

# 계획서

3조 강은영, 강청우, 곽원석, 김민철, 최승희

## 주제

- 유형별 임대주택 설계 시 단지 내 적정 주차 수요를 예측

주차수요 예측 AI 경진대회  
정형 | 한국토지주택공사 | 주차수요 | MAE  
상금 : 총 1,300만원  
2021.06.10 ~ 2021.07.30 18:00 + Google Calendar  
273명 D-44 참여중

대회안내 데이터 코드 공유 토크 대회문의 리더보드 팀 제출

## 대회 주요 일정



## 배경 및 목적

- 차량대수의 증가는 공동주택의 주차부족 문제를 야기시켜 이중주차, 보도위 주차, 주택가 이면도로의 밤샘 주차 등에 의해 긴급차량 통행의 어려움과 주차공간 확보를 위한 주민 갈등을 야기 시키는 사회문제이다.
- 일반적으로 시간의 경과에 따라 세대별 차량보유율이 증가하지만 주차면수는 그 주차수요에 맞추어 탄력적으로 증가하지 못한다.
- 세대에 필요한 주차공간을 충분하게 공급하기 위해서는 정확한 주차수요의 예측이 중요하다.

## 데이터 유무 및 확보전략

-주어진 데이터 (train, test, age\_gender\_info)

```
In [4]: train.shape, test.shape
```

```
Out[4]: ((2952, 15), (1022, 14))
```

```
In [5]: train.head()
```

```
Out[5]:
```

	단지코 드	총세 대수	임대건 물구분	지역	공급 유형	전용면 적	전용면적 별세대수	공가 수	자격 유형	임대보증 금	임대료	도보 10분거리 내 지하철 역 수(환승노선 수 반영)	도보 10분거리 내 버스정류장 수	단지내주차 면수	등록차량 수
0	C2483	900	아파트	경상 북도	국민 임대	39.72	134	38.0	A	15667000	103680	0.0	3.0	1425.0	1015.0
1	C2483	900	아파트	경상 북도	국민 임대	39.72	15	38.0	A	15667000	103680	0.0	3.0	1425.0	1015.0
2	C2483	900	아파트	경상 북도	국민 임대	51.93	385	38.0	A	27304000	184330	0.0	3.0	1425.0	1015.0
3	C2483	900	아파트	경상 북도	국민 임대	51.93	15	38.0	A	27304000	184330	0.0	3.0	1425.0	1015.0
4	C2483	900	아파트	경상 북도	국민 임대	51.93	41	38.0	A	27304000	184330	0.0	3.0	1425.0	1015.0

```
In [6]: test.head()
```

```
Out[6]:
```

	단지코 드	총세대 수	임대건물 구분	지역	공급 유형	전용면 적	전용면적 별세대수	공가 수	자격 유형	임대보증 금	임대료	도보 10분거리 내 지하철 역 수(환승노선 수 반영)	도보 10분거리 내 버 스정류장 수	단지내주차 면수
0	C1072	754	아파트	경기 도	국민 임대	39.79	116	14.0	H	22830000	189840	0.0	2.0	683.0
1	C1072	754	아파트	경기 도	국민 임대	46.81	30	14.0	A	36048000	249930	0.0	2.0	683.0
2	C1072	754	아파트	경기 도	국민 임대	46.90	112	14.0	H	36048000	249930	0.0	2.0	683.0
3	C1072	754	아파트	경기 도	국민 임대	46.90	120	14.0	H	36048000	249930	0.0	2.0	683.0
4	C1072	754	아파트	경기 도	국민 임대	51.46	60	14.0	H	43497000	296780	0.0	2.0	683.0

### 결측치 확인

```
In [7]: train.isna().sum()
```

```
Out[7]:
단지코드          0
총세대수          0
임대건물구분      0
지역              0
공급유형          0
전용면적          0
전용면적별세대수  0
공가수            0
자격유형          0
임대보증금        569
임대료            569
도보 10분거리 내 지하철역 수(환승노선 수 반영)  211
도보 10분거리 내 버스정류장 수              4
단지내주차면수    0
등록차량수        0
dtype: int64
```

```
In [8]: test.isna().sum()
```

```
Out[8]:
단지코드          0
총세대수          0
임대건물구분      0
지역              0
공급유형          0
전용면적          0
전용면적별세대수  0
공가수            0
자격유형          2
임대보증금        180
임대료            180
도보 10분거리 내 지하철역 수(환승노선 수 반영)  42
도보 10분거리 내 버스정류장 수              0
단지내주차면수    0
dtype: int64
```

```
In [10]: agegender_info.isna().sum()
```

```
Out[10]:
지역          0
10대미만(여자)  0
10대미만(남자)  0
10대(여자)     0
10대(남자)     0
20대(여자)     0
20대(남자)     0
30대(여자)     0
30대(남자)     0
40대(여자)     0
40대(남자)     0
50대(여자)     0
50대(남자)     0
60대(여자)     0
60대(남자)     0
70대(여자)     0
70대(남자)     0
80대(여자)     0
80대(남자)     0
90대(여자)     0
90대(남자)     0
100대(여자)    0
100대(남자)    0
dtype: int64
```



- 공공 데이터와 같이 누구나 얻을 수 있고 법적 제약이 없는 외부 데이터 허용
- 대회에서 제공하는 데이터사용, 필요시 공공데이터를 수집하여 활용  
(공공데이터 포털, 통계청 등)

## 사용 예상 기술

(머신러닝 알고리즘 - 회귀)

- KNeighborsRegressor
- LinearRegression
- Ridge
- Lasso
- ElasticNet
- SVR
- DecisionTreeRegressor
- RandomForestRegressor
- AdaBoostRegressor
- GradientBoostingRegressor
- LGBMRegressor
- XGBRegressor

학습을 위한 전처리 후 위의 알고리즘들을 사용하여 하이퍼 파라미터 튜닝, 최적의 모델에 데이터 학습하고 검증

추가적으로 점수를 올리기위한 데이터 추가 및 가공

## to-do list

1. 데이터 탐색
2. 전처리
3. 상관관계 분석
4. 분석 모델 선정
5. 분석 모델 평가
6. 추가 데이터 수집
7. 모델 선정