|  |
| --- |
| **제10회「2022 빅콘테스트」데이터 분석 계획서** |

|  |  |  |  |
| --- | --- | --- | --- |
|  | \* 해당란에 ☑표시 | | |
| **참가리그** | ☑데이터분석리그 | | |
| **세부분야** | □ 이노베이션분야 ☑ 데이터분석분야 | | |
| **세부부문**  \*해당시 체크 | □ 루키부문 □ 퓨처스부문 ☑ 챔피언부문  \*데이터분석분야에 한함(선택) | | |
| **개인/팀여부** | □ 개인 ☑ 팀(총 4명) | **개인/팀명** | 고수찜닭 |
| **지도교사명** |  | | |
| **대표ID** | joel0839@naver.com | | |

※ **5장 내외로** 목차는 준수하여 자유롭게 작성

|  |  |
| --- | --- |
| **분석 주제명** | 용인시 전기차 완속 충전소 입지 선정 모델 |
| **분석 배경** | 가파르게 증가하는 전기자동차 수요에 비해 전기차 충전소 보급률은 한없이 부족하다. 우리나라의 전기차 충전소는 전기차 2.4대당, 1개만큼 설치되어 있어 수치상 여유가 있다고 느껴지지만 실상은 그렇지 못하다. 이용자들이 원할 때 이용할 가까운 충전소가 없거나, 대기 시간이 길어 원하는 시간대에 충전하지 못하는 등, 전기차 사용자들은 여전히 많은 불편함을 겪고 있다. 늘어나는 전기차 이용자 및 예비 이용자들의 불편함을 해결하고, 전기차 보급률을 높이기 위해서는 보다 적당한 장소에 더 많은 충전소를 설치할 필요가 있다. 이에 환경부에서는 2025년까지 급속 충전소는 1만2천기, 완속 충전소는 50만기 설치를 목표로 하고 있다.    자료 출처)  https://www.mk.co.kr/news/society/view/2022/04/364109/  이번에 제시하는 입지 선정 모델은 가장 많은 설치를 진행할 완속 충전소를 대상으로 한다. 완속 충전소의 경우 완충까지 5시간 이상이 필요해 거주자와 활동자에 적합하다. 때문에 이번 분석의 타겟은 용인시 거주자와 직장에 약 7시간 가량 머물러 있는 사람들의 수요로 두고 진행하였다. |
| **분석 내용**  **요약** | 1. 수요의 경우 거주자, 활동자의 이용을 모두 고려한다.  2. 공급의 경우 설치가 예정되어 있는 아파트 주차장, 설치 되어있는 충전소 위치를 이용한다.  3. EDA 결과 이용자 개인의 특성은 설치에 영향을 주지 않는다고 판단하였다.  4. 클러스터링을 통해 수요를 묶은 후 지표에 따라 후보군을 최대한 선정, 이후 선형 계획법을 진행 후 수익성과 타당성을 판단하고자 한다. |
| **분석방법 및**  **계획** | 1. 추가 데이터  용인시 아파트 정보 목록(용인시 사이트)  용인시 전기 충전소 정보(공공데이터 포탈)  공공 주차장 데이터(용인시 공공데이터)  읍면동별 세대 및 인구(통계청)  2. 통계·분석 기법, 방법론  1) 선형계획법(LP)   1. 최대화 문제  * Maximize (최대화) * Subject to (제약 조건식)  1. 최소화 문제   - Minimize (최소화)  - Subject to (제약 조건식)  주어진 용인시 제약 조건으로 예산과 수익성 등을 고려하여 설치 가능한 충전소 최대 개수를 구한다. 구한 추가 설치 충전소의 수를 이용하여 최대지역 커버문제를 해결한다. 충전소 위치 후보군 중 일정 거리 내 포인트가 커버된다면 최종 후보지가 된다.    2) K-means Clustering  k-means Clustering 알고리즘은 클러스터링 방법 중 분할법에 속한다. 분할법은 주어진 데이터를 여러 파티션으로 나누는 방법이다. 예를 들어 n개의 데이터 오브젝트를 입력 받았다고 가정하면, 이 때 분할법은 입력 데이터를 n보다 작거나 같은 k개의 그룹으로 나누는데, 각 그룹은 클러스터를 형성한다. 다시 말해, 데이터를 한 개 이상의 데이터 오브젝트로 구성된 k개의 그룹으로 나누는 것이다.    자료 출처)  https://ko.wikipedia.org/wiki/K-%ED%8F%89%EA%B7%A0\_%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98  https://www.linkedin.com/pulse/finding-optimal-number-clusters-k-means-through-elbow-asanka-perera  3) Greedy 알고리즘  탐욕 알고리즘은 말 그대로 선택의 순간마다 당장 눈앞에 보이는 최적의 상황만을 쫓아 최종적인 해답에 도달하는 방법이다. 탐욕 알고리즘은 여러 경우 중 하나를 결정해야 할 때마다 그 순간에 최적이라고 생각되는 것을 선택해 나가는 방식으로 진행하여 최종적인 해답에 도달한다.  3-1) 집합 커버 문제(Set cover problem)  전체 집합 U와, 그 부분 집합 Si들을 원소로 갖는 집합 F가 있다고 하자.  F의 원소인 Si들을 합집합하여 전체 집합 U와 같아지는 집합 Si들의 최소 개수를 찾는 문제이다. F에 있는 집합들의 모든 조합을 하나씩 찾을 경우 만약 F의 원소의 개수가 n개이면 나올 수 있는 경우의 수는 (2n-1)개이고, n이 증가함에 따라 이 방식으로 최적해를 찾는 것은 실질적으로 불가능하다.  따라서 최적해 대신 최적해에 근접한 근사해(Approximation Solution)을 찾아야 한다. U에 있는 원소를 가장 많이 커버하는 Si를 반복적으로 찾아서 포함시키는 방식을 사용한다.   * 수도 코드   SetCover(Set u, Set f)  {  Set cover;  while (u.size() != 0)  {  u의 원소들을 가장 많이 포함하는 집합 s\_i를 f에서 선택한다.  u에서 s\_i를 뺀다. (차집합)  s\_i를 cover에 추가한다.  s\_i를 f에서 제거한다.  }  return cover;  }    자료 출처)  https://caniro.tistory.com/242  3. 방법  1) 수요: 빅콘테스트에서 제공한 ev\_app\_resident, ev\_app\_activit 데이터내 count\_cust column, 미래 전기차 수요치와 한 대당 전기차 충전 수요 값을 활용한 충전 수요 보정치  2) 공급: 설치가 예정되어 있는 아파트 주차장, 설치되어 있는 충전소  3) 선형계획법을 통해 주어진 용인시 제약 조건으로 설치 가능한 충전소 최대 개수를 구한다.  4) 공급과 예상 수요가 발생하는 위치(경도와 위도) 지도 위에 표시하여 밀집도를 파악한다.  5) 공급이 커버하지 못하는 예상수요를 타겟으로 하고 해당 위치에 대한 데이터 프레임을 생성한다.  6) 타겟에 대해 clustering을 진행한다.  7) 각 cluster 내 예상 수요를 집합 커버 문제로 정의한다. (초기에 주어진 수요의 위치는 50m\*50m 격자 중심이다. 이것을 각 노드로 가정하고, 예상 수요를 노드 내 value로 가정한다.)  8) 집합 커버 문제에 따라 최대 예상 수요를 커버하는 지점에 충전소 위치를 결정한다.  9) 모델을 통해 나온 결과는 예상 수요, 즉 사업성 만을 판단해 선정된 위치이다. 추가 지수를 활용해 사업성 이외의 타당성, 공익을 기준으로 평가와 필터링을 진행한다.  추가 지수 예시) 접근성  4. 시각화  기존 충전소 위치와 모델링을 통해 구한 후보군 위치, 최종 선정 입지를 leaflet.js 라이브러리를 활용해 지도 위에 그려 웹 서비스로 제공할 예정이다. |
| **분석결과 활용 및 시사점** | 1. 활용 방안  전기차 이용자 수요를 바탕으로 최적화한 모델을 제시하여 향후 충전소 설치 계획에 참고할 후보군 선정에 도움이 될 것으로 기대된다.  용인시 인접 도시의 데이터를 이용하면 보다 적합한 모델링이 가능할 것으로 예상된다.  2. 적용대상  용인시를 비롯한 타 도시의 전기차 입지 선정 모델  3. 결과 적용시 기대효과 및 시사점  충전소 최적 입지 선정 이후 친환경 자동차의 이용 증가에 도움이 될 수 있다.  용인시민 생활환경 향상 도모와 전기자동차 산업 성장촉진에 기여하여 친환경 전기자동차 및 스마트그리드 연계 녹색 인프라를 구축할 수 있다.  사업적 기준(월평균 이용 횟수 15회 이상, 총 이용시간 30000분 이상 예측)을 달성하여 수익이 생기도록 하겠다. |

※ 제출자료는 평가에 반영 예정