

# 데이터 EDA

## 3번 용인시 거주자의 타지역 전기차 앱실행 고객수

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

df = pd.read_csv("../data/3.ev_app_activity_resident.csv", encoding = "utf-8")
df
```

```
Out[1]:
```

	base_dt	dow	mega_cd	mega_nm	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	
0	20220611	6	41	경기도	4146	용인시	41461253	모현읍	86695054	1
1	20220605	7	11	서울특별시	1168	강남구	11680580	삼성1동	83591688	1
2	20220602	4	11	서울특별시	1120	성동구	11200670	성수2가1동	83463765	1
3	20220608	3	41	경기도	4146	용인시	41463510	신갈동	84951117	1
4	20220608	3	41	경기도	4150	이천시	41500330	호법면	92774456	1
...	...	...	...	...	...	...	...	...	...	...
293166	20220609	4	41	경기도	4146	용인시	41465530	신봉동	83943267	1
293167	20220626	7	41	경기도	4146	용인시	41465540	죽전1동	85719164	1
293168	20220607	2	41	경기도	4113	성남시	41131625	위례동	85863462	1
293169	20220619	7	11	서울특별시	1168	강남구	11680640	역삼1동	82791718	1
293170	20220605	7	41	경기도	4146	용인시	41463516	영덕1동	83879146	1

293171 rows × 16 columns

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 293171 entries, 0 to 293170
Data columns (total 16 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   base_dt     293171 non-null  int64
1   dow         293171 non-null  int64
2   mega_cd     293171 non-null  int64
3   mega_nm     293171 non-null  object
4   ccw_cd      293171 non-null  int64
5   ccw_nm      293171 non-null  object
6   adng_cd     293171 non-null  int64
7   adng_nm     293171 non-null  object
8   cell_id     293171 non-null  int64
9   cell_xcrd   293171 non-null  float64
10  cell_ycrd   293171 non-null  float64
11  gender      293171 non-null  object
12  age         293171 non-null  int64
13  app_web     293171 non-null  object
14  time_zone   293171 non-null  int64
15  count_cust  293171 non-null  float64
dtypes: float64(3), int64(8), object(5)
memory usage: 35.8+ MB
```

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

Out[3]:

	base_dt	dow	mega_cd	ccw_cd	adng_cd	cell_id
<b>count</b>	2.931710e+05	293171.000000	293171.000000	293171.000000	2.931710e+05	2.931710e+05
<b>mean</b>	2.022062e+07	3.909449	37.597419	3802.462464	3.802770e+07	8.512723e+07
<b>std</b>	8.704484e+00	1.899829	9.490771	945.279777	9.453587e+06	3.278407e+06
<b>min</b>	2.022060e+07	1.000000	11.000000	1111.000000	1.111052e+07	1.754859e+07
<b>25%</b>	2.022061e+07	2.000000	41.000000	4122.000000	4.122054e+07	8.392714e+07
<b>50%</b>	2.022062e+07	4.000000	41.000000	4146.000000	4.146352e+07	8.495104e+07
<b>75%</b>	2.022062e+07	5.000000	41.000000	4146.000000	4.146551e+07	8.637392e+07
<b>max</b>	2.022063e+07	7.000000	44.000000	4482.000000	4.482535e+07	1.430279e+08

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

Out[4]:

	base_dt	dow	mega_cd	mega_nm	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	
<b>171915</b>	20220603	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>48914</b>	20220603	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>2209</b>	20220608	3	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>6257</b>	20220608	3	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>114821</b>	20220609	4	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>32305</b>	20220610	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>77833</b>	20220610	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>89225</b>	20220610	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>113111</b>	20220610	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>144638</b>	20220610	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>144161</b>	20220610	5	43	충청북도	4311	청주시	43113747	가경동	93845061	1
<b>38204</b>	20220611	6	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>149040</b>	20220612	7	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>92015</b>	20220612	7	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>214939</b>	20220612	7	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>60240</b>	20220613	1	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>230665</b>	20220613	1	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>125000</b>	20220613	1	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>263288</b>	20220614	2	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>138100</b>	20220615	3	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>14474</b>	20220615	3	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>261404</b>	20220616	4	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>86839</b>	20220617	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>219071</b>	20220617	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>37593</b>	20220618	6	43	충청북도	4311	청주시	43113747	가경동	93845061	1
<b>46150</b>	20220620	1	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>139125</b>	20220620	1	43	충청북도	4311	청주시	43113747	가경동	94149054	1

	base_dt	dow	mega_cd	mega_nm	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	
<b>89033</b>	20220622	3	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>152330</b>	20220624	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>273894</b>	20220624	5	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>135363</b>	20220626	7	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>280253</b>	20220626	7	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>39324</b>	20220627	1	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>35571</b>	20220627	1	43	충청북도	4311	청주시	43113747	가경동	94149054	1
<b>193922</b>	20220601	3	41	경기도	4167	여주시	41670250	가남읍	97254162	1
<b>22310</b>	20220601	3	41	경기도	4167	여주시	41670250	가남읍	97382180	1
<b>26311</b>	20220601	3	41	경기도	4167	여주시	41670250	가남읍	97846220	1
<b>83668</b>	20220601	3	41	경기도	4167	여주시	41670250	가남읍	97382180	1
<b>148312</b>	20220601	3	41	경기도	4167	여주시	41670250	가남읍	97910182	1
<b>2324</b>	20220601	3	41	경기도	4167	여주시	41670250	가남읍	97382180	1
<b>37701</b>	20220602	4	41	경기도	4167	여주시	41670250	가남읍	97382180	1
<b>286443</b>	20220602	4	41	경기도	4167	여주시	41670250	가남읍	97910182	1
<b>89890</b>	20220602	4	41	경기도	4167	여주시	41670250	가남읍	97382180	1
<b>13569</b>	20220602	4	41	경기도	4167	여주시	41670250	가남읍	97382180	1
<b>281543</b>	20220603	5	41	경기도	4167	여주시	41670250	가남읍	97302217	1
<b>171638</b>	20220603	5	41	경기도	4167	여주시	41670250	가남읍	97910182	1
<b>71565</b>	20220603	5	41	경기도	4167	여주시	41670250	가남읍	97846220	1
<b>225007</b>	20220603	5	41	경기도	4167	여주시	41670250	가남읍	97382180	1
<b>135321</b>	20220603	5	41	경기도	4167	여주시	41670250	가남읍	97846220	1
<b>258341</b>	20220604	6	41	경기도	4167	여주시	41670250	가남읍	97846220	1

```
In [5]: df.loc[df.ccw_nm == "용인시"]
```

```
## 용인시 데이터가 2/3인데? 타지역이라는 이야기의 정의가...?
```

Out[5]:

	base_dt	dow	mega_cd	mega_nm	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	
<b>0</b>	20220611	6	41	경기도	4146	용인시	41461253	모현읍	86695054	1
<b>3</b>	20220608	3	41	경기도	4146	용인시	41463510	신갈동	84951117	1
<b>6</b>	20220608	3	41	경기도	4146	용인시	41463535	보라동	85175019	1
<b>7</b>	20220607	2	41	경기도	4146	용인시	41463510	신갈동	84887106	1
<b>8</b>	20220614	2	41	경기도	4146	용인시	41465550	죽전2동	84967209	1
...	...	...	...	...	...	...	...	...	...	...
<b>293164</b>	20220626	7	41	경기도	4146	용인시	41463516	영덕1동	84103141	1
<b>293165</b>	20220623	4	41	경기도	4146	용인시	41463586	상하동	85814987	1
<b>293166</b>	20220609	4	41	경기도	4146	용인시	41465530	신봉동	83943267	1
<b>293167</b>	20220626	7	41	경기도	4146	용인시	41465540	죽전1동	85719164	1
<b>293170</b>	20220605	7	41	경기도	4146	용인시	41463516	영덕1동	83879146	1

192557 rows × 16 columns

## 1) 날짜

```
In [6]: ## 날짜열
df.base_dt.unique()
### 30일 치
```

Out[6]: 30

```
In [7]: ## 날짜열
df.base_dt.value_counts()
```

```
Out[7]: 20220608    10834
        20220603    10788
        20220610    10735
        20220621    10728
        20220628    10697
        20220622    10576
        20220602    10475
        20220629    10402
        20220620    10267
        20220607    10190
        20220624    10161
        20220609    10122
        20220623    10014
        20220627    10012
        20220615     9975
        20220613     9915
        20220630     9764
        20220611     9684
        20220614     9628
        20220617     9627
        20220604     9562
        20220601     9540
        20220625     9500
        20220618     9300
        20220605     9027
        20220619     8663
        20220616     8562
        20220612     8499
        20220626     8275
        20220606     7649
Name: base_dt, dtype: int64
```

## 2) 요일

- 월 ~ 일
- 1 ~ 7

```
In [8]: ## 요일 열
        df.dow.value_counts()
```

```
Out[8]: 3    51327
        4    48937
        5    41311
        2    41243
        6    38046
        1    37843
        7    34464
Name: dow, dtype: int64
```

## 3) 시군구코드

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

```
Out[9]: 4146      192557
        4113      18487
        4111      14276
        1168       8973
        4159       8351
        ...
        4471         1
        4217         1
        2871         1
        4219         1
        4378         1
Name: ccw_cd, Length: 111, dtype: int64
```

## 4) 시군구명

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

```
Out[10]: 용인시      192557
        성남시      18487
        수원시      14276
        강남구       8973
        화성시       8351
        ...
        금산군         1
        동해시         1
        강화군         1
        태백시         1
        증평군         1
Name: ccw_nm, Length: 107, dtype: int64
```

## 5) 행정동코드

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

```
Out[11]: 41461253      9106
        41461520      8636
        41465540      8320
        41463520      8170
        41461250      7344
        ...
        11170660         1
        42760360         1
        41173578         1
        41171580         1
        41285551         1
Name: adng_cd, Length: 1005, dtype: int64
```

```
In [12]: df.adng_cd.nunique()
```

```
Out[12]: 1005
```

## 6) 행정동명

```
In [13]: df.adng_nm.value_counts()
```

```
Out[13]: 모현읍      9106
역삼동      8636
죽전1동     8320
구갈동      8170
포곡읍      7344
...
선부3동      1
박달1동      1
월곡1동      1
파평면      1
백석1동      1
Name: adng_nm, Length: 974, dtype: int64
```

## 7) 셀ID

- 지형지물정보 셀 ID

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

```
Out[14]: 19241
```

## 8) Cell center X

- 셀 중심점 Y 좌표 위도

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

```
Out[15]: 13839
```

```
In [16]: df.cell_xcrd.describe()
```

```
Out[16]: count    293171.000000
mean        127.116569
std          0.115833
min         124.714325
25%         127.074356
50%         127.110140
75%         127.160350
max         129.161670
Name: cell_xcrd, dtype: float64
```

## 9) Cell center Y

- 셀 중심점 X 좌표 경도

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

```
Out[17]: 16140
```

```
In [18]: df.cell_ycrd.describe()
```



```
Out[18]: count      293171.000000
         mean        37.303852
         std         0.124673
         min         36.011906
         25%         37.251510
         50%         37.292080
         75%         37.334835
         max         38.300770
         Name: cell_ycrd, dtype: float64
```

## 10) 성별

```
In [19]: df.gender.value_counts()
```

```
Out[19]: MALE      191383
         FEMALE    101788
         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 [20]: df.age.value_counts()
         ### 의문 20세미만도 포함을 시켜야하나?
```

```
Out[20]: 10      39513
         8       37933
         11      35167
         9       33500
         12      27003
         14      22809
         7       22529
         13      20451
         6       18836
         5       15856
         4       8850
         3       7428
         2       3044
         1        252
         Name: age, dtype: int64
```

## 12) 앱웹

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

```
Out[21]: array(['EV라운지', '하이차저', '일렉배리', 'EV Infra', '파워큐브차저', 'G차저', '채비',
        '모두의충전', 'KEPCO PLUG', '차지비', '에버온', '환경부 통합 누리집', 'Epit',
        'evWhere', '에너지플러스EV', '테슬라', '해피차저'], dtype=object)
```

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

```
Out[22]: EV라운지      243715
하이차저      42475
파워큐브차저      1678
G차저          1590
EV Infra       1544
차지비         483
에버온         384
KEPCO PLUG     285
환경부 통합 누리집 256
모두의충전     247
일렉배리       237
Epit           143
채비           104
테슬라         14
evWhere        13
에너지플러스EV    2
해피차저        1
Name: app_web, dtype: int64
```

```
In [23]: df.app_web.nunique()
```

```
Out[23]: 17
```

## 13) 시간

- 1 0시~6시
- 2 6시~9시
- 3 9시~12시
- 4 12시~18시
- 5 18시~21시
- 6 21시~24시

```
In [24]: df.time_zone.value_counts()
```

```
Out[24]: 4    92380
3    52899
5    50873
2    39183
6    32426
1    25410
Name: time_zone, dtype: int64
```

## 14) 실행자수

```
In [25]: df.count_cust.describe()
```

```
Out[25]: count      293171.000000
         mean         6.495661
         std         4.306967
         min         3.011000
         25%         3.802000
         50%         4.507000
         75%         8.052000
         max        177.332000
         Name: count_cust, dtype: float64
```

```
In [26]: df.loc[df.count_cust >= 100]
```

	base_dt	dow	mega_cd	mega_nm	ccw_cd	ccw_nm	adng_cd	adng_nm	cell_id	
<b>18747</b>	20220605	7	41	경기도	4146	용인시	41463572	동백1동	86503006	1
<b>44139</b>	20220608	3	41	경기도	4146	용인시	41465510	풍덕천1동	84567216	1
<b>260233</b>	20220608	3	41	경기도	4146	용인시	41465510	풍덕천1동	84567216	1

## 15) 특별시, 도

```
In [27]: df.mega_nm.value_counts()
```

```
Out[27]: 경기도      255586
         서울특별시    32797
         충청남도      1644
         인천광역시    1218
         충청북도      1156
         대전광역시      484
         강원도        286
         Name: mega_nm, dtype: int64
```

```
In [ ]:
```