

데이터 EDA

1번 용인시 거주자의 전기차 앱실행 고객수

```
In [1]: # 데이터 로드
import pandas as pd

df = pd.read_csv("../data/1.ev_app_resident.csv", encoding = "utf-8")
df
```

```
Out[1]:
```

	base_dt	dow	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	cell_xcrd	cell_ycrd
0	20220605	7	4146	용인시	41463572	동백1동	86135034	127.152250	37.286300
1	20220625	6	4146	용인시	41465580	상현2동	84375196	127.090080	37.309536
2	20220614	2	4146	용인시	41461250	포곡읍	87958893	127.216600	37.274303
3	20220619	7	4146	용인시	41465530	신봉동	84119243	127.080970	37.323480
4	20220624	5	4146	용인시	41465580	상현2동	84423187	127.091780	37.306840
...
292038	20220605	7	4146	용인시	41465570	상현1동	84071205	127.079384	37.304993
292039	20220607	2	4146	용인시	41461510	중앙동	87286847	127.193080	37.234580
292040	20220611	6	4146	용인시	41461259	남사읍	86630714	127.170300	37.156097
292041	20220618	6	4146	용인시	41461360	양지면	89926672	127.286095	37.230280
292042	20220618	6	4146	용인시	41465510	풍덕천1동	84343235	127.088860	37.326210

292043 rows × 14 columns

```
In [2]: ## 데이터 정보
df.info()
### 결측치 없다.
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 292043 entries, 0 to 292042
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   base_dt     292043 non-null  int64
1   dow         292043 non-null  int64
2   ccw_cd      292043 non-null  int64
3   ccw_nm      292043 non-null  object
4   adng_cd     292043 non-null  int64
5   adng_nm     292043 non-null  object
6   cell_id     292043 non-null  int64
7   cell_xcrd   292043 non-null  float64
8   cell_ycrd   292043 non-null  float64
9   gender      292043 non-null  object
10  age         292043 non-null  int64
11  app_web     292043 non-null  object
12  time_zone   292043 non-null  int64
13  count_cust  292043 non-null  float64
dtypes: float64(3), int64(7), object(4)
memory usage: 31.2+ MB
```

```
In [3]: df.describe()
```

Out[3]:

	base_dt	dow	ccw_cd	adng_cd	cell_id	cell_xcrd	cell_ycrd
count	2.920430e+05	292043.000000	292043.0	2.920430e+05	2.920430e+05	292043.000000	292043.000000
mean	2.022062e+07	3.910503	4146.0	4.146370e+07	8.560274e+07	127.133481	3.910503
std	8.706172e+00	1.900620	0.0	1.559666e+03	1.611870e+06	0.056916	1.900620
min	2.022060e+07	1.000000	4146.0	4.146125e+07	8.274341e+07	127.032210	3.910503
25%	2.022061e+07	2.000000	4146.0	4.146351e+07	8.443909e+07	127.092210	3.910503
50%	2.022062e+07	4.000000	4146.0	4.146357e+07	8.509495e+07	127.115350	3.910503
75%	2.022062e+07	5.000000	4146.0	4.146553e+07	8.637495e+07	127.160500	3.910503
max	2.022063e+07	7.000000	4146.0	4.146559e+07	9.383021e+07	127.423720	3.910503

```
In [4]: df.sort_values(by=["adng_nm", "base_dt", "time_zone", "age", "cell_id"]).head(50)
```

Out[4]:

	base_dt	dow	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	cell_xcrd	cell_ycrd
235000	20220601	3	4146	용인시	41463520	구갈동	84935083	127.110010	37.274452
177177	20220601	3	4146	용인시	41463520	구갈동	84999101	127.112210	37.284374
183152	20220601	3	4146	용인시	41463520	구갈동	84983098	127.111660	37.282570
182513	20220601	3	4146	용인시	41463520	구갈동	85143080	127.117320	37.278980
116127	20220601	3	4146	용인시	41463520	구갈동	85207063	127.119600	37.273130
145325	20220601	3	4146	용인시	41463520	구갈동	85207065	127.119600	37.274033
281292	20220601	3	4146	용인시	41463520	구갈동	84951083	127.110565	37.274902
278826	20220601	3	4146	용인시	41463520	구갈동	84967084	127.111130	37.275806
185628	20220601	3	4146	용인시	41463520	구갈동	85255066	127.121280	37.275840
254224	20220601	3	4146	용인시	41463520	구갈동	84999096	127.112220	37.282120
90721	20220601	3	4146	용인시	41463520	구갈동	85239075	127.120700	37.279440
54653	20220601	3	4146	용인시	41463520	구갈동	85255058	127.121300	37.272232
44796	20220601	3	4146	용인시	41463520	구갈동	85335065	127.124090	37.277653
79578	20220601	3	4146	용인시	41463520	구갈동	85447058	127.128040	37.277664
201288	20220601	3	4146	용인시	41463520	구갈동	85591031	127.133156	37.269566
80090	20220601	3	4146	용인시	41463520	구갈동	84999100	127.112210	37.283924
103193	20220601	3	4146	용인시	41463520	구갈동	85079083	127.115060	37.278522
115357	20220601	3	4146	용인시	41463520	구갈동	85143067	127.117350	37.273120
16646	20220601	3	4146	용인시	41463520	구갈동	85207061	127.119606	37.272230
123604	20220601	3	4146	용인시	41463520	구갈동	85255058	127.121300	37.272232
172393	20220601	3	4146	용인시	41463520	구갈동	85527034	127.130900	37.269108
201770	20220601	3	4146	용인시	41463520	구갈동	84935093	127.109985	37.278957
263214	20220601	3	4146	용인시	41463520	구갈동	85127075	127.116770	37.276276
16922	20220601	3	4146	용인시	41463520	구갈동	85255058	127.121300	37.272232
205885	20220601	3	4146	용인시	41463520	구갈동	85431052	127.127490	37.274506
215816	20220601	3	4146	용인시	41463520	구갈동	85495041	127.129760	37.271360
121878	20220601	3	4146	용인시	41463520	구갈동	85607030	127.133720	37.269570
85493	20220601	3	4146	용인시	41463520	구갈동	85015096	127.112785	37.282574
79522	20220601	3	4146	용인시	41463520	구갈동	85063093	127.114480	37.282578
66508	20220601	3	4146	용인시	41463520	구갈동	85079089	127.115050	37.281227
81935	20220601	3	4146	용인시	41463520	구갈동	85127079	127.116750	37.278076
82775	20220601	3	4146	용인시	41463520	구갈동	85159078	127.117880	37.278534
208329	20220601	3	4146	용인시	41463520	구갈동	85447045	127.128070	37.271805
252568	20220601	3	4146	용인시	41463520	구갈동	85495041	127.129760	37.271360
186742	20220601	3	4146	용인시	41463520	구갈동	85079090	127.115040	37.281677
214156	20220601	3	4146	용인시	41463520	구갈동	85207061	127.119606	37.272230

	base_dt	dow	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	cell_xcrd	cell_ycrd
95440	20220601	3	4146	용인시	41463520	구갈동	85255073	127.121260	37.278996
226101	20220601	3	4146	용인시	41463520	구갈동	85271074	127.121826	37.279896
1870	20220601	3	4146	용인시	41463520	구갈동	85527053	127.130860	37.277670
214708	20220601	3	4146	용인시	41463520	구갈동	85079084	127.115060	37.278973
123886	20220601	3	4146	용인시	41463520	구갈동	85111071	127.116210	37.274020
240987	20220601	3	4146	용인시	41463520	구갈동	84935085	127.110000	37.275352
147928	20220601	3	4146	용인시	41463520	구갈동	84999101	127.112210	37.284374
161866	20220601	3	4146	용인시	41463520	구갈동	85127085	127.116745	37.280780
250841	20220601	3	4146	용인시	41463520	구갈동	85239058	127.120735	37.271782
290166	20220601	3	4146	용인시	41463520	구갈동	85239061	127.120730	37.273132
248461	20220601	3	4146	용인시	41463520	구갈동	85159061	127.117920	37.270870
202463	20220601	3	4146	용인시	41463520	구갈동	85047092	127.113914	37.281673
54944	20220601	3	4146	용인시	41463520	구갈동	85143080	127.117320	37.278980
60942	20220601	3	4146	용인시	41463520	구갈동	85271074	127.121826	37.279896

1) 날짜

```
In [5]: ## 날짜열
df.base_dt.unique()
### 30일치
```

```
Out[5]: array([20220605, 20220625, 20220614, 20220619, 20220624, 20220610,
        20220612, 20220620, 20220609, 20220617, 20220615, 20220621,
        20220616, 20220606, 20220608, 20220613, 20220607, 20220618,
        20220601, 20220622, 20220611, 20220629, 20220627, 20220628,
        20220623, 20220602, 20220626, 20220604, 20220603, 20220630],
        dtype=int64)
```

```
In [6]: ## 날짜열
df.base_dt.value_counts()
```

```
Out[6]: 20220608    10757
        20220603    10753
        20220621    10677
        20220610    10667
        20220628    10656
        20220622    10542
        20220602    10420
        20220629    10377
        20220620    10245
        20220607    10128
        20220624    10119
        20220609    10055
        20220623     9974
        20220627     9961
        20220615     9952
        20220613     9889
        20220630     9741
        20220611     9641
        20220614     9585
        20220617     9562
        20220604     9551
        20220601     9519
        20220625     9488
        20220618     9251
        20220605     8998
        20220619     8640
        20220612     8518
        20220616     8501
        20220626     8264
        20220606     7612
Name: base_dt, dtype: int64
```

2) 요일

- 월 ~ 일
- 1 ~ 7

```
In [7]: ## 요일 열
        df.dow.value_counts()
```

```
Out[7]: 3    51147
        4    48691
        5    41101
        2    41046
        6    37931
        1    37707
        7    34420
Name: dow, dtype: int64
```

3) 시군구코드

```
In [8]: df.ccw_cd.value_counts()
        # 제거해도 무방할수도?
```

```
Out[8]: 4146    292043
Name: ccw_cd, dtype: int64
```

4) 시군구명

```
In [9]: df.ccw_nm.value_counts()
```

```
Out[9]: 용인시      292043
        Name: ccw_nm, dtype: int64
```

5) 행정동코드

```
In [10]: df.adng_cd.value_counts()
```

```
Out[10]: 41463520      13819
         41465540      13640
         41461520      13456
         41465590      12615
         41465560      12160
         41465520      11760
         41465570      11491
         41463510      11129
         41463560      11029
         41465530      10955
         41463535      10520
         41465510      10395
         41463590       9684
         41463572       9627
         41461253       9371
         41463570       9325
         41461530       9048
         41465580       8940
         41461250       8872
         41463586       8809
         41463575       7431
         41463516       7264
         41463517       6726
         41461510       6232
         41461360       6091
         41465550       5808
         41463577       5296
         41461259       5285
         41461256       5053
         41463550       4954
         41463540       4527
         41463530       4488
         41461540       3020
         41461340       2177
         41461350       1046
        Name: adng_cd, dtype: int64
```

```
In [11]: df.adng_cd.nunique()
```

```
Out[11]: 35
```

6) 행정동명

```
In [12]: df.adng_nm.value_counts()
```

```
Out[12]:
```

구갈동	13819
죽전1동	13640
역삼동	13456
성북동	12615
동천동	12160
풍덕천2동	11760
상현1동	11491
신갈동	11129
구성동	11029
신봉동	10955
보라동	10520
풍덕천1동	10395
보정동	9684
동백1동	9627
모현읍	9371
마북동	9325
유림동	9048
상현2동	8940
포곡읍	8872
상하동	8809
동백2동	7431
영덕1동	7264
영덕2동	6726
중앙동	6232
양지면	6091
죽전2동	5808
동백3동	5296
남사읍	5285
이동읍	5053
서농동	4954
기흥동	4527
상갈동	4488
동부동	3020
원삼면	2177
백암면	1046

Name: adng_nm, dtype: int64

7) 셀ID

- 지형지물정보 셀 ID

```
In [13]: df.cell_id.unique()
```

```
Out[13]: 12381
```

8) Cell center X

- 셀 중심점 Y 좌표 위도

```
In [14]: df.cell_xcrd.unique()
```

```
Out[14]: 7919
```

```
In [15]: df.cell_xcrd.describe()
```

```
Out[15]: count      292043.000000
         mean       127.133481
         std        0.056916
         min       127.032210
         25%       127.092210
         50%       127.115350
         75%       127.160500
         max       127.423720
         Name: cell_xcrd, dtype: float64
```

9) Cell center Y

- 셀 중심점 X 좌표 경도

```
In [16]: df.cell_ycrd.unique()
```

```
Out[16]: 9371
```

```
In [17]: df.cell_ycrd.describe()
```

```
Out[17]: count      292043.000000
         mean       37.283523
         std        0.042515
         min       37.087600
         25%       37.258533
         50%       37.284970
         75%       37.318108
         max       37.370174
         Name: cell_ycrd, dtype: float64
```

10) 성별

```
In [18]: df.gender.value_counts()
```

```
Out[18]: MALE      190474
         FEMALE    101569
         Name: gender, dtype: int64
```

11) 연령

- 1 10세미만
- 2 10세이상-15세미만
- 3 15세이상-20세미만
- 4 20세이상-25세미만
- 5 25세이상-30세미만
- 6 30세이상-35세미만
- 7 35세이상-40세미만
- 8 40세이상-45세미만
- 9 45세이상-50세미만
- 10 50세이상-55세미만
- 11 55세이상-60세미만
- 12 60세이상-66세미만
- 13 66세이상-70세미만
- 14 70세 이상


```
In [19]: df.age.value_counts()
#### 의문 20세미만도 포함을 시켜야하나?
```

```
Out[19]: 10    39480
         8     37380
         11    35061
         9     32969
         12    26936
         14    22767
         7     22244
         13    20430
         6     18820
         5     16132
         4      9069
         3      7426
         2      3077
         1       252
Name: age, dtype: int64
```

12) 앱웹

```
In [20]: df.app_web.unique()
```

```
Out[20]: array(['EV라운지', '파워큐브차저', '하이차저', 'G차저', '모두의충전', 'EV Infra',
               '차지비',
               'Epit', '에버온', '일렉배리', '채비', 'KEPCO PLUG', '환경부 통합 누리집', '해
               피차저',
               'evWhere', '테슬라', '에너지플러스EV'], dtype=object)
```

```
In [21]: df.app_web.value_counts()
```

```
Out[21]: EV라운지          242568
         하이차저          42469
         파워큐브차저        1674
         G차저            1594
         EV Infra         1555
         차지비            484
         에버온           385
         KEPCO PLUG        291
         환경부 통합 누리집    256
         모두의충전         247
         일렉배리          241
         Epit             143
         채비             106
         테슬라            14
         evWhere           13
         에너지플러스EV         2
         해피차저           1
Name: app_web, dtype: int64
```

```
In [22]: df.app_web.nunique()
```

```
Out[22]: 17
```

13) 시간

- 10시~6시
- 6시~9시
- 9시~12시

- 4 12시~18시
- 5 18시~21시
- 6 21시~24시

In [23]: `df.time_zone.value_counts()`

Out[23]:

```

4    91537
3    52764
5    50688
2    39157
6    32383
1    25514
Name: time_zone, dtype: int64

```

14) 실행자수

In [24]: `df.count_cust.describe()`

Out[24]:

```

count    292043.000000
mean         6.570716
std         4.424778
min         3.011000
25%         3.802000
50%         4.535000
75%         8.098000
max        177.332000
Name: count_cust, dtype: float64

```

In [25]: `df.loc[df.count_cust >= 100]`

Out[25]:

	base_dt	dow	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	cell_xcrd	cell_ycrd
9367	20220605	7	4146	용인시	41463572	동백1동	86503006	127.16524	37.284084
63453	20220608	3	4146	용인시	41465510	풍덕천1동	84567214	127.09678	37.323082
150449	20220608	3	4146	용인시	41465510	풍덕천1동	84567214	127.09678	37.323082