

(직접분석 보고서)

| | | |
|--------|----|--|
| 팀 명 | | 토양환경중요하조 |
| 과제명 | | 미션 : <input type="checkbox"/> 미래가치 <input type="checkbox"/> 경제활력 <input checked="" type="checkbox"/> 민생·안전 |
| | | 폐휴업 주유소 분석 및 예측 모델을 활용한 토양오염 예방 |
| 활용 데이터 | 공공 | - 2021 서울특별시 교통량 조사자료.pdf - 2020 서울특별시 교통량 조사자료.pdf - 2019 서울특별시 교통량 조사자료.pdf - 2018 서울특별시 교통량 조사자료.pdf - 2017 서울특별시 교통량 조사자료.pdf - 국내 전국 개별 주유소 판매가격 일간.csv - 2021 토양안심주유소 지정현황.csv - 2023 한국석유공사_알뜰주유소 현황.csv - 오피넷(https://www.opinet.co.kr/searRgSelect.do) |
| | 민간 | |

과제 개요(150자)

주유소 폐휴업 후 시설 및 오염도 조사의 관리가 미흡한 곳이 많다. 방치된 폐휴업 주유소의 석유 저장 시설은 토양오염의 원인이 된다. 하지만 정기적인 토양오염검사의 시행 및 세부적인 규정이 확립되고 있지 않다. 폐휴업한 주유소의 토양오염 원인은 시설물 부식으로 인한 누출 등 관리 부주의가 대부분이다. 따라서 본 분석은 파이썬 LogisticRegression 모델을 통해 폐업할 주유소를 사전에 미리 예측해 토양오염에 대비를 하도록 하는 데에 목적이 있다.

활용 데이터 및 분석도구

1. 활용 데이터1: 서울특별시 교통량 조사자료

가. 자료 제목/자료 출처

- 1) 2021 서울특별시 교통량 조사자료.pdf (2021.01.01.~2021.12.31.)/서울시 교통정보과(무료)
- 2) 2020 서울특별시 교통량 조사자료.pdf (2020.01.01.~2020.12.31.)/서울시 교통정보과(무료)
- 3) 2019 서울특별시 교통량 조사자료.pdf (2019.01.01.~2019.12.31.)/서울시 교통정보과(무료)
- 4) 2018 서울특별시 교통량 조사자료.pdf (2018.01.01.~2018.12.31.)/서울시 교통정보과(무료)
- 5) 2017 서울특별시 교통량 조사자료.pdf (2017.01.01.~2017.12.31.)/서울시 교통정보과(무료)

나. 내용 및 특징

- 1) 서울시 도로별 교통량의 현황 표기
- 2) 조사지점 표기(총 135개 지점): 도심 24개소, 시계 27개소, 교량 20개소, 간선 45개소, 도시고속 9개소

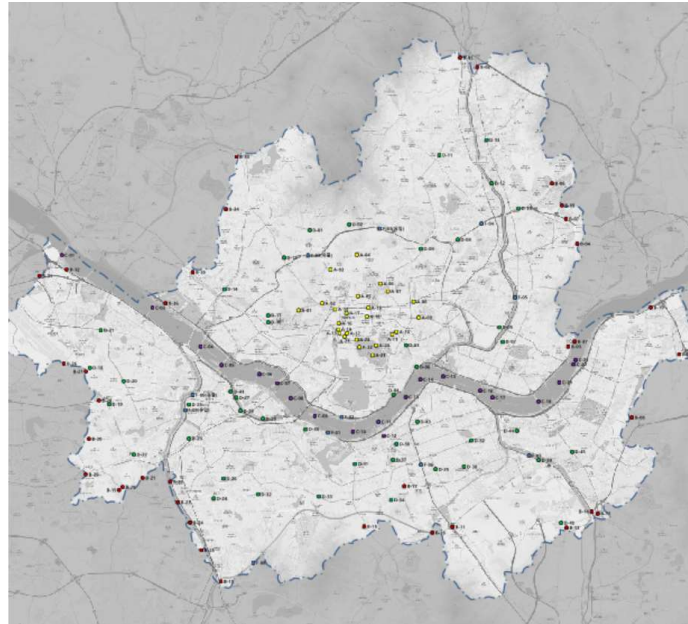
3) 조사 지점에 차량감지기(LOOP, 지자기, 영상감지기) 및 교통량 제어기 설치 후 통신망을 통한 온라인 조사 실시 자료

다. 수집방안

1) 서울시 주유소별 위치 파악 후 해당 도로의 교통량 수집

2) 2017년부터 2021년까지 총 5개년의 데이터 수집

라. 자료 예시: 서울특별시 지도를 활용한 교통량 표기 자료



2. 활용 데이터2: 전국 주유소 판매가격 조사자료

가. 자료 제목/자료 출처

1) 국내 전국 개별 주유소 판매가격 일간.csv (2020.01.~2021.12.)/한국석유공사(무료)

나. 내용 및 특징

- 1) 서울시 주유소별 고급휘발유, 휘발유, 경유, 실내등유 판매 가격 월간 평균 표기
- 2) 주유소명, 지역, 주소, 기간, 평균 가격 등의 열을 사용한 표기
- 3) 휴업, 폐업 등의 이유로 일부 월의 평균 가격이 누락되어 있는 주유소 존재
- 4) 고급휘발유와 실내등유의 경우, 판매를 하지 않았다면 값은 0으로 표기
- 5) 4개의 파일(2020년 1월~6월, 2020년 7월~12월, 2021년 1월~6월, 2021년 7월~12월) 사용

다. 수집 방안

- 1) 고급휘발유와 실내등유의 경우, 판매하지 않는 주유소가 많아 가격이 아닌 판매 여부에 관한 자료 수집
- 2) 휘발유와 경유는 주유소별로 2020년 1월~ 6월 가격 평균(값1)과 2020년 7월~12월 가격 평균

(값2)의 평균을 2020년 평균 가격으로, 2021년 1월~6월 가격 평균(값3)과 2021년 7월~12월 가격 평균(값4)의 평균을 21년 평균 가격으로 고려하여 자료 활용

- 3) 각각의 값(값1, 값2, 값3, 값4)을 구하는 과정에서 6개(1월~6월 또는 7월~12월)의 데이터가 필요하나, 휴업이나 폐업 등의 이유로 6개 미만의 값만 있는 주유소가 존재하여, 해당 경우엔 존재하는 데이터만으로 평균을 구해 사용 -> 만약 어떤 주유소가 6월의 평균 가격 데이터가 누락되어 1~5월 데이터만 존재할 시, (1~5월 가격의 합/5)를 평균값으로 사용

라. 자료 예시(일부 발췌): 2020년 1월~ 6월 휘발유, 경유, 실내등유 표기 자료

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|------------------------|-----------------|---------|-----------------------|-----------|-------|------|---------|---------|---------|---------|---|
| 1 | 번호 | 지역 | 상호 | 주소 | 기간 | 상표 | 셀프여부 | 고급휘발유 | 휘발유 | 경유 | 실내등유 | |
| 2 | 기준 : 월간(202001~202006) | | | | | | | | | | | |
| 3 | A0006039 | 서울 강남구 (유)동하석유 | 힐탑셀프주유소 | 서울 강남구 논현로 640 | 2020년 01월 | SK에너지 | 셀프 | 1802 | 1657 | 1495 | 0 | |
| 4 | A0006039 | 서울 강남구 (유)동하석유 | 힐탑셀프주유소 | 서울 강남구 논현로 640 | 2020년 02월 | SK에너지 | 셀프 | 1795.38 | 1637.97 | 1483.97 | 0 | |
| 5 | A0006039 | 서울 강남구 (유)동하석유 | 힐탑셀프주유소 | 서울 강남구 논현로 640 | 2020년 03월 | SK에너지 | 셀프 | 1741.26 | 1570.61 | 1431.13 | 0 | |
| 6 | A0006039 | 서울 강남구 (유)동하석유 | 힐탑셀프주유소 | 서울 강남구 논현로 640 | 2020년 04월 | SK에너지 | 셀프 | 1617 | 1425.33 | 1290 | 0 | |
| 7 | A0006039 | 서울 강남구 (유)동하석유 | 힐탑셀프주유소 | 서울 강남구 논현로 640 | 2020년 05월 | SK에너지 | 셀프 | 1565 | 1350.81 | 1197.74 | 0 | |
| 8 | A0006039 | 서울 강남구 (유)동하석유 | 힐탑셀프주유소 | 서울 강남구 논현로 640 | 2020년 06월 | SK에너지 | 셀프 | 1616.33 | 1417.33 | 1246.33 | 0 | |
| 9 | A0000525 | 서울 강남구 (주)만정에너지 | 삼보주유소 | 서울 강남구 봉은사로 433 (삼성동) | 2020년 01월 | GS칼텍스 | 일반 | 2289 | 1999 | 1903.65 | 1358.74 | |
| 10 | A0000525 | 서울 강남구 (주)만정에너지 | 삼보주유소 | 서울 강남구 봉은사로 433 (삼성동) | 2020년 02월 | GS칼텍스 | 일반 | 2287.48 | 1997.48 | 1863.69 | 1359 | |
| 11 | A0000525 | 서울 강남구 (주)만정에너지 | 삼보주유소 | 서울 강남구 봉은사로 433 (삼성동) | 2020년 03월 | GS칼텍스 | 일반 | 2253.39 | 1963.39 | 1806.61 | 1359 | |
| 12 | A0000525 | 서울 강남구 (주)만정에너지 | 삼보주유소 | 서울 강남구 봉은사로 433 (삼성동) | 2020년 04월 | GS칼텍스 | 일반 | 2156.53 | 1862.73 | 1692 | 1359 | |
| 13 | A0000525 | 서울 강남구 (주)만정에너지 | 삼보주유소 | 서울 강남구 봉은사로 433 (삼성동) | 2020년 05월 | GS칼텍스 | 일반 | 2080.55 | 1790.55 | 1611.19 | 1359 | |
| 14 | A0000525 | 서울 강남구 (주)만정에너지 | 삼보주유소 | 서울 강남구 봉은사로 433 (삼성동) | 2020년 06월 | GS칼텍스 | 일반 | 2079 | 1832.33 | 1652.33 | 1359 | |
| 15 | A0001219 | 서울 강남구 (주)보성 | 세곡주유소 | 서울 강남구 현릉로 731 (세곡동) | 2020년 01월 | SK에너지 | 셀프 | 0 | 1539.52 | 1379.52 | 0 | |

3. 활용 데이터3: 토양안심주유소 조사자료

가. 자료 제목/자료 출처

- 1) 2021 토양안심주유소 지정현황(2007~2021) / 환경부(무료)

나. 내용 및 특징

- 1) 전국 토양안심주유소 지정현황 표기
2) 주유소명, 소재지, 대표자, 완공검사일, 지정일자, 직영 / 자영 여부, 정유사의 정보 포함

다. 수집 방안

- 1) 2007년부터 2022년까지의 전국 토양안심주유소 지정 현황 데이터 수집
2) 필터를 이용해 서울 소재 주유소의 주유소명 및 정유사 추출 자료
3) 서울시 전 주유소명과 대조 후 토양안심주유소 여부 판단한 정보 포함

라. 자료 예시(일부 발췌): 2007년 토양안심주유소 지정 주유소 현황

| | A | B | C | D | E | F | G | H |
|----|--------|---------------------|-------------------------------|--------|------------|------------|--------|--------|
| 1 | 순번+A3C | 주유소명 | 소재지 | 대표자 | 완공검사일 | 지정일자 | 직영, 자영 | 정유사 |
| 2 | 1 | 현대오일뱅크(주)직영 인천대교주유소 | 인천광역시 연수구 동춘동 913-4 | 대표이사 | | 2007.06.07 | 직영 | 현대오일뱅크 |
| 3 | 2 | SK네트웍스(주)둔전주유소 | 경기도 용인시 처인구 포곡동 둔전리 52-1 | 이창규 | | 2007.06.07 | 직영 | SK네트웍스 |
| 4 | 3 | 현대오일뱅크(주)직영 구로셀프주유소 | 서울시 구로구 시흥대로 531(구로동) | 문종박 | 2007.05.14 | 2007.08.31 | 직영 | 현대오일뱅크 |
| 5 | 4 | SK에너지(주)봉수대로주유소 | 인천광역시 서구 석남동 224-11 | 대표이사 | 2006.10.26 | 2007.09.07 | 직영 | SK에너지 |
| 6 | 5 | 오케이에너지(주)청계동주유소 | 경기도 의왕시 청계동 996-11 | 송주섭 | 2007.03.23 | 2007.09.17 | 자영 | SK에너지 |
| 7 | 6 | SK네트웍스(주)솔빛나루주유소 | 경기도 화성시 10용사로 650(반송동 241) | 조경목 | 2007.09.12 | 2007.10.23 | 직영 | SK에너지 |
| 8 | 7 | 서울석유(주)장충주유소 | 서울시 중구 장충동 1가 31-1 | 김수남 | 2007.06.28 | 2007.10.29 | 자영 | SK에너지 |
| 9 | 8 | GS칼텍스 세창주유소 | 서울시 성북구 길음동 21-73 | 허세홍,김택 | 2007.01.11 | 2007.11.15 | 자영 | GS칼텍스 |
| 10 | 9 | 정릉주유소 | 서울시 성북구 정릉동 404-4 | 김희경,김택 | 2005.12.21 | 2007.11.15 | 직영 | GS칼텍스 |
| 11 | 10 | 효성주유소 | 서울시 성북구 하월곡동 67-49 | 김치명 | 2007.06.26 | 2007.11.15 | 직영(임대) | GS칼텍스 |
| 12 | 11 | 대한주유소 | 경기도 시흥시 수인로 3293(신천동 83-65) | 박성호 | 2007.05.31 | 2007.12.10 | 직영 | GS칼텍스 |
| 13 | 12 | 지에스칼텍스(주)부천고속주유소 | 경기도 부천시 오정구 신흥로 4324(내동 72) | 허세홍 | 2007.03.27 | 2007.12.10 | 직영 | GS칼텍스 |
| 14 | 13 | GS칼텍스(주)대창주유소 | 경기도 수원시 팔달구 우만동 489-14 | 최선태 | 2007.05.01 | 2007.12.21 | 자영 | GS칼텍스 |
| 15 | 14 | GS칼텍스(주)세교주유소 | 경기도 평택시 동삭로 224(세교동) | 허동수 | 2006.07.10 | 2007.12.21 | 직영 | GS칼텍스 |
| 16 | 16 | 태원유통(주)도곡주유소 | 경기도 남양주시 와부읍 도곡리1040-8 | 황성환 | 2005.07.28 | 2007.12.27 | 자영 | S-Oil |
| 17 | 17 | 에스제이에너지산업(주)개포주유소 | 서울시 강남구 양재대로 339 (개포동 1164-2) | 박종호 | 2007.09.12 | 2007.12.27 | 자영 | GS칼텍스 |

4. 활용 데이터4: 알뜰주유소 조사자료

가. 자료 제목/자료 출처

1) 2023 한국석유공사_알뜰주유소 현황_20230125.csv / 한국석유공사(무료)

나. 내용 및 특징

- 1) 현재 영업 중인 전국 알뜰주유소 현황정보
- 2) 자영알뜰주유소, EX알뜰주유소(한국도로공사 운영), NH알뜰주유소(농협 운영)로 구분
- 3) 사업자 보고를 통한 데이터 수집
- 4) 상호, 주소, 상표, 셀프 여부의 정보 포함

다. 수집 방안

- 1) 전국 알뜰주유소 운영 현황 데이터 수집
- 2) 필터를 이용해 서울 소재 주유소의 주유소명 추출
- 3) 서울시 전 주유소명과 대조 후 알뜰주유소 여부 판단

라. 자료 예시(일부 발췌): 알뜰주유소 표기 자료

| | A | B | C | D |
|----|---------------|------------------------|------|------|
| 1 | 상호 | 주소 | 상표 | 셀프여부 |
| 2 | 비금농협주유소 | 전남 신안군 비금면 덕산리 142-2 | 농협알뜰 | N |
| 3 | 도초농협주유소 화도지점 | 전남 신안군 도초면 발매리 1-1 | 농협알뜰 | N |
| 4 | 남신안농협주유소 | 전남 신안군 하의면 웅곡리 242-6 | 농협알뜰 | N |
| 5 | 서진도농협 조도지점주유소 | 전남 진도군 조도면 창유리 225-1 | 농협알뜰 | N |
| 6 | 신안농협 자은지점주유소 | 전남 신안군 자은면 구영리 375-1 | 농협알뜰 | N |
| 7 | 남신안농협신의지점주유소 | 전남 신안군 신의면 상태동리 375-2 | 농협알뜰 | N |
| 8 | 신안농협주유소 | 전남 신안군 암태면 단고리 26-2 | 농협알뜰 | N |
| 9 | 임자농협주유소 | 전남 신안군 임자면 진리 431-29 | 농협알뜰 | N |
| 10 | 북신안농협주유소 | 전남 신안군 증도면 증동리 1569-1 | 농협알뜰 | N |
| 11 | 서진도농협 지산지점주유소 | 전남 진도군 지산면 인지리 309-26 | 농협알뜰 | N |
| 12 | 대정농협무릉주유소 | 제주 서귀포시 대정읍 무릉리 2954-8 | 농협알뜰 | N |
| 13 | 서진도농협광석지소 | 전남 진도군 임회면 삼막리238 | 농협알뜰 | N |
| 14 | 대정농협주유소 | 제주 서귀포시 대정읍 상모리 3715-1 | 농협알뜰 | N |
| 15 | 한림농협주유소 | 제주 제주시 한림읍 한림리 1570 | 농협알뜰 | N |

5. 활용 데이터5: 세차장, 경정비, 편의점, 24시간 운영 여부 조사 자료

가. 자료 제목/자료 출처

- 1) 지역_위치별(주유소)/한국석유공사opinet
(<https://www.opinet.co.kr/searRgSelect.do>)/한국석유공사 (무료)

나. 내용 및 특징

- 1) 현재 영업 중인 서울특별시 주유소 중에 세차장, 경정비, 편의점, 24시간 운영하는 주유소에 관한 정보 포함
- 2) 지역, 상호, 주소, 상표, 셀프 여부의 정보 포함
- 3) 석유사업자가 보고한 자료를 토대로 자료 수집

다. 수집 방안

- 1) 세차장, 경정비, 편의점, 24시간 운영 여부 데이터 수집
- 2) 세차장, 경정비 및 편의점 설치 여부, 24시간 운영 중인 데이터만 표기되어 있으므로
데이터에 없는 주유소는 각각의 해당 시설에 대한 데이터 값 0으로 표기

라. 자료 예시(일부 발췌): 서울시 강남구 세차장 설치된 주유소 자료

| | A | B | C | D | E | F | G | H | I | J |
|----|-------------|-----------------|-------------------|--------|---------------|------|-------|------|------|-------------|
| 1 | 지역_위치별(주유소) | | | | | | | | | |
| 2 | | | | | | | | | | (단위 : 원/리터) |
| 3 | 지역 | 상호 | 주소 | 상표 | 전화번호 | 셀프여부 | 고급휘발유 | 휘발유 | 경유 | 실내등유 |
| 4 | 서울특별시 | (주)보성 세곡주유소 | 강남구 현대로 731 (세 | SK에너지 | 02-445-6870 | Y | - | 1582 | 1472 | - |
| 5 | 서울특별시 | 방죽주유소 | 강남구 방고개로 215 (방 | GS칼텍스 | 02-459-3434 | Y | - | 1597 | 1479 | - |
| 6 | 서울특별시 | 자곡셀프주유소 | 강남구 방고개로 120 (자 | SK에너지 | 02-445-5841 | Y | - | 1599 | 1489 | - |
| 7 | 서울특별시 | 오일뱅크(주)직영 산성셀프 | 서울 강남구 현대로 730 | 현대오일뱅크 | 02-2226-4963 | Y | 1867 | 1599 | 1499 | - |
| 8 | 서울특별시 | 오일뱅크(주)직영 도곡셀프 | 강남구 남부순환로 2718 (도 | 현대오일뱅크 | 02-529-5101 | Y | 1822 | 1629 | 1534 | - |
| 9 | 서울특별시 | (주)중앙에너지비스 수서지 | 강남구 광명로 202 (수 | SK에너지 | 070-8707-4582 | Y | 1867 | 1635 | 1497 | - |
| 10 | 서울특별시 | 오일뱅크(주)직영 유진 | 강남구 논현로 152 (도 | 현대오일뱅크 | 02-3462-5215 | N | 1837 | 1644 | 1419 | 1570 |
| 11 | 서울특별시 | 오일플러스 셀프 | 강남구 남부순환로 2651 (오 | SK에너지 | 02-3462-5100 | Y | 1846 | 1644 | 1549 | - |
| 12 | 서울특별시 | (주)성진도곡주유소 | 강남구 도곡로 162 (도 | S-OIL | 02-2058-3024 | N | 1880 | 1664 | 1579 | - |
| 13 | 서울특별시 | (주)에너지(주)진달래주유소 | 서울 강남구 도곡로 206 | SK에너지 | 02-3462-0018 | N | 1849 | 1664 | 1579 | - |
| 14 | 서울특별시 | 삼성동주유소 | 강남구 테헤란로 619 (삼 | 현대오일뱅크 | 02-508-6572 | Y | 1929 | 1689 | 1599 | - |
| 15 | 서울특별시 | SK서광주유소 | 서울 강남구 역삼로 142 | SK에너지 | 02-562-4855 | Y | 1850 | 1695 | 1595 | - |

6. 분석도구

가. 파이썬 활용

창의성

- 1. 다양한 데이터 소스 활용: 교통 데이터, 지리 정보, 경유 및 휘발유의 가격 등과 같은 데이터를 통합하여 기존에 없던 폐업할 주유소 예측 모델을 개발하여 토양오염 대비 전략을 수립한다. 다양한 데이터 소스를 활용하여, 보다 포괄적이고 정확한 예측과 대비 방안을 제시할 수 있다.
- 2. 인공지능 및 머신러닝 활용: 기존 연구들은 통계 모델을 사용하여 예측을 수행하였다. 본 과제의 독창적인 특징은 기계학습 기술을 활용하는 것이다. LogisticRegression를 활용해 기존에 없는 새로운 예측 모델을 개발하고 토양오염 대비 전략을 수립할 수 있다.

3. 지역별 맞춤형 접근 가능: 폐업 주유소와 토양오염은 지역에 따라 다른 패턴과 특징을 가질 수 있다. 본 과제는 분석 지역을 서울시로 한정하여 진행하였지만, 서울시뿐 아니라 다른 지역에도 분석 모델을 활용할 수 있다. 분석 모델을 지역별 맞춤형으로도 적용 가능하다. 예를 들어, 특정 지역의 교통량, 판매가격 데이터를 사용하여 해당 분석 모델을 적용한다면, 해당 지역의 폐업 주유소를 예측할 수 있고 토양오염 대비 전략 또한 수립 가능하다.

4. 예방적 접근 방식 제안: 선행 연구 대부분은 문제가 발생한 후에 대응하는 방식이다. 본 과제는 선행되었던 연구들과 달리, 후속처리 접근 방식뿐만 아니라 예방적인 접근 방식 또한 고려하여 주유소로 인한 토양오염 문제를 미리 예방하는 분석 모델과 해결방안을 제시한다.

적합성

1. 분석 수행 시 활용한 데이터의 융합 및 활용성

가. 본 모델 구성에 사용된 데이터: 15가지(서울특별시의 연도별 교통량과 유류별 판매가격, 실내등유 및 고급 휘발유 판매 여부, 셀프, 24시간, 세차장, 편의점, 경정비 시설, 알뜰주유소, 토양안심주유소 여부, 주유소명, 주소, 영업 상태)

나. 영업 상태를 제외한 데이터들을 영업 상태에 영향을 주는 요인들로 판단하여 데이터 간의 관계나 다른 정보를 이용해 새로운 정보(열)를 생성 후 활용하였다.

1) 구별 교통량의 평균값: 연도별로 나뉜 교통량을 평균값으로 대체해 사용

2) 구별 폐업 및 휴업 주유소와 영업 중 주유소와의 비율 ($(\#휴업 \text{ 및 } \#폐업) / (\#전체 \text{ 주유소})$):

다른 특징과 구별 폐업, 휴업 주유소 비율 간 상관관계 규정

3) (주유소 유류 가격) - (구별 유류 가격 평균): 구별 가격의 상대성에 의해 구별 평균가에 대한 가격 차를 이용

다. 서울시 구별 평균 소득, 차량의 가격, 소비 여력에 대한 정도의 차이를 인식해 데이터들을 구별로 그룹화한 후 모델 학습에 사용하였다.

라. 보편적인 특징으로 분석을 수행하였기에 수집 데이터를 서울시로 한정하지 않고 전국 시도별 데이터로 확장 및 추가 시, 전국 주유소에 대한 휴업/폐업 위험성 판단이 가능하다.

2. 분석 방법 및 사용 모델의 내용에 대한 구체적인 설명 (분석 모델의 구현 예시 시각화 결과 및 설명)

가. 분석방법: 분석에 사용한 정보는 서울특별시의 연도별 교통량과 유류별 판매가격, 실내등유 및 고급 휘발유 판매 여부 등으로 구성된다. 2년 내 폐업/휴업한 주유소에 대한 정보와 현재 영업 중인 주유소에 대한 정보를 이용해 폐업/휴업에 관한 각 변수의 영향력을 상관계수로 평가하였고

기계학습을 이용해 제작한 모델의 적합성을 교차검증을 이용해 검정하였다.

나. 분석모델: 로지스틱 회귀(Logistic regression)을 이용해 폐업/휴업 가능성(0과 1 사이의 값을 가지는 확률)을 예측하고 특정 기준에 따라 잘 운영되고 있는 주유소, 관심 주유소, 심각 주유소로 구분. 각 주유소는 해당 결과에 따라 적절한 대응을 할 수 있도록 결과를 도출하였다.

다. 로지스틱 회귀(Logistic regression): 종속 변수가 이진(binary) 형태인 경우를 다루는 통계적 분석 방법이다. 주로 분류(classification) 문제에서 사용되며, 입력 변수들과 해당 종속 변수 간의 관계를 0과 1 사이의 값으로 압축하여 확률로 해석할 수 있게 하여 관계 모델링에 적합한 모델이다. 제작 모델의 경우 휴업/폐업을 동일한 클래스(1)로 설정하고 운영 중인 주유소를 다른 클래스(0)로 설정하여 분류문제로 해결하려 하고, 다른 변수들과 관계를 통해 확률을 계산하고자 하므로 가장 적합한 방식이라고 생각하였다.

3. 분석과정

가. 데이터 전처리

- 1) 결측값(수집되지 못한 데이터) 처리
- 2) 교통량 결측값: 같은 주소이면 같은 교통량으로 기록되어 있기 때문에 주소별(구별)로 그룹화한 후 결측값 대체하였다.
- 3) 유가(휘발유/경유) 결측값: 구별 2021년과 2020년의 유가 차이를 계산, 2021년의 유가가 2020년의 유가보다 높기 때문에, 결측값의 위치(2021년, 2020년)에 따라 이전/이후의 데이터에서 빼거나 더하는 방식으로 결측값 대체하였다.

나. 데이터 스케일링 (Min-max 스케일링)

- 1) 데이터 스케일링은 입력 변수들의 범위를 일정하게 조정해 모델의 학습과 범위가 크게 다른 데이터들의 예측에 대한 영향을 줄이고 일반화 성능을 향상시켰다.
- 2) Min-max 스케일링: 데이터의 범위를 [0, 1] 또는 다른 원하는 범위로 조정하는 데이터 스케일링 방법. 주어진 변수의 최솟값을 0으로, 최댓값을 1로 매핑하여 데이터를 변환한다.
이를 통해 변수들 간의 범위를 동일하게 만들고, 변수의 상대적인 중요도를 고려하였다.

다. 새로운 열 정의 및 상관관계 분석

1) traffic_mean(평균 교통량) : 연도별 전체 교통량에 대한 평균

2) proportion(비율) : 전체 주유소 중 휴업/폐업 주유소 수에 대한 비율

3) diesel_2021_diff, diesel_2020_diff, gasoline_2021_diff, gasoline_2020_diff :

(주유소 유가) - (구별 유가 평균), 유가의 절대적 크기보다는 해당 구의 평균 유가에 대한 상대적인 차이가 해석에 유용할 것으로 판단하였다.

4) 상관관계 분석 (피어슨 상관계수): 두 변수 간의 선형 관계의 강도와 방향성을 나타낸다.

값이 1에 가까울수록 양의 선형 관계가 강하고, 값이 -1에 가까울수록 음의 선형 관계가 강하다.

값이 0에 가까울수록 두 변수 간에는 선형 관계가 거의 없다고 볼 수 있다.

라. 결과: 시각화 결과(그림 10)

4. 로지스틱 회귀 모델 제작 및 결과 해석

가. Logistic Regression 라이브러리를 이용하여 분류 모델을 제작하였다.

나. 주유소명, 주소, 도로번호를 제외한 입력 변수로 상태 변수를 예측하였다.

다. 전체 데이터셋을 5등분하여 5번의 학습과정 동안 각 데이터셋을 검증용으로 사용하여 모델의 정확도를 검증하였고, 이에 대한 결과는 시각화 결과(그림 11)이다.

라. 영업 중(0), 휴업/폐업(1)로 분류하고 휴업/폐업(1)에 가까울 확률이 결과로 도출하였다.

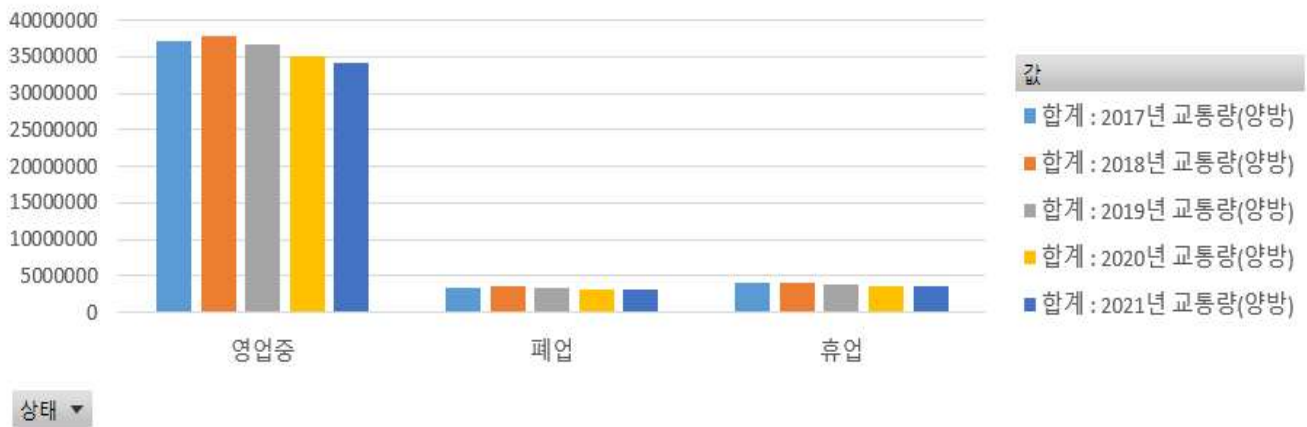
특정 기준(잘 운영되는 주유소 < 0.02, 0.4 < 관심 주유소 < 0.5, 0.5 < 심각 주유소)에 따라 주유소들을 분류하고 주유소명을 출력하였다.

5. 시각화 결과

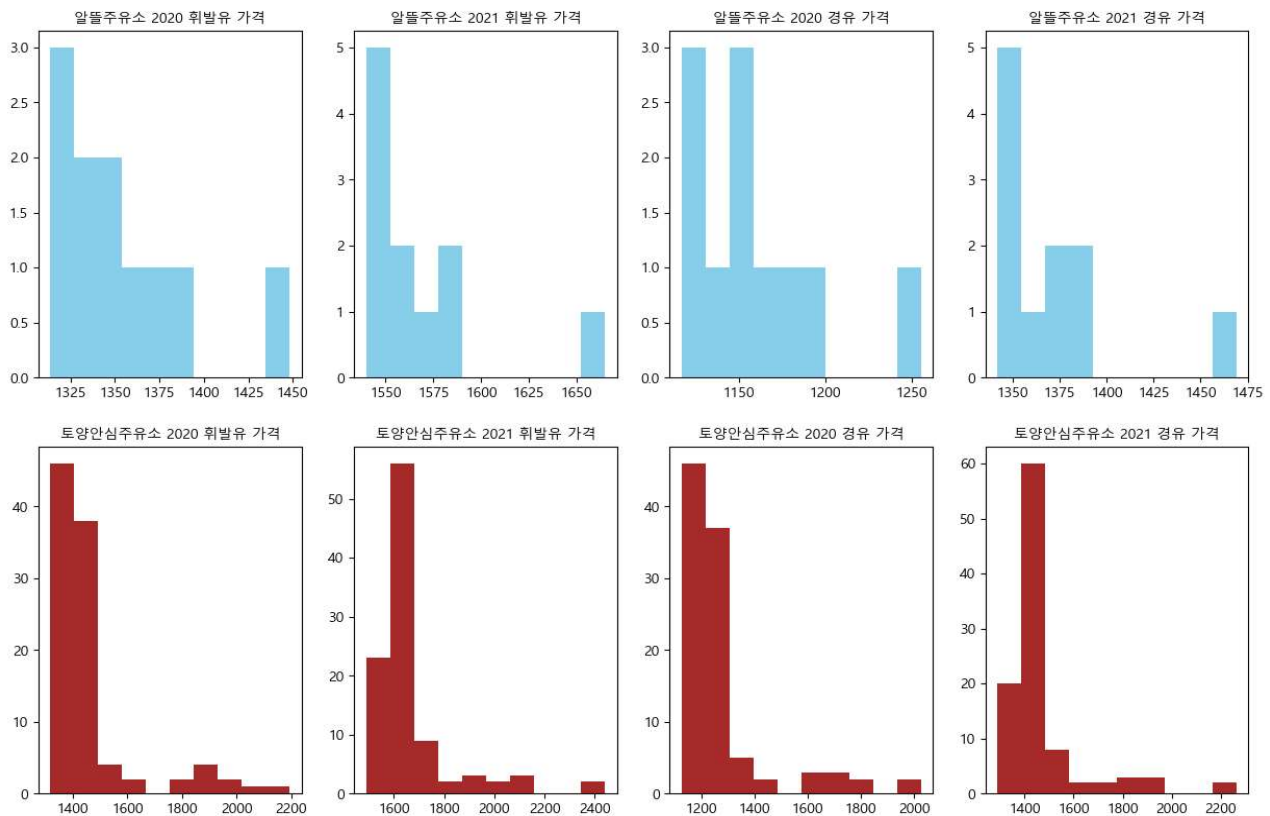
가. 그림 1. 서울시 연도별 교통량

합계 : 2017년 교통량(양방) 합계 : 2018년 교통량(양방) 합계 : 2019년 교통량(양방) 합계 : 2020년 교통량(양방) 합계 : 2021년 교통량(양방)

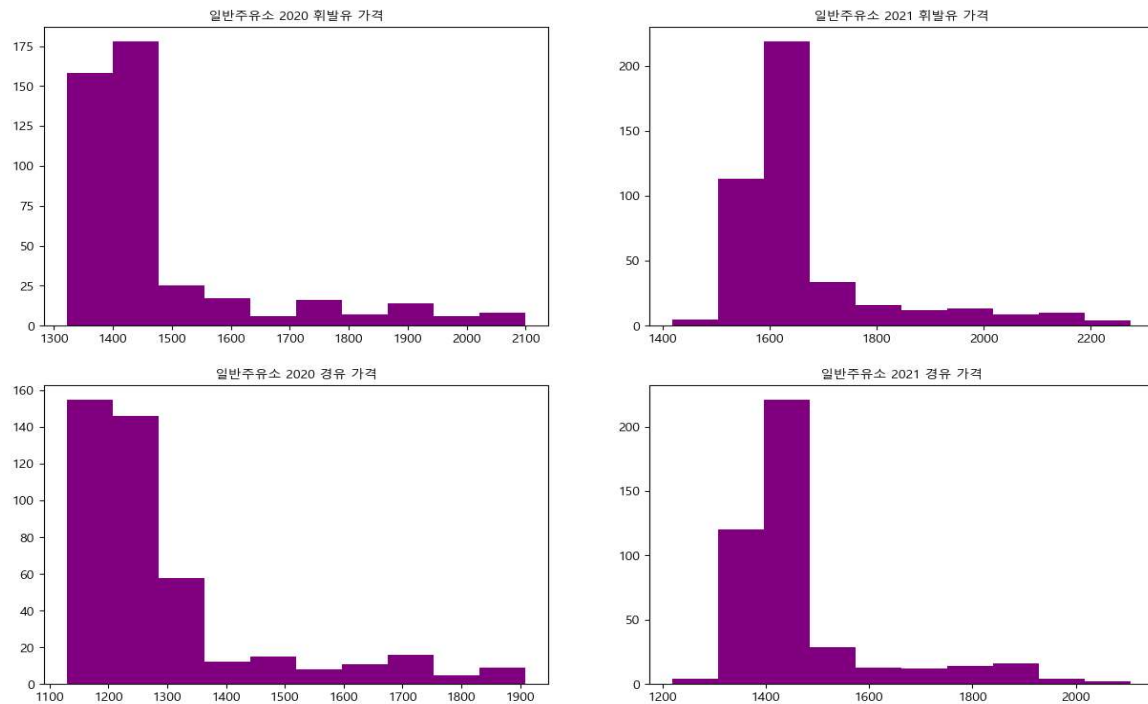
서울특별시 연도별 교통량



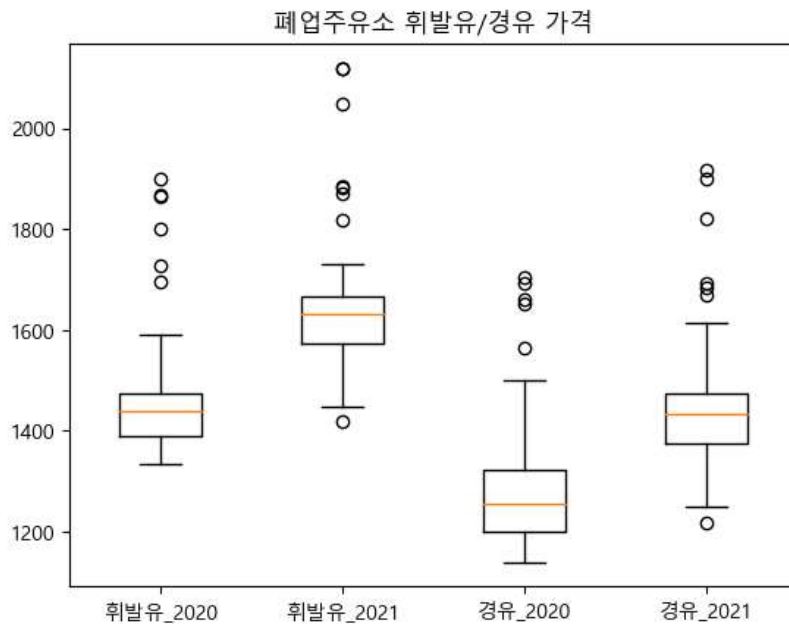
나. 그림 2. 알뜰주유소와 토양안심주유소의 휘발유 / 경유 가격 비교: 알뜰주유소의 휘발유, 경유 가격이 훨씬 싸며 토양안심주유소는 일반 주유소와 비슷하다.



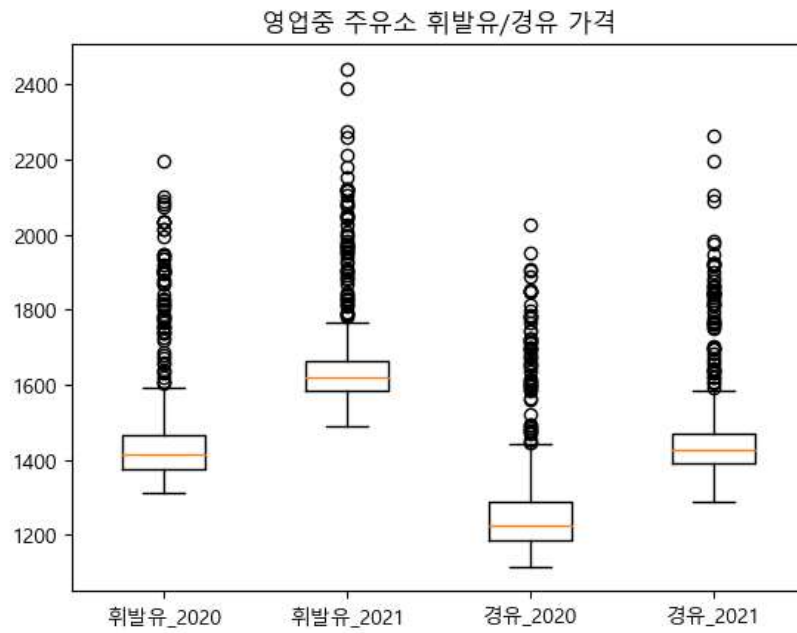
다. 그림 3. 현재 일반주유소의 휘발유 가격, 경유 가격 자료 (토양안심주유소, 알뜰주유소 제외)



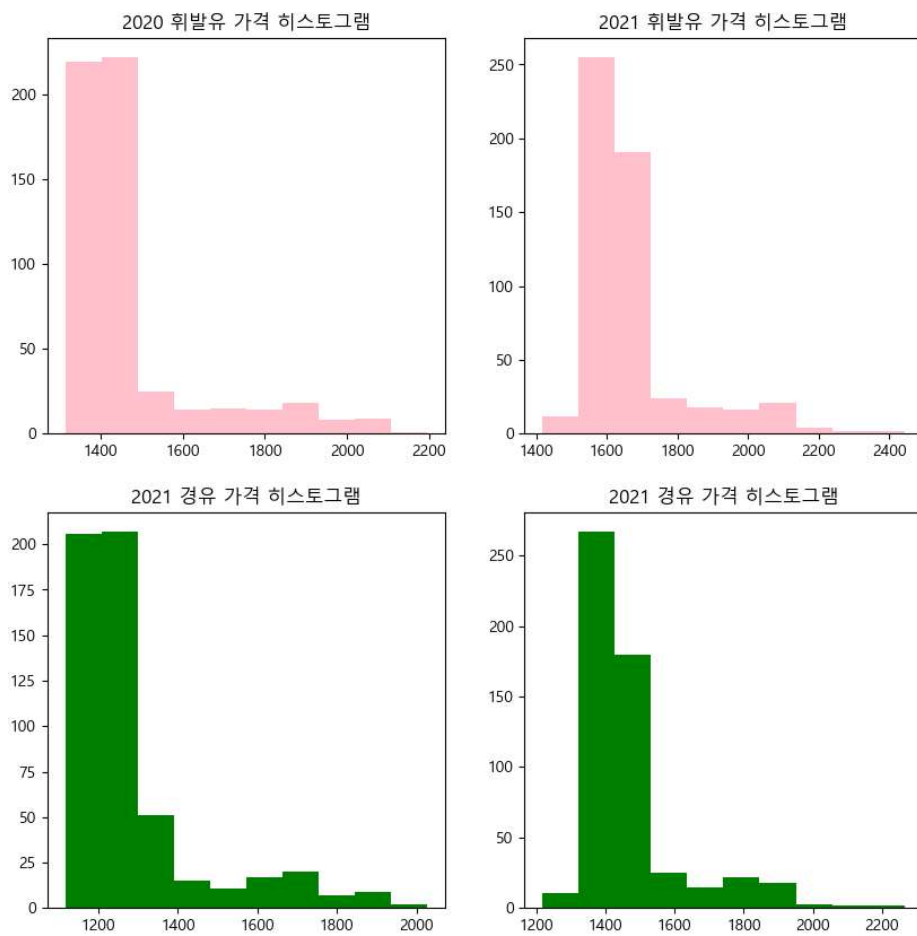
라. 그림 4. 폐업한 주유소의 휘발유, 경유 가격



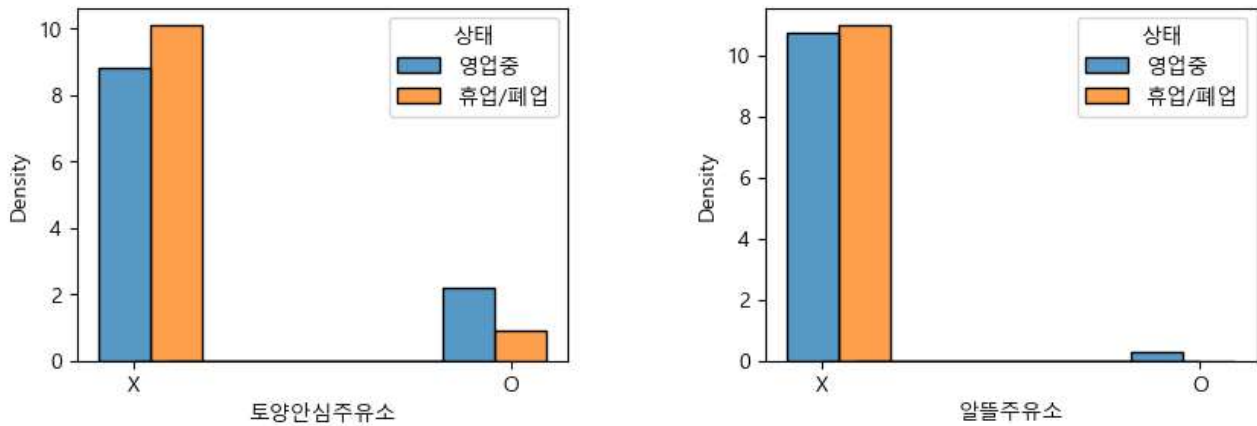
마. 그림 5. 영업 중인 주유소의 휘발유와 경유 가격



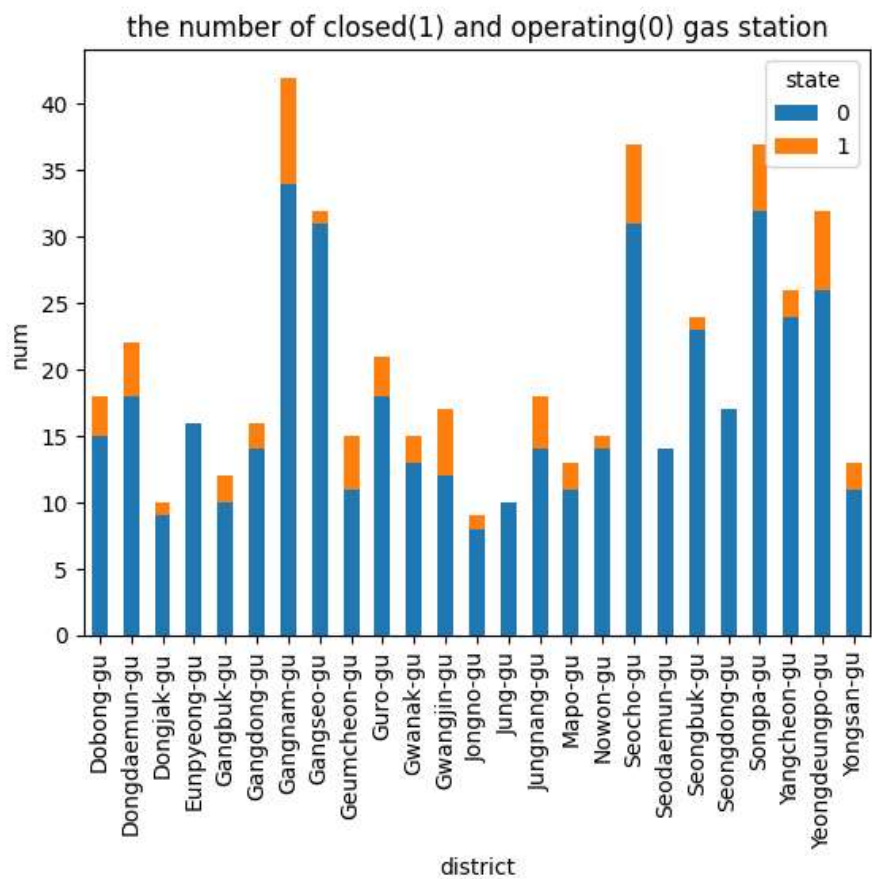
바. 그림 6. 2020년, 2021년 휘발유, 경유 가격 분포(전체 주유소) : 2021년 휘발유와 경유 가격 모두 약 200원 상승



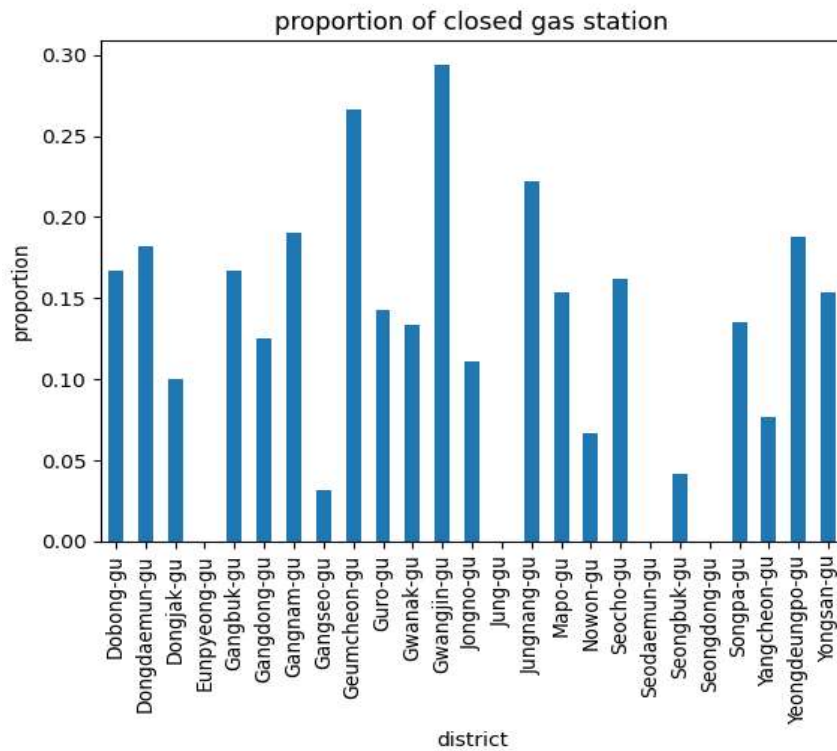
사. 그림 7. 토양안심주유소 및 알뜰주유소 폐업 비율. 토양안심주유소 및 알뜰주유소는 아닌 주유소보다 휴폐업률이 낮다.



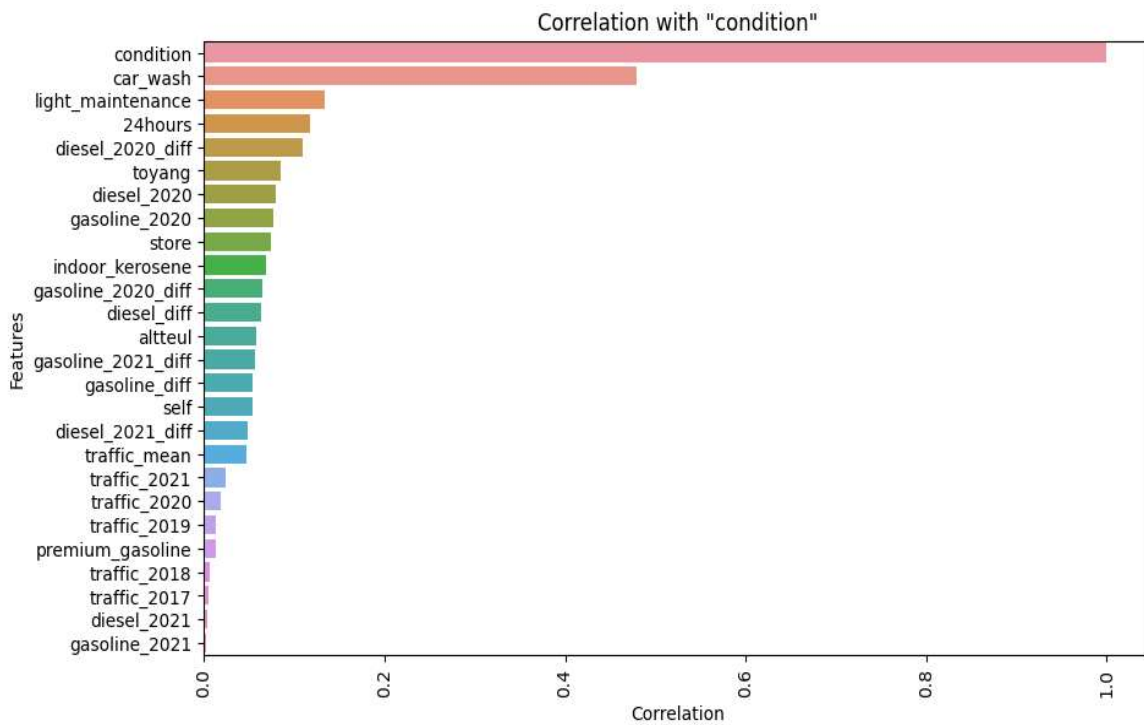
아. 그림 8. 구별 폐업, 휴업(1) / 운영 중(0) 주유소 수



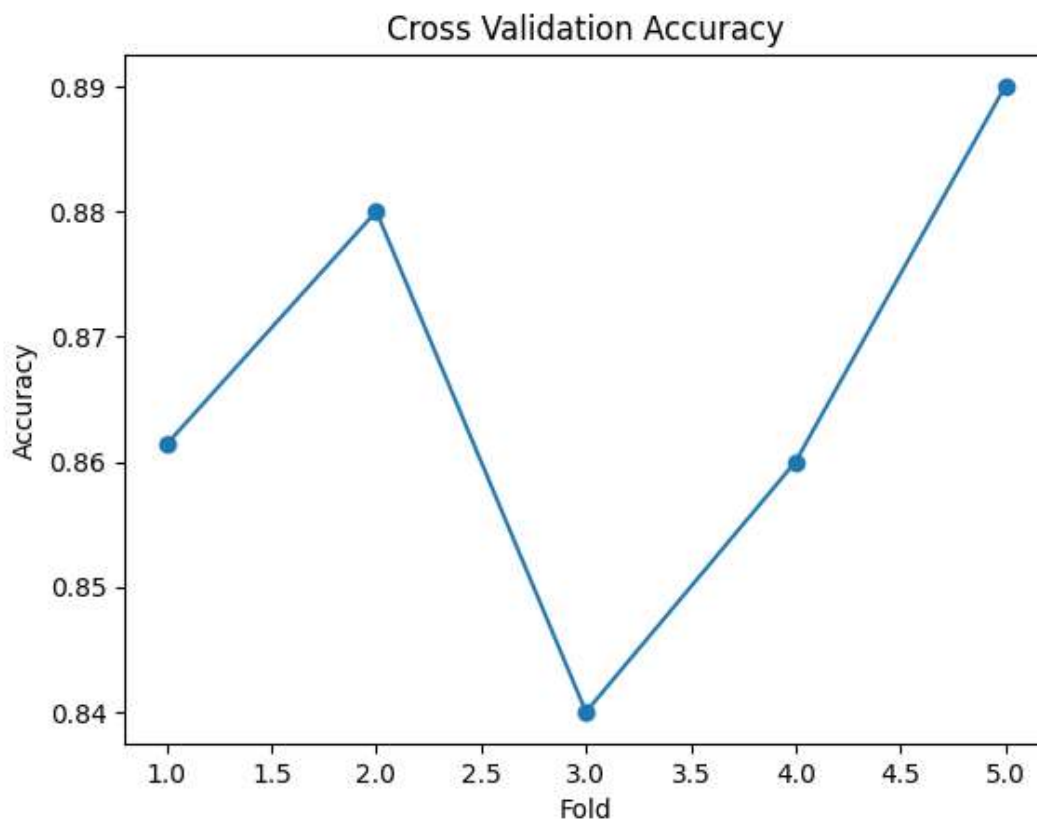
자. 그림 9. 구별 폐업, 휴업 주유소 비율



차. 그림 10. 운영 상태(condition)과 다른 변수 간 상관관계



카. 그림 11. Cross Validation Accuracy(교차검증을 통한 모델의 적합성 점수)



Mean Accuracy: 0.866277227722722

타. 그림 12. 잘 운영되는 주유소명, 관심 주유소명, 심각 주유소명 순

name_good

| | |
|-----|---------------------|
| 27 | 세화주유소 |
| 42 | GS칼텍스(주)도봉주유소 |
| 126 | GS칼텍스(주)서부주유소 |
| 136 | 박물관주유소 |
| 141 | SK에너지(주)가양주유소 |
| 184 | 세원주유소 |
| 242 | (주)신용구도일주유소화랑로 |
| 243 | (주)명연에너지효성주유소 |
| 297 | 형산석유(주)원주유소 |
| 302 | 양천구주유소 |
| 319 | 구인주유소 |
| 465 | (주)삼표에너지정릉주유소 |
| 471 | 원천주유소 |
| 486 | (주)에이치앤디이서울만남의광장주유소 |

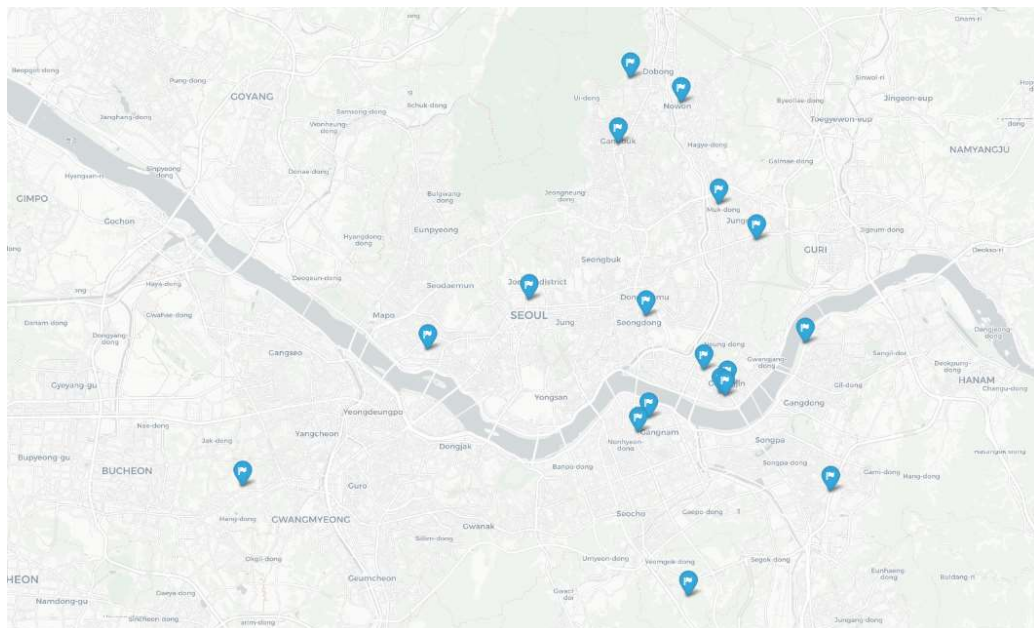
name_interest

64 (주)남경석유세영주유소
 70 현대오일뱅크(주)직영명일셀프주유소
 142 마곡대성주유소
 151 (주)중앙에너지동교지점
 279 현대오일뱅크(주)직영하계주유소
 284 GS칼텍스(주)홍제동주유소
 288 연희주유소
 308 (주)한진김포공항제2주유소
 327 구로주유소
 329 한일주유소
 381 난곡주유소
 415 일원주유소
 449 송파알찬주유소(셀프)
 468 (주)우남에너지문화사랑주유소
 484 현대오일뱅크(주)중랑교주유소
 495 현대오일뱅크(주)직영양천현대셀프주유소

name_serious

79 거여셀프주유소(셀프)
 95 크로바주유소
 156 (주)안국에너지삼보주유소
 161 마포주유소
 191 대성석유(주)동호주유소
 286 sk삼미상사(주)서부주유소
 287 현대오일뱅크(주)직영연세주유소
 290 (주)가재울뉴타운주유소
 385 현대오일뱅크(주)직영관악셀프주유소
 440 대교주유소

파. 그림 13. 서울 내 휴업 주유소 위치



활용성

주유소는 폐업 신고 시 토양오염 조사를 받고 시설물 철거 및 정화 등의 조치를 취해야 한다. 하지만 주유소 폐업 시 1개 주유소당 평균 1억 5000만 원의 비용 탓에 폐업 상태 그대로 방치되는 경우가 상당하다. 방치된 주유소의 시설물 부식은 토양으로의 누유로 이어져 토양오염의 원인이 된다. 휴폐업 주유소의 시설물 방치로 인한 토양오염을 방지하기 위해 직접적으로 지원할 수 있는 구체적인 방안이 마련되지 않은 상황이다. 또한 기존 일반 주유소에서 토양안심주유소로 전환하는 비용이 약 1억 원 가량 발생하여 전환을 꺼리고 있는 자영업자가 많다.

따라서 주유소가 폐휴업을 하고 난 이후에 지원을 해주는 것뿐만 아니라 해당 분석 모델을 통해 주유소 폐휴업의 원인을 분석하고 주유소가 폐휴업하는 것을 방지하고자 한다. 머신러닝을 활용하여 폐휴업 위험 주유소를 선정하고 해당 주유소에 대해 우선적으로 각종 지원을 해준다면 폐휴업 주유소로 인한 토양오염이 확연히 줄어들 것이라 예상한다. 해당 분석 모델은 교통량, 기름 판매 가격, 부가 시설 유무, 알뜰주유소 여부, 토양안심주유소 여부가 주유소의 폐휴업에 주요한 영향을 끼칠 것이라고 보았다.

1. 이미 폐휴업한 주유소 해결방안: 일본의 주유소 폐업 지원 사례 벤치마킹 (철거 비용 지원, 이자보급사업 등)

가. 지하저장탱크 철거 비용 지원

지하저장탱크는 유류를 저장하기 위해 지하에 매설한 저장 탱크이며, 주유소 철거 비용의 대부분을 차지한다. 일본은 토양오염 방지를 위해 지하저장탱크의 철거 비용(약 7000만원)을 일정 상한액 내에서 3분의 2 정도를 보조해주는 식으로 주유소의 폐업을 지원하고 있다.

우리나라 또한 지하저장탱크 철거 비용의 약 3분의 2(약 2000만원)를 보조한다면 주유소 철거 비용 부담이 확실히 줄어들 것이다.

나. 이자보급사업

토양오염도검사를 성실하게 받아온 영업자들을 대상으로, 해당 영업자가 주유소의 폐업에 필요한 사업자금을 금융기관으로부터 차입했을 경우 그 이자의 일부를 지원하는 제도를 도입한다.

2. 토양안심주유소 전환 해결방안

토양안심주유소란 기름 유출 사고를 대비하여 휘발유나 경유가 유출되더라도 주변 토양이 오염되지 않게 여러 장치를 갖춘 주유소를 의미한다. 기존 일반 주유소들을 토양안심주유소로 대체한다면 기름 유출 사고를 미리 예방할 수 있으며 친환경 사업장으로서 이미지 개선과 브랜드 가치 상승이 가능하다.

가. 토양안심주유소 전환 비용 지원

분석 결과 토양안심주유소의 폐업률은 아닌 주유소보다 현저히 낮다. 주유소 영업자들은 기존 주유소에서 토양안심주유소로 전환하는 비용(약 1억 원)이 부담되어 전환을 꺼린다.

분석 모델을 통해 예측한 폐업 위기 주유소에 우선적으로 전환 비용(약 1억 원)의 일부를 지원한다. 토양오염도검사를 성실하게 받아온 영업자일수록 지원 비율을 확대한다.

1) 추가 비용 지원

토양안심주유소로 전환할 시, 전기 충전소 설치 비용을 지원해준다.

(전기 충전소 한 대당 설치 비용 약 3000만원, 최대 50%까지 지원)

3. 토양오염도검사 강화

가. 토양오염도검사 실시

현행법상 주유소 설치 후 15년간은 5년에 한 번, 이후에는 2년에 한 번 토양오염조사 결과를 제출하도록 되어있지만 방치된 주유소들 중 절반 가량은 토양오염 조사에서 누락되어 있어 예상되는 피해는 더욱 크다. 따라서 영업 중인 주유소의 토양오염도검사를 꾸준히 진행해야 한다.

특히, 분석 모델을 통해 예측한 폐업 위기 주유소를 대상으로는 더욱 면밀한 토양오염도검사를 시행한다. 규칙적이며 정확한 오염검사를 통해 폐업 후에 야기될 수 있는 토양오염의 정도 감소 효과를 기대할 수 있다.

나. 토양오염도 검사 비용 지원과 과태료 인상

토양오염도검사를 성실하게 받은 영업자들을 대상으로 토양오염도검사 비용(약 100만원/회)의 일부를 지원 해준다. 토양오염도 검사를 받지 않은 주유소에 과태료(현재 200만원 이하)를 인상한다.

정책 활용

전국 주유소가 고유가와 주유소 간 경쟁 심화, 자동차 시장의 변화로 인해 감소하고 있다. 경영난과 폐업 비용을 감당할 수 없어 휴업하는 주유소가 증가하면서 환경오염 문제를 일으키고 있다. 앞으로 휘발유나 경유를 사용하지 않는 전기차와 수소차가 빠르게 보급될수록 이러한 문제는 더욱 심각해질 전망이다. 이에 폐휴업하는 주유소에 대한 정부 차원의 정책 개선이 필요하다.

1. 주유소 폐휴업에 관한 정책

관련 정책(일부 발췌): <석유 및 석유대체연료 사업법> 제12조(사업의 개시·휴업 및 폐업의 신고)

① 석유정제업자·석유수출입업자 또는 석유판매업자는 그 사업의 등록 또는 신고를 한 날부터 대통령령으로 정하는 기간 이내에 사업을 시작하여야 한다.

② 석유정제업자·석유수출입업자·국제석유거래업자 또는 석유판매업자는 그 사업을 개시·휴업 또는 폐업하였을 때에는 산업통상자원부령으로 정하는 바에 따라 산업통상자원부장관, 시·도지사 또는 시장·군수·구청장에게 신고하여야 한다. <개정 2013. 3. 23., 2017. 4. 18.>

③ 제2항에 따른 신고가 신고서의 기재사항 및 첨부서류에 흠이 없고, 법령 등에 규정된 형식상의 요건을 충족하는 경우에는 신고서가 접수기관에 도달된 때에 신고된 것으로 본다. <신설 2017. 12. 12.>

④ 시·도지사 또는 시장·군수·구청장은 제2항에 따라 석유판매업자의 휴업 또는 폐업 신고를 받은 때에는 그 내용을 산업통상자원부장관, 환경부장관 및 소방청장에게 통보하여야 한다. <신설 2020. 2. 4.>

폐휴업 주유소를 철거하지 않은 채 방치하면 기존에 남아 있던 기름과 이미 지하로 침투된 기름이 그대로 남아 있으면서 토양오염을 진행시키고 심지어 지하수까지 침투하게 된다. 관련 현행 법안을 살펴본 바, 주유소 사업에 관하여 폐휴업할 시에 공공기관에 신고 절차만 정책 지침에 표기되어 있을 뿐, 폐휴업 후에 시설 조치나 폐유 조치에 관한 조항은 확인할 수 없었다. 주유소 폐휴업에 관한 별도의 세부적인 법 조항이나 규칙이 없어 영업자들은 폐휴업한 주유소를 쉽게 방치한다. 따라서 폐휴업 주유소에 관하여 주유소 폐휴업 신고 후 지하저장탱크 철거 등의 비용을 지원하거나 토양안심주유소로 전환하게 하여 폐휴업 신고 절차만 거친 후 무작위로 방치되는 폐휴업 주유소 수를 줄여야만 한다.

2. 토양안심주유소로의 전환에 관한 정책

관련 정책(일부 발췌): <토양안심주유소 지정 및 운영·관리 지침>

⑤ 토양안심주유소 지정

가. 종합검토

(1) 서류검토 및 현장 확인 결과를 종합하여 토양안심주유소 지정요건에 적합한 경우에는 토양안심주유소로 지정

(2) 유역(지방)환경청장은 토양안심주유소로 지정한 때에는 관계기관(환경부장관, 관할 시장·군수·구청장 및 신청인)에게 지정 사실을 통보

※ 지정 사실을 통보할 때 토양오염검사면제승인신청서도 같이 제출할 수 있음

나. 토양안심주유소 지정서 및 현판

(1) 지정서는 '별지2', 현판은 '별지3'의 규격에 따라 '토양안심주유소 지정서' 및 '현판'을 발급·제작하여 지급

(2) 지정서 번호는 유역(지방)환경청 별로 일련번호 부여

토양안심주유소를 설치했을 때 주유소 폐업률이 확연히 낮아지는 것을 확인할 수 있다. 기본 주유소를 토양안심주유소로 전환하여 주유소가 폐업하는 것을 방지할 수 있고, 이는 토양환경오염을 예방한다.

3. 토양환경보전에 관한 정책

관련 정책(일부 발췌): <토양환경보전법> 제10조의4(오염토양의 정화책임 등)

① 다음 각 호의 어느 하나에 해당하는 자는 정화책임자로서 제11조제3항, 제14조제1항, 제15조제1항·제3항 또는 제19조제1항에 따라 토양정밀조사, 오염토양의 정화 또는 오염토양 개선사업의 실시(이하 "토양정화등"이라 한다)를 하여야 한다. <개정 2017. 11. 28.>

③ 시·도지사 또는 시장·군수·구청장은 제11조제3항, 제14조제1항, 제15조제1항·제3항 또는 제19조제1항에 따라 토양정화등을 명할 수 있는 정화책임자가 둘 이상인 경우에는 대통령령으로 정하는 바에 따라 해당 토양오염에 대한 각 정화책임자의 귀책정도, 신속하고 원활한 토양정화의 가능성 등을 고려하여 토양정화등을 명하여야 하며, 필요한 경우에는 제10조의9에 따른 토양정화자문위원회에 자문할 수 있다.

⑤ 국가 및 지방자치단체는 다음 각 호의 어느 하나에 해당하는 경우에는 제11조제3항, 제14조제1항, 제15조제1항·제3항 또는 제19조제1항에 따라 토양정화등을 하는 데 드는 비용(제4항에 따른 구상권 행사를 통하여 상환받을 수 있는 비용 및 토양정화등으로 인한 해당 토지 가액의 상승분에 상당하는 금액은 제외한다. 이하 같다)의 전부 또는 일부를 대통령령으로 정하는 바에 따라 지원할 수 있다. <개정 2017. 11. 28.>

토양환경보전법을 살펴본 바, 토양오염을 발생시킨 자에게 오염토양개선사업을 실시해야 한다는 조항이 명시되어 있지만 폐휴업 주유소와 주요소 영업자에 관해서는 해당 법 조항의 효력이 잘 나타나지 않는 듯하다. 따라서 주유소 관련 토양환경보전법이 개별적으로 마련되어야 한다.

4. 알뜰주유소가 일반주유소에 미치는 영향에 관한 정책

일반주유소와 비교했을 때, 알뜰주유소는 폐업률이 현저히 낮다. 이용자들이 주유 가격이 싼 알뜰주유소를 많이 이용하므로 알뜰주유소는 폐업을 거의 하지 않고 영업을 지속할 수 있었다. 이용자들의 알뜰주유소 이용률이 높다는 것은 알뜰주유소가 아닌 일반주유소 이용률이 낮다는 것을 의미한다. 따라서 알뜰주유소가 아닌 일반주유소에 대한 이용률 증가를 위해 알뜰주유소 가격과 일반주유소와의 가격 차이를 줄여 일반주유소에 대한 이용률을 증가시킬 수 있다. 일반주유소에 대한 이용률 증가는 주유소 폐업률과 휴업률을 감소시킬 수 있고, 이는 토양오염 감소로 이어질 수 있다.

기대효과

주유소 폐업 시, 영업자들은 위험물안전관리법과 토양환경보전법을 준수하여 위험물 저장 시설의 철거 및 토양 정화의 책임이 있다. 2019년 한국주유소협회의 주장에 따르면, 300평 규모의 주유소의 폐업 시 드는 비용은 위험물 철거비용 7000만원, 토양정화비용 7000만원 정도이다. 영업자들은 폐업 시 발생할 상당한 폐업 비용을 감당하지 못하고 폐업 대신 휴업을 선택하는 경우가 많다.

방치된 휴업 주유소에서 폐기물이 방치되고 누출되어 토양오염을 일으키고 해당 주유소의 토양오염은 인근 환경에 악영향을 끼친다. 주유소에서 배출될 수 있는 오염원으로는 지하저장탱크에서의 기름 유출과 배관부식이며, 오염예상물질로는 BTEX, TPH, PAHs, PCP 등 발암물질이 있다. 또한 위에서도 언급한 바와 같이, 현행법상 주유소 설치 후 15년간은 5년에 한번, 이후에는 2년에 한번 토양오염조사 결과를 제출하도록 되어있지만 방치된 주유소들 중 절반 가량은 토양오염조사에서 누락되어 있어 예상되는 피해는 더욱 크다.

따라서 본 프로젝트에서는 휴업 및 폐업 주유소의 수를 최소화하여 환경오염의 문제점을 해결하기 위해 분석 모델을 제작했다. 해당 분석모델은 서울시 주유소의 특징과 주변의 환경을 분석하여 폐업 및 휴업의 가능성이 높은 주유소를 구별해 낼 수 있으며 이에 대한 기대효과는 다음과 같다.

1. 효율적인 운영 관리

주유소 영업자들이 모델을 활용한 결과를 직접 확인할 수 있도록 공개하여 운영자 자신의 주유소가 폐휴업 예상 주유소에 해당하는지에 대해 인지하도록 한다. 채택된 주유소의 운영자는 해당 주유소의 특징을 분석해 폐업 및 휴업의 가능성을 최소화할 것이다. 이는 영업자가 운영 비용, 인력 및 자원을 효율적으로 분배하게 하여 주유소 운영에 긍정적인 효과를 낳고 주유소 폐휴업 비율을 감소시킬 것이다.

2. 휴업 위기 주유소 관리

분석 모델을 활용하여 현재 영업 중인 주유소 중 폐휴업이 예상되는 주유소들을 선별해보았다. 현재 서울시 주유소 중 '남경석유세영주유소', '마곡대성주유소' 등이 휴업 관심 주유소로 선정되었고 '거여셀프주유소', '크로바 주유소', '마포주유소' 등이 심각 주유소로 선정되었다. 해당 주유소를 포함한 휴업 관심 주유소, 심각 주유소들은 분석 모델을 이용해 개선점을 찾고 휴업 가능성을 최소화할 수 있으며 현 운영 방식에 대한 경각심을 가질 수 있다.

3. 환경오염의 최소화

방치된 휴업 주유소는 토양오염의 심각한 문제를 야기하고, 주유소 주변으로 오염이 확산되어 주변 인근 지역의 과도한 복구 비용 발생뿐만 아니라 유출된 기름이 지하수로 흘러 들어가 식수까지 오염시킨 사례가 있다. 이런 휴업 주유소의 경우 가짜 석유 업자들이 싼 값에 임대해 범죄에 악용하는 일도 적지 않다. 해당 문제가 발생할 경우 철거 비용을 상회하는 사회적 비용이 들 수 있다. 즉 휴업 가능성을 분석하고 시설 개조 및 개선 등의 비용을 지원하거나 철거 보조금 지원제도를 마련하여 환경오염을 최소화할 필요가 있다.

4. 정책 개발 및 개선

해당 모델은 주유소 운영상태에 대한 실시간 데이터를 분석할 수 있으므로, 정부나 지자체는 이를 활용해 주유소 관련 정책을 개발하거나 개선할 수 있다. 예를 들어 폐업 및 휴업 주유소의 경향을 파악해 기존 주유소를 융복합 스테이션으로 전환하거나 폐업 주유소를 철거하는 등의 주유소 업계 지원 정책을 수립하거나 주유소 운영 상태에 따른 교통 인프라 조정을 통해 환경오염을 최소화할 수 있다.

참고 (적합성 - 분석과정 코드 첨부)

```
path = './최종최종.csv'💡
```

df

[illegible]

데이터 전처리 및 상관관계 분석

1. 중복값 및 2020년 이전 폐업 주유소 삭제
2. 결측값 처리
3. 상관관계 분석
4. 데이터 스케일링

```
df.isnull().sum()
```

```
상태          0
주유소명      0
주소          0
도로 번호     0
2021년 교통량(양방)  0
2020년 교통량(양방)  0
2019년 교통량(양방)  0
2018년 교통량(양방)  0
2017년 교통량(양방) 23
휘발유_2020    7
경유_2020      7
휘발유_2021    7
경유_2021      7
실내등유 판매 여부      6
고급휘발유 판매 여부    6
셀프          6
24시간        6
세차장        6
편의점        6
경정비        6
알뜰주유소     4
토양안심주유소    0
dtype: int64
```

```
#주유소명 중복값 삭제
df.drop_duplicates(subset=['주유소명'], inplace=True)
#판매가가 없는 행 삭제
to_remove = ['서강주유소', '삼동주유소', '(주)미래아스팔트신우주유소', 'SK네트웍스(주)삼성주유소', 'SK네트웍스(주)동우주유소']
df = df[~df['주유소명'].isin(to_remove)]
```

```
df.isnull().sum()
```

```
상태          0
주유소명      0
주소          0
도로 번호     0
2021년 교통량(양방)  0
2020년 교통량(양방)  0
2019년 교통량(양방)  0
2018년 교통량(양방)  0
2017년 교통량(양방) 21
휘발유_2020    1
경유_2020      1
휘발유_2021    1
경유_2021      1
실내등유 판매 여부      0
고급휘발유 판매 여부    0
셀프          0
24시간        0
세차장        0
편의점        0
경정비        0
알뜰주유소     0
토양안심주유소    0
dtype: int64
```



```
#교통량의 ,를 지우고 numerical type으로 변환
```

```
df['2017년 교통량(양방)'] = df['2017년 교통량(양방)'].str.replace(',', '')
df['2017년 교통량(양방)'] = pd.to_numeric(df['2017년 교통량(양방)'], errors='coerce').astype('Int64')
df['2018년 교통량(양방)'] = df['2018년 교통량(양방)'].str.replace(',', '')
df['2018년 교통량(양방)'] = pd.to_numeric(df['2018년 교통량(양방)'], errors='coerce').astype('Int64')
df['2019년 교통량(양방)'] = df['2019년 교통량(양방)'].str.replace(',', '')
df['2019년 교통량(양방)'] = pd.to_numeric(df['2019년 교통량(양방)'], errors='coerce').astype('Int64')
df['2021년 교통량(양방)'] = df['2021년 교통량(양방)'].str.replace(',', '')
df['2021년 교통량(양방)'] = pd.to_numeric(df['2021년 교통량(양방)'], errors='coerce').astype('Int64')
df.dtypes
```

```
상태                object
주유소명            object
주소                object
도로 번호            object
2021년 교통량(양방)  Int64
2020년 교통량(양방)  Int64
2019년 교통량(양방)  Int64
2018년 교통량(양방)  Int64
2017년 교통량(양방)  Int64
회발유_2020          float64
경유_2020            float64
회발유_2021          float64
경유_2021            float64
실내등유 판매 여부   float64
고급회발유 판매 여부 float64
셀프                object
24시간              object
세차장              object
편의점              object
경정비              object
알뜰주유소          object
토양안심주유소       object
dtype: object
```

```
#교통량 결측값 처리 : 같은 주소 = 같은 교통량
```

```
df['주소'] = df['주소'].str.split().str[1:2].str.join(' ')
df['2017년 교통량(양방)'] = df.groupby('주소')['2017년 교통량(양방)'].transform(lambda x: x.fillna(x.mean().round()))
```

```
df['2017년 교통량(양방)'].isnull().sum()
```

```
0
```

```
#기름값 결측값 처리 : 구별 차이를 이용해 이전/이후 년도의 기름값에서 빼거나 더함
```

```
df['회발유_2020'].fillna(0,inplace = True)
df['경유_2020'].fillna(0,inplace = True)
df['회발유_2021'].fillna(0,inplace = True)
df['경유_2021'].fillna(0,inplace = True)
```

```
#영업종: 0 / 휴업, 폐업 : 1
```

```
df.loc[df['상태'] == '영업중', '상태'] = 0
df.loc[df['상태'] == '휴업', '상태'] = 1
df.loc[df['상태'] == '폐업', '상태'] = 1
```

```
df['상태'].value_counts()
```

```
0    436
1     65
```

```
Name: 상태, dtype: int64
```

```

: df.loc[df['주소'] == '강서로', '주소'] = '강서구'
df.loc[df['주소'] == '중앙구상봉로', '주소'] = '중앙구'
df.loc[df['주소'] == '성산로', '주소'] = '서대문구'
df.loc[df['주소'] == '통일로', '주소'] = '서대문구'
df.loc[df['주소'] == '모래내로', '주소'] = '서대문구'
df.loc[df['주소'] == '세검정로', '주소'] = '서대문구'
df.loc[df['주소'] == '연희로', '주소'] = '서대문구'
address_dict = {
    '강남구': 'Gangnam-gu',
    '서초구': 'Seocho-gu',
    '송파구': 'Songpa-gu',
    '영등포구': 'Yeongdeungpo-gu',
    '강서구': 'Gangseo-gu',
    '양천구': 'Yangcheon-gu',
    '성북구': 'Seongbuk-gu',
    '동대문구': 'Dongdaemun-gu',
    '구로구': 'Guro-gu',
    '도봉구': 'Dobong-gu',
    '정남구': 'Jungnang-gu',
    '광진구': 'Gwangjin-gu',
    '성동구': 'Seongdong-gu',
    '강동구': 'Gangdong-gu',
    '은평구': 'Eunpyeong-gu',
    '관악구': 'Gwanak-gu',
    '금천구': 'Geumcheon-gu',
    '노원구': 'Nowon-gu',
    '마포구': 'Mapo-gu',
    '서대문구': 'Seodaemun-gu',
    '용산구': 'Yongsan-gu',
    '강북구': 'Gangbuk-gu',
    '중구': 'Jung-gu',
    '동작구': 'Dongjak-gu',
    '종로구': 'Jongno-gu'
}

df['주소'] = df['주소'].map(address_dict)
df['주소'].value_counts()

```

```

Gangnam-gu      42
Seocho-gu       37
Songpa-gu       37
Yeongdeungpo-gu 32
Gangseo-gu      32
Yangcheon-gu    26
Seongbuk-gu     24
Dongdaemun-gu   22
Guro-gu         21
Dobong-gu       18
Jungnang-gu     18
Gwangjin-gu     17
Seongdong-gu    17
Gangdong-gu     16
Eunpyeong-gu    16
Gwanak-gu       15
Geumcheon-gu    15
Nowon-gu        15
Seodaemun-gu    14
Yongsan-gu      13
Mapo-gu         13
Gangbuk-gu      12
Jung-gu         10
Dongjak-gu      10
Jongno-gu       9
Name: 주소, dtype: int64

```

```
#구별 영업중 / 폐업, 휴업 주유소 수
```

```
counts = df.groupby('주소')['상태'].value_counts().unstack().fillna(0)
counts.value_counts
```

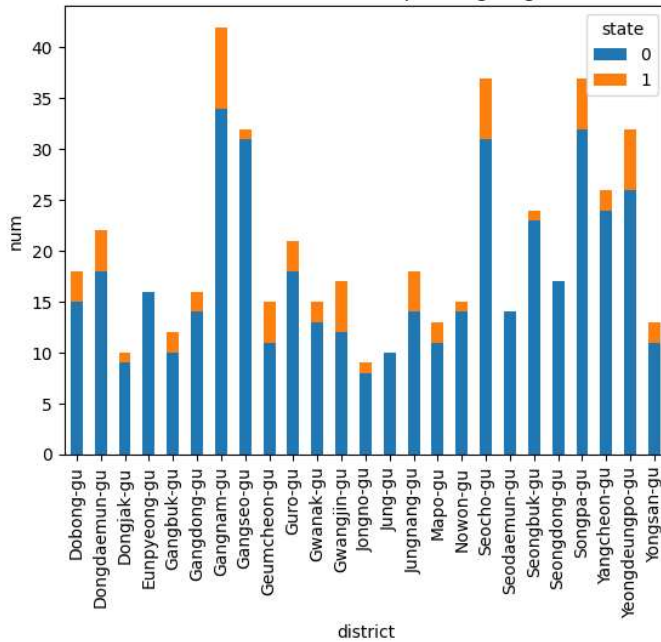
```
<bound method DataFrame.value_counts of 상태          0    1
주소
Dobong-gu      15.0  3.0
Dongdaemun-gu  18.0  4.0
Dongjak-gu      9.0  1.0
Eunpyeong-gu   16.0  0.0
Gangbuk-gu     10.0  2.0
Gangdong-gu    14.0  2.0
Gangnam-gu     34.0  8.0
Gangseo-gu     31.0  1.0
Geumcheon-gu   11.0  4.0
Guro-gu        18.0  3.0
Gwanak-gu      13.0  2.0
Gwangjin-gu    12.0  5.0
Jongno-gu       8.0  1.0
Jung-gu        10.0  0.0
Jungnang-gu    14.0  4.0
Mapo-gu        11.0  2.0
Nowon-gu       14.0  1.0
Seocho-gu      31.0  6.0
Seodaemun-gu   14.0  0.0
Seongbuk-gu    23.0  1.0
Seongdong-gu   17.0  0.0
Songpa-gu      32.0  5.0
Yangcheon-gu   24.0  2.0
Yeongdeungpo-gu 26.0  6.0
Yongsan-gu     11.0  2.0>
```

```
counts = df.groupby('주소')['상태'].value_counts().unstack().fillna(0)
```

```
# 그래프 표시
```

```
counts.plot(kind='bar', stacked=True)
plt.xlabel('district')
plt.ylabel('num')
plt.title('the number of closed(1) and operating(0) gas station')
plt.legend(title='state', loc='upper right')
plt.show()
```

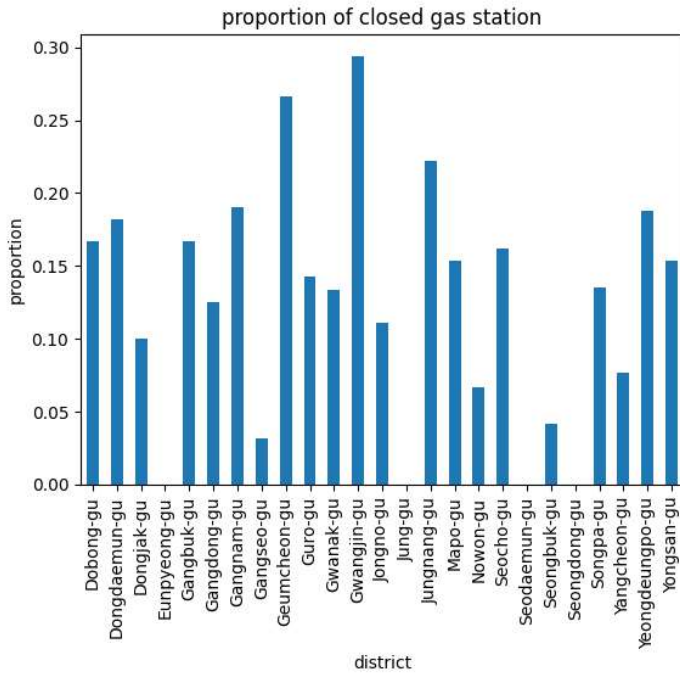
the number of closed(1) and operating(0) gas station



```

#지역별 폐업, 휴업 비율을 새로운 열 proportion으로 정의
counts['proportion'] = counts[1] / (counts[0] + counts[1])
df = df.merge(counts['proportion'], left_on='주소', right_index=True, how='left')
# 그래프 표시
counts['proportion'].plot(kind='bar')
plt.xlabel('district')
plt.ylabel('proportion')
plt.title('proportion of closed gas station')
plt.show()

```



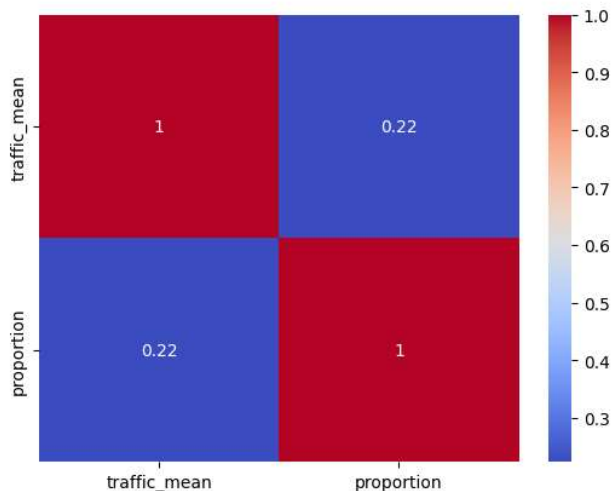
```
#교통량은 평균 교통량으로 계산하여 지역별로 비교
#연도별 전체 교통량을 평균내서 새로운 열 traffic_mean으로 정의
traffic_volume2017 = df.groupby('주소')['2017년 교통량(양방)'].mean()
traffic_volume2018 = df.groupby('주소')['2018년 교통량(양방)'].mean()
traffic_volume2019 = df.groupby('주소')['2019년 교통량(양방)'].mean()
traffic_volume2020 = df.groupby('주소')['2020년 교통량(양방)'].mean()
traffic_volume2021 = df.groupby('주소')['2021년 교통량(양방)'].mean()
mean_traffic_volume = (traffic_volume2017 + traffic_volume2018 + traffic_volume2019 + traffic_volume2020 + traffic_volume2021) / 5
mean_traffic_volume.name = 'traffic_mean'
df = df.merge(mean_traffic_volume, left_on='주소', right_index=True, how='left')
```

```
#교통량 별 영업 / 휴업, 폐업 확인
merged_df = pd.concat([mean_traffic_volume, counts['proportion']], axis=1)

# 상관관계 계산
correlation = merged_df.corr()

# 상관관계를 히트맵으로 시각화
sns.heatmap(correlation, annot=True, cmap='coolwarm')
```

<Axes: >



```
df.loc[df['셀프'] == 'Y', '셀프'] = '1'
df.loc[df['셀프'] == 'N', '셀프'] = '0'

df.loc[df['24시간'] == 'Y', '24시간'] = '1'
df.loc[df['24시간'] == 'N', '24시간'] = '0'

df.loc[df['세차장'] == 'Y', '세차장'] = '1'
df.loc[df['세차장'] == 'N', '세차장'] = '0'

df.loc[df['편의점'] == 'Y', '편의점'] = '1'
df.loc[df['편의점'] == 'N', '편의점'] = '0'

df.loc[df['경정비'] == 'Y', '경정비'] = '1'
df.loc[df['경정비'] == 'N', '경정비'] = '0'

df.loc[df['알뜰주유소'] == 'Y', '알뜰주유소'] = '1'
df.loc[df['알뜰주유소'] == 'N', '알뜰주유소'] = '0'

df.loc[df['토양안심주유소'] == 'Y', '토양안심주유소'] = '1'
df.loc[df['토양안심주유소'] == 'N', '토양안심주유소'] = '0'
```

```
#diff = (기름값) - (구별 평균 기름값)
diesel_2021 = df[df['경유_2021']!=0].groupby('주소')['경유_2021'].mean()
diesel_2021.name = 'diesel_2021'

gasoline_2021 = df[df['휘발유_2021']!=0].groupby('주소')['휘발유_2021'].mean()
gasoline_2021.name = 'gasoline_2021'

diesel_2020 = df[df['경유_2020']!=0].groupby('주소')['경유_2020'].mean()
diesel_2020.name = 'diesel_2020'

gasoline_2020 = df[df['휘발유_2020']!=0].groupby('주소')['휘발유_2020'].mean()
gasoline_2020.name = 'gasoline_2020'
```

```
diesel_diff = diesel_2021-diesel_2020
diesel_diff
```

```
주소
Dobong-gu      196.710717
Dongdaemun-gu  195.411728
Dongjak-gu     177.837417
Eunpyeong-gu   211.150722
Gangbuk-gu     205.854008
Gangdong-gu    208.291042
Gangnam-gu     165.552677
Gangseo-gu     209.825469
Geumcheon-gu   197.820455
Guro-gu        199.387351
Gwanak-gu      174.952747
Gwangjin-gu    196.975889
Jongno-gu      125.163942
Jung-gu        160.593000
Jungnang-gu    202.944620
Mapo-gu        144.036666
Nowon-gu       186.204055
Seocho-gu      152.135608
Seodaemun-gu   184.142789
Seongbuk-gu    223.059011
Seongdong-gu   192.027451
Songpa-gu      186.823768
Yangcheon-gu   206.450952
Yeongdeungpo-gu 157.359995
Yongsan-gu     116.886218
dtype: float64
```

```
gasoline_diff = gasoline_2021 - gasoline_2020
gasoline_diff
```

```
주소
Dobong-gu      204.742083
Dongdaemun-gu  201.706098
Dongjak-gu     204.830917
Eunpyeong-gu   217.962128
Gangbuk-gu     212.370357
Gangdong-gu    222.063438
Gangnam-gu     193.334936
Gangseo-gu     215.706068
Geumcheon-gu   199.818313
Guro-gu        202.530843
Gwanak-gu      197.563985
Gwangjin-gu    201.589391
Jongno-gu      156.235238
Jung-gu        159.277250
Jungnang-gu    211.796759
Mapo-gu        163.349263
Nowon-gu       194.023722
Seocho-gu      183.557473
Seodaemun-gu   188.228529
Seongbuk-gu    226.411267
Seongdong-gu   205.485990
Songpa-gu      201.102803
Yangcheon-gu   211.521964
Yeongdeungpo-gu 165.208923
Yongsan-gu     133.033654
dtype: float64
```

```
df = df.merge(diesel_2021, on='주소', how='left')
df['diesel_2021_diff'] = df['경유_2021'] - df['diesel_2021']

df = df.merge(diesel_2020, on='주소', how='left')
df['diesel_2020_diff'] = df['경유_2020'] - df['diesel_2020']

df = df.merge(gasoline_2021, on='주소', how='left')
df['gasoline_2021_diff'] = df['휘발유_2021'] - df['gasoline_2021']

df = df.merge(gasoline_2020, on='주소', how='left')
df['gasoline_2020_diff'] = df['휘발유_2021'] - df['gasoline_2020']

diesel_diff.name = 'diesel_diff'
gasoline_diff.name = 'gasoline_diff'

df = df.merge(diesel_diff, on='주소', how='left')
df = df.merge(gasoline_diff, on='주소', how='left')
```



```
def fill_missing_2021(row):
    if row['경유_2021'] == 0:
        return row['경유_2020'] + row['diesel_diff']
    return row['경유_2021']

def fill_missing_2020(row):
    if row['경유_2020'] == 0:
        return row['경유_2021'] - row['diesel_diff']
    return row['경유_2020']

def f_m_2021(row):
    if row['휘발유_2021'] == 0:
        return row['휘발유_2020'] + row['gasoline_diff']
    return row['휘발유_2021']

def f_m_2020(row):
    if row['휘발유_2020'] == 0:
        return row['휘발유_2021'] - row['gasoline_diff']
    return row['휘발유_2020']

df['경유_2021'] = df.apply(fill_missing_2021, axis=1)
df['경유_2020'] = df.apply(fill_missing_2020, axis=1)

df['휘발유_2021'] = df.apply(f_m_2021, axis=1)
df['휘발유_2020'] = df.apply(f_m_2020, axis=1)
```

```
df['상태'] = df['상태'].astype(int)
df['셀프'] = df['셀프'].astype(int)
df['24시간'] = df['24시간'].astype(int)
df['세차장'] = df['세차장'].astype(int)
df['편의점'] = df['편의점'].astype(int)
df['경정비'] = df['경정비'].astype(int)
df['알뜰주유소'] = df['알뜰주유소'].astype(int)
df['토양안심주유소'] = df['토양안심주유소'].astype(int)
```

```
df.drop('diesel_2020', axis = 1, inplace = True)
df.drop('diesel_2021', axis = 1, inplace = True)
df.drop('gasoline_2020', axis = 1, inplace = True)
df.drop('gasoline_2021', axis = 1, inplace = True)
```

```
df.rename(columns={'상태': 'condition', '2021년 교통량(양방)': 'traffic_2021', '주유소명': 'name', '주소': 'address', '도로 번호': 'road_num',
                  '2020년 교통량(양방)': 'traffic_2020', '2019년 교통량(양방)': 'traffic_2019', '2018년 교통량(양방)': 'traffic_2018',
                  '2017년 교통량(양방)': 'traffic_2017'}, inplace = True)
df.rename(columns={'경정비': 'light_maintenance', '알뜰주유소': 'altteul', '토양안심주유소': 'toyang', '세차장': 'car_wash', '편의점': 'store',
                  '실내등유 판매 여부': 'indoor_kerosene', '고급휘발유 판매 여부': 'premium_gasoline'}, inplace = True)
df.rename(columns={'경유_2020': 'diesel_2020', '경유_2021': 'diesel_2021', '휘발유_2020': 'gasoline_2020', '휘발유_2021': 'gasoline_2021',
                  '24시간': '24hours', '셀프': 'self'}, inplace=True)
```

```
df.drop('proportion', axis = 1, inplace = True)
```

```
#데이터 스케일링 (Min Max scaling)
cols_to_scale = ['traffic_2021', 'traffic_2020', 'traffic_2019', 'traffic_2018', 'traffic_2017',
                 'gasoline_2020', 'diesel_2020', 'gasoline_2021', 'diesel_2021',
                 'indoor_kerosene', 'premium_gasoline', 'traffic_mean',
                 'diesel_2021_diff', 'diesel_2020_diff', 'gasoline_2021_diff',
                 'gasoline_2020_diff', 'diesel_diff', 'gasoline_diff']
```

```
# Min-Max 스케일러 생성
scaler = MinMaxScaler()
```

```
# 선택한 열 스케일링
df[cols_to_scale] = scaler.fit_transform(df[cols_to_scale])
```



```
correlation = df.corr()
corr = correlation['condition'].abs().sort_values(ascending = False)
corr
```

```
<ipython-input-175-e7ce823eb04d>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version,
it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
correlation = df.corr()
```

```
condition      1.000000
car_wash       0.479716
light_maintenance 0.133369
24hours        0.117796
diesel_2020_diff 0.109384
toyang         0.085361
diesel_2020     0.079010
gasoline_2020   0.076492
store          0.074507
indoor_kerosene 0.068761
gasoline_2020_diff 0.064836
diesel_diff     0.063382
altteul        0.057851
gasoline_2021_diff 0.056910
gasoline_diff   0.054376
self           0.053349
diesel_2021_diff 0.048586
traffic_mean    0.047529
traffic_2021    0.023758
traffic_2020    0.018398
traffic_2019    0.013268
premium_gasoline 0.012878
traffic_2018    0.006938
traffic_2017    0.005589
diesel_2021     0.003552
gasoline_2021   0.002722
Name: condition, dtype: float64
```

```
X = df.drop({'condition', 'name', 'address', 'road_num'}, axis = 1)
y = df['condition']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

```
model = LogisticRegression()
scores = cross_val_score(model, X, y, cv=5)
```

```
for fold, score in enumerate(scores):
    print(f'Fold {fold+1}: Accuracy = {score}')
```

```
mean_accuracy = scores.mean()
print('Mean Accuracy:', mean_accuracy)
```

```
Fold 1: Accuracy = 0.8613861386138614
Fold 2: Accuracy = 0.88
Fold 3: Accuracy = 0.84
Fold 4: Accuracy = 0.86
Fold 5: Accuracy = 0.89
Mean Accuracy: 0.866277227227222
```

```
model.fit(X,y)
y_pred = model.predict(X)
probabilities = model.predict_proba(X)
```

```
indices_good = np.where((probabilities[:, 1] < 0.02))[0]
indices_interest = np.where((probabilities[:, 1] >= 0.4) & (probabilities[:, 1] < 0.5))[0]
indices_serious = np.where((probabilities[:, 1] >= 0.5) & (probabilities[:, 1] < 1))[0]
name_good = df.loc[(df['condition'] == 0) & (df.index.isin(indices_good)), 'name']
name_interest = df.loc[(df['condition'] == 0) & (df.index.isin(indices_interest)), 'name']
name_serious = df.loc[(df['condition'] == 0) & (df.index.isin(indices_serious)), 'name']
```

name_good

```

27          세화주유소
42      GS칼텍스(주)도봉주유소
126     GS칼텍스(주)서부주유소
136          박물관주유소
141      SK에너지(주)가양주유소
184          세원주유소
242      (주)신용구도일주유소화랑로
243      (주)명연에너지효성주유소
297          형산석유(주)원주유소
302          양천구주유소
319          구민주유소
465      (주)삼표에너지정릉주유소
471          원천주유소
486      (주)에이치앤디이서울만남의광장주유소
Name: name, dtype: object

```

name_interest

```

64      (주)남경석유세영주유소
70      현대오일뱅크(주)직영명일셀프주유소
142          마곡대성주유소
151      (주)중앙에너지동교지점
279      현대오일뱅크(주)직영하계주유소
284      GS칼텍스(주)홍제동주유소
288          연화주유소
308      (주)한진김포공항제2주유소
327          구로주유소
329          한일주유소
381          난곡주유소
415          일원주유소
449          송파알찬주유소(셀프)
468      (주)우남에너지문화사랑주유소
484      현대오일뱅크(주)중랑교주유소
495      현대오일뱅크(주)직영양천현대셀프주유소
Name: name, dtype: object

```

name_serious

```

79          거여셀프주유소(셀프)
95          크로바주유소
156      (주)안국에너지삼보주유소
161          마포주유소
191      대성석유(주)종호주유소
286      sk삼미상사(주)서부주유소
287      현대오일뱅크(주)직영연세주유소
290      (주)가재울뉴타운주유소
385      현대오일뱅크(주)직영관악셀프주유소
440          대교주유소
Name: name, dtype: object

```