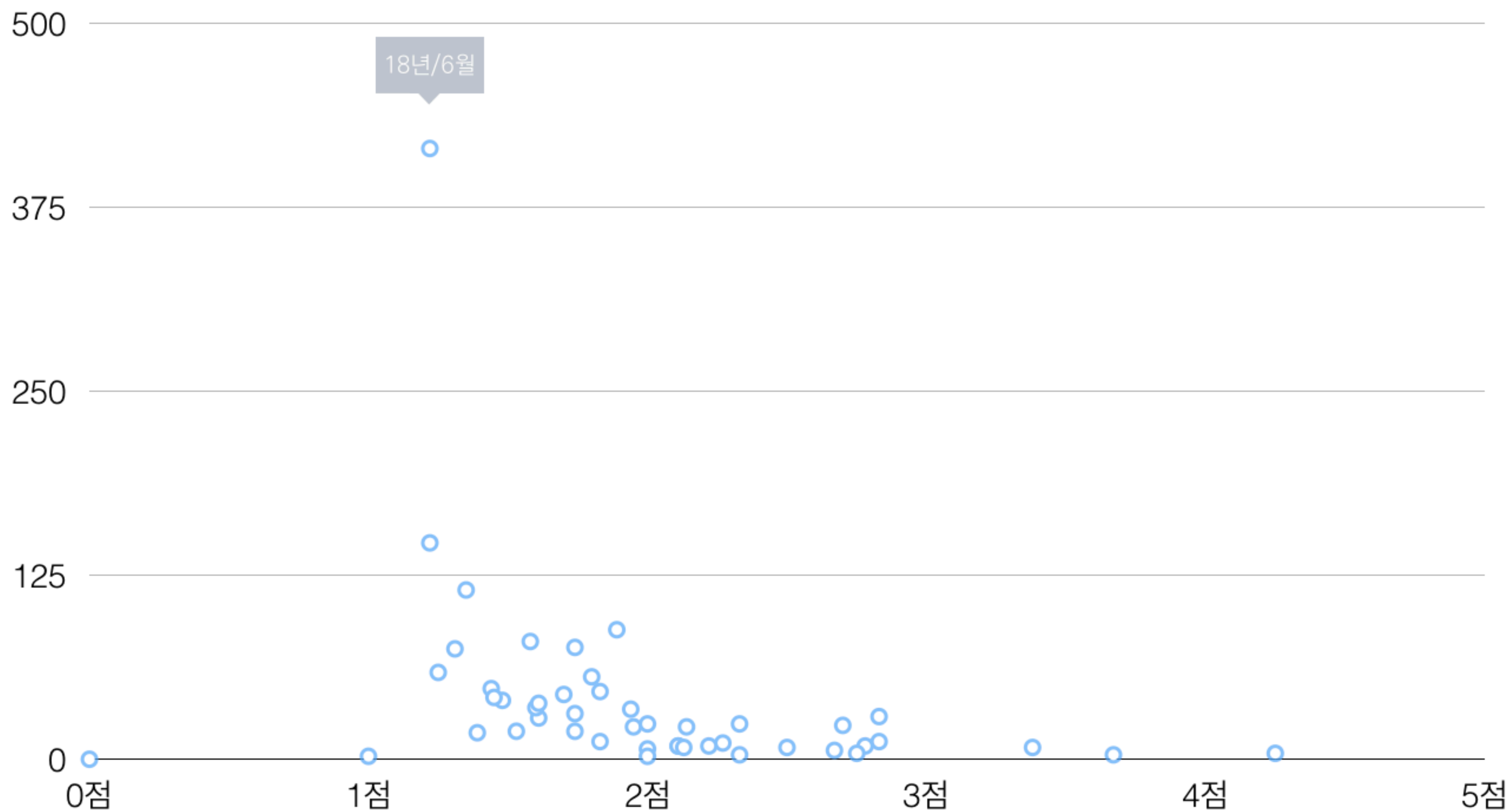
# GOOGLE PLAYSTORE REVIEW ANALYSIS

월별 리뷰수



# GOOGLE PLAYSTORE REVIEW ANALYSIS

리뷰 수 - 평점(평균) Scatter (단위: 월)







## 형용사 추출

상위 30개 항목

[('안되고', 117), ('입니다', 94), ('아니', 88), ('어떻게', 85), ('안됨', 61), ('좋아요', 48), ('좋은', 46), ('이런', 45), ('좋은데', 44), ('있는', 42), ('없고', 36), ('안되네요', 36), ('같은', 35), ('없는', 32), ('안되서', 32), ('없네요', 32), ('많고', 32), ('같습니다', 31), ('없음', 29), ('느리고', 29), ('같아요', 25), ('불편합니다', 24), ('새로', 23), ('있습니다', 23), ('불편하고', 22), ('아니고', 22), ('있는데', 22), ('있어요', 20), ('느려', 20), ('안되요', 20)]

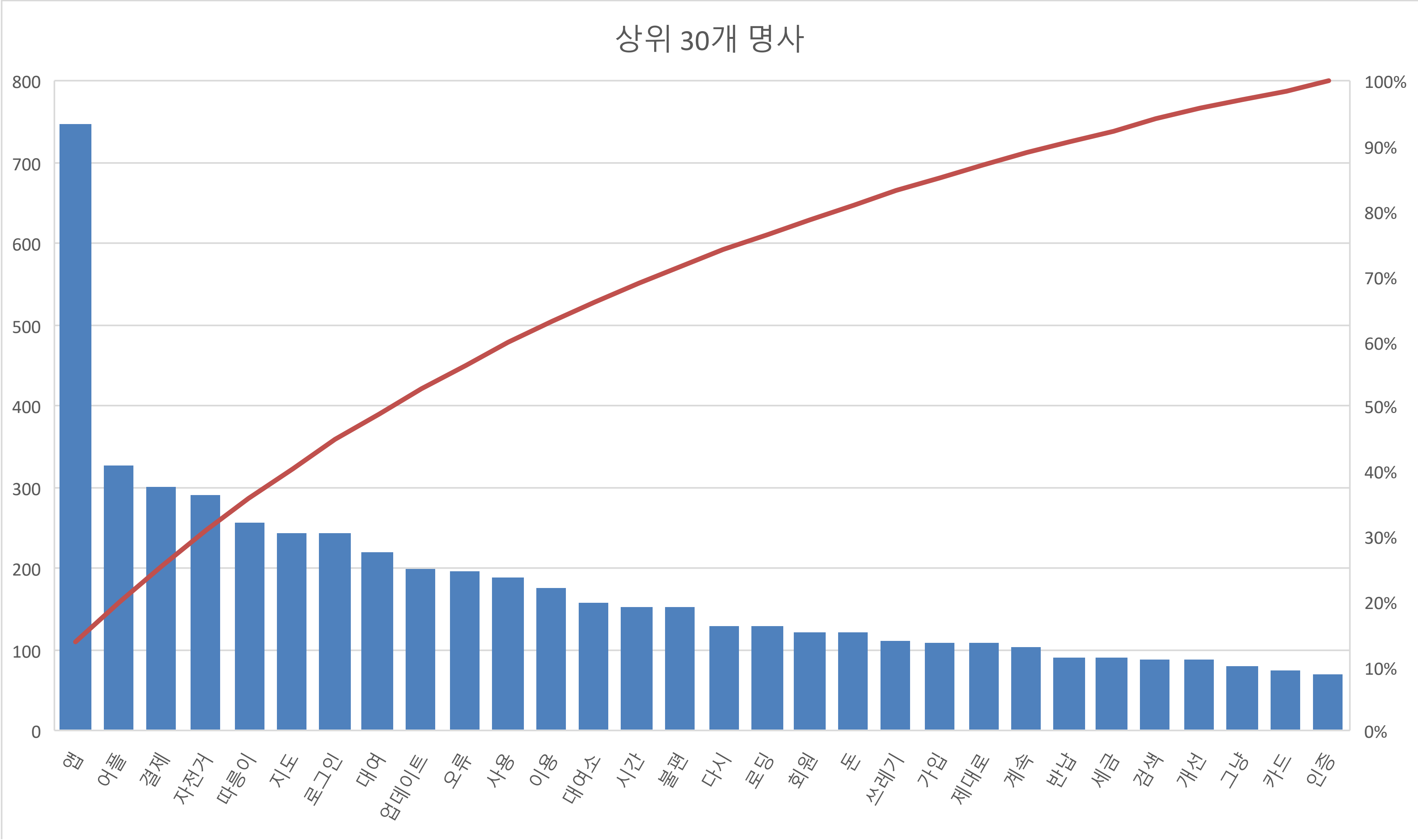


## 명사 추출

상위 70개 항목

[('앱', 749), ('어플', 329), ('결제', 301), ('자전거', 293), ('따릉이', 258), ('지도', 246), ('로그인', 245), ('대여', 221), ('왜', 213), ('업데이트', 201), ('오류', 198), ('진짜', 191), ('사용', 190), ('이용', 177), ('최악', 163), ('대여소', 159), ('시간', 154), ('불편', 153), ('다시', 131), ('로딩', 131), ('회원', 123), ('돈', 122), ('쓰레기', 113), ('제대로', 109), ('가입', 109), ('계속', 105), ('정말', 96), ('반납', 92), ('세금', 92), ('검색', 90), ('개선', 89), ('그냥', 80), ('카드', 75), ('인증', 71), ('사람', 71), ('기능', 70), ('아이디', 70), ('제발', 70), ('시스템', 69), ('위치', 68), ('화면', 68), ('번호', 66), ('개발자', 64), ('실행', 60), ('비밀번호', 58), ('생각', 57), ('서울시', 57), ('처음', 56), ('자체', 55), ('짜증', 54), ('타고', 54), ('개발', 53), ('이용권', 51), ('말', 51), ('공무원', 50), ('한번', 50), ('버전', 50), ('문제', 49), ('게', 49), ('무슨', 49), ('자동로그인', 48), ('전화', 48), ('찾기', 48), ('하니', 47), ('임', 46), ('비번', 46), ('지금', 46), ('서비스', 45), ('관리', 45)]

구글 플레이스토어 PARETO





# 데이터 마이닝

Python Library(Beutifulsoup4)



## 시민의견수렴 게시판

총 게시글 수: 8,618개

URL: <https://www.bikeseoul.com:446/customer/opinionBoard/opinionBoardList.do>

2016.11.13 부터 2019.08.03 까지 '시민의견수렴' 게시판에  
게시된 모든 게시글을 파이썬 BeautifulSoup4 라이브러리를 사  
용하여 크롤링함.  
약 8000천개의 데이터를 작성자 ID, 게시날짜, 제목, 내용으로  
수집함.

# 시민의견수렴 게시판 분석

월별 게시글 수



Highly increasing the number of reviews

2015

2016

2017

2018

2019



## 형용사 추출

상위 40개 항목

[('입니다', 1789), ('있는', 1272), ('있습니다', 1101), ('있', 1091), ('없', 921), ('같습니다', 743), ('많은', 631), ('있는데', 520), ('좋겠습니다', 476), ('어떻게', 461), ('없습니다', 453), ('없는', 408), ('이런', 372), ('없네요', 354), ('같은', 340), ('같은데', 304), ('빠른', 294), ('있고', 280), ('있어', 280), ('안되고', 276), ('좋은', 273), ('많습니다', 273), ('아닌', 272), ('있으면', 270), ('없어', 259), ('없고', 250), ('같아요', 247), ('있게', 210), ('있을', 175), ('안되네요', 158), ('안됩니다', 156), ('많은데', 156), ('안되는', 156), ('있다고', 154), ('궁금합니다', 153), ('있는지', 150), ('좋을', 148), ('어떤', 139), ('많고', 130), ('있나요', 130)]



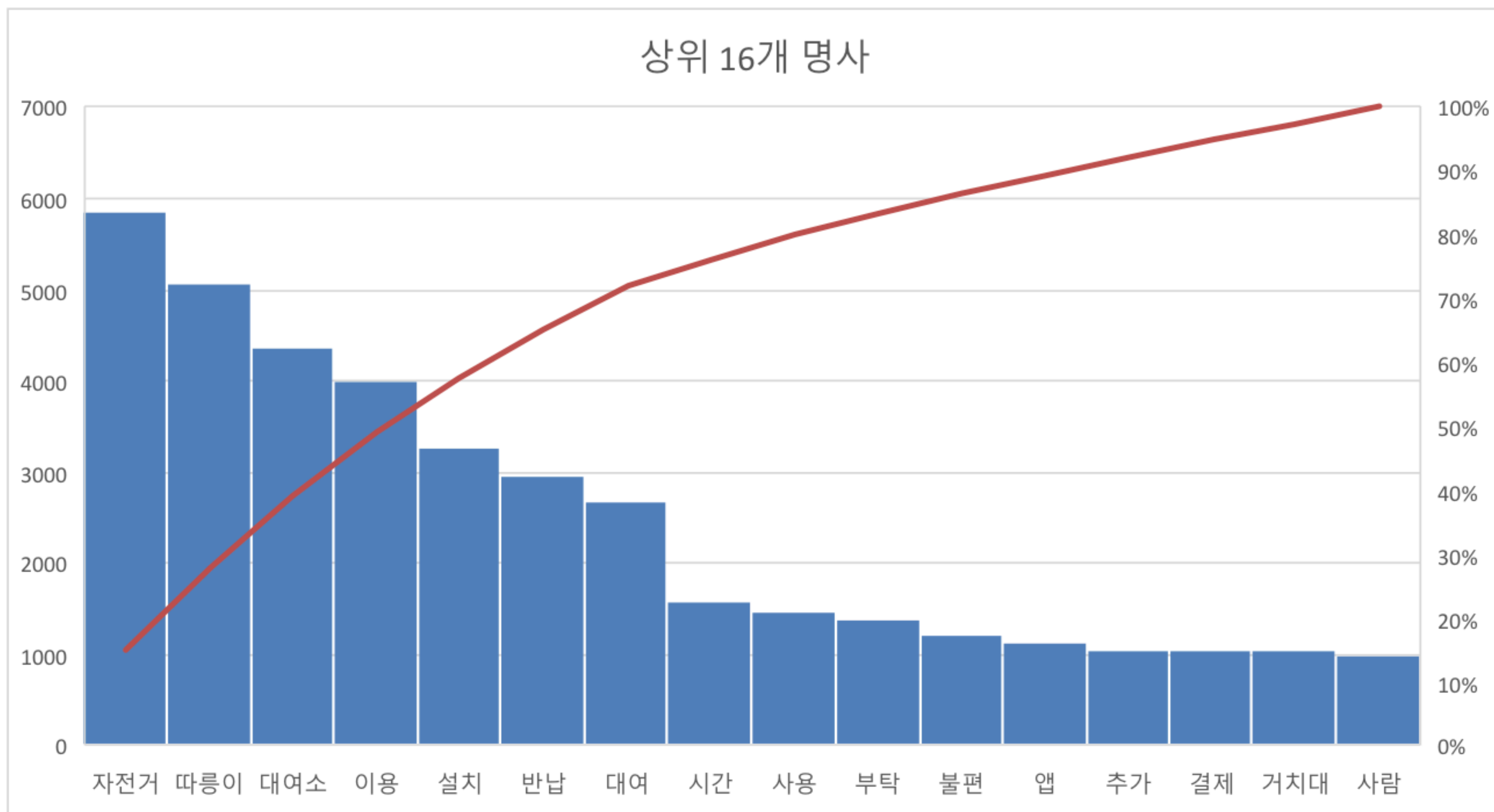
## 명사 추출

상위 70개 항목

[('자전거', 5858), ('따릉이', 5085), ('대여소', 4369), ('이용', 3995), ('설치', 3278), ('반납', 2967), ('대여', 2677), ('시간', 1587), ('사용', 1475), ('부탁', 1379), ('불편', 1207), ('앱', 1132), ('추가', 1055), ('결제', 1048), ('거치대', 1039), ('사람', 998), ('답변', 993), ('확인', 942), ('가능', 934), ('생각', 847), ('고장', 827), ('지도', 802), ('근처', 797), ('출구', 790), ('시민', 770), ('다른', 760), ('다시', 736), ('처리', 728), ('글', 714), ('경우', 708), ('위치', 672), ('아파트', 653), ('계속', 648), ('왜', 645), ('문제', 628), ('정기권', 622), ('장애', 619), ('주변', 617), ('전화', 589), ('타고', 588), ('번호', 584), ('분', 569), ('연결', 567), ('건의', 567), ('이용자', 563), ('못', 554), ('카드', 554), ('지역', 545), ('주시', 542), ('하나', 536), ('통신', 534), ('개선', 533), ('등록', 529), ('오늘', 525), ('요청', 519), ('환불', 509), ('신고', 501), ('거리', 498), ('이용권', 495), ('현재', 492), ('사거리', 489), ('안장', 488), ('다가', 481), ('공원', 475), ('언제', 470), ('오류', 468), ('서울시', 467), ('주민', 465), ('요금', 461), ('어플', 458), ('정도', 453), ('필요', 448), ('문의', 448), ('개', 434), ('바로', 427), ('방법', 425), ('전', 422), ('거나', 418), ('서울', 418), ('단지', 404), ('센터', 404), ('안내', 404), ('이상', 403), ('구매', 399), ('항상', 393), ('정말', 390), ('지금', 388), ('아침', 386), ('말', 383), ('문자', 381), ('밋', 380), ('공지', 379), ('길', 378), ('의견', 373), ('버스', 371), ('때문', 371), ('자주', 370), ('상태', 368), ('가요', 367), ('검색', 365)]



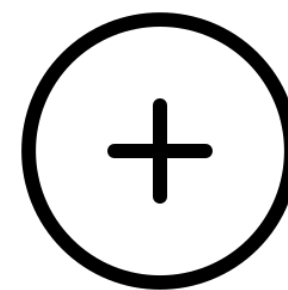
## 시민의견수렴 게시판 파레토 차트





Google Store

빈도 수 상위 70개의 명사



시민 의견수렴

빈도 수 상위 70개의 명사





## 시민의견수렴, 구글플레이 리뷰 공통 단어

각 상위 70개 항목을 추출 28개의 공통 단어가 수집됨.  
(단순히 공통된 항목, 빈도 순 아님)

['자전거', '따릉이', '대여소', '이용', '반납', '대여', '시간', '사용',  
'불편', '앱', '결제', '사람', '생각', '지도', '다시', '위치', '계속',  
'문제', '전화', '타고', '번호', '카드', '개선', '등록', '이용권', '오  
류', '서울시', '어플']

# 데이터 마이닝

Python Library(Selenium, BeautifulSoup4)



## INSTAGRAM

총 게시물 수: 1,258개

URL: <https://www.instagram.com/explore/tags/따릉이/?hl=ko>

2019.7.22 부터 2019.08.11 까지 ‘#따릉이’(이)가 사용된 게시물과 위치태그가 사용된 1,258개의 게시물 텍스트를 크롤링함.

크롤링 된 데이터를 기반으로 게시물 URL, 게시 날짜, 텍스트 순으로 정리함.

트위터에 비해 업로드 되는 게시물 수가 매우 많기 때문에 여러 SNS 중 인스타그램을 선정함.

하지만, 트위터와 달리 ‘API’ 사용이 제한 되었기 때문에 하루에 최신순 게시물 200개를만 크롤링이 가능함.

## 인스타그램 형용사 추출

SNS 특성상 표준어 사용이 적기 때문에 형태소 분석에 어려움을 겪음.

1

[('좋은', 25), ('있는', 22), ('좋다', 19), ('좋아', 19), ('없는', 15), ('좋아하는', 15), ('좋아요', 15), ('같은', 14), ('입니다', 13), ('시원한', 13), ('많은', 13), ('더운', 10), ('같다', 10), ('같아서', 10), ('덥고', 10), ('신나게', 9), ('있어', 9), ('좋', 8), ('더워서', 8), ('덥다', 8)]

2

('ㅋㅋ', 38), ('ㅋㅋㅋ', 27), ('ㅎㅎ', 24), ('ㅠㅠ', 19), ('ㅋㅋㅋㅋ', 19), ('ㅋ', 17), ('ㅎㅎㅎ', 12), ('ㅌㅌ', 12), ('ㅋㅋㅋㅋㅋ', 10), ('ㅠ', 9), ('—', 8), ('ㅌ', 7), ('ㅎ', 6), ('ㅋㅋㅋㅋㅋㅋ', 5), ('ㄹ', 4), ('ㅋㅋㅋㅋㅋㅋㅋㅋ', 4), ('ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ', 4), ('ㄱㄱ', 3), ('ㅎㅎㅎ', 2), ('ㅠㅌ', 2)





## 인스타그램 해시태그 추출

‘#따릉이’ 해시태그와 함께 사용된 해시태그를 추출함.  
상위 50개 항목

[('#따릉이', 861), ('#한강', 232), ('#자전거', 186), ('#일상', 107), ('#운동', 86), ('#서울', 64), ('#라이딩', 52), ('#daily', 47), ('#seoul', 39), ('#좋아요', 35), ('#한강공원', 35), ('#데일리', 33), ('#운동하는여자', 31), ('#소통', 30), ('#퇴근길', 30), ('#맞팔', 30), ('#야경', 28), ('#hanriver', 26), ('#selfie', 25), ('#일상스타그램', 24), ('#다이어트', 24), ('#운동스타그램', 22), ('#korea', 22), ('#여의도', 21), ('#여의도한강공원', 20), ('#한강라이딩', 20), ('#선팔', 19), ('#팔로우', 18), ('#퇴근', 18), ('#자전거그램', 18), ('#산책', 16), ('#여름', 16), ('#주말', 15), ('#셀스타그램', 15), ('#데이트', 15), ('#새벽', 15), ('#뚝섬유원지', 15), ('#셀카', 14), ('#셀피', 14), ('#running', 14), ('#bicycle', 14), ('#diet', 14), ('#좋아요반사', 13), ('#자전거라이딩', 13), ('#자전거타기', 12), ('#중랑천', 12), ('#f4f', 12), ('#workout', 12), ('#서울자전거', 12), ('#잠실', 11)]



종류와 갯수 사용이 매우 다양함, 빈도수를 추출하기 어렵음.

[('🚲', 147), (' ', 64), ('👩\u200d♀️', 40), ('❤️', 27), ('😭', 21), ('💕', 21), ('👩\u200d♀️', 21), ('•', 21), ('👍', 17), ('💧', 16), ('👩\u200d♂️', 15), ('😄', 15), ('😂', 14), ('♥️', 14), ('😊', 13), ('😟', 12), ('💙', 12), ('💚', 11), ('😡', 11), ('😞', 10), ('✅', 10), ('🔥', 9), ('🚴', 9), ('♡', 8), ('ㄨ', 8), ('🚴‍♀️', 7), ('🎵', 7), ('😊', 7), ('👁️', 7), ('😓', 7), ('🙏', 7), ('🌙', 7), ('👩\u200d♂️', 6), ('🙏', 6), ('😬', 6), ('✨', 6), ('😐', 6), ('😇', 6), ('👤', 6), ('😜', 6), ('🙄', 6), ('"', 6), ('👍', 5), ('😠', 5), ('😣', 5), ('❤️', 5), ('😋', 5), ('💖', 5), (' ', 5), ('😢', 5), ('😬', 5), ('💜', 5), ('☔', 5), ('😎', 5), ('🍏', 4), ('❤️', 4), ('💧', 4), ('👉👉', 4), ('😞', 4), ('🙏', 4), ('✅', 4), ('👫', 4), ('♥️', 4), ('👤', 4), ('👉', 4), ('👩\u200d♀️', 4), ('😭', 4), (' ', 4), ('📺', 4), ('♥️', 4), ('🏠', 4), ('☀️', 4), ('😂😂', 4), ('😞', 4), ('👉', 4), ('㊦', 4), ('🚲', 4), ('➡️', 4), ('😊', 3), ('😱', 3), ('✌️', 3), ('⋯', 3), ('👉', 3), ('😓', 3), ('❗', 3), ('😊', 3), ('🚲\u200d♀️ 🚲\u200d♂️', 3), ('😊', 3), ('🇰🇷', 3), ('☁️', 3), ('😊', 3), ('🚶\u200d♂️', 3), ('📺', 3), ('!!!', 3), ('👉', 3), ('🌿', 3), ('👩\u200d♀️', 3), ('3', 3), ('🚲\u200d♂️ 🚲\u200d♀️', 3), ('😞', 3), ('👩\u200d♀️', 3), ('예', 3), ('🌈', 3), ('🚲\u200d♂️ 🚲\u200d♂️', 2), ('🍽️', 2), ('✉️', 2), ('📺', 2), ('🐾', 2), ('의', 2), ('🌿', 2), ('☀️', 2), ('😞', 2), ('❤️❤️❤️💙', 2), ('🐼', 2), ('🍺', 2), ('👏👏👏', 2), ('👩\u200d♀️', 2), ('🇪🇸', 2), ('※', 2), ('🏹🏹🏹', 2), ('🚶\u200d♀️ 🚶\u200d♂️', 2), ('🏊\u200d♂️', 2), ('☀️', 2), ('😊', 2), ('짬', 2), ('😄😄', 2), ('😞', 2), ('👏', 2), ('😡', 2), ('㊦', 2), ('📺💙', 2), ('👏', 2), ('💕', 2), ('😞', 2), ('😞😞', 2), ('😞💕', 2), ('😊', 2), ('✅', 2), ('☆', 2), ('😞', 2), ('👏', 2), ('❗', 2), ('㊦', 2), ('😭😭', 2), ('👩\u200d♀️ 🙌', 2), ('😞', 2), ('🚶\u200d♂️', 2), ('🚲💕', 2), ('썰', 2), ('👉', 2), ('권', 2), ('☀️', 2), ('😄😄😄', 2), ('😄😄😄', 2), ('⚡', 2), ('👩\u200d♀️', 2), ('🔪', 2), ('🏊', 2), ('💩', 2), ('👐', 2), ('👤', 2), ('ㄴ', 2), ('🍎', 2), ('ㄴ', 2), ('😞😞', 2), ('🚶\u200d♂️', 2), ('🐸🐸', 2), ('!⚠️', 2), ('🚴🚴🚴🚴', 2), ('ㄴ', 2), ('😄😄', 2), ('💕', 2), ('^', 2), ('🍉', 1), ('🌹🌹', 1), ('👩\u200d♂️', 1), ('[E]', 1), ('😞😞', 1), ('😞😞', 1), ('😄😄😄😄😄', 1), ('👐👐', 1), ('👐👐', 1), ('👐👐', 1), ('👐💕', 1), ('😓💧', 1), ('♪', 1), ('🚲👐', 1), ('👩\u200d♂️ 🚲\u200d♂️', 1), ('봳는데', 1), ('😄😄😄😄😄', 1), ('🎉', 1), ('👏', 1), ('😊💧💧', 1), ('👩\u200d♀️ 🚶\u200d♀️', 1), ('👩\u200d♀️ 📺\u200d♀️ 🚲\u200d♀️ 📺\u200d♀️', 1), ('💕', 1), ('🎵🎵', 1), ('🍏💖💙', 1), ('😄😄😄😄😄', 1), ('`ω`', 1), ('👩\u200d♂️ 🚲\u200d♀️ 🐶🐶💕', 1), ('😊❤️', 1), ('👩\u200d♀️ 🚲\u200d♀️ 🚲\u200d♀️', 1), ('😞', 1), ('🚴', 1), ('🍦🍷🍷🍷🍷🍷', 1), ('😞', 1), ('タルンイ', 1), ('バンポ大橋', 1), ('夜景', 1), ('😄😄😄', 1), ('...', 1), ('👐', 1), ('👤', 1), ('👩\u200d♀️ 🚲\u200d♀️ 🚲\u200d♀️', 1), ('📺💡📺🚲', 1), ('🚶\u200d♀️ 🚲\u200d♂️ 🚲\u200d♀️ 📺\u200d♀️ 📺\u200d♂️ 👐', 1), ('👩\u200d♀️', 1), ('😞😞', 1), ('😞❤️', 1), ('ㄨ', 1), ('🌶️', 1), ('💕🌸🌸', 1), ('🍦', 1), ('👤', 1), ('ㅅ', 1), ('ㅇㅅㅇ', 1), ('ノ今', 1), ('📺\u200d♂️', 1), ('👩\u200d♂️', 1), ('👩\u200d♀️', 1), ('🚲\u200d♂️', 1), ('🏕️', 1), ('👏👏👏👏👏👏👏👏👏👏👏👏👏👏', 1), ('👏', 1), ('🚶\u200d♂️ 🚶\u200d♂️ 🚶\u200d♂️', 1), ('👍', 1), ('🐘', 1), ('🐦', 1), ('🐮', 1), ('👍👍👍', 1), ('☁️', 1), ('인덱', 1), ('👩\u200d♂️', 1), ('💊', 1), ('☀️', 1), ('18', 1), ('👍', 1), ('👩\u200d♂️ 🚲\u200d♀️', 1), ('😞😞😞', 1), ('📺📺📺📺📺', 1), ('💕', 1), ('🌸🌸🌸', 1), ('💪', 1), ('👐', 1), ('👩\u200d♂️', 1), ('📺', 1), ('🍷🌿', 1), ('•Д•', 1), ('😄😄😄', 1), ('🚲', 1), ('👐👐', 1), ('😞😞', 1), ('から', 1), ('自転車', 1), ('で', 1), ('帰り道', 1), ('📺👁️✨', 1), ('👩\u200d♀️ 🚲\u200d♀️', 1), ('📺', 1), ('💪', 1), ('💩', 1), ('🚶\u200d♀️', 1), ('😄', 1), ('🚲😞', 1), ('📺', 1), ('😞', 1), ('🇰🇷🇺🇸', 1), ('🌶️', 1), ('🍉☀️', 1), ('🎨', 1), ('🌿', 1), ('👩\u200d♀️ 🚲\u200d♂️ 🚲\u200d♀️ 🚲\u200d♂️', 1), ('♡♡', 1), ('🚲🏊', 1), ('マカロンも溶ける夏。夏本番 🍷🍷🍷', 1), ('😄😄', 1), ('👩\u200d♀️ 🚶\u200d♀️', 1), ('👍', 1)]



## 인스타그램 명사 추출

‘#따릉이’ 해시태그와 함께 사용된 해시태그를 추출함.

상위 50개 항목

[('따릉이', 1541), ('한강', 608), ('자전거', 576), ('운동', 273), ('그램', 242), ('타고', 214), ('오늘', 205), ('서울', 199), ('일상', 197), ('스타', 189), ('라이딩', 172), ('공원', 153), ('시간', 113), ('집', 98), ('나', 85), ('다이어트', 85), ('여의도', 82), ('내', 78), ('퇴근', 75), ('날', 70), ('맛팔', 70), ('daily', 70), ('길', 67), ('것', 66), ('날씨', 65), ('다가', 65), ('사진', 63), ('더', 63), ('소통', 62), ('이', 61), ('맛집', 61), ('데일리', 61), ('비', 58), ('데이트', 56), ('야경', 56), ('때', 55), ('생각', 55), ('저녁', 54), ('카페', 53), ('하늘', 53), ('퇴근길', 53), ('잠실', 53), ('행복', 53), ('어서', 52), ('새벽', 52), ('고', 51), ('밤', 51), ('여름', 51), ('중', 50), ('선팔', 50), ('뚝섬', 50), ('주말', 49), ('타', 49), ('오랜만', 49), ('다리', 45), ('하루', 45), ('시원', 45), ('여자', 44), ('사람', 44), ('어요', 43), ('후', 43), ('진짜', 43), ('동하', 41), ('요', 41), ('산책', 41), ('어제', 40), ('수', 39), ('다시', 39), ('우리', 38), ('반포', 37), ('기분', 37), ('면서', 37), ('라면', 36), ('서울시', 36), ('바람', 35), ('여행', 35), ('기록', 35), ('그', 34), ('거', 34), ('또', 33), ('러닝', 33), ('안', 33), ('언니', 33), ('말', 33), ('휴가', 33), ('땀', 32), ('아침', 32), ('기', 32), ('간', 31), ('요즘', 31), ('하나', 30), ('팔로우', 30), ('달리', 30), ('왜', 29), ('그냥', 29), ('힐링', 29), ('저', 29), ('친구', 28), ('감성', 28), ('시작', 28)]