

AI 20기 1팀

## 1차 프로젝트 주제

자동차 등록 현황  
데이터 수집 및 기업 FAQ

AI Family 20기 1팀 머선일이조

김효빈 활수현 김지은 정소영 강민지

PLAY DATA

# Index



# 조사할 데이터

## 01 자동차등록현황(성별/연령별)



1 자료갱신일 : 2025-01-09

조사기간 : 2015 ~ 2024

출처 : 서울특별시, 자동차등록현황보고

## 02 자동차등록현황(지역별)



2 자료갱신일 : 2025-08

조사기간 : 2015~ 2024

출처 : 「국토교통통계연보」, 국토교통통계누리

## 03 자동차등록현황(연료별)



3 자료갱신일 : 2025-01-09

조사기간 : 2015.12 ~ 2024.12

출처 : 서울특별시, 자동차등록현황보고

## 04 기업 FAQ



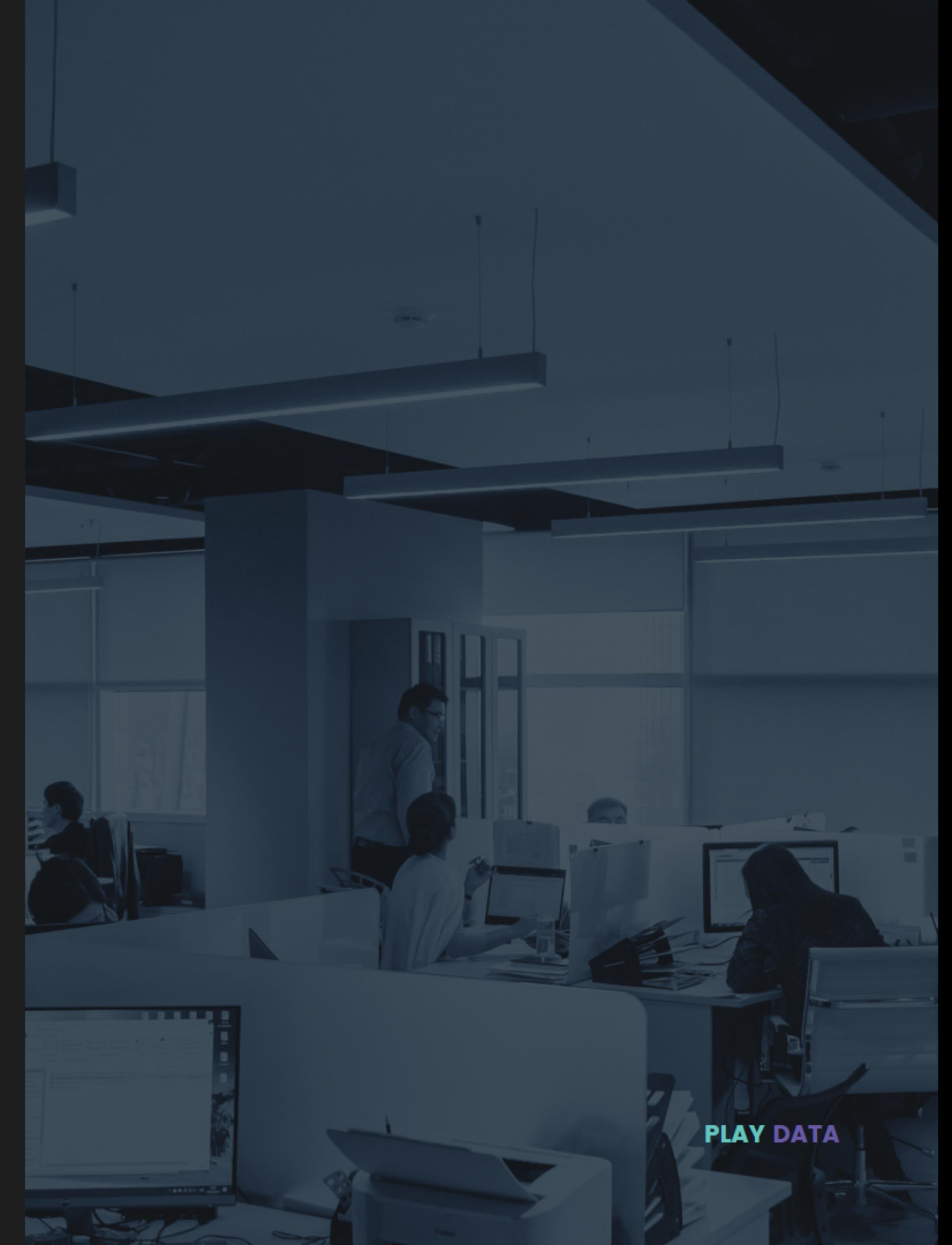
3 현대 :<https://www.hyundai.com/kr/ko/e>

기아 :<https://www.kia.com/kr>

쉐보레 :<https://www.chevrolet.co.kr/>

BUSINESS

# 1. 테이블 생성하기



# 테이블 설정

성별 / 연령별 자동차 등록 데이터 : gender\_age\_car

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
id	INT	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
year	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
gender	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
age_group	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
car_count	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

연료별 자동차 등록 데이터 : fuel\_car

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
id	INT	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
ym	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
car_type	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
fuel_type	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
car_count	INT	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					

지역별 자동차 등록 데이터 : car\_region

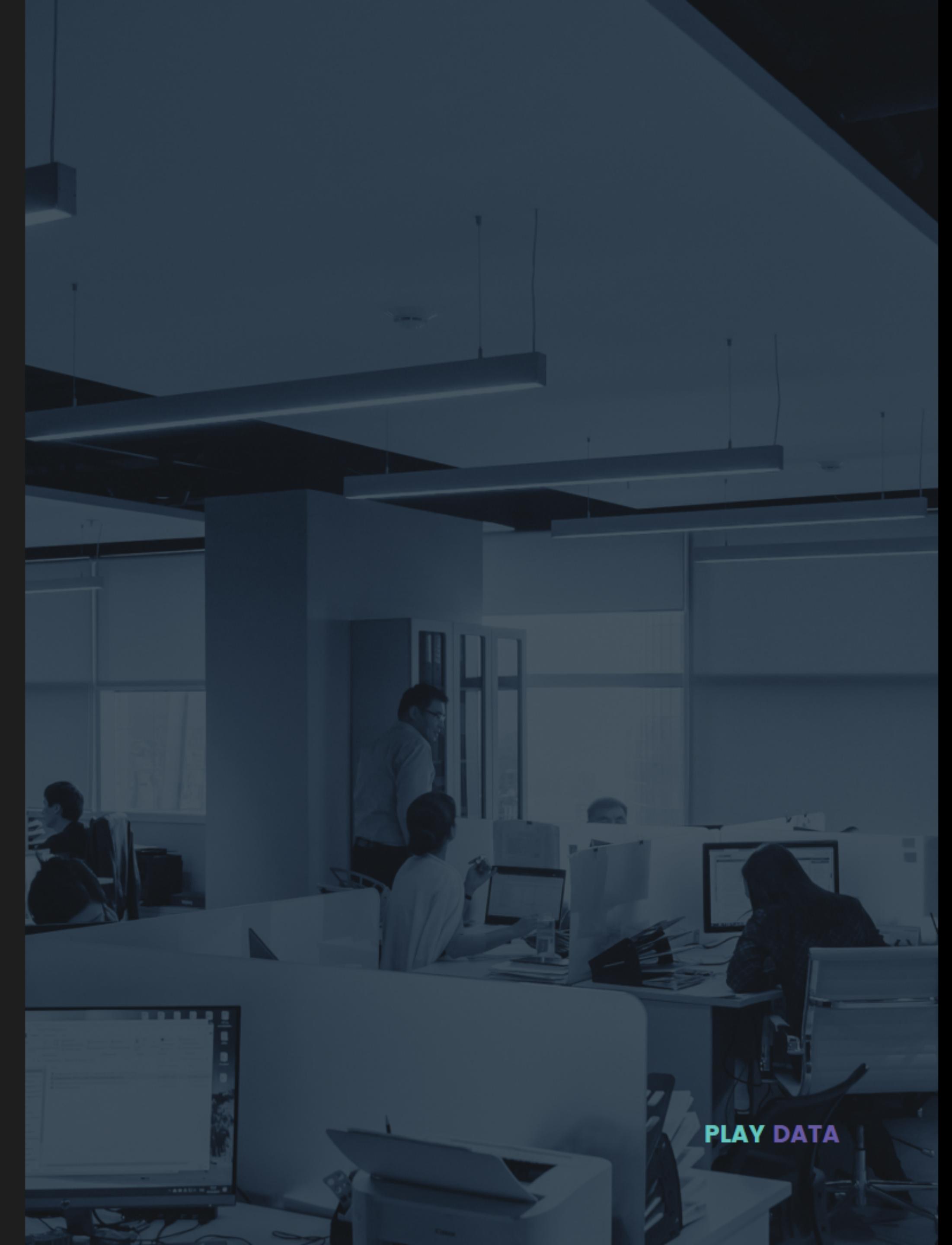
Column	Datatype	PK	NN	UQ	B...	UN	ZF	AI	G	Default / Expression
id	INT	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
ym	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
region	VARCHAR(5)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
passenger_to...	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
bus_total	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
truck_total	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
special_total	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
total_count	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
<click to edit>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

FAQ : faq

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
id	INT	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
question	VARCHAR(500)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
answer	LONGTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

BUSINESS

## 2. 데이터 수집 크롤링



# 웹 크롤링

## 01 환경준비

```
# pip install selenium webdriver-manager
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from webdriver_manager.chrome import ChromeDriverManager
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys # 엔터 키등을 입력하기 위해 사용
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support.ui import Select
from selenium.webdriver.support import expected_conditions as EC
from bs4 import BeautifulSoup
import time

# 웹드라이버를 자동으로 설치하고 최신버전을 유지
service = Service(ChromeDriverManager().install())
driver = webdriver.Chrome(service=service)

driver.get('https://stat.esouls.go.kr/statHtml/statHtml.do?orgId=201&tblId=DT_201')
driver.maximize_window()
print('브라우저가 성공적으로 열렸습니다.')
time.sleep(3)
```

## 02 브라우저 작동 자동화

```
# 시점 -> 시작 지점 -> 2015.12 -> 2024.12
# 시점 클릭
point_click = driver.find_element(By.ID, 'tabTimeText')
point_click.click()
time.sleep(2)

# 2015.01 클릭
# start_point_click = driver.find_element(By.XPATH, '//*[@id="timeM"]/h2/select[1]/option[109]')
# start_point_click.click()
# select 요소 찾기
select_element = driver.find_element("css selector", "select.box") # class="box"인 select
# select 객체 생성
select = Select(select_element)
# 방법 1: value 값으로 선택
select.select_by_value("201512") # 2015.12
time.sleep(3)

# 2024.12 클릭
y2020_m01_click = driver.find_element(By.XPATH, '//*[@id="timeM"]/h2/select[1]/option[120]')
y2020_m01_click.click()
time.sleep(3)

# 조회 v 클릭
y2025_m01_click = driver.find_element(By.XPATH, '//*[@id="tabTimeText"]/button[2]')
y2025_m01_click.click()
time.sleep(10)
```

### 03 데이터 추출

```
# 해당 정보를 담은 문자열 추출
soup = BeautifulSoup(driver.page_source, 'html.parser')
fuel_car_datas = soup.select('#mainTable > tbody > tr')

# 행 별로 저장
total_rows = []
for i, row in enumerate(fuel_car_datas[:5]): # 상위 5행만 처리 range(0,4)
    tds = row.select('td')
    total_rows.append([td.text.replace('\xa0', '') for td in tds])
```

### 04 데이터 전처리

```
car_types = ['승용차', '승합차', '화물차', '특수차']
fuel_list = ['휘발유', '경유', 'LPG', '전기', 'CNG', '하이브리드', '수소', '기타', '소계']

# 쿼리를 담을 리스트
queries = []

start_year = 2024
start_month = 12

for idx, row in enumerate(zip(*[data[3:] for data in total_rows[1:]])):
    # 총 몇 개 월이 지나갔는지 계산 (9개 단위로 1개월 감소)
    months_ago = idx // 9

    # 연도 계산
    year = start_year
    month = start_month - months_ago

    # month가 0 이하이면 연도 감소 + 월 12부터 다시 시작
    while month <= 0:
        month += 12
        year -= 1

    # ym 문자열
    ym = f'{year}{month:02d}'
```

### 05 MySQL DB 연결 후 데이터 추가

```
import pymysql
from dotenv import load_dotenv
import os

# .env 로드
load_dotenv()

# sautoreg_kr
def get_connection():
    return pymysql.connect(
        host = os.getenv('DB_HOST'),
        user = os.getenv('DB_USER'),
        password = os.getenv('DB_PASSWORD'),
        database= 'autoreg_kr'
    )
conn = get_connection()
cursor = conn.cursor()

# 모든 쿼리를 한 번에 실행

table_name = "fuel_car"
sql = f"INSERT INTO {table_name} (ym, car_type, fuel_type, car_count) VALUES (%s, %s, %s, %s)"
cursor.executemany(sql, queries)

# 커밋
conn.commit()

# 연결 종료
cursor.close()
conn.close()
```

# 엑셀 크롤링

## 01 환경준비

```
# pip install SQLAlchemy
import pandas as pd
from sqlalchemy import create_engine
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from webdriver_manager.chrome import ChromeDriverManager
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import Select
import pymysql
from dotenv import load_dotenv
import os
from sqlalchemy import create_engine

load_dotenv(".env")
# 2. 환경 변수 확인
DB_HOST = os.getenv('DB_HOST')
DB_USER = os.getenv('DB_USER')
DB_PASSWORD = os.getenv('DB_PASSWORD')
DB_NAME = os.getenv('DB_NAME')
```

## 02. 엑셀 파일 조회

```
# 1. 엑셀 파일 읽기
file_path = "data/자동차등록현황보고_자동차등록대수현황 시도별 (201501 ~ 201812).xlsx"
# 헤더 없이 처음부터 읽기
df = pd.read_excel(file_path,header=0)
```

## 03 엑셀파일 내부에 데이터 수집을 원하는 열 지정

```
# 컬럼 이름 직접 지정 (원하는 이름으로)
df.columns = ['ym', 'region', 'passenger_total', 'bus_total', 'truck_total', 'special_total', 'total_count']
# 필요한 열만 선택 (이미 지정했으므로 생략 가능)
df = df[['ym', 'region', 'passenger_total', 'bus_total', 'truck_total', 'special_total', 'total_count']]
print(df) # 데이터 확인
```

## 04 MySQL DB 연결 후 데이터 추가

```
def get_connection():
    return pymysql.connect(
        host=os.getenv('DB_HOST'),
        user=os.getenv('DB_USER'),
        password=os.getenv('DB_PASSWORD'),
        database='autoreg_kr',
        charset='utf8mb4'
    )

# 1. SQLAlchemy 엔진 생성
engine = create_engine(
    f"mysql+pymysql://{os.getenv('DB_USER')}:{os.getenv('DB_PASSWORD')}@{os.getenv('DB_HOST')}/autoreg_kr?charset=utf8mb4"
)

# 2. 테이블 이름 지정
table_name = "car_reg_region"

# 3. 데이터프레임을 DB로 저장
df.to_sql(
    name=table_name,
    con=engine,
    if_exists='append', # 기존 테이블 유지 후 추가
    index=False
)

print("데이터베이스 저장 완료!")
```

# 데이터 추가된 모습

성별 / 연령별 자동차 등록 데이터 : gender\_age\_car

	id	ym	region	passenger_total	bus_total	truck_total	special_total	total_count
▶	4081	2015-01	서울	2518714	149504	347568	5902	3021688
	4082	2015-01	제주	297785	20670	69755	786	388996
	4083	2015-01	경남	1173999	63720	274860	6144	1518723
	4084	2015-01	경북	907784	55167	292604	5850	1261405
	4085	2015-01	전남	581324	41747	218789	5777	847637
	4086	2015-01	충남	679795	46079	200462	3389	929725
	4087	2015-01	충북	522263	34041	143678	2970	702952
	4088	2015-01	강원	493577	33219	141290	2058	670144
	4089	2015-01	전북	595286	36725	176139	2701	810851
	4090	2015-01	세종	55459	3224	11738	182	70603
	4091	2015-01	울산	419138	17270	68958	2314	507680
	4092	2015-01	대전	509858	25901	85190	1756	622705
	4093	2015-01	광주	476844	24677	88511	1988	592020

▶ count(\*)  
▶ 190

지역별 자동차 등록 데이터 : car\_region

	id	ym	region	passenger_total	bus_total	truck_total	special_total	total_count
▶	4081	2015-01	서울	2518714	149504	347568	5902	3021688
	4082	2015-01	제주	297785	20670	69755	786	388996
	4083	2015-01	경남	1173999	63720	274860	6144	1518723
	4084	2015-01	경북	907784	55167	292604	5850	1261405
	4085	2015-01	전남	581324	41747	218789	5777	847637
	4086	2015-01	충남	679795	46079	200462	3389	929725
	4087	2015-01	충북	522263	34041	143678	2970	702952
	4088	2015-01	강원	493577	33219	141290	2058	670144

▶ count(\*)  
▶ 2040

# 데이터 추가된 모습

연료별 자동차 등록 데이터 : fuel\_car

	id	ym	car_type	fuel_type	car_count
▶	1	202412	승용차	휘발유	1638268
	2	202412	승합차	휘발유	1233
	3	202412	화물차	휘발유	3249
	4	202412	특수차	휘발유	37
	5	202412	승용차	경유	619097
	6	202412	승합차	경유	62546
	7	202412	화물차	경유	257311
	8	202412	특수차	경유	10246
	9	202412	승용차	LPG	180620

COUNT(\*)
3924

print(len(queries))
3924

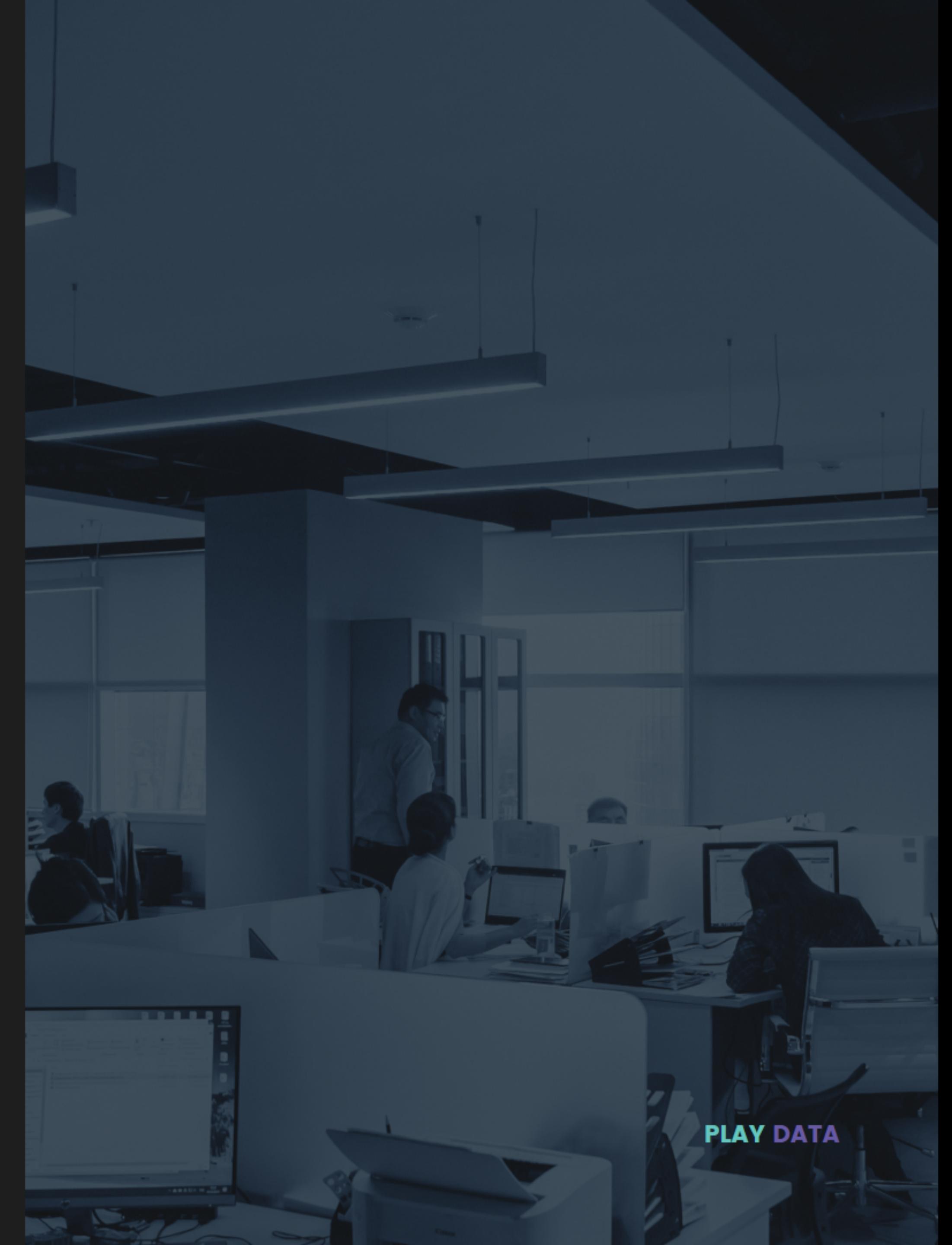
FAQ : faq

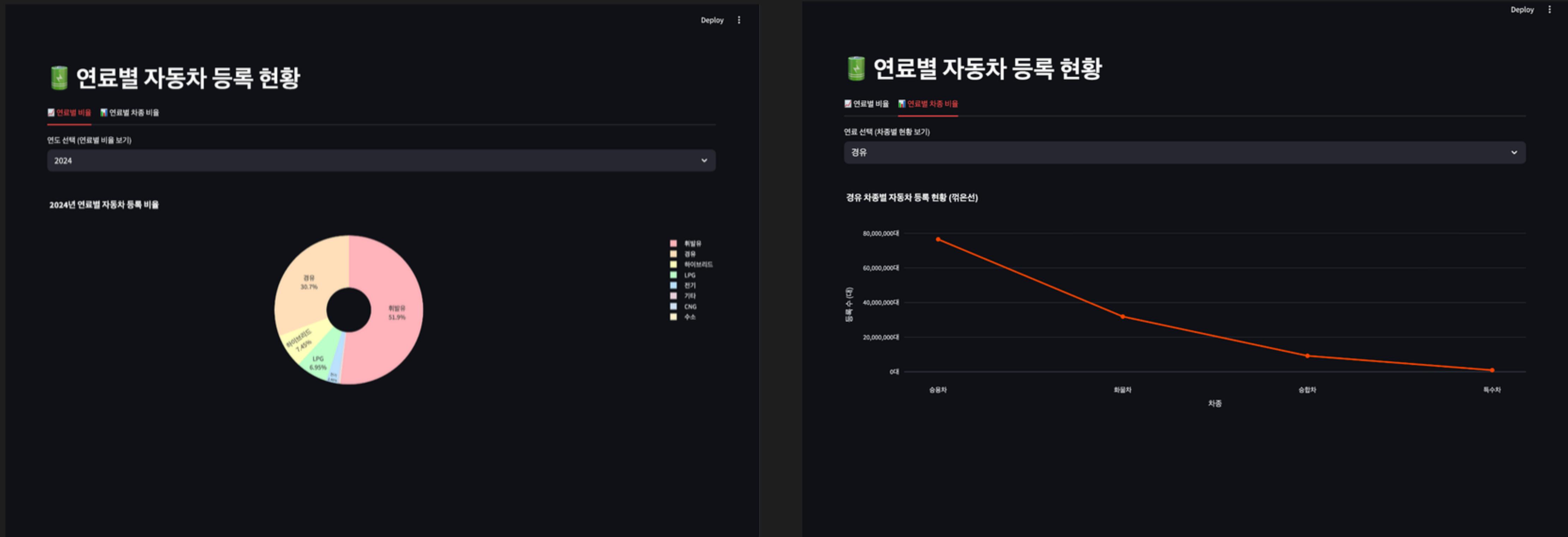
	id	question	answer
▶	59	에어컨 성능을 향상시키는 방법 안내	• 에어컨 작동은 1단부터 작동하기 보다는 최고단수부터 작동하는 것이 좋습니다. ■ 에어컨...
	60	직영 서비스센터 운영 시간은 어떻...	직영 서비스센터의 업무시간은 평일(월~금) 08:30~17:30까지이며, 중식시간은 12:30~...
	61	전자식 파킹 브레이크(EPB) 기능이...	전자 파킹 브레이크(EPB : Electronic Parking Brake)는 페달 또는 레버로 케이블을 당겨 ...
	62	버튼시동 스마트 키 차량의 경우 중...	버튼시동 스마트 키 차량의 경우 D/R/N 단 위치에서 시동을 끌 경우 ACC 상태가 되어 전원...
	63	스마트 키 배터리 교환방법에 대해 ...	스마트 키의 사용횟수와 기간에 따라 배터리 성능이 저하되었을 경우 배터리 교환이 필요합...
	64	내비게이션 업데이트는 어디서, 어...	내비게이션 업데이트는 기아멤버스 홈페이지에 접속하시어 이용하실 수 있습니다. 업데이트 ...
*	65	주행중 시동이 꺼지면 어떻게 해야...	시동꺼짐 현상은 그 원인이 다양하기 때문에 전문정비사를 통해 시동꺼짐 현상이 재현되어야...
		NULL	NULL

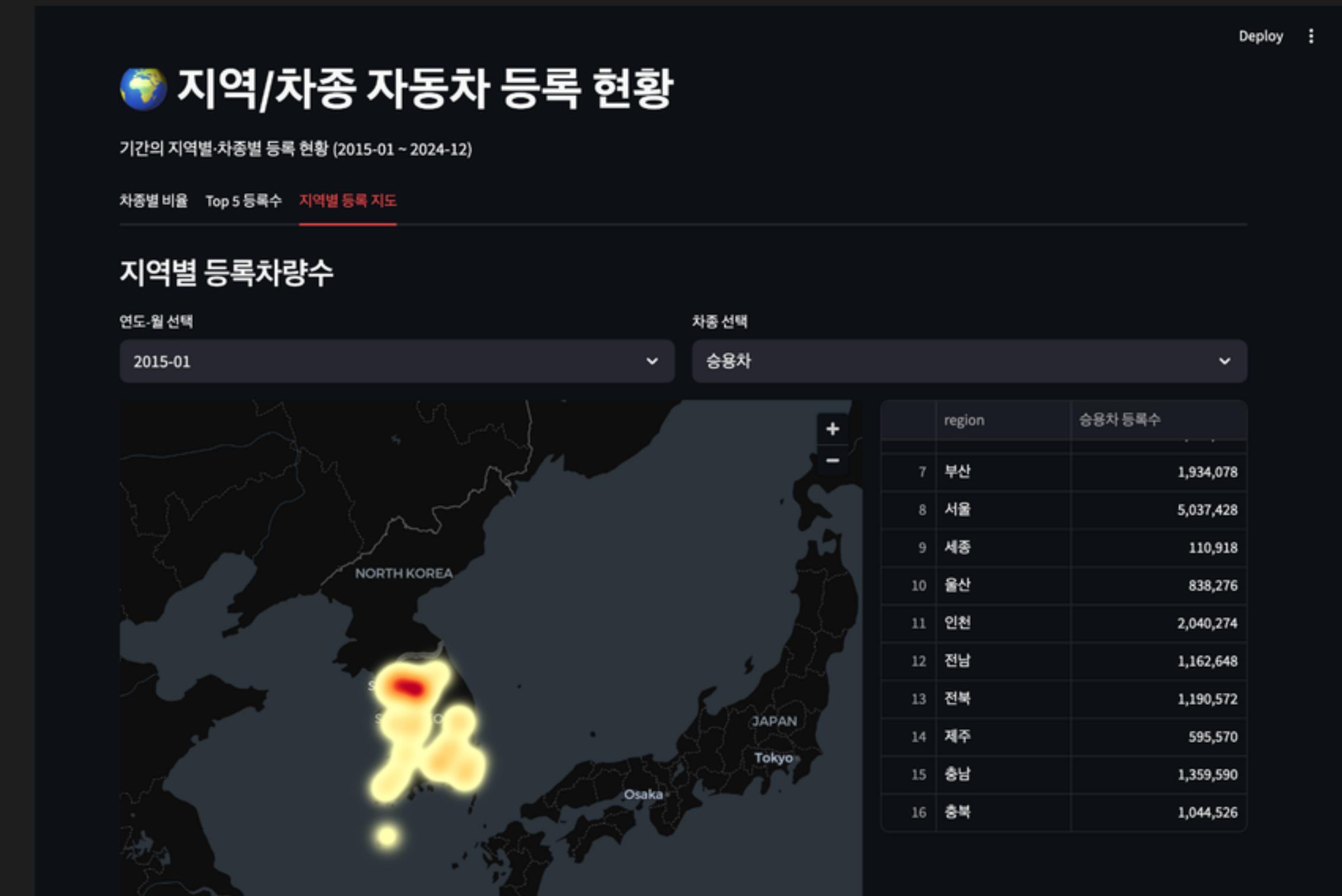
count(\*)
7

BUSINESS

# 3. Streamlit 구현하기







Deploy :

## FAQ

직영 서비스센터 운영 시간은 어떻게 됩니까?

전자식 파킹 브레이크(EPB) 기능이 무엇인가요?

버튼시동 스마트 키 차량의 경우 충립 주차는 어떻게 합니까?

스마트 키 배터리 교환방법에 대해 알려주세요

내비게이션 업데이트는 어디서, 어떻게 하나요?

주행중 시동이 끄지면 어떻게 해야하나요?

에어컨 성능을 향상시키는 방법 안내



# 머선일이조 맡은 역할

## 김효빈

성별 / 연령별 자동차  
데이터 크롤링  
Streamlit 최종 구현  
코드 오류 수정



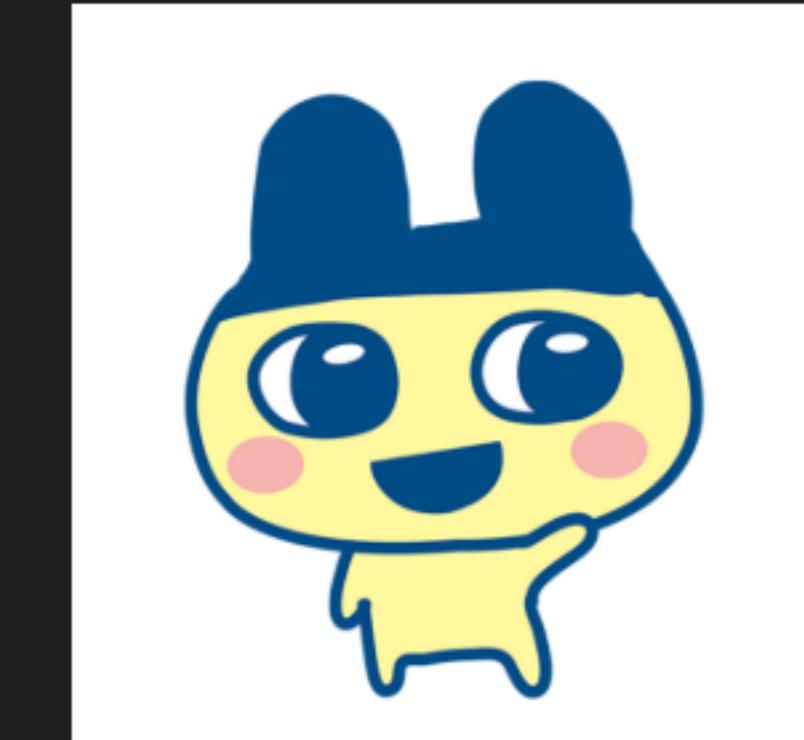
## 황수현

지역별 자동차 데이터 크롤링  
테이블 생성  
Git main 꾸미기  
Streamlit 구현



## 김지은

연료별 자동차 데이터 크롤링  
feat. 이규명 강사님  
PPT 제작



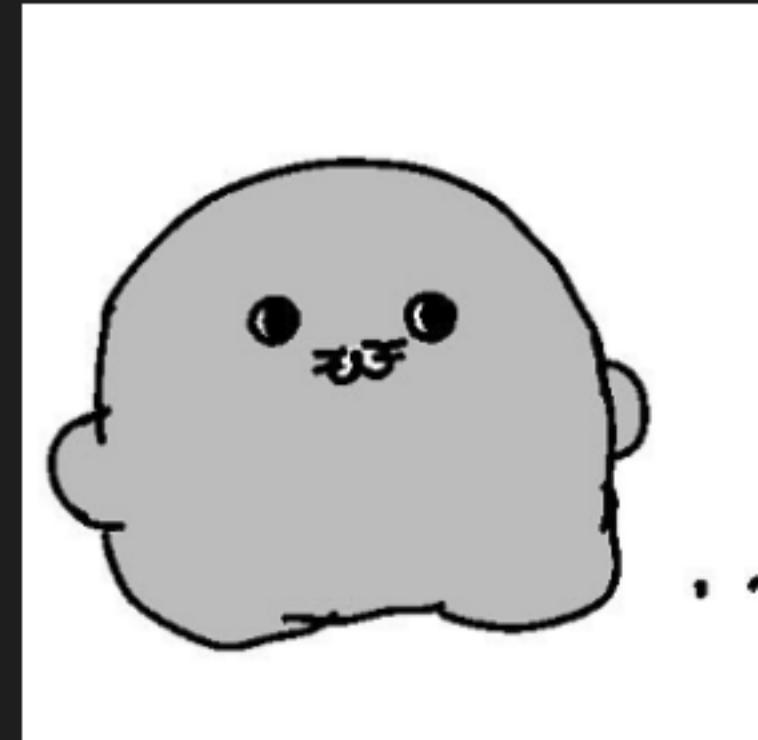
## 정소영

FAQ 데이터 크롤링 : 기아 쉐보레  
발표 / Git main 꾸미기  
FAQ 테이블 최종 수정  
Streamlit 구현



## 강민지

FAQ 데이터 크롤링 : 현대  
Streamlit 구현  
산출물 자료 제작  
Git main 꾸미기



발표를 들어주셔서 감사합니다.

Thank you