

집밥콩떡

ON?! Or OFF??

PBL ML과정 MINI Project



(05725-00037)

청구서의 고객성명이 실제 전기사용자와 다른 경우 14일 이내 **현전에 전기사용 계약변경 신청**을 하여야 합니다. 변경시 반드시 **현전** 또는 고객센터(123)으로 연락주시기 바랍니다.

홍길동 고객님의 2020년 11월분 전기요금 자동납부 청구서

● 자동이체 통장 잔액이 부족한 경우 잔액만큼만 인출됩니다.

전기요금 복지할인제도를 운영하고 있으니 장애의 정도가 심한 장애인, 3급이상 상이유공자, 독립유공자, 기초생활수급자, 사회복지시설 등에 해당되는 경우 신청하시어 할인혜택을 받으시기 바랍니다. (청구서 뒷면 참조)

참구내역

기본 요금	7,300
전 력 량 요금	81,800
-개편전요금	84,300
-환경비용차감	-2,500
기후환경요금	2,650
연료비조정액	-1,500
복 지 할 인	-10,000
전 기 요 금 계	80,250
부 가 가 치 세	8,025
전 력 기 금	2,960
원 단 위 절 사	5
당 월 요 금 계	91,230

고 객 사 항

계 약 종 별	주택용전력
검 침 일	05
계 량 기 번 호	1111111111
계 량 기 배 수	1
계 약 전 력	3
가 구 수	1
T V 대 수	1

년 일 일 현재 미납내역

미납월	금 액
계	미납요금 없음

청구금액 91,230 원

전기사용장소 대전광역시 금천구 홍길동로 44

청 구 금 액 91,230 원

고 객 번 호 01 2345 6789 납기일 2020년 11월 25일

사 2020년 10월 05일 ~ 2020년 11월 04일

전자 세금 계산서

공급자등록번호	120-82-00052
공급받는자등록번호	세금계산서 미발급
공 급 가 액	80,250 원
세 액	8,025 원
각 성 일 자	2020년 11월 18일

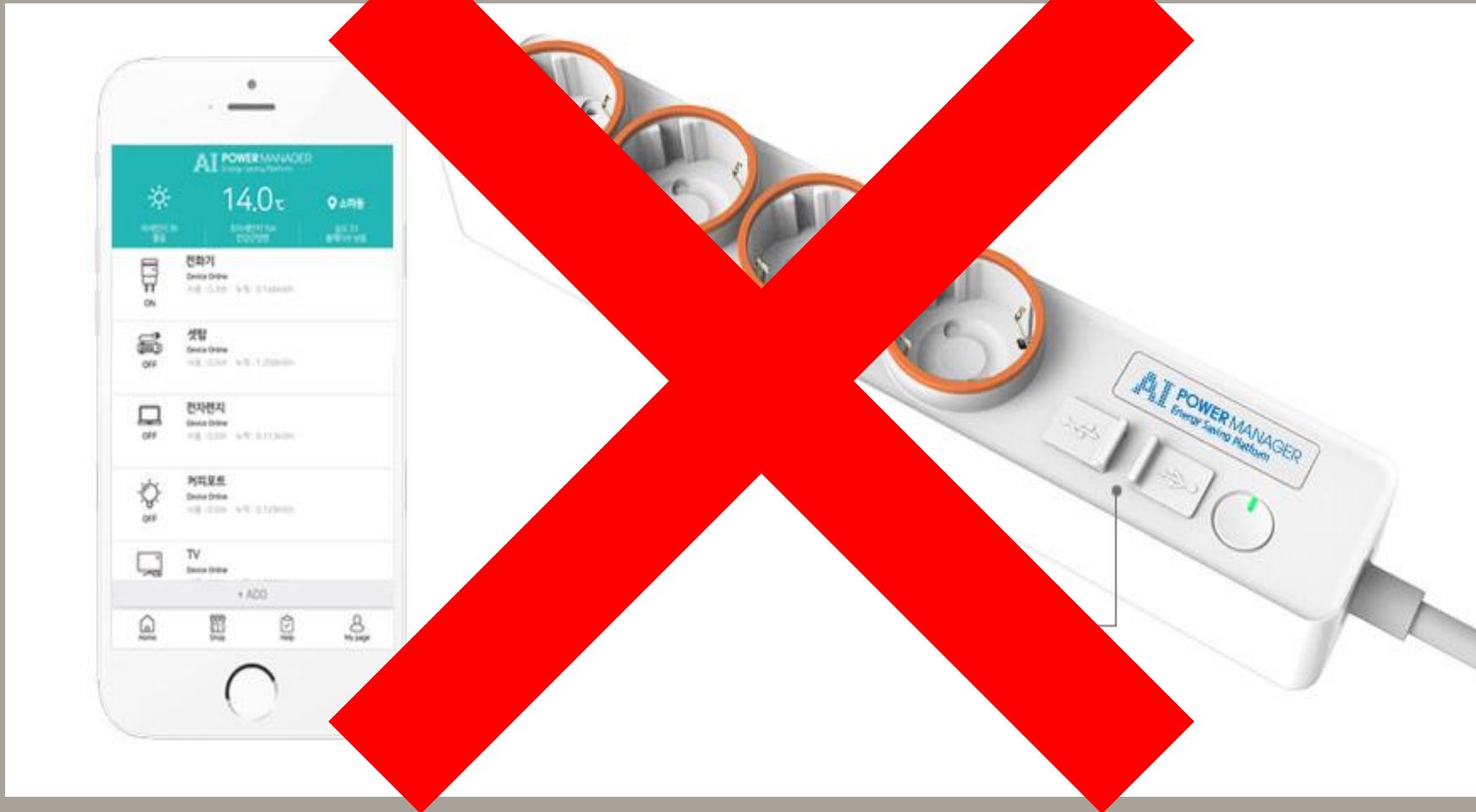
사 용 량 비 교

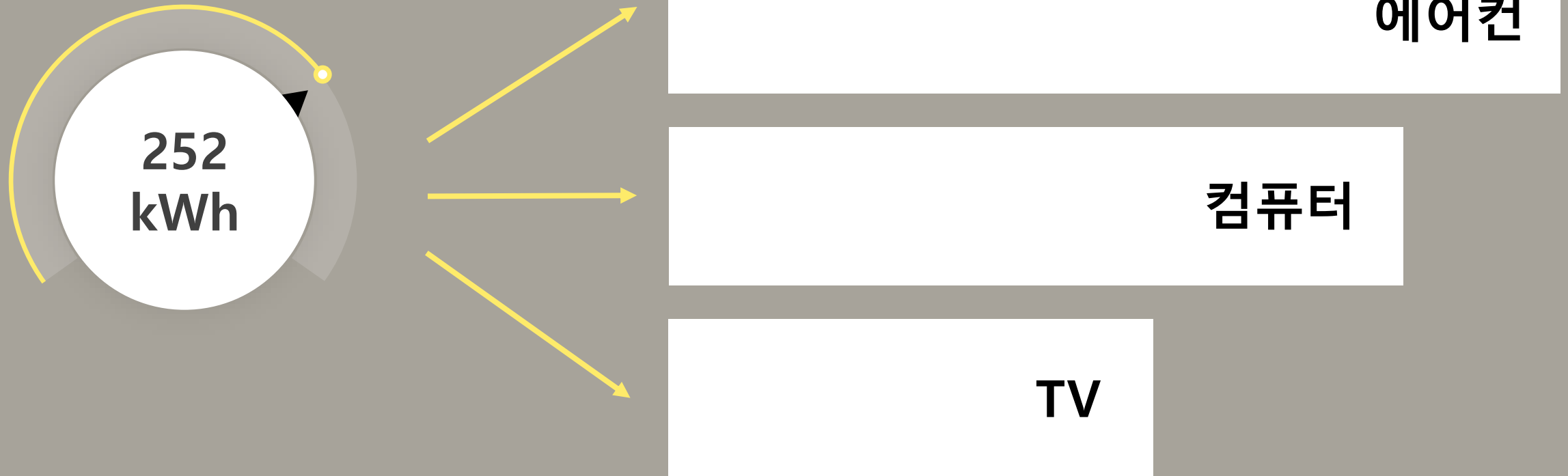
당 월	500 kWh
전 월	263 kWh
전 년 동 월	260 kWh

이 청구서는 부가가치세법 시행령 제50조 제1항에 따라 내부세무서에게
이송된
신청하여 등록된 사업자가 발행하는 전자세금계산서입니다.
유급액을 청구하지 않습니다.

한국전력공사

- 식비 12% 120,000원 >
- 여행/숙박 10% 100,300원 >
- 교통 5% 50,000원 >





1

데이터 탐색

데이터를 살펴보자

2

데이터 전처리

데이터를 씹고 뜯자

3

데이터 모델링

데이터를 요리하자

4

데이터 분석

데이터를 맛보자

5

결론 도출

제 의견은요...

데이터 전처리

```
# 1번 가구 필터
```

```
ele_num = ele_drop['H_ID'] == 1
```

```
# 전력총합, 에어컨 테이블 2개로 분리
```

```
temp_1 = elenum['PointName'] == '전력총합'
```

```
temp_2 = elenum['PointName'] == '에어컨'
```

```
ele_total = elenum[temp_1]
```

```
ele_se = elenum[temp_2]
```

전력총합과 에어컨의 교집합만 MERGE (NULL값 제거)

ele_merge = pd.merge(ele_total, ele_se, on = 'ST_DATE')

전력총합

에어컨

H_ID	ST_DATE	PointName	H00_1	H00_2	H00_3	H00_4	H00_5	H00_6	H01_1	...	H22_3	H22_4	H22_5	H22_6	H23_1	H23_2	H23_3	H23_4
0	2011-12-10	전력총합	0.033	0.030	0.028	0.029	0.027	0.027	0.027	...	0.060	0.023	0.047	0.051	0.043	0.031	0.030	0.019
1	2011-12-11	전력총합	0.018	0.022	0.020	0.021	0.010	0.009	0.030	...	0.038	0.039	0.041	0.040	0.048	0.085	0.062	0.055
2	2011-12-12	전력총합	0.013	0.022	0.020	0.021	0.014	0.010	0.009	...	0.064	0.096	0.062	0.026	0.020	0.011	0.010	0.015
3	2011-12-13	전력총합	0.020	0.009	0.009	0.019	0.021	0.022	0.014	...	0.063	0.060	0.062	0.046	0.033	0.034	0.082	0.046
4	2011-12-14	전력총합	0.008	0.011	0.023	0.021	0.018	0.000	0.027	...	0.045	0.089	0.040	0.061	0.044	0.032	0.044	0.043
...
133552	2012-10-10	전력총합	0.007	0.017	0.020	0.020	0.020	0.033	0.007	...	0.037	0.028	0.025	0.041	0.036	0.036	0.037	0.048
133556	2012-10-11	전력총합	0.064	0.071	0.031	0.025	0.042	0.010	0.007	...	0.045	0.054	0.039	0.055	0.047	0.060	0.058	0.055
133561	2012-10-12	전력총합	0.057	0.044	0.026	0.044	0.020	0.021	0.021	...	0.007	0.041	0.018	0.020	0.012	0.006	0.013	0.019
133565	2012-10-13	전력총합	0.008	0.008	0.046	0.018	0.018	0.009	0.007	...	0.043	0.019	0.020	0.016	0.006	0.007	0.018	0.019
133570	2012-10-14	전력총합	0.007	0.048	0.019	0.018	0.011	0.009	0.011	...	0.051	0.046	0.043	0.048	0.033	0.023	0.025	0.022

232 rows × 147 columns

94 rows × 293 columns

H_ID	ST_DATE	PointName	H00_1	H00_2	H00_3	H00_4	H00_5	H00_6	H01_1	...	H22_3	H22_4	H22_5	H22_6	H23_1	H23_2	H23_3	H23_4
71056	2012-07-07	에어컨	0.000	0.0	0.000	0.0	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000	0.000	0.0	0.001	0.000
71061	2012-07-10	에어컨	0.034	0.0	0.000	0.0	0.000	0.000	0.000	...	0.000	0.000	0.000	0.001	0.000	0.0	0.000	0.000
71066	2012-07-11	에어컨	0.000	0.0	0.000	0.0	0.000	0.000	0.000	...	0.000	0.000	0.001	0.011	0.000	0.0	0.000	0.000
71071	2012-07-12	에어컨	0.000	0.0	0.000	0.0	0.000	0.000	0.001	...	0.001	0.000	0.000	0.000	0.000	0.0	0.000	0.000
71076	2012-07-13	에어컨	0.000	0.0	0.000	0.0	0.001	0.000	0.000	...	0.000	0.000	0.000	0.000	0.000	0.0	0.000	0.000
...
133551	2012-10-10	에어컨	0.000	0.0	0.000	0.0	0.000	0.001	0.000	...	0.000	0.000	0.000	0.000	0.000	0.0	0.000	0.001
133555	2012-10-11	에어컨	0.000	0.0	0.000	0.0	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000	0.000	0.0	0.000	0.000
133560	2012-10-12	에어컨	0.000	0.0	0.001	0.0	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000	0.001	0.0	0.000	0.000
133564	2012-10-13	에어컨	0.000	0.0	0.000	0.0	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000	0.000	0.0	0.000	0.000
133568	2012-10-14	에어컨	0.000	0.0	0.000	0.0	0.000	0.000	0.000	...	0.000	0.001	0.000	0.000	0.000	0.0	0.000	0.000

95 rows × 147 columns

```
cnt = 3
cnt1 = 149

while cnt < 146 :
    cnt = cnt + 1
    cnt1 = cnt1 + 1
    plus = ele_merge.iloc[:, [cnt, cnt1]]
    ele_x = ele_x.append(plus)
```

ele_x

	x	y
0	0.024	0.001
1	0.036	0.000
2	0.035	0.000
3	0.032	0.011
4	0.003	0.012
...
168	0.051	0.004
169	0.051	0.004
170	0.051	0.004
171	0.051	0.004
172	0.051	0.004

24912 rows × 2 columns

10분 간
총 전력량

ele_x			
	x	y	hour
0	0.024	0.001	0
1	0.036	0.000	0
2	0.035	0.000	0
3	0.032	0.011	0
4	0.003	0.012	0
...
168	0.051	0.004	23
169	0.051	0.004	23
170	0.051	0.004	23
171	0.051	0.004	23
172	0.051	0.004	23

24912 rows × 3 columns

10분 간
에어컨 전력량

```
for i in range(ele_x['x'].count()) :  
    if ele_y.iloc[i] > 0.02 :  
        ele_Y.iloc[i] = 1  
    else :  
        ele_Y.iloc[i] = 0
```



On/Off

판단기준량

```
ele_Y.value_counts()
```

```
0.0    21976  
1.0     2936  
Name: y, dtype: int64
```

ele_Y

```
0      0.0  
1      0.0  
2      0.0  
3      1.0  
4      1.0  
...  
168     0.0  
169     0.0  
170     0.0  
171     0.0  
172     0.0
```

Name: y, Length: 24912, dtype: float64

```
from sklearn.ensemble import RandomForestClassifier  
clf = RandomForestClassifier(min_samples_split=8)  
clf = clf.fit(X_train, y_train)
```

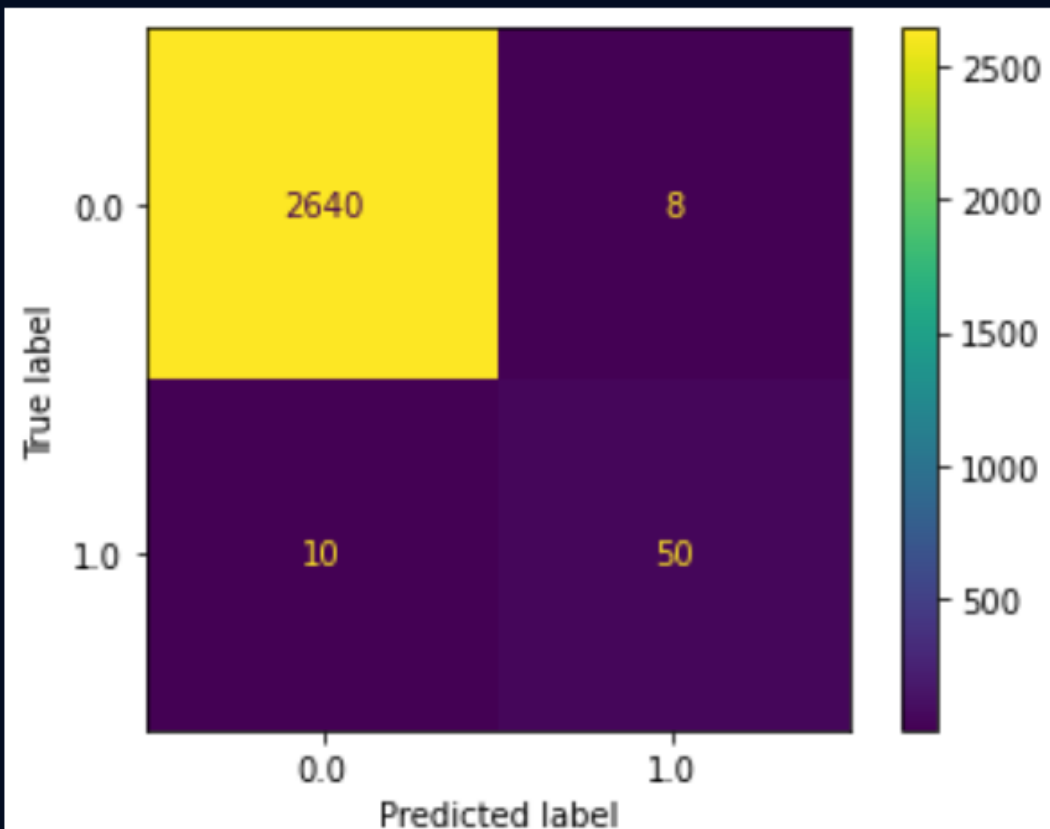
```
> ▶ ≡ M↓
```

```
pred = clf.predict(X_test)
```

```
print(accuracy_score(pred, y_test), precision_score(pred, y_test), recall_score(pred,  
y_test))
```

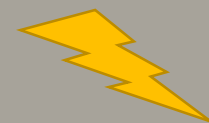
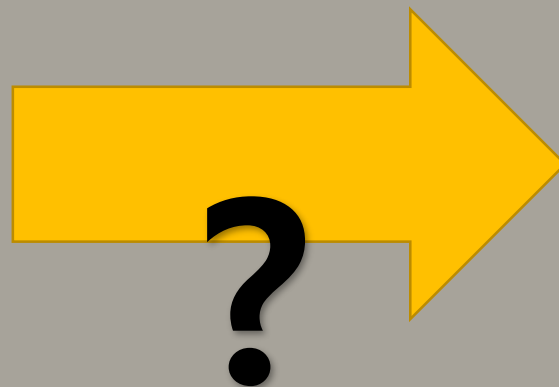
```
0.9933530280649926 0.8333333333333334 0.8620689655172413
```

```
from sklearn.metrics import plot_confusion_matrix  
import matplotlib.pyplot as plt  
plot_confusion_matrix(clf, X_test, y_test)  
plt.show()
```

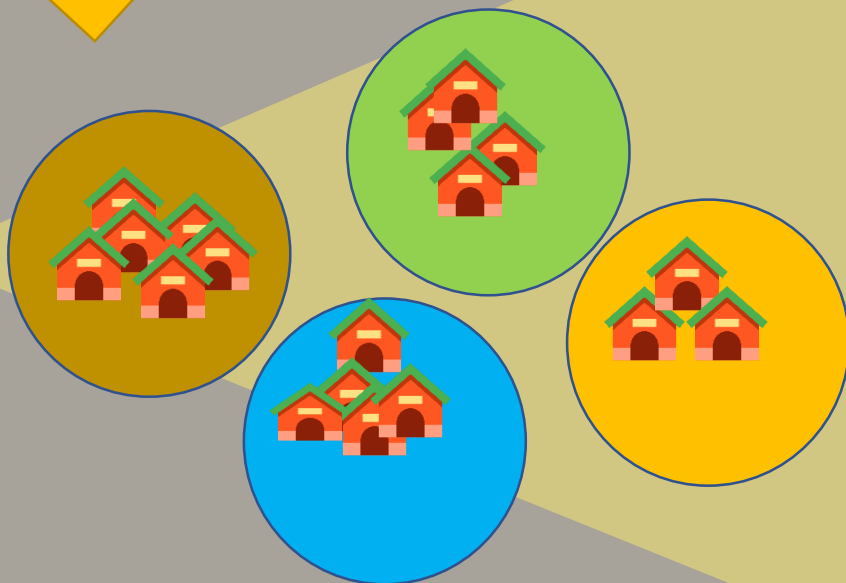
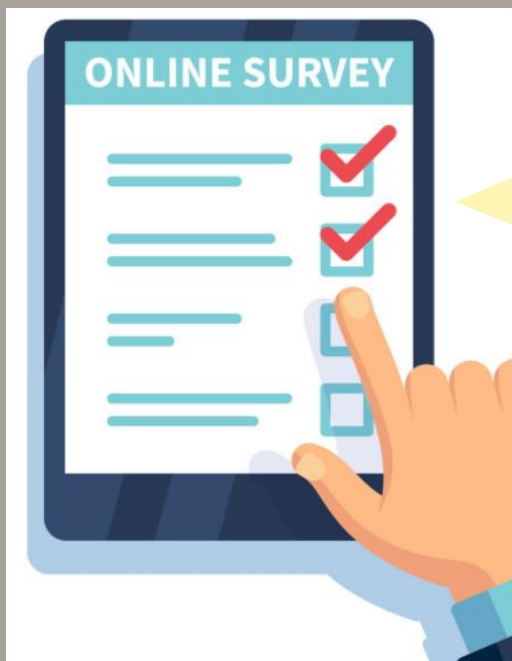
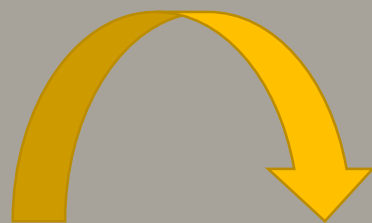


1) 결과 해석 (결과해석과 결과활용 방안)

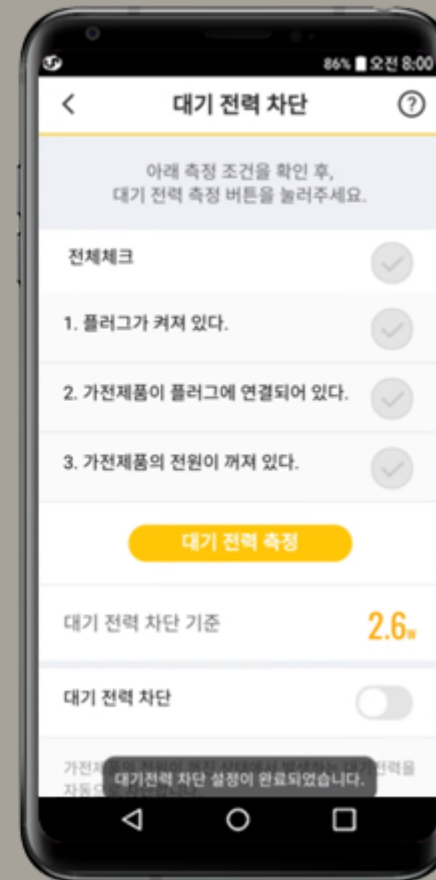
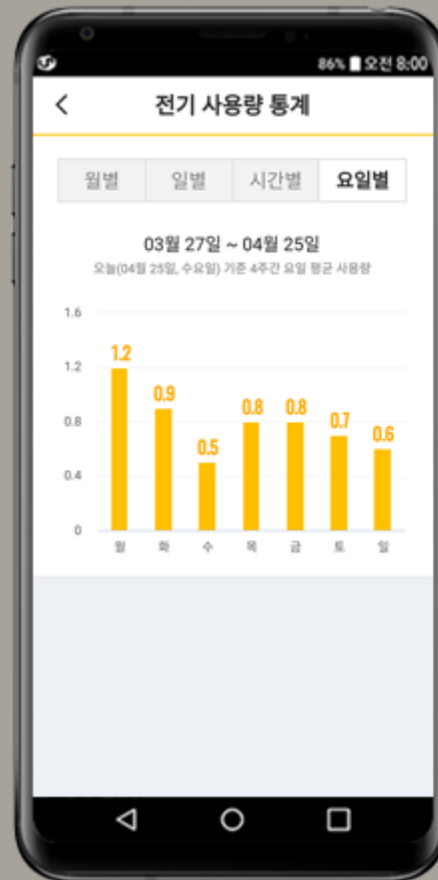
시간 별 on/off
0
0
0
1
1
1
...



시간 별 사용량
10
20
20
50
50
50
...



H_ID	HA	ST_DATE	APT_TF	M2	PersonCou
4	전력총합	2012-07-01	비아파트	20평~29평	4인이상
5	전력총합	2012-07-01	비아파트	20평~29평	4인이상
6	전력총합	2012-07-01	아파트	20평~29평	4인이상
8	TV	2012-07-01	아파트	30평~39평이상	4인이상
8	김치냉장고	2012-07-01	아파트	30평~39평이상	4인이상
8	세탁기	2012-07-01	아파트	30평~39평이상	4인이상
8	전력총합	2012-07-01	아파트	30평~39평이상	4인이상
8	전자렌지	2012-07-01	아파트	30평~39평이상	4인이상
8	정수기	2012-07-01	아파트	30평~39평이상	4인이상
19	TV	2012-07-01	비아파트	1평~19평	3인
19	냉장고	2012-07-01	비아파트	1평~19평	3인



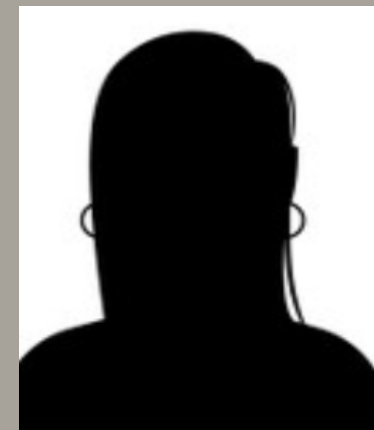
1) 분석 후기 (좋았던 점, 아쉬운 점)

TGIF !



직접 말할예정

김태규 최고!



데이콘 에서 제공
되는 데이터가 정
말 깔끔했었다는
느낌이 들었고,
앞으로 산학 프로
젝트가 설렘 반
걱정 반입니다.