

1. 분석 개요

본 분석은 공공데이터포털의 부산광역시 버스정류장 위치 데이터를 활용하여, 부산 시내 전체 정류장 정보를 수집하고 이를 전처리한 뒤, 공간기반 클러스터링을 통해 주요 정류장 밀집지역을 시각적으로 파악하고자 한다. 이는 향후 특정 지역 기반 교통 인프라 분석, 혼잡도 예측, 이벤트 최적 위치 선정 등에 활용할 수 있는 기초 데이터 분석이다.

2. 데이터 수집 및 전처리

	bstopid	bstopnm	lon	lat
0	167970102	영주삼거리	129.033322	35.115356
1	169310303	영주삼거리	129.033030	35.115283
2	167970301	시민아파트	129.031749	35.115140
3	167840102	시민아파트	129.032160	35.114921
4	167970302	중앙공원.민주공원입구	129.029762	35.114487

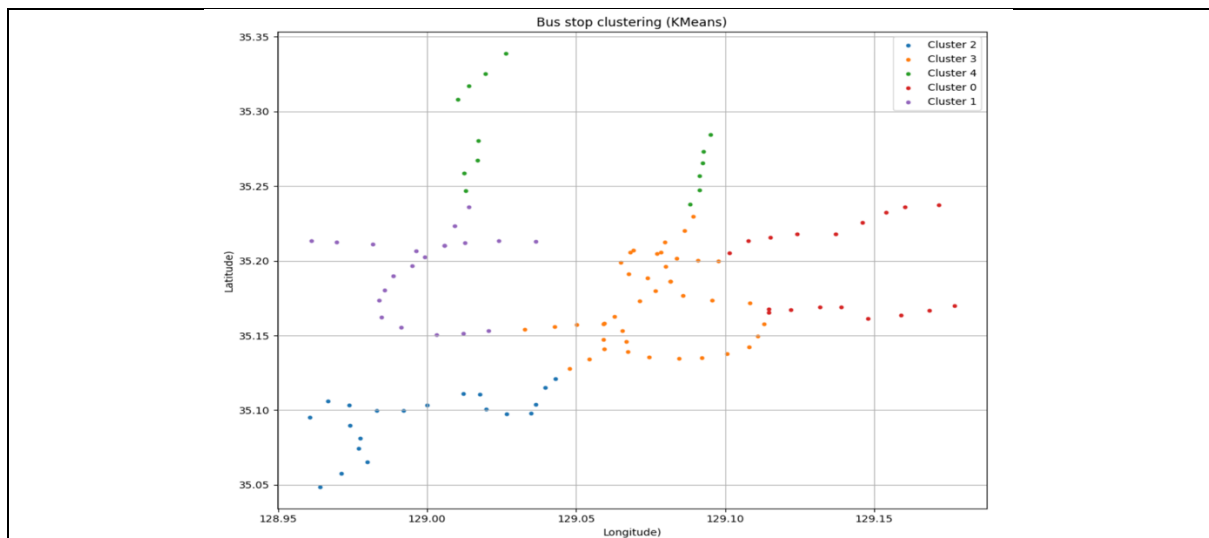
본 프로젝트에서는 부산광역시 내 전체 버스 정류장에 대한 위치 정보를 확보하기 위해 공공데이터포털에서 제공하는 부산BIMS버스정류장정보 API를 활용하였다. 데이터 수집은 Python의 Requests 라이브러리를 이용해 API를 호출하고, XML 포맷으로 응답되는 데이터를 파싱하는 방식으로 진행하였다.

수집된 데이터에서 고유 식별자 bstopid, 정류장 이름 bstopnm, 정류장 위치를 나타내는 경도와 위도를 추출하고, 이 중 경도와 위도는 문자열 형태로 제공되었기 때문에, 이를 지도 시각화 및 클러스터링 분석에 활용할 수 있도록 float타입으로 변환하는 전처리를 수행하였다.

이후, 데이터 정합성을 확보하기 위해 결측치가 포함된 행을 제거하고, 정류장 이름과 좌표가 동일한 중복 데이터를 제거하였다. 중복 제거는 버스 정류장 이름과 위도, 경도 의 조합을 기준으로 수행되었으며, 실제로 동일한 이름의 정류장이 도로 양방향에 존재하거나, GPS 오차로 인해 동일 정류장이 중복 표기된 경우가 다수 확인되었다.

이러한 전처리 과정을 거쳐 생성된 데이터는 busan_cleaned_bus_stops.csv파일로 저장되었고 최종적으로 확보된 정류장 수는 3000개로, 이는 분석 시점 기준으로 부산광역시 내 운영 중인 정류장의 대부분을 포함한다.

3. 클러스터링 분석



부산 전체 버스 정류장의 위도(lat)와 경도(lon)를 이용해 KMeans 알고리즘으로 공간적 군집화를 수행하였다. 클러스터 수는 시각적으로 확인하기 위해 5 개로 설정하였으며, 추후 Elbow Method 등을 통해 최적화할 수 있다. 분석 목적은 서로 가까운 위치의 정류장을 묶어 교통 밀집 지역을 파악하고, 지역별 특성을 비교하여 향후 노선 재설계나 수요 예측에 활용하기 위함이다. 결과적으로 각 정류장은 5 개의 클러스터 중 하나로 분류되었으며, 지도 시각화를 통해 군집 간 공간적 차이를 확인할 수 있었다.

4. 시각화 (벡스코 반경 정류장 기준)



클러스터링 결과는 folium 라이브러리를 활용해 지도 위에 시각화하였다. 각 정류장은 클러스터별로 색상이 지정된 원형 마커(CircleMarker)로 표시되며, 지도 중심은 부산 전체 정류장의 평균 좌표로 설정하고 zoom_start=12 를 적용하여 전체 분포를 한눈에 확인할 수 있도록 구성하였다. 해당 결과는 busan_busstop_map.html 파일로 저장되어 웹 브라우저에서 직관적으로 확인 가능하다. 또한, 벡스코(BEXCO)는 부산을 대표하는 전시·컨벤션 중심지로 대중교통 이용 수요가 높은 핵심 지역이다. 이에 따라 벡스코를 중심으로 반경 500m 이내 정류장을 필터링하여 별도로 시각화하였다. 중심 좌표는 벡스코 위치(35.16834, 129.1368)로 설정되었으며, 반경 내 정류장은 거리 정보와 함께 마커로 표시된다. 이를 통해 행사장 인근의

교통 인프라 밀집도와 접근성을 직관적으로 파악할 수 있으며, 향후 참가자 이동 경로 분석 및 안내 시스템 설계 등에 활용할 수 있다.

5. 결론 향후 및 계획

이번 분석은 부산광역시 전체 버스 정류장 위치 데이터를 기반으로 정류장 간의 공간적 분포를 파악하고, 밀집 지역을 시각적으로 식별함으로써 교통 인프라 분석의 기초 자료를 마련하는 데 목적이 있었다. 특히 벡스코(BEXCO) 주변 정류장 분포를 별도로 시각화하여, 행사장 인근의 대중교통 접근성과 정류장 밀도를 보다 직관적으로 확인할 수 있었다.

향후에는 분석의 정밀도와 실효성을 높이기 위해 클러스터 수를 자동으로 결정할 수 있는 Elbow Method 나 Silhouette Score 등을 적용하고자 한다. 또한 단순 위치 정보 외에 정류장 이용량, 유동 인구 데이터와 같은 다양한 변수를 결합함으로써 군집 분석의 현실 반영 수준을 한층 강화할 계획이다. 더 나아가, 정류장 간의 거리 기반 네트워크 구조나 혼잡도 예측 모델을 도입하여 교통 수요 관리를 위한 실질적인 인사이트를 도출할 수 있도록 분석을 확장할 예정이다.

이와 함께, 추후 확보될 행사 참가자 데이터와 본 분석 결과를 연계함으로써 사전 등록자와 현장 방문자의 공간적 패턴을 비교하거나, 대중교통 접근성을 고려한 추천 이동 경로를 설계하는 등 실무 활용이 가능한 방향으로 발전시킬 수 있을 것이다.

이번 중간 분석은 전체적인 프레임 설정과 기초 데이터 구축을 중심으로 이루어진 것으로, 이후 실데이터를 기반으로 한 시뮬레이션 및 정책 제안의 출발점이 될 수 있다는 점에서 의미가 있다. 교통 운영, 이벤트 기획, 도시계획 등 다양한 응용 분야에서 실질적 활용 가능성을 확인했다는 점에서 그 의의를 찾을 수 있다.