2024년 학교 자율교육과정(수업량 유연화)

빅데이터 분석을 통한 전기차 수요와 인프라의 상관관계 분석

요약 설명

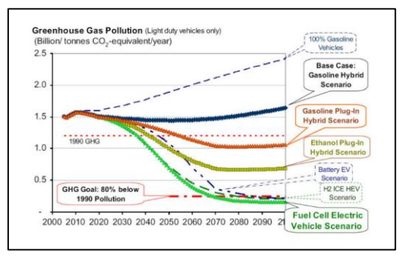
현재 점점 더 심화되어가는 지구온난화의 주범으로 꼽히는 온실가스의 배출을 줄이기 위해 전인류적으로 노력하고 있다. 이산화탄소는 현재 온실가스의 약 88.6%를 차지하는 대표적인 지구온난화의 주범이라고 할 수 있다. 이러한 이산화탄소의 배출은 약 40%는 화석연료 기반의 승용차, 약 34%는 화물차량에서 발생하여 전체의 74%의 높은 비중을 차지하고 있다. 따라서 현재 정책적으로도 전기차 산업을 육성하려 노력하고 있지만, 대표적으로 지원금을 주는 형태의 지원과 현재의 전기차 인프라가 전기차 구매의 상관관계는 어떤 경향을 보이는지 분석하고자 한다.

창원과학고

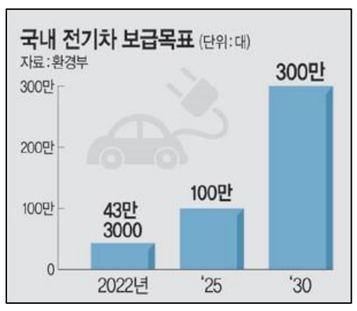
2219 한 OO
2310 이 OO
2403 김 OO

연구 내용

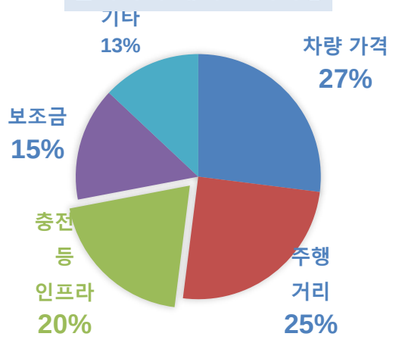
1. 주제선정 및 목적



구동방식에 따른 C02배출



전기차 구매시 고려사항



전기차 구매 장려를 위해 인프라 구축이 필수적이다.

2. 연구 내용 및 방법

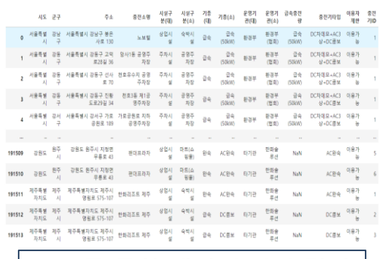
데이터 전처리 과정

DATA 공공데이터포털 .GO.KR

공공데이터 포털 한국환경공단_환경부 전기차 충전기 충전량 자료 수집

```
pip install pandas
pip install geokakao

import pandas as pd
import geokakao as gk

# Address to coordinates
df = pd.read_csv("data/EV.csv", encoding = "cp949")
Gyeongsangnadostation = df[df["시도"].str.contains("경상남도")]
Gyeongsangnadostation = Gyeongsangnadostation.reset_index(drop = True)

gk.add_coordinates_to_dataframe(Gyeongsangnadostation, "시도")
Gyeongsangnadostation.to_csv("Gyeongsangnadostation_coord.csv", index =
# Runtime of the code: 60 min

# Dataframe creation & edit
df_coord = pd.read_csv("Gyeongsangnadostation_coord.csv")
df
```

데이터 전처리를 위한 코드 작성



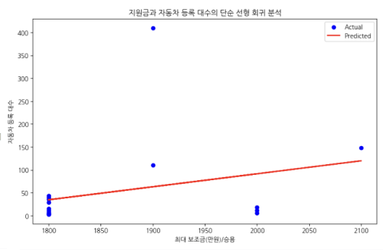
코드를 통한 데이터 인식 결과

3. 연구 결과

경남 지역 보조금과 접수대수의 상관관계 분석

```
pip install pandas
pip install matplotlib
pip install korean-lunar-calendar

import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import font_manager as fm
import seaborn as sns

df = pd.read_csv("data/money.csv", encoding = "cp949")
gyeongnam_data = df[df['시도'] == '경남']
correlation = gyeongnam_data['최대 보조금(만원)/승용'].corr(gyeongnam_data['접수대수'])
print(f"경남 지역 보조금과 접수대수의 상관계수: {correlation}")
plt.rcParams['font.family'] = 'NanumGothic'

plt.figure(figsize=(10, 6))
sns.scatterplot(x='최대 보조금(만원)/승용', y='접수대수', data=gyeongnam_data)
plt.title('경남 지역 보조금과 수요의 상관관계')
plt.xlabel('최대 보조금(만원)/승용')
plt.ylabel('접수대수')
plt.show()
```

상관계수와 그래프 출력



코드를 통해 출력한 그래프

경남 지역 보조금과 접수대수의 상관계수: 0.2427055710

경남 지역 보조금과 접수대수의 결정계수: 0.0795963207

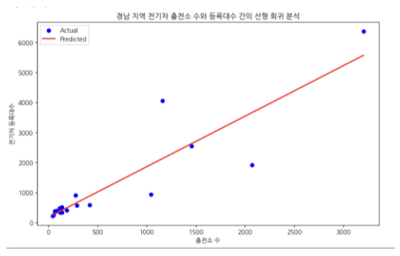
결정계수가 1과 떨어짐

보조금과 전기차 수요 사이, 낮은 상관관계가 있는것으로 판단된다.

충전소 수와 전기차 수요의 상관관계분석

```
pip install pandas
pip install matplotlib
pip install korean-lunar-calendar

import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import font_manager as fm
import seaborn as sns

mf = pd.read_csv("data/money.csv", encoding = "cp949")
ef = pd.read_csv("data/EV.csv", encoding = "cp949")
gyeongnam_edata = ef[ef['시도'] == '경남']
demand_data = mf[mf['시도'] == '경남']
charging_counts = gyeongnam_edata['군구'].value_counts().reset_index()
charging_counts.columns = ['군구','충전소 수']
demand_sums = demand_data.groupby('군구')['접수대수'].sum().reset_index()
print(demand_sums)

merged_data = pd.merge(charging_counts,demand_sums, on='군구')
print(merged_data)

correlation = merged_data['충전소 수'].corr(merged_data['접수대수'])
print(f"충전소 수와 전기차 수요의 상관계수: {correlation}")

plt.rcParams['font.family'] = 'NanumGothic'
plt.figure(figsize=(10, 6))
sns.scatterplot(x='충전소 수', y='접수대수', data=merged_data)
plt.title('충전소 수와 전기차 수요의 상관관계')
plt.xlabel('충전소 수')
plt.ylabel('전기차 수요')
plt.show()
```

상관계수와 그래프 출력



코드를 통해 출력한 그래프

충전소 수와 전기차 수요의 상관계수: 1.6803283457

충전소 수와 전기차 수요의 상관계수: 0.8089353943

결정계수가 1과 가까움

충전소 수와 전기차 수요 사이, 높은 상관관계가 있는것으로 판단된다.

빅데이터의 시각화

```
import pandas as pd
import numpy as np
import folium
import json

# EV 데이터 로드
df = pd.read_csv('data/EV.csv', encoding='cp949')

# 경상남도에 해당하는 데이터 필터링
Gyeongsangnamdo_station = df[df['시도'].str.contains('경상남도')]
Gyeongsangnamdo_station = Gyeongsangnamdo_station.reset_index(drop=True)

# 이미 좌표 추가 (실제 데이터가 없으면 이 데이터로 대체)
Gyeongsangnamdo_station['위도'] = np.random.uniform(low=34.5, high=35.7, size=len(Gyeongsangnamdo_station))
Gyeongsangnamdo_station['경도'] = np.random.uniform(low=127.5, high=129.5, size=len(Gyeongsangnamdo_station))

# 업데이트된 데이터셋을 CSV로 저장
Gyeongsangnamdo_station.to_csv('data/Gyeongsangnamdo_station_coord.csv', index=False, encoding='utf-8')

# 업데이트된 데이터셋 로드
df_coord = pd.read_csv('data/Gyeongsangnamdo_station_coord.csv')

# 경상남도 지역의 GeoJSON 데이터 로드
geo_path = 'data/gngeo.geojson'
with open(geo_path, encoding='utf-8') as f:
    geo_str = json.load(f)

# 군구별 충전소 수 집계
num_gu = df_coord.groupby('군구').size().reset_index(name='충전소수')
```

시각화할 데이터 처리 코드

```
# 데이터 값의 범위 확인
num_gu_min = num_gu['충전소수'].min()
num_gu_max = num_gu['충전소수'].max()
print(num_gu_min, num_gu_max)

# 초기 지도 설정
loc = [35.2378, 128.6917]
m = folium.Map(location=loc, zoom_start=10)

# 적절한 threshold_scale 설정
threshold_scale = [num_gu_min, 10, 50, 100, 500, 1000, 1500, num_gu_max]

# Choropleth 레이어 추가
folium.Choropleth(
    geo_data=geo_str,
    data=num_gu,
    columns=["군구", "충전소수"],
    fill_color="PuRd",
    threshold_scale=threshold_scale,  # 데이터 범위에 따라 색상 조정
    key_on="feature.properties.sggnm",  # 올바른 키 설정
    legend_name="충전소 개수"
).add_to(m)
```

지도데이터 표시 코드

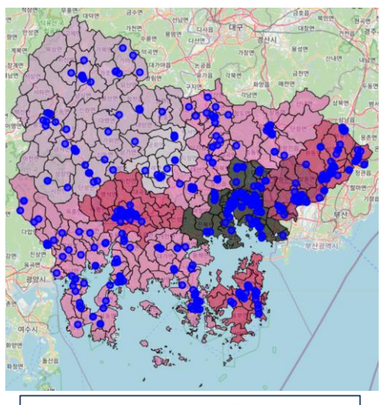
```
# 충전소 타입에 따라 다른 색상 설정하는 함수
def color_select(charger_type):
    return 'red' if charger_type == '급속' else 'blue'

csv_path = 'data/EVcenter.csv'
try:
    df_csv = pd.read_csv(csv_path, encoding='cp949')  # 인코딩을 cp949로 설정

    # 지도에 모든 충전소 위치 추가
    for index, row in df_csv.iterrows():
        folium.CircleMarker(
            location=[row["위도"], row["경도"]],  # 충전소 위치
            radius=5,
            color=color_select(row["충전기타입"]),  # 충전기 타입에 따른 색상 선택
            fill=True,
            fill_color=color_select(row["충전기타입"]),  # 채우기 색상 설정
            fill_opacity=0.6,
            popup=folium.Popup(row["충전소명"], parse_html=True)  # 팝업에 충전소 이름 표시
        ).add_to(m)

    # HTML 파일로 지도 저장
    m.save('Gyeongsangnamdo_EV_Stations.html')
```

충전소 위치 표시 코드

```
{
  "type": "Feature",
  "properties": {
    "OBJECTID": 3091,
    "adm_nm": "경상남도 진주시 문산읍",
    "adm_cd": "3803011",
    "adm_cd2": "4817025000",
    "sgg": "48170",
    "sido": "48",
    "sidonm": "경상남도",
    "sggnm": "진주시"
  },
  "geometry": {
    "type": "MultiPolygon",
    "coordinates": [
      [
        [
          [
            128.2473736540518
            35.12846902841578
          ],
          [
            128.2466943036763
            35.1272552269718
          ],
```

geojson 파일 형식 로드

- car
- EV
- EVcar
- EVcenter
- EVcenter.json
- gngeo.geojson



충전소 데이터 시각화 결과

특정 동에 모인 충전소

↓

시 전체 충전소로 분석됨

↓

데이터 수집 단위가 시

↓

구역 단위 세분화 필요!

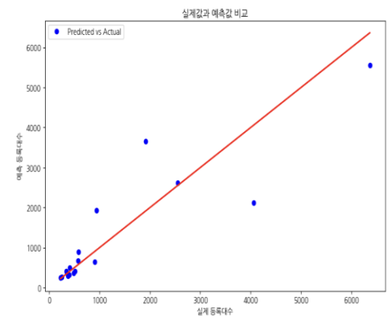
보조금 데이터 시각화 결과

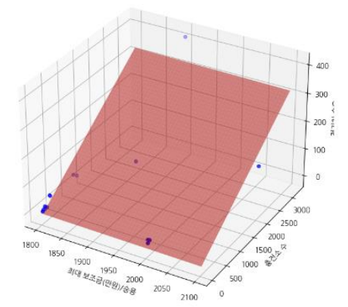
다중선형회귀 활용 데이터 분석

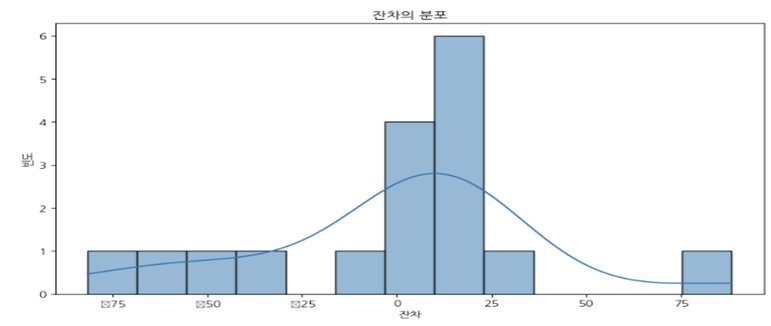
```
# 회귀 분석을 위한 데이터 준비
X = merged_data[['최대 보조금(만원)/승용', '충전소 수']]
y = merged_data['접수대수']

# 데이터 형식 확인
print("\nX (independent variables):")
print(X.head())
print("\ny (dependent variable):")
print(y.head())

# 선형 회귀 모델 생성 및 학습
model = LinearRegression()
model.fit(X, y)

# 예측
y_pred = model.predict(X)

# 회귀 결과 출력
print(f"회귀계수(기울기): {model.coef_}")
print(f"절편: {model.intercept_}")
print(f"결정계수(R^2): {r2_score(y, y_pred)}")

# 시각화
plt.rcParams['font.family'] = 'NanumGothic'
plt.figure(figsize=(10, 6))

# 예측값과 실제값 비교
plt.scatter(y, y_pred, color='blue', label='Predicted vs Actual')
plt.plot([y.min(), y.max()], [y.min(), y.max()], color='red', lw=2)
plt.title('실제값과 예측값 비교')
plt.xlabel('실제 접수대수')
plt.ylabel('예측 접수대수')
plt.legend()
plt.show()
```

```
# 잔차 분석
residuals = y - y_pred
plt.figure(figsize=(10, 6))
sns.histplot(residuals, kde=True)
plt.title('잔차의 분포')
plt.xlabel('잔차')
plt.ylabel('빈도')
plt.show()

# 3D 그래프 시각화
fig = plt.figure(figsize=(12, 8))
ax = fig.add_subplot(111, projection='3d')

# 산점도
ax.scatter(merged_data['최대 보조금(만원)/승용'], merged_data['충전소 수'], y, color='blue'

# 회귀 평면
x_surf, y_surf = np.meshgrid(np.linspace(X.iloc[:, 0].min(), X.iloc[:, 0].max(), 100),
                             np.linspace(X.iloc[:, 1].min(), X.iloc[:, 1].max(), 100))
z_surf = model.intercept_ + model.coef_[0] * x_surf + model.coef_[1] * y_surf

ax.plot_surface(x_surf, y_surf, z_surf, color='red', alpha=0.5)

ax.set_title('최대 보조금, 충전소 수와 전기차 수요의 회귀 분석')
ax.set_xlabel('최대 보조금(만원)/승용')
ax.set_ylabel('충전소 수')
ax.set_zlabel('전기차 수요')

plt.legend()
plt.show()

except FileNotFoundError as e:
    print("파일을 찾
    print(e)
except Exception as e
    print("데이터를
    print(e)
```

다중선형회귀 출력 코드







코드를 통해 출력한 그래프

4. 결론 및 적용

지역별 보조금과 전기차 수요 사이 상관관계는 없는것으로 보인다.

지역별 충전소 수와 전기차 수요 사이 유의미한 상관관계가 있다고 보인다.

빅데이터 시각화 결과 데이터의 정확성을 위해 지역 세분화가 필요하다.

다중선형회귀를 통해 지원금 대비 충전소에 의한 영향이 크다고 보인다.

효과 및 전망

경상남도에 설치되어있는 충전소와 전기차 수요의 상관관계를 조사해본 결과 상관계수가 약 0.9로 1에 가까운 값이 나와 높은 상관관계를 보인다 판단할 수 있었고, 경남 지역의 보조금과 전기차 수요의 상관관계를 조사해본 결과 결정계수가 약 0.24로 1과는 다소 떨어진 값이 나오게 되어 현재 가지고 있는 자료만으로 지자체의 지원금이 전기차 수요에 유의미한 영향을 주는것으로 판단을 내리기엔 다소 어려움이 존재한다고 판단된다. 처음 조사했던 설문조사의 결과 전기차 구매를 꺼리는 이유의 35.7%를 차지하는 충전 인프라 부족의 문제가 빅데이터 분석결과 충전 인프라가 전기차의 수요에 높은 상관관계를 보인다는 것을 판단할 수 있었다. 그러나 설문조사 결과 전기차 구매를 하게된 이유의 17%를 차지해 4번째 요인이였던 보조금의 경우 현재의 경상남도 자료로써는 보조금이 전기차의 수요에 유의미한 영향을 주었다고 영향을 내리기엔 무리가 있다고 판단이 된다. 그러나 표본을 전국으로 확대하여 빅데이터 분석을 진행한다면 보조금과 전기차 수요의 상관관계를 보다 정밀하게 밝혀낼 수 있을것으로 판단된다.