

도시의 재구성: 사람들은 무엇에 이끌리는가

삼성KPMG4기팀 : UrbanVision

김다은, 김민지, 박서형, 지준희

1. 프로젝트 기안서

1-1. 주제 선정 배경

-UrbanVision은 ‘사람들은 도시의 어떤 요소에 이끌리게 되는가?’라는 질문에 대한 답변을 찾기위한 목적을 가졌다. 한국 내에서의 시도별 발달 정도의 불균형이 심화되고 있는 상황에 서울시의 자치구를 비교분석하여 도시가 성공하기 위한 구조적 요인을 파악하고자 한다. 이를 통해 지역 경제를 활성화 가능한 도시 구조 예측 및 개발 전략을 수립하며 나아가 도시 간의 가지는 연결성을 볼 수 있다.

- 현재 대한민국이 가진 큰 사회 문제는 도시의 불균형과 인구감소라고 볼 수 있다. 이 둘은 개별의 사회문제가 아닌, 연관된 사회문제이며 2023 이데일리 포럼에서는 도시의 불균형과 도시국가의 성질, 서울 집중형의 발전이 현재의 인구감소를 야기했다고 한다. 따라서 사회문제에 대한 아이디어, 정책 제안과 같은 공모전에 참여할 때 근거로 삼을 데이터가 될 수 있다.

1-2. 마스터 플랜

-주거, 삶이라는 측면에서 도시가 가지는 혹은 가져야하는 요소들을 파악하여 서울시내의 각 구별 요소 정도를 분석한다. 분석 결과를 시각화하여 불균형의 정도를 확인하고 요소들의 분포를 확인해 도시가 필요로 하는 성분을 알 수 있다.

-도시 매력에 핵심이 되는 중심 key값을 찾아 어떤 요소가 가장 도시에 영향을 크게 미치는 지 파악하여 가장 잘 발달된 도시의 해당 키값과 다른 도시들을 비교하여 부족한 부분을 확인한다.

-머신러닝을 통해 주거 측면에서 어떤 요소들이 가장 필요한지 확인하여 해당 값을 백분위로 표시한 뒤 합쳐 각 자치구에 반영, 시각화로 표출한다.

2. 프로젝트 개발환경

2-1. 작업 라이브러리

- 운영체제 : Windows 10 Pro (64비트)
- 데이터 분석 : Python 3.11.11
- 시각화 : Matplotlib, Seaborn, Plotly, Folium, Geopandas
- 머신러닝 : XGBoost, LightGBM, Scikit-learn, statsmodel

2-2. 플랫폼 아키텍처

데이터 분석 (Data Analysis)

-Python 3.11.11, Pandas(Python)(2.2.2), Numpy(1.26.4)

데이터 시각화 (Data Visualization)

-Matplotlib(3.10.0), Seaborn(0.13.2), Plotly(5.24.1), Folium(0.19.4), Geopandas(1.0.1)

머신러닝 (Machine Learning)

-XGBoost(2.1.3), LightGBM(4.5.0), Scikit-learn(1.6.0), statsmodel(0.14.4)

3. 프로젝트 추진체계

3-1. 팀 소개

팀장: 김다은, 팀원: 김민지, 박서형, 지준희

3-2. 일정 계획

Task	Time Schedule						
	1/17(Fri)	1/19(Sun)	1/20(Mon)	1/21(Tue)	1/22(Wed)	1/23(Thu)	1/24(Fri)
Data Preprocessing							Due Date Presentation
Data cleaning							
Data Analysis							
Data Visualization							
Machine Learning							
Final Adjustments							

4. 프로젝트 데이터 셋

4-1. 원본 데이터

서울시(기준)

인구 등록인구	직장 직장인구, 유동인구	인프라 학교, 학원, 병원
교통 지하철 역, 버스 정류장	상권 외식, 소매, 전문직, 기타	부동산 가격, 거래량

4-2. 데이터 전처리



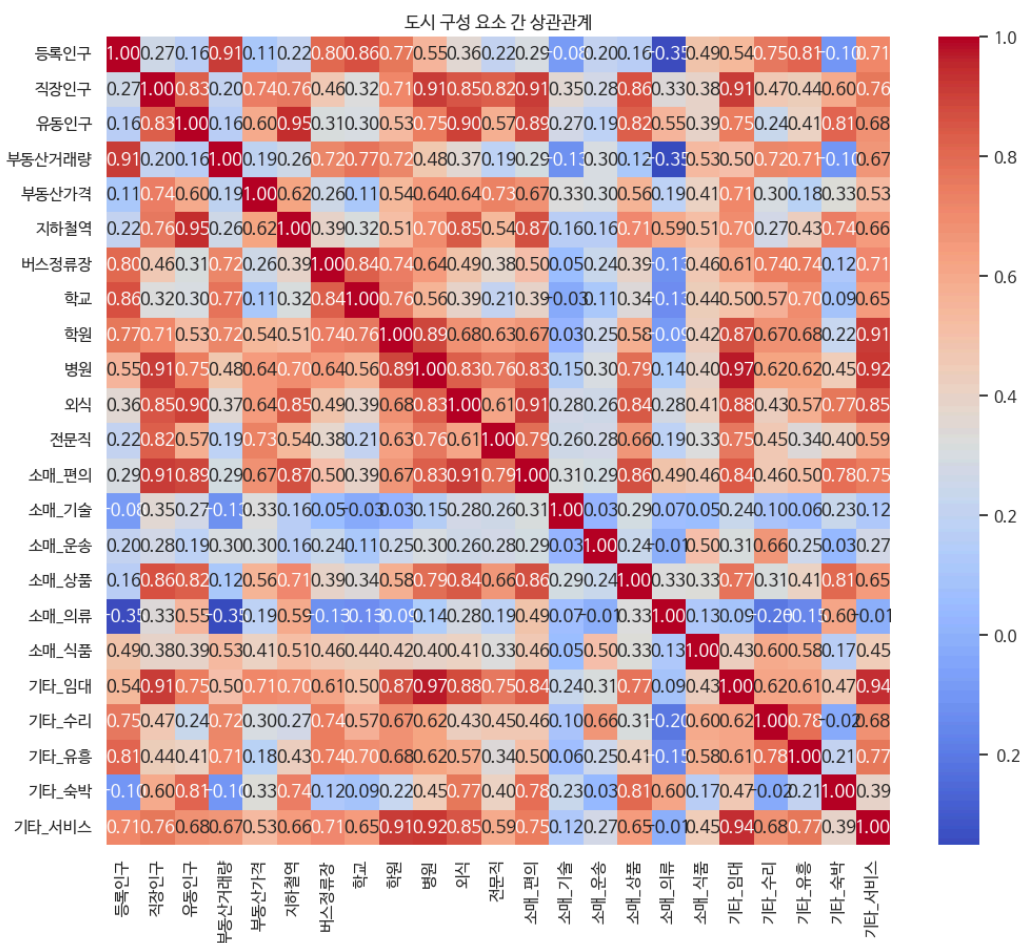
-서울시, 통계청 등 관공서 데이터를 위주로 가져와 전처리작업을 진행했고
자치구명을 통해 합치는 작업을 수행했다.

자치구명	등록인구	직장인	총유동인구	총유입유동인구	총유출유동인구	거래된부동산건물면적(㎡)	지하철역	버스정류장	병원	소매	외식	학원	기타	전문직	학교	
강남구	564280	1121201	503302811	255145874	248156937	4961	2337.323	43	424	3004	23012	13967	4624	12222	3026	85
강동구	469464	75000	123064048	60868155	62195893	5277	1290.044	18	302	916	9206	5317	1879	4765	170	64
강북구	289678	28246	86536158	42536410	43999748	2571	683.6846	12	195	501	5649	3839	687	3106	118	38
강서구	563515	149719	177323432	87454747	89868685	6058	876.226	19	335	987	12622	7284	2120	6282	412	82
관악구	496469	72879	136579127	66984626	69594501	3220	828.5015	10	259	739	7609	5867	1406	4619	217	59
광진구	349307	81443	126469385	62491444	63977941	3273	1236.279	11	174	617	8145	5343	1147	4114	210	46
구로구	412441	139820	159862041	79433353	80428688	3597	814.0908	12	254	630	11840	4968	1191	4083	402	60
금천구	239577	126426	64103637	32557824	31545813	2340	826.6018	4	168	374	9119	3803	592	3047	405	35
노원구	498358	56701	159311071	78891021	80420050	4620	917.5643	16	390	805	8048	4586	2152	3865	120	100
도봉구	306926	36388	69991699	34360861	35630838	2771	649.0097	8	232	396	5128	2998	943	2508	192	46
동대문구	359219	80350	119079063	58798436	60280627	3887	1092.453	11	258	638	10518	5042	1077	3741	189	50
동작구	387792	78066	193600630	96558084	97042546	3963	1299.739	20	141	624	6570	4284	1386	3264	149	51
마포구	373874	142802	288018206	146093517	141924689	4833	1414.108	29	252	851	11550	10020	1859	6551	531	52
서대문구	319749	66165	77063767	38617106	38446661	3486	1112.833	9	211	479	6117	4536	1143	3087	82	43
서초구	412611	576938	212438876	106438151	106000725	4199	2304.642	23	325	1484	14936	7420	2726	6581	4444	57
성동구	282385	80690	150399028	75939296	74459732	3570	1744.107	17	187	516	9594	4789	996	3432	359	39
성북구	435492	110756	98814397	48530761	50283636	4209	980.0545	13	282	601	7680	4842	1482	3816	83	62
송파구	657991	399728	260133120	129690032	130443088	6277	1651.266	32	439	1326	18490	8743	2814	7518	993	96
양천구	435867	139602	54510743	27278389	27232354	4257	1165.709	6	257	728	7406	4024	2634	3946	345	64
영등포구	397514	397723	260703845	131452268	129251577	4233	1317.82	26	291	833	12496	8018	1342	5393	865	47
용산구	218370	176381	120116395	60858726	59257669	2215	1991.513	14	231	363	9575	5470	624	2943	183	38
은평구	466809	48907	103631922	51044053	52587869	4859	840.352	13	323	749	7922	4779	1529	4285	122	68
종로구	150011	234085	230827930	115833890	114994040	1274	1062.745	17	189	491	13256	7015	636	3685	447	47
중구	131589	334297	355879391	179054428	176824963	1246	1300.944	38	164	596	25062	6756	416	3683	754	36
중랑구	386131	39323	81478637	40161641	41316996	3295	863.4356	14	357	598	7450	4392	911	3889	83	48

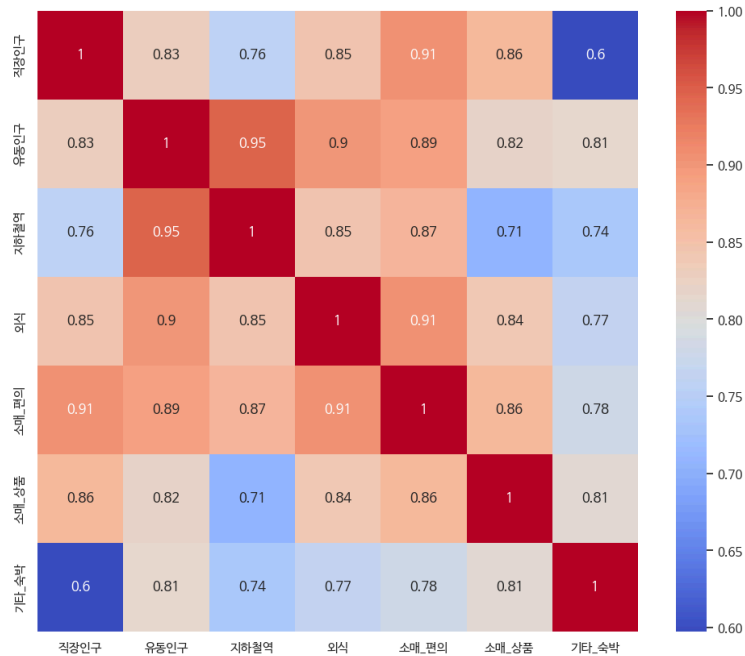
-상관계수에 필요한 수치를 나타내기 위해 크기의 정도를 반영했다.

5. 프로젝트 진행 과정

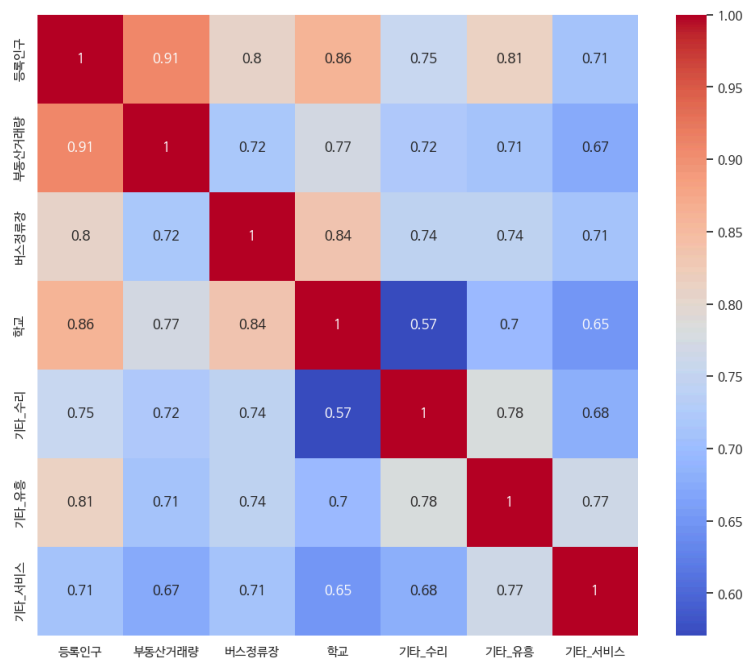
5-1. 상관관계 히트맵



- 전처리된 데이터를 바탕으로 상관관계수 제작, **corr**을 이용하여 제작하여 다양한 데이터가 서로 어떤 상관성을 가지는 지 확인했다.
- 머신러닝을 통해 유동인구, 등록인구를 기준으로 어떤 값이 제일 영향이 큰지 파악하여 **0.7**이상을 뽑아냈다

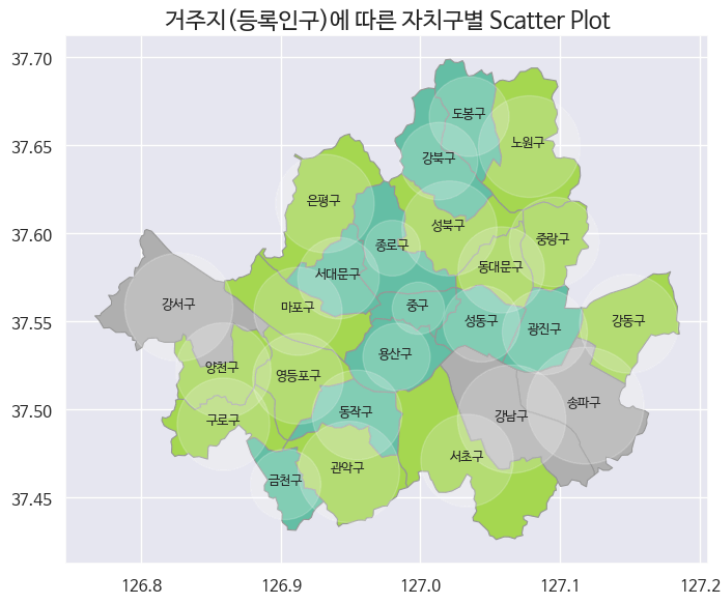


-유동인구 중심 히트맵



-등록인구 중심 히트맵

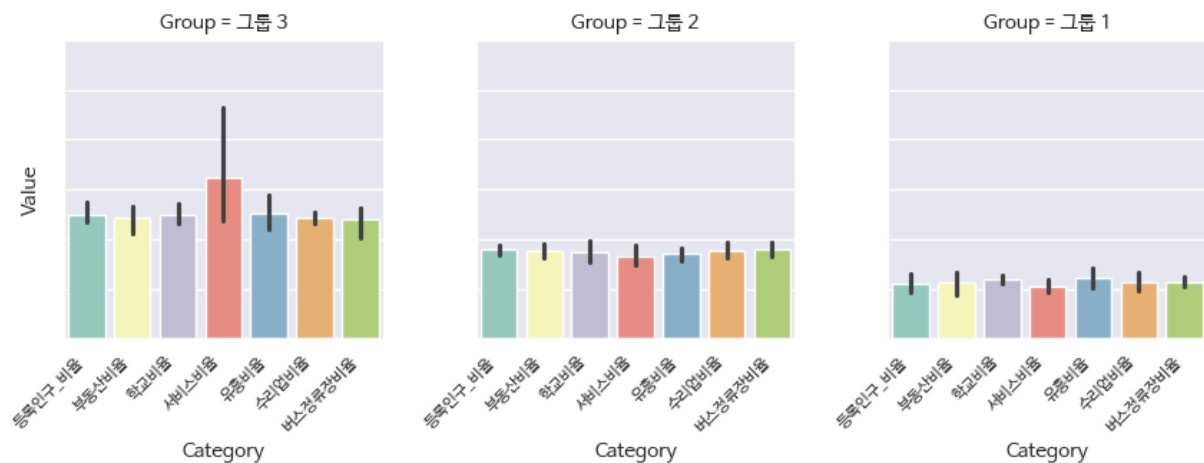
5-2. 유사도 클러스터



- 유사도 클러스터

등록인구에 따른 유사도를 머신러닝을 통해 이끌었고 3개의 그룹으로 나뉘게 되었다.

- 클러스터 간 비교 (Bar-plot)



- 그룹별 시각화 진행

6. 프로젝트 결과 보고

6-1. 활용 방안

- 각 요소가 고르게 분포한 곳일수록 주거적 측면에서 사람들을 이끌었다.
- 인구가 적고 부동산거래량이 적은 곳은 상업지구로 활동할 가능성이 높다.
하지만 낙후된 지역일 수도 있다. 주거적 측면으로만 보기에 다른 측면의 파악이 어렵다.
- 인구에 비해 서비스, 유흥 등 소비 관련 항목의 비율이 낮다면 해당 항목과 관련된 서비스의 창업을 고려할만한 지역으로 분류도 가능하다.
- 주거지역과 상업지역은 밀접한 연관이 있다.
- 주거지역의 농도가 짙은 곳 주변엔 농도가 옅은 지역이 발생한다.
- 농도가 짙은 곳은 신도시 혹은 재개발로 밀집을 해결하려는 정부의 정책을 찾기가 가능

6-2. 개선점

- 인과관계에 대한 분석이 미흡, 더 다양한 변수들을 넣었다면 정확도가 올라가며 만족스러운 결과를 얻을 수 있었을 것 같다. 여러 관점으로 도시를 바라보는 시간을 가지며 1차원적인 요소뿐만 아니라 자치구들에 대한 대중의 생각, 그린벨트 지역, 도시의 문화적 요소와 같은 변수와 변수가 이끌어내는 요소들을 고려했다면 결과 역시 다르게 나왔을 것이다. 또한 고려하지 못한 요소들 역시 상당부분 존재하여 데이터 마이닝 단계에서 이를 확인 및 점검할 수 있었다면 하는 아쉬운 부분 역시 존재했다. 그 지역만이 가지는 요소(공공청사, 문화재 등)를 생각하여 필요한 정도의 가중치를 지역마다 반영했다면 더 좋았기 때문이다. 프로젝트 초기에 상관계수로 도시의 요소를 바라본다면 핵심이 되는 key값이 존재할 것이라고 생각했지만 핵심으로 삼을 key값이 존재하지 않아 클러스터링을 시도하는 결과를 가져왔다. 하지만 클러스터링을 하며 주거와 생활을 분리하는 계기가 되어 발전으로 이어진 것은 전화위복이 되었다.

이 후 주거라는 측면으로 방향을 다시 잡아 머신러닝을 통해 어떤 요소가 주거에 강한 영향을 미치는지 파악하여 결과물을 도출하게 되었다. 더 다양한 경험을 통해 필요한 데이터를 추출하는 능력을 키우고 어떤 통계법을 적용시켜야 하는지 판단하는 통계 지식을 배워 수행한다면 보다 훌륭한 프로젝트를 진행 가능할 것이다.