

데이터 EDA

2번 용인시 거주자의 전기차 앱실행 고객수

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

df = pd.read_csv("../data/2.ev_app_activity.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	20220614	2	4146	용인시	41461256	이동읍	87958702	127.216920	37.188220
1	20220613	1	4146	용인시	41463570	마북동	85431116	127.127350	37.303352
2	20220602	4	4146	용인시	41463575	동백2동	86198990	127.154590	37.268280
3	20220606	1	4146	용인시	41461259	남사읍	86854741	127.178110	37.174600
4	20220608	3	4146	용인시	41465540	죽전1동	85159183	127.117645	37.325855
...
264708	20220621	2	4146	용인시	41461253	모현읍	86023090	127.148200	37.308372
264709	20220628	2	4146	용인시	41465530	신봉동	84103247	127.080400	37.324825
264710	20220626	7	4146	용인시	41463510	신갈동	84375154	127.090180	37.290607
264711	20220620	1	4146	용인시	41465590	성북동	83895245	127.073105	37.318040
264712	20220601	3	4146	용인시	41463520	구갈동	85111078	127.116196	37.277176

264713 rows × 14 columns

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 264713 entries, 0 to 264712
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   base_dt    264713 non-null  int64
 1   dow        264713 non-null  int64
 2   ccw_cd     264713 non-null  int64
 3   ccw_nm     264713 non-null  object
 4   adng_cd    264713 non-null  int64
 5   adng_nm    264713 non-null  object
 6   cell_id    264713 non-null  int64
 7   cell_xcrd  264713 non-null  float64
 8   cell_ycrd  264713 non-null  float64
 9   gender     264713 non-null  object
10  age        264713 non-null  int64
11  app_web    264713 non-null  object
12  time_zone  264713 non-null  int64
13  count_cust 264713 non-null  float64
dtypes: float64(3), int64(7), object(4)
memory usage: 28.3+ MB
```

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

Out[3]:

	base_dt	dow	ccw_cd	adng_cd	cell_id	cell_xcrd	
count	2.647130e+05	264713.000000	264713.0	2.647130e+05	2.647130e+05	264713.000000	264713.0
mean	2.022062e+07	3.916963	4146.0	4.146337e+07	8.586557e+07	127.142797	3.916963
std	8.717480e+00	1.905967	0.0	1.569749e+03	1.919799e+06	0.067744	1.905967
min	2.022060e+07	1.000000	4146.0	4.146125e+07	8.269542e+07	127.030500	3.916963
25%	2.022061e+07	2.000000	4146.0	4.146152e+07	8.447099e+07	127.093290	3.916963
50%	2.022062e+07	4.000000	4146.0	4.146354e+07	8.519117e+07	127.119030	3.916963
75%	2.022062e+07	6.000000	4146.0	4.146551e+07	8.691885e+07	127.180100	3.916963
max	2.022063e+07	7.000000	4146.0	4.146559e+07	9.378225e+07	127.422020	3.916963

```
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
225894	20220601	3	4146	용인시	41463520	구갈동	84935083	127.110010	37.274452
196287	20220601	3	4146	용인시	41463520	구갈동	84999101	127.112210	37.284374
32744	20220601	3	4146	용인시	41463520	구갈동	85207063	127.119600	37.273130
113822	20220601	3	4146	용인시	41463520	구갈동	84951083	127.110565	37.274902
134217	20220601	3	4146	용인시	41463520	구갈동	84967084	127.111130	37.275806
194829	20220601	3	4146	용인시	41463520	구갈동	85175082	127.118430	37.280790
13238	20220601	3	4146	용인시	41463520	구갈동	85255066	127.121280	37.275840
21827	20220601	3	4146	용인시	41463520	구갈동	85335050	127.124120	37.270890
160479	20220601	3	4146	용인시	41463520	구갈동	84983101	127.111650	37.283920
15959	20220601	3	4146	용인시	41463520	구갈동	85111088	127.116170	37.281680
210227	20220601	3	4146	용인시	41463520	구갈동	85207077	127.119570	37.279438
54112	20220601	3	4146	용인시	41463520	구갈동	85447058	127.128040	37.277664
20688	20220601	3	4146	용인시	41463520	구갈동	85591031	127.133156	37.269566
133052	20220601	3	4146	용인시	41463520	구갈동	85175082	127.118430	37.280790
252643	20220601	3	4146	용인시	41463520	구갈동	85207061	127.119606	37.272230
219627	20220601	3	4146	용인시	41463520	구갈동	85239079	127.120690	37.281246
55804	20220601	3	4146	용인시	41463520	구갈동	85527034	127.130900	37.269108
86354	20220601	3	4146	용인시	41463520	구갈동	85527044	127.130875	37.273617
108045	20220601	3	4146	용인시	41463520	구갈동	84935093	127.109985	37.278957
222317	20220601	3	4146	용인시	41463520	구갈동	85127077	127.116760	37.277176
43648	20220601	3	4146	용인시	41463520	구갈동	85447054	127.128044	37.275860
126206	20220601	3	4146	용인시	41463520	구갈동	85031097	127.113340	37.283474
45226	20220601	3	4146	용인시	41463520	구갈동	85063093	127.114480	37.282578
238809	20220601	3	4146	용인시	41463520	구갈동	85079089	127.115050	37.281227
86611	20220601	3	4146	용인시	41463520	구갈동	85127079	127.116750	37.278076
163214	20220601	3	4146	용인시	41463520	구갈동	85447045	127.128070	37.271805
229058	20220601	3	4146	용인시	41463520	구갈동	84999085	127.112250	37.277160
24482	20220601	3	4146	용인시	41463520	구갈동	85207061	127.119606	37.272230
165876	20220601	3	4146	용인시	41463520	구갈동	85255059	127.121290	37.272686
82380	20220601	3	4146	용인시	41463520	구갈동	85255071	127.121270	37.278090
150946	20220601	3	4146	용인시	41463520	구갈동	85527053	127.130860	37.277670
12004	20220601	3	4146	용인시	41463520	구갈동	85079080	127.115070	37.277172
198534	20220601	3	4146	용인시	41463520	구갈동	85079084	127.115060	37.278973
164699	20220601	3	4146	용인시	41463520	구갈동	85111071	127.116210	37.274020
93053	20220601	3	4146	용인시	41463520	구갈동	84999101	127.112210	37.284374
236913	20220601	3	4146	용인시	41463520	구갈동	85095075	127.115640	37.275370

	base_dt	dow	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	cell_xcrd	cell_ycrd
133581	20220601	3	4146	용인시	41463520	구갈동	85239058	127.120735	37.271782
209477	20220601	3	4146	용인시	41463520	구갈동	85607031	127.133710	37.270020
209521	20220601	3	4146	용인시	41463520	구갈동	84967084	127.111130	37.275806
166837	20220601	3	4146	용인시	41463520	구갈동	85031089	127.113365	37.279870
14688	20220601	3	4146	용인시	41463520	구갈동	85159081	127.117874	37.279884
159180	20220601	3	4146	용인시	41463520	구갈동	85175082	127.118430	37.280790
237705	20220601	3	4146	용인시	41463520	구갈동	85335050	127.124120	37.270890
145448	20220601	3	4146	용인시	41463520	구갈동	85527050	127.130870	37.276320
84729	20220601	3	4146	용인시	41463520	구갈동	85111088	127.116170	37.281680
251174	20220601	3	4146	용인시	41463520	구갈동	85175080	127.118440	37.279890
202190	20220601	3	4146	용인시	41463520	구갈동	85207077	127.119570	37.279438
16730	20220601	3	4146	용인시	41463520	구갈동	85447058	127.128040	37.277664
20145	20220601	3	4146	용인시	41463520	구갈동	85063074	127.114525	37.274014
199030	20220601	3	4146	용인시	41463520	구갈동	85175069	127.118460	37.274930

1) 날짜

```
In [5]: ## 날짜 열
df.base_dt.nunique()
#### 30일 치
```

```
Out[5]: 30
```

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

```
Out[6]: 20220603    9787
        20220608    9675
        20220610    9673
        20220628    9633
        20220621    9552
        20220622    9457
        20220629    9416
        20220602    9353
        20220620    9214
        20220624    9195
        20220607    9187
        20220627    9074
        20220609    9045
        20220623    9015
        20220613    9007
        20220615    8850
        20220630    8808
        20220601    8779
        20220611    8755
        20220604    8707
        20220614    8632
        20220625    8601
        20220617    8534
        20220618    8532
        20220605    8237
        20220619    7890
        20220612    7873
        20220626    7620
        20220616    7562
        20220606    7050
Name: base_dt, dtype: int64
```

2) 요일

- 월 ~ 일
- 1 ~ 7

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

```
Out[7]: 3    46177
        4    43783
        5    37189
        2    37004
        6    34595
        1    34345
        7    31620
Name: dow, dtype: int64
```

3) 시군구코드

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

```
Out[8]: 4146    264713
Name: ccw_cd, dtype: int64
```

4) 시군구명

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

```
Out[9]: 용인시      264713
        Name: ccw_nm, dtype: int64
```

5) 행정동코드

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

```
Out[10]: 41463550      16892
         41461253      14135
         41465540      10591
         41461520      10136
         41463540       9694
         41463520       9683
         41461250       9641
         41463570       9427
         41461259       9078
         41463590       8945
         41463510       8417
         41461530       8142
         41465590       8137
         41465570       8101
         41465560       7920
         41463535       7861
         41465520       7628
         41463516       7364
         41463560       7355
         41461360       7320
         41465530       7200
         41465510       7103
         41461256       6796
         41461510       6482
         41463575       6184
         41463517       5559
         41463586       5242
         41463572       5238
         41465550       5236
         41461350       4751
         41465580       4383
         41463577       4344
         41463530       3715
         41461340       3085
         41461540       2928
        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]:
```

서농동	16892
모현읍	14135
죽전1동	10591
역삼동	10136
기흥동	9694
구갈동	9683
포곡읍	9641
마북동	9427
남사읍	9078
보정동	8945
신갈동	8417
유림동	8142
성북동	8137
상현1동	8101
동천동	7920
보라동	7861
풍덕천2동	7628
영덕1동	7364
구성동	7355
양지면	7320
신봉동	7200
풍덕천1동	7103
이동읍	6796
중앙동	6482
동백2동	6184
영덕2동	5559
상하동	5242
동백1동	5238
죽전2동	5236
백암면	4751
상현2동	4383
동백3동	4344
상갈동	3715
원삼면	3085
동부동	2928

Name: adng_nm, dtype: int64

7) 셀ID

- 지형지물정보 셀 ID

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

```
Out[13]: 13742
```

8) Cell center X

- 셀 중심점 Y 좌표 위도

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

```
Out[14]: 9086
```

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

```
Out[15]: count      264713.000000
         mean       127.142797
         std        0.067744
         min       127.030500
         25%       127.093290
         50%       127.119030
         75%       127.180100
         max       127.422020
         Name: cell_xcrd, dtype: float64
```

9) Cell center Y

- 셀 중심점 X 좌표 경도

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

```
Out[16]: 10634
```

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

```
Out[17]: count      264713.000000
         mean       37.272698
         std        0.050756
         min       37.086697
         25%       37.239090
         50%       37.276700
         75%       37.315323
         max       37.370174
         Name: cell_ycrd, dtype: float64
```

10) 성별

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

```
Out[18]: MALE      165260
         FEMALE    99453
         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    34579
        11    31928
         8    29967
         9    28455
        12    25838
        14    21443
        13    20714
         7    18880
         6    18233
         5    15996
         4     8232
         3     7166
         2     3023
         1       259
        Name: age, dtype: int64
```

12) 앱웹

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

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

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

```
Out[21]: EV라운지          221980
        하이차저          37037
        G차저            1534
        파워큐브차저        1251
        EV Infra          1247
        차지비           387
        환경부 통합 누리집    301
        KEPCO PLUG         247
        에버온            231
        모두의충전          190
        Epit              122
        일렉배리           112
        채비              56
        테슬라             13
        evWhere            3
        해피차저            1
        에너지플러스EV        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    84145
3    47692
5    45752
2    34366
6    28633
1    24125
Name: time_zone, dtype: int64
```

14) 실행자수

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

Out[24]:

```

count    264713.000000
mean         6.690827
std         4.712933
min         3.001000
25%         3.791000
50%         4.620000
75%         8.133000
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
16948	20220605	7	4146	용인시	41463572	동백1동	86503006	127.165240	37.284084
36189	20220623	4	4146	용인시	41463550	서농동	83943045	127.075294	37.229260
69013	20220614	2	4146	용인시	41463550	서농동	83943045	127.075294	37.229260
70980	20220608	3	4146	용인시	41465510	풍덕천1동	84567216	127.096770	37.323982
104908	20220608	3	4146	용인시	41465510	풍덕천1동	84567216	127.096770	37.323982