

주제: 유형별 임대주택 설계 시 단지 내 적정 주차 수요를 예측



<배경>

아파트 단지 내 필요한 주차대수는 ①법정주차대수 ②장래주차수요 중 큰 값에 따라 결정하게 되어 있어, 정확한 ②장래주차수요의 산정을 필요로 합니다.

현재 ②장래주차수요는 ‘주차원단위’와 ‘건축연면적’을 기초로 하여 산출되고 있으며, ‘주차원단위’는 신규 건축예정 부지 인근의 유사 단지를 피크 시간대 방문하여 주차된 차량대수를 세는 방법으로 조사하고 있습니다.

이 경우 인력조사로 인한 오차발생, 현장조사 시점과 실제 건축시점과의 시간차 등의 문제로 과대 또는 과소 산정의 가능성을 배제할 수 없습니다.

<목적>

 유형별 임대주택 설계 시 단지 내 적정  주차 수요를 예측

▼ test 데이터 예시

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	단지코드	총세대수	임대건물	지역	공급유형	전용면적	세대수	공가수	자격유형	임대보증금	임대료	지하철역	버스정류장	단지내주차면수
2	C1072	754	아파트	경기도	국민임대	39.79	116	14	H	22830000	189840	0	2	683

▼ test 데이터 예측 결과 예시

	A	B
1	단지코드	차량수
2	C1072	922
3	C1128	1350
4	C1456	579
5	C1840	311
6	C1332	957
7	C1563	1145
8	C1794	982
9	C1640	98
10	C1377	78

1. test 데이터를 통해 훈련된 모델로 ‘등록 차량수’ 예측
2. 모델의 성능과 정확도 향상, 손실(loss) 감소 목적 포함

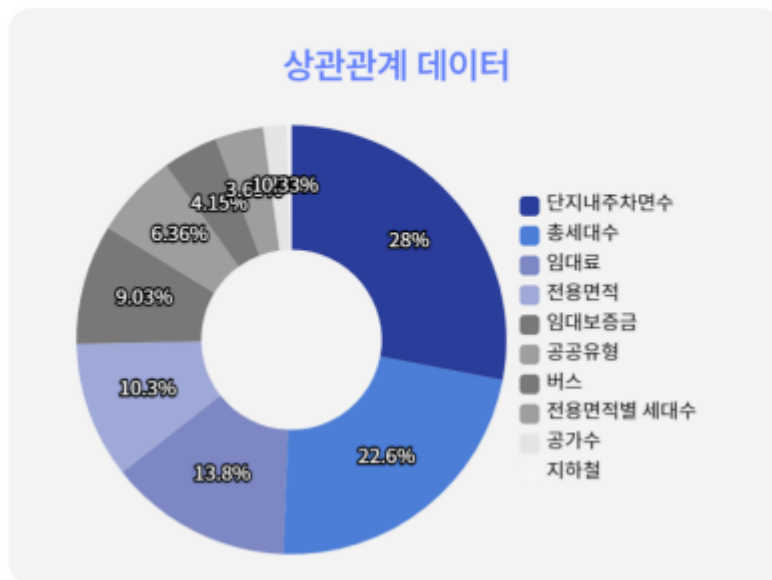
1. 데이터 탐색

<https://dacon.io/competitions/official/235745/data>

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	단지코드	총세대	입대건물구지역		공급유형	전용면적	전용면적별	공가수	자격유형	임대보증금	임대료	도보	도보	단지내주차면수	등록차량수
2	C2483	900	아파트	경상북도	국민임대	39.72	134	38 A		15667000	103680	0	3	1425	1015
3	C2483	900	아파트	경상북도	국민임대	39.72	15	38 A		15667000	103680	0	3	1425	1015
4	C2483	900	아파트	경상북도	국민임대	51.93	385	38 A		27304000	184330	0	3	1425	1015
5	C2483	900	아파트	경상북도	국민임대	51.93	15	38 A		27304000	184330	0	3	1425	1015
6	C2483	900	아파트	경상북도	국민임대	51.93	41	38 A		27304000	184330	0	3	1425	1015
7	C2483	900	아파트	경상북도	국민임대	51.95	89	38 A		27304000	184330	0	3	1425	1015
8	C2483	900	아파트	경상북도	국민임대	51.95	135	38 A		27304000	184330	0	3	1425	1015
9	C2483	900	아파트	경상북도	국민임대	59.88	86	38 A		30357000	214270	0	3	1425	1015
10	C2515	545	아파트	경상남도	국민임대	33.48	276	17 A		9216000	82940	0	3	624	205
11	C2515	545	아파트	경상남도	국민임대	39.6	60	17 A		12672000	107130	0	3	624	205
12	C2515	545	아파트	경상남도	국민임대	39.6	20	17 A		12672000	107130	0	3	624	205
13	C2515	545	아파트	경상남도	국민임대	46.9	38	17 A		18433000	149760	0	3	624	205
14	C2515	545	아파트	경상남도	국민임대	46.9	19	17 A		18433000	149760	0	3	624	205
15	C2515	545	아파트	경상남도	국민임대	51.97	106	17 A		23042000	190090	0	3	624	205
16	C2515	545	아파트	경상남도	국민임대	51.97	26	17 A		23042000	190090	0	3	624	205
17	C1407	1216	아파트	대전광역시	국민임대	30.95	288	13 A		15620000	127350	1	1	1285	1064
18	C1407	1216	아파트	대전광역시	국민임대	30.99	68	13 A		15620000	127350	1	1	1285	1064
19	C1407	1216	아파트	대전광역시	국민임대	30.99	34	13 A		15620000	127350	1	1	1285	1064
20	C1407	1216	아파트	대전광역시	국민임대	41.11	148	13 A		27637000	164610	1	1	1285	1064
21	C1407	1216	아파트	대전광역시	국민임대	41.39	74	13 A		27637000	164610	1	1	1285	1064
22	C1407	1216	아파트	대전광역시	국민임대	41.58	70	13 A		27637000	164610	1	1	1285	1064
23	C1407	1216	아파트	대전광역시	국민임대	41.58	70	13 A		27637000	164610	1	1	1285	1064
24	C1407	1216	아파트	대전광역시	국민임대	46.36	170	13 A		36048000	198250	1	1	1285	1064

- 등록차량수와 상관계 분석

i 데이터 전처리 - 상관계 분석



2. 오류 데이터 삭제

```
# 데이터 오류로 인한 삭제
error_data = ['C2085', 'C1397', 'C2431', 'C1649', 'C1095', 'C2051', 'C1218', 'C1894', 'C2483', 'C1502', 'C1988']
for error in error_data:
    error_index = train[train['단지코드'] == error].index
    train.drop(error_index, inplace=True)

# 중복값 제거
train = train.drop_duplicates()
test = test.drop_duplicates()

# 아파트이면서 공공분양이 아닌 데이터만 사용
아파트 단지내 주차장에 대해 알아보기 위해 상가 데이터 제거.
공공 분양은 임대료와 임대보증금이 없기 때문에 제거.
train = train[(train.임대건물구분 == '아파트') & (train.공급유형 != '공공분양')]
test = test[(test.임대건물구분 == '아파트') & (test.공급유형 != '공공분양')]
```

3. 결측치 확인

train.info()

train 데이터에서는 임대보증금, 임대료, 지하철역 수, 버스정류장 수에 NULL값 존재

test.info()

test 데이터에서는 자격유형, 임대보증금, 임대료, 지하철역 수에 NULL값 존재

4. 결측치 처리

1) 임대보증금과 임대료 타입 float로 변환

'-'(하이픈)이 들어간 데이터를 없애고 데이터 타입을 실수형으로 변환

```
train.loc[train.임대보증금=='-', '임대보증금'] = np.nan
test.loc[test.임대보증금=='-', '임대보증금'] = np.nan
train['임대보증금'] = train['임대보증금'].astype(float)
test['임대보증금'] = test['임대보증금'].astype(float)
train.loc[train.임대료=='-', '임대료'] = np.nan
test.loc[test.임대료=='-', '임대료'] = np.nan
train['임대료'] = train['임대료'].astype(float)
test['임대료'] = test['임대료'].astype(float)
```

2) 공급유형이 장기전세면 임대료 = 0

```
train.loc[(train.공급유형=='장기전세')&(train.임대료.isnull()), '임대료']=0
```

3) 지하철역과 버스 정류장의 NULL값의 경우, 0으로 판단, 밑에서 컬럼명 바꿈(지하철, 버스)

```
train['도보 10분거리 내 지하철역 수(환승노선 수 반영)'].fillna(0, inplace=True)
train['도보 10분거리 내 버스정류장 수'].fillna(0, inplace=True)
```

```
test['도보 10분거리 내 지하철역 수(환승노선 수 반영)'].fillna(0, inplace=True)
test['도보 10분거리 내 버스정류장 수'].fillna(0, inplace=True)
```

4) test의 자격유형에 있는 2개 결측치 채우기

```
test.loc[(test.단지코드=='C2411') & (test.자격유형.isnull()), '자격유형'] = 'A'
test.loc[(test.단지코드=='C2253') & (test.자격유형.isnull()), '자격유형'] = 'C'
```

5) 임대보증금 / 임대료 결측치 찾아서 채우기

[Train] C1786 | 강원도 | 행복주택 | 총 480세대 : 춘천거두2 행복주택
춘천거두2 행복주택 임대조건 :

<https://www.myhome.go.kr/hws/portal/sch/selectRsdRcritNtcDetailView.do?pblancId=1425>

[Train] C1326 | 부산광역시 | 국민임대 | 총 1934세대 : 부산정관 7단지 A-1블록
부산정관 7단지 A-1BL 국민임대 모집공고 :

<https://www.myhome.go.kr/hws/portal/sch/selectRsdRcritNtcDetailView.do?pblancId=915>

[Train] C2186 | 대구광역시 | 국민임대 | 총 924세대 : 대구연경 A-2블록 국민임대주택

대구연경 A-2블록 국민임대주택 입주자모집 안내 :

<https://m.blog.naver.com/PostView.nhn?isHttpsRedirect=true&blogId=sugar9025404&logNo=221577519654&categoryNo=59&proxyReferer=>

[Test] C2152 | 강원도 | 영구임대 | 총 120세대 : 화천신읍 공공실버주택
화천신읍 공공실버주택 예비입주자 모집 :

<https://www.myhome.go.kr/hws/portal/sch/selectRsdRcritNtcDetailView.do?pblancId=9070>

[Test] C1267 | 경상남도 | 행복주택 | 총 675세대 : 창원가포 A-1블록 행복주택
창원가포 A-1블록 행복주택 입주자 모집공고 :

<https://www.myhome.go.kr/hws/portal/sch/selectRsdRcritNtcDetailView.do?pblancId=6373>

[Test] C1006 | 대전광역시 | 영구임대 | 총 1505세대 : 대전둔산1단지
대전둔산1 영구임대주택 그린리모델링세대 입주자 자격완화 모집 :

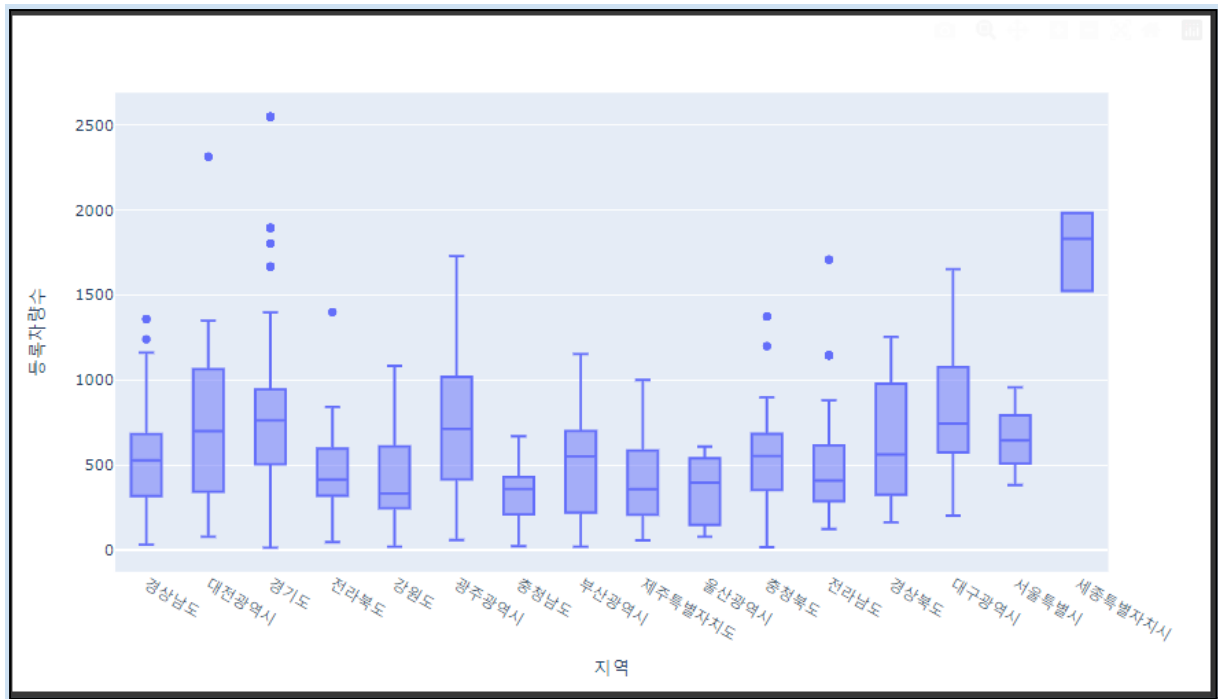
<https://www.myhome.go.kr/hws/portal/sch/selectRsdRcritNtcDetailView.do?pblancId=8399>

5. 파생변수 생성

=====

지역의 피처가 16개로 비교적 많기 때문에

밑에 지역 - 등록차량수의 중앙값을 기준으로 5개 그룹으로 통합



n그룹(5개)

강원도, 제주특별자치도, 충청남도, 울산광역시 >> 0

전라남도,전라북도, 경상남도 >> 1

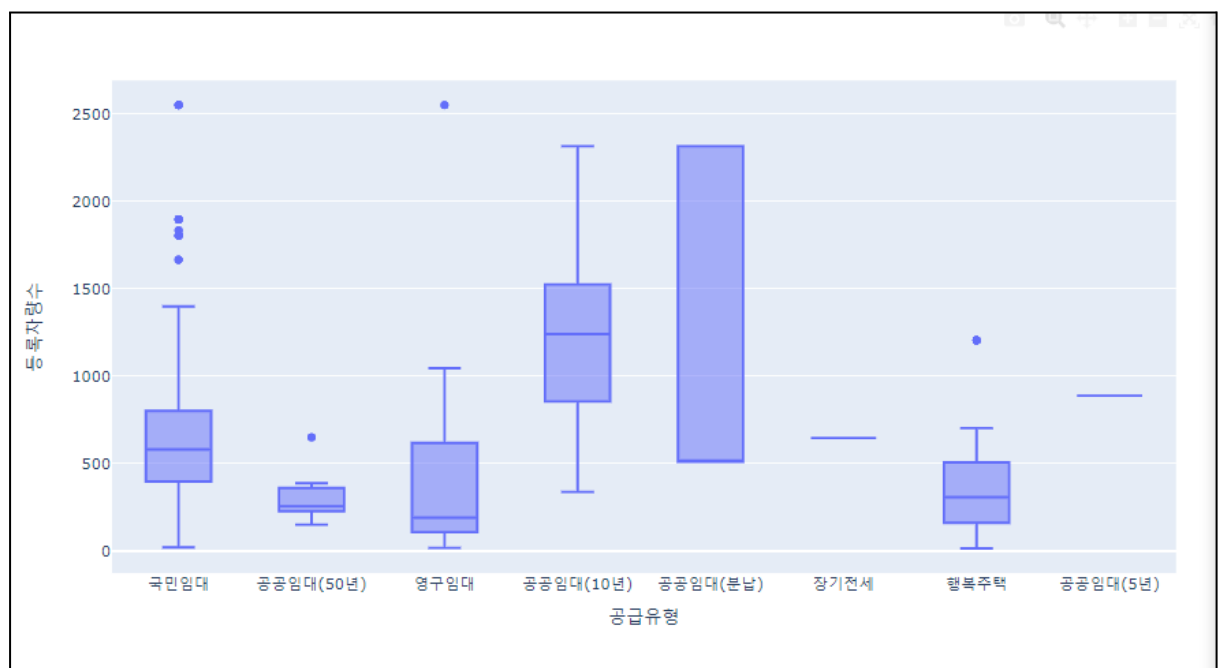
부산광역시, 충청북도, 경상북도 >> 2

대전광역시, 광주광역시, 서울특별시 >> 3

경기도, 대구광역시, 세종특별자치시 >> 4

=====

공급유형의 피쳐도 10개로 비교적 많기 때문에
그룹으로 통합



=====

자격유형그룹화

주거가능자리

자격유형

자격유형그룹

- 1
- 2
- 3
- 4
- 5

범주형 데이터를 수치형 데이터로 변환하기 위한 원-핫 인코딩 결과

[illegible]

7. 훈련 모델 선정

pycaret을 이용한 다양한 모델 확인

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
et	Extra Trees Regressor	10.0199	1198.6530	31.8151	0.9926	0.1024	0.0344
xgboost	Extreme Gradient Boosting	28.0903	2851.9743	52.8277	0.9824	0.1521	0.0819
rf	Random Forest Regressor	29.9223	3259.0749	56.6379	0.9799	0.1615	0.0884
dt	Decision Tree Regressor	13.2708	3643.8881	59.0516	0.9775	0.1321	0.0378
catboost	CatBoost Regressor	39.4561	3699.6117	60.2848	0.9773	0.1870	0.1157
lightgbm	Light Gradient Boosting Machine	43.3005	4522.7895	66.8946	0.9719	0.1948	0.1243
gbr	Gradient Boosting Regressor	81.0017	12697.4905	112.4160	0.9218	0.2773	0.2120
ada	AdaBoost Regressor	137.3587	29593.4900	171.9110	0.8188	0.4409	0.4479
ridge	Ridge Regression	134.8746	35605.8614	188.4694	0.7834	0.4404	0.3199
lr	Linear Regression	135.0288	35608.1357	188.4761	0.7834	0.4537	0.3214
br	Bayesian Ridge	134.4792	35668.8338	188.6333	0.7831	0.4198	0.3142
llar	Lasso Least Angle Regression	134.7040	35848.0928	189.1271	0.7820	0.4369	0.3106
lasso	Lasso Regression	134.7040	35848.0928	189.1271	0.7820	0.4369	0.3106
en	Elastic Net	140.8105	39149.0852	197.6221	0.7624	0.4173	0.3287
huber	Huber Regressor	217.2652	108193.1933	311.5834	0.3686	0.5589	0.5604
knn	K Neighbors Regressor	244.4454	121696.0742	347.8657	0.2620	0.6630	0.8633

상위 모델 5가지를 뽑아 평가점수 비교한 후



최종 모델 선정 - **ExtraTreesRegressor**

```
from sklearn.ensemble import ExtraTreesRegressor
```

```
model = ExtraTreesRegressor()
```

```
scores = cross_validate(model, train_input, train_target, return_train_score=True, n_jobs=-1) #
```

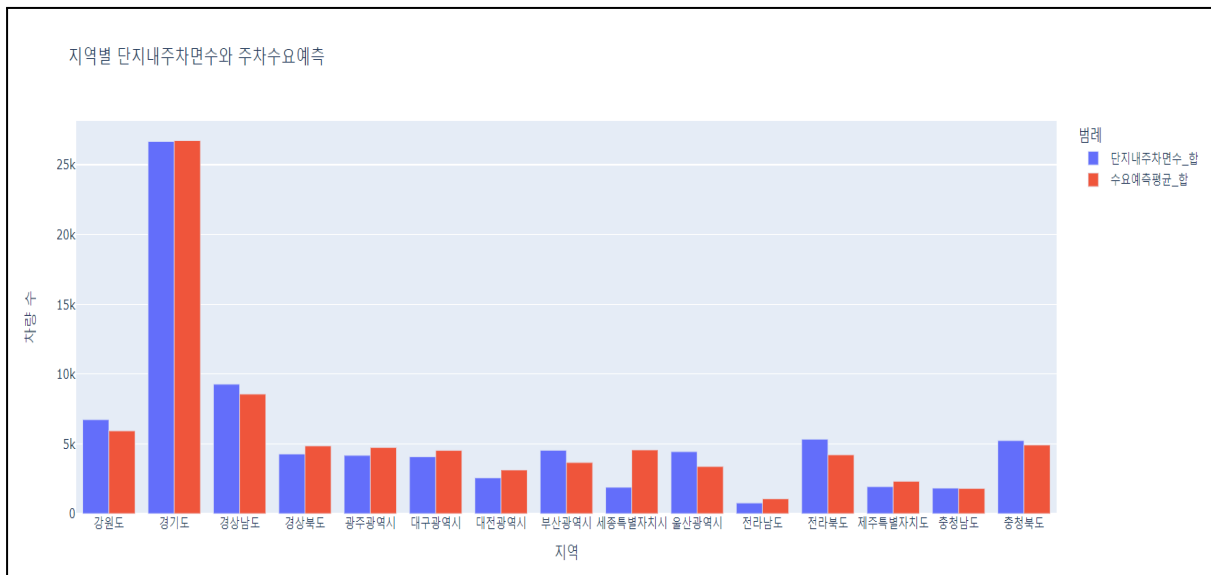
교차검증

```
model.fit(train_input, train_target)
```

8. test데이터 최종 예측

지역별 단지내주차면수, 예측차량수 비교

```
pred = model.predict(test)
```



➡ 대부분의 지역들이 예측한 주차수요보다 주차가능자리가 부족하다는 것을 확인.

<결론>

주차 수요 예측 모델을 훈련시킨 데이터는 단지코드, 총세대수, 임대건물, 지역, 공급유형, 전용면적, 세대수, 공가수, 자격유형, 임대료, 보증금, 가까운 대중교통의 특성으로 이루어져 있음.

이러한 특성들을 기반으로 주차 수요 예측 모델은 아파트 주차 수요에 영향을 미치는 다양한 요소들을 고려하며, 장래적으로 예상되는 차량 주차 공간의 수요를 예측함. 이 모델을 활용해 예측된 주차 수요에 대한 인사이트를 분석해보면, 아파트 단지 규모, 인구 밀도 등이 아파트 주차 수요에 큰 영향을 미치는 것으로 보임.

(임대 아파트 데이터를 사용했기 때문에 모든 지역의 아파트 주차수요를 일반화하기에는 오류가 있기 때문에 추후 더 많은 데이터를 가지고 분석할 필요가 있어보임.)

주차 수요를 예측한 결과, 해당 지역의 아파트들 대부분은 주차 수요를 충족시키지 못할 것으로 예측됨. 따라서, 해당 지역의 아파트 단지들은 현재 주차 공간 부족 문제를 해결하기 위한 적극적인 대책이 필요가 있음

해결방안으로 특정 지역에서 주차 공간 부족 문제가 심각하다면, 해당 지역의 아파트 건설 관련 정책을 개선하여 주차 공간을 더 확보할 수 있도록 지원해야할 것임.

또한, 주차 수요 예측 모델은 건설사와 입주자 모두에게 중요한 정보를 제공할 수 있음. 건설사는 아파트 건설 시 주차 공간 비율을 결정할 때 이 모델을 참고하여 불필요한 주차공간 비용을 줄일 수 있고, 입주자는 해당 아파트 단지의 주차 공간 수요에 대한 정보를 미리 알고 선택할 수 있을 것임.

이러한 인사이트들을 바탕으로 정부와 지방 자치 단체, 건축업계 등의 협력이 필요할 것이고 이를 통해 주민들의 생활 편의성을 높이고, 도시의 교통체증 문제를 해결해야 할 것임.