**1. 워드클라우드**

라이더 안전에 대한 이슈들을 알기 위해 ‘라이더 안전’ 키워드로 기사제목을 크롤링 하였다.

테이블이(가) 표시된 사진

자동 생성된 설명

총 161 건의 기사 제목이 수집되었고 수집된 기사 제목에서 한글 이외의 문자들을 제외하였다.

텍스트이(가) 표시된 사진

자동 생성된 설명

한 글자 키워드와 불용어, 빈도수가 5 이하인 단어들 그리고 불필요한 단어들을 제거하였다.

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

워드클라우드 결과를 통해서 라이더의 안전 문제에 대한 키워드가 기사에 많이 언급되고 있는 것을 알 수 있었고, 프로젝트의 기획 단계에서 방향성을 잡는데 도움이 되었다.

**2. 보험 데이터 전처리 (위도, 경도, 거리, 라이더 등급)**

임의로 설계된 라이더의 데이터에서 출발지, 도착지의 위도, 경도 값과 배달 거리를 구한다.

테이블이(가) 표시된 사진

자동 생성된 설명

rank에 1km 당 alertCount(위험신호 수)가 많을 수록 높은 점수를 산정하였다.

user별로 누적된 rank의 평균값으로 rank를 A/B/C/D/E 등급으로 나누어 라이더를 평가하였다.

테이블이(가) 표시된 사진

자동 생성된 설명테이블이(가) 표시된 사진

자동 생성된 설명

이 데이터를 보험사에 전달해주면 보험사에서는 라이더의 등급에 따라 보험비에 차등을 주어 청구할 수 있고, 라이더는 보험비를 적게 내기 위해 안전운전을 하는데 더 노력할 수 있을 것 이라고 생각한다.

**3. 아쉬운 점**

텍스트이(가) 표시된 사진

자동 생성된 설명



데이터에서 상관관계를 알아보기 위해 거리와 위험신호 횟수, 나이와 위험신호 횟수의 상관계수와 p-value를 구해보았다. 하지만 임의로 설계한 데이터이기때문에 변수들간의 상관관계를 알 수 없었다.

실제 데이터였다면 라이더 등급에 영향을 많이 주는 변수들을 활용하여 다양한 예측 기법들로 라이더의 등급을 예측하는 과정을 진행해 보고 싶었지만 이번 프로젝트에서는 진행해보지 못했던 점이 아쉽게 느껴진다.

**4. 참고자료**

워드클라우드

<https://lovit.github.io/nlp/2018/04/17/word_cloud/>

<https://pythonkim.tistory.com/82>

라이더 평가 데이터

<https://github.com/yjw5344/Geocoder-Python>

<https://wikidocs.net/91661>

<https://pypi.org/project/geocoder/0.6.0/>

<https://mizykk.tistory.com/21>

<https://pypi.org/project/haversine/>