

# Culture Data Competition

[개요]

## 0. Setting

### Library Call

In [1]:

```
1 # Library Call
2 import glob
3 import pandas as pd
4 import numpy as np
5 import matplotlib
6 import matplotlib.pyplot as plt
7 import seaborn as sns
8 import plotly.graph_objects as go
9 import plotly.express as px
10 import time
11 import datetime as dt
12 # !pip install folium
13 import folium
14 from wordcloud import WordCloud
15
16 # 한글 폰트 패치.
17 matplotlib.rcParams['font.family']='Malgun Gothic'
18 matplotlib.rcParams['axes.unicode_minus'] = False
19
20 # 그래프에 retina display 적용
21 %config InlineBackend.figure_format = 'retina'
22
23 # ggplot style skima
24 plt.style.use("ggplot")
25
26 # seaborn plot style definition
27 sns.set_style("whitegrid")
28 sns.set_context("talk")
29
30 # 경고문 처리
31 import warnings
32 warnings.filterwarnings('ignore')
```

### Visualization Font Setting

In [2]:

```
1 import platform
2
3 def get_font_family():
4     system_name = platform.system()
5
6     if system_name == "Darwin" :
7         font_family = "AppleGothic"
8     elif system_name == "Windows":
9         font_family = "Malgun Gothic"
10    else:
11        !apt-get install fonts-nanum -qq > /dev/null
12        !fc-cache -fv
13
14        import matplotlib as mpl
15        mpl.font_manager._rebuild()
16        findfont = mpl.font_manager.fontManager.findfont
17        mpl.font_manager.findfont = findfont
18        mpl.backends.backend_agg.findfont = findfont
19
20        font_family = "NanumBarunGothic"
21    return font_family
22
23 %config InlineBackend.figure_format = 'retina'
24
25 plt.rc("font", family=get_font_family())
26 plt.rc("axes", unicode_minus=False)
27
28 get_font_family()
```

Out[2]:

'Malgun Gothic'

## User Function Definition

In [3]:

```
1 # Merge Same DataFrame Format
2 def Multiple_Data_Load(flist):
3     df = []
4     for file in flist:
5         ele = pd.read_csv(file)
6         df.append(ele)
7     return pd.concat(df)
8
9 # Return Col names by Metadata Seat
10 def return_col(meta):
11     df = pd.read_excel('Metadata/'+meta, index_col='순서')
12     col = df['컬럼한글명'].values
13     return col
14
15 # Weekday List Definition
16 weekday_label = np.array(['월요일', '화요일', '수요일', '목요일', '금요일', '토요일', '일요일'])
17
18 # return MissingValue Existence
19 def MissingValue_Existence(df):
20     value = df.isna().sum().sum()
21     return f"MissinValue's counts : {value}"
22
23 # Display WordCloud
24 def display_word_cloud(data, width=1200, height=500):
25     word_draw = WordCloud(
26         font_path = "malgun",
27         width=width, height=height,
28         background_color="white",
29         random_state=42
30     )
31     word_draw.generate(data)
32
33     plt.figure(figsize=(15, 7))
34     plt.imshow(word_draw)
35     plt.axis("off")
36     plt.show()
```

In [ ]:

1

# 1. Culture Data EDA & Visualization

## 1. 문화누리 키워드 검색량 데이터

[출처] : 데이터마케팅코리아

[Data Description]

- '문화누리카드', '문화누리' 키워드에 대한 모바일검색량, PC검색량 등이 포함된 일별 검색량 데이터
- 데이터 수집 기간 : 2021-11-01 ~ 2022-03-31 (5 months, by a day)

[Applications]

- 문화누리카드, 문화누리 대한 소비자의 일별 검색 트렌드 분석

## [Insight Goal]

- 문화누리카드의 일별 검색 트렌드를 분석함으로써 최근 N년 간 문화누리카드의 관심도 변화와 인식 등에 대해 분석하고자함. 또한, 모바일 매채별, 날짜별로 검색량을 그룹화하여 특정 카테고리에 따라 어떠한 양상을 보이는지 확인하고자함. 이를 통해, 문화누리 카드의 현주소를 파악하고, 트렌드 추세에 맞는 아이디어를 기획하고자 함.

## Data Load

In [4]:

```
1 # Data Load
2 keyword_flist = sorted(glob.glob('data/DM_CLTUR_KWRD_SCCNT_RESULT*.csv'))
3 keyword_df = Multiple_Data_Load(keyword_flist)
4 keyword_df.columns=return_col('문화누리 키워드 검색량 데이터_컬럼정의서.xls')
5 print('keyword_df.shape :',keyword_df.shape)
6 keyword_df.head()
```

keyword\_df.shape : (304, 6)

Out[4]:

	일련	검색어	Mobile 검색량	PC 검색량	검색량 합계	검색량 날짜
0	4258	문화누리	254	207	461	2021-11-01
1	4259	문화누리	300	190	490	2021-11-02
2	4260	문화누리	263	169	432	2021-11-03
3	4261	문화누리	224	198	422	2021-11-04
4	4262	문화누리	263	192	455	2021-11-05

## Data EDA & Preprocessing

In [5]:

```
1 # Data Information
2 keyword_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 304 entries, 0 to 61
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   일련            304 non-null   int64
1   검색어          304 non-null   object
2   Mobile 검색량   304 non-null   int64
3   PC 검색량       304 non-null   int64
4   검색량 합계     304 non-null   int64
5   검색량 날짜     304 non-null   object
dtypes: int64(4), object(2)
memory usage: 16.6+ KB
```

In [6]:

```
1 # Check MissingValue Existence
2 keyword_df.dropna(inplace=True)
```

Out[6]:

"MissinValue's counts : 0"

In [7]:

```
1 # Feature Selection
2 keyword_df.drop('일련', inplace=True, axis=1)
```

In [8]:

```
1 # Drop Duplicates
2 keyword_df.drop_duplicates(inplace=True)
```

In [9]:

```
1 # Datetime 파생변수 생성
2 keyword_df['검색량 날짜'] = pd.to_datetime(keyword_df['검색량 날짜'])
3 keyword_df.sort_values('검색량 날짜', inplace=True)
4
5 keyword_df['검색월'] = keyword_df['검색량 날짜'].dt.month
6 keyword_df['검색일'] = keyword_df['검색량 날짜'].dt.day
7 keyword_df['검색년도'] = keyword_df['검색량 날짜'].dt.year
8 keyword_df['검색요일'] = weekday_label[keyword_df['검색량 날짜'].dt.weekday.values]
```

In [10]:

```
1 # ReIndexing
2 keyword_df.index = keyword_df['검색량 날짜']
3 keyword_df.drop('검색량 날짜', inplace=True, axis=1)
4 keyword_df.index.name = 'date'
```

In [11]:

```
1 keyword_df.head(4)
```

Out[11]:

	검색어	Mobile 검색량	PC 검색량	검색량 합계	검색월	검색일	검색년도	검색요일
date								
2021-11-01	문화누리	254	207	461	11	1	2021	월요일
2021-11-01	문화누리카드	4307	1560	5867	11	1	2021	월요일
2021-11-02	문화누리카드	3830	1335	5165	11	2	2021	화요일
2021-11-02	문화누리	300	190	490	11	2	2021	화요일

In [12]:

```
1 # 검색어 비율
2 keyword_df['검색어'].value_counts(normalize=True)
```

Out[12]:

```
문화누리      0.5
문화누리카드   0.5
Name: 검색어, dtype: float64
```

In [13]:

```
1 # Keyword에 따라 데이터 분리
2 key_a = keyword_df[keyword_df['검색어'] == '문화누리']
3 key_b = keyword_df[keyword_df['검색어'] == '문화누리카드']
4
5 print('key_a : 문화누리 : ',key_a.shape)
6 print('key_b : 문화누리카드 : ',key_b.shape)
```

```
key_a : 문화누리 : (151, 8)
key_b : 문화누리카드 : (151, 8)
```

In [14]:

```
1 # 문화누리 관련 키워드 총합 데이터프레임
2 key_total = key_a[['Mobile 검색량','PC 검색량','검색량 합계']] + key_b[['Mobile 검색량','PC 검색량','검색량 합계']]
3 key_total = pd.concat([key_total, key_a.iloc[:,4:]], axis=1)
4 print('key_total : ',key_total.shape)
5 key_total.head()
```

key\_total : (151, 7)

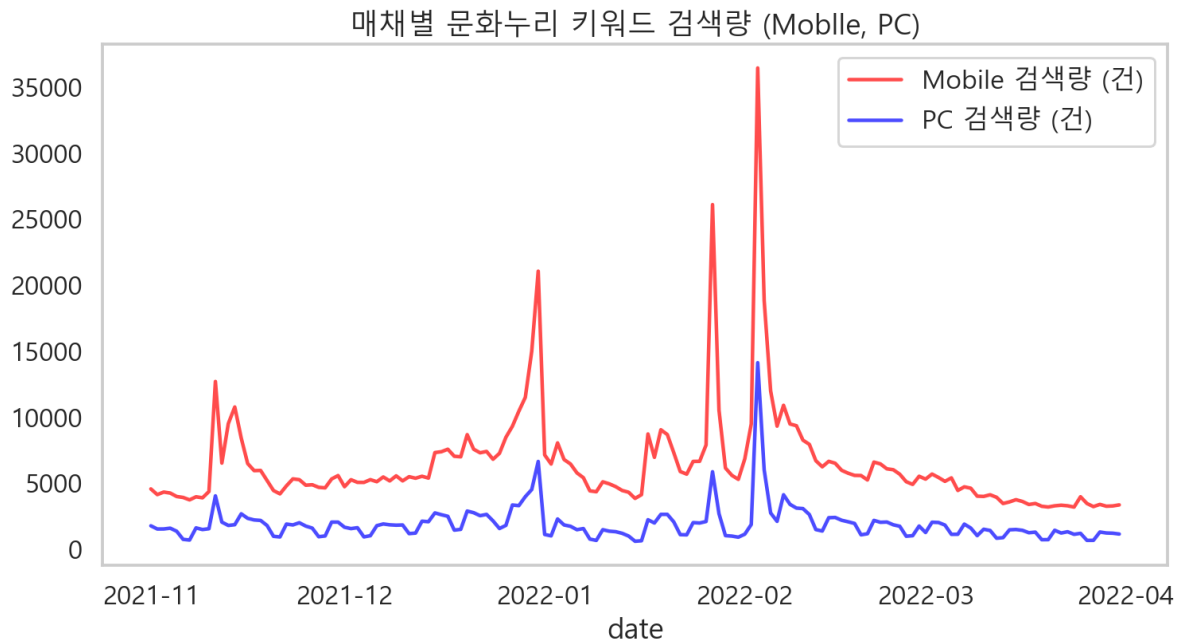
Out[14]:

	Mobile 검색량	PC 검색량	검색량 합계	검색월	검색일	검색년도	검색요일
date							
2021-11-01	4561	1767	6328	11	1	2021	월요일
2021-11-02	4130	1525	5655	11	2	2021	화요일
2021-11-03	4324	1528	5852	11	3	2021	수요일
2021-11-04	4247	1587	5834	11	4	2021	목요일
2021-11-05	3987	1356	5343	11	5	2021	금요일

## Insight Inference & Visualize

In [15]:

```
1 plt.figure(figsize=(12,6))
2 sns.lineplot(x = key_total.index, y = key_total['Mobile 검색량'], label='Mobile 검색량 (건)', c
3 sns.lineplot(x = key_total.index, y = key_total['PC 검색량'], label='PC 검색량 (건)', color='b
4 plt.title('매채별 문화누리 키워드 검색량 (Mobile, PC)')
5 plt.ylabel('')
6 plt.legend(loc='best')
7 plt.grid(False)
8 plt.show()
```



In [16]:

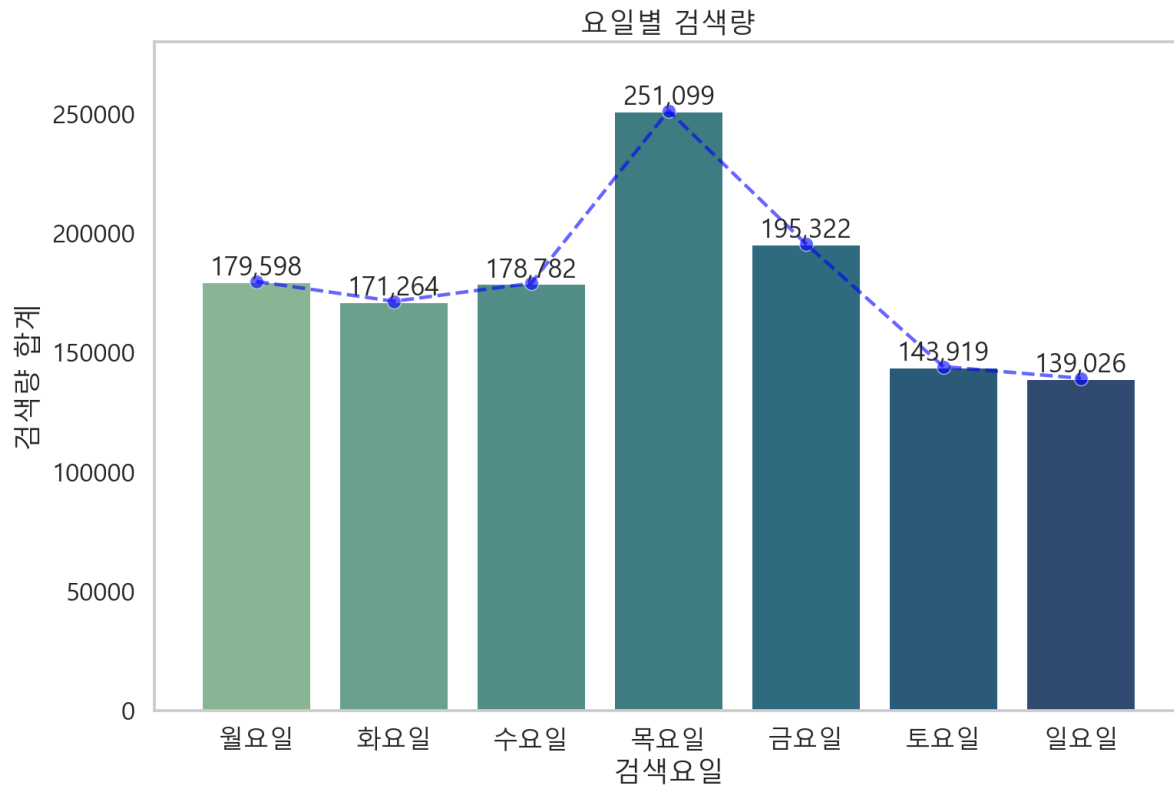
```
1 plt.figure(figsize=(12,6))
2 sns.lineplot(x = key_total.index, y = key_total['검색량 합계'], label='총 검색량', color='g', &
3 plt.title('문화누리 관련 키워드 총 검색량 변화추이')
4 plt.legend(loc='best')
5 plt.show()
```





In [17]:

```
1 keyword_weekday = keyword_df.groupby('검색요일')[['검색량 합계']].sum().reindex(weekday_label)
2 plt.figure(figsize=(12,8))
3 sns.barplot(keyword_weekday.index, keyword_weekday['검색량 합계'], palette = 'crest')
4 sns.lineplot(keyword_weekday.index, keyword_weekday['검색량 합계'], color='b', linestyle='--',a
5 sns.scatterplot(keyword_weekday.index, keyword_weekday['검색량 합계'], color='b', alpha=0.6)
6 plt.title('요일별 검색량')
7 for i,bar in enumerate(keyword_weekday['검색량 합계']):
8     plt.text(weekday_label[i],bar, '{:,d}'.format(bar),ha='center',va='bottom',size=17)
9 plt.ylim(0,280000)
10 plt.grid(False)
11 plt.show()
```



In [18]:

```
1 keyword_weekday[ '검색량 합계' ]
```

Out[18]:

검색요일

월요일 179598

화요일 171264

수요일 178782

목요일 251099

금요일 195322

토요일 143919

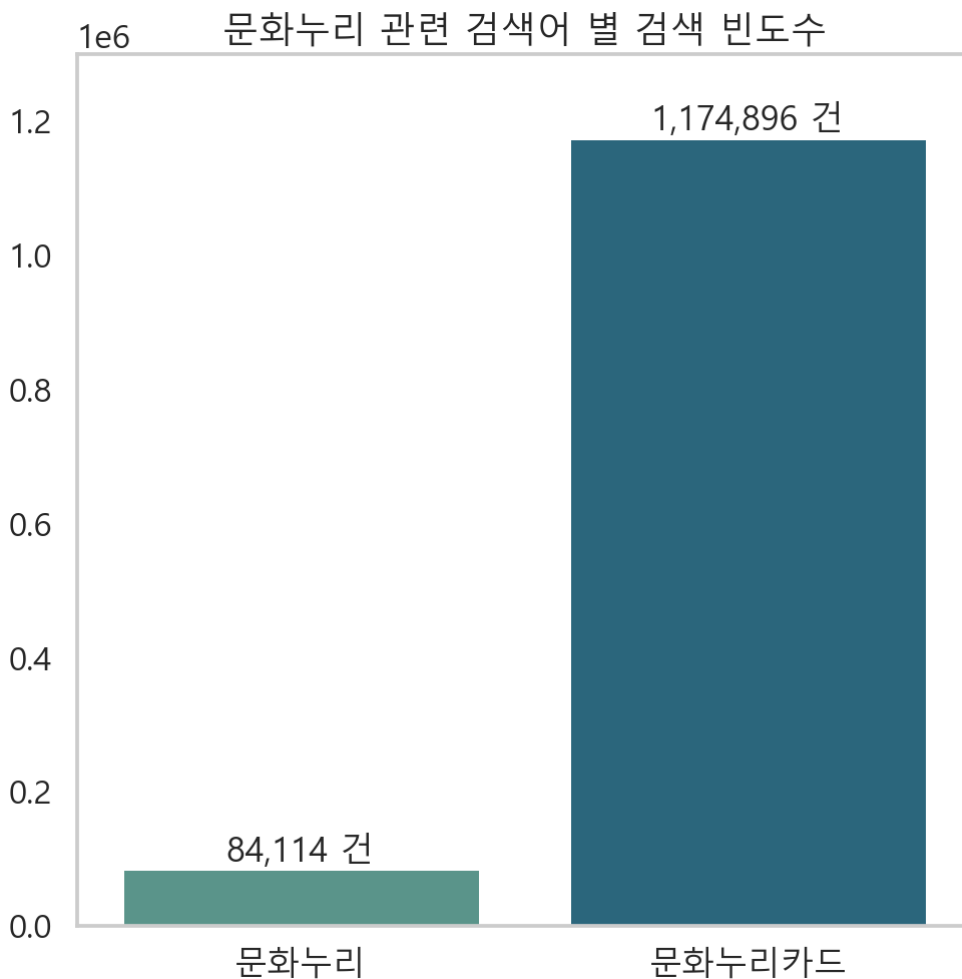
일요일 139026

Name: 검색량 합계, dtype: int64

In [19]:

```
1 plt.figure(figsize=(8,8))
2 sns.barplot(['문화누리', '문화누리카드'], [np.sum(key_a['검색량 합계']), np.sum(key_b['검색량 합계'])])
3 print([np.sum(key_a['검색량 합계']), np.sum(key_b['검색량 합계'])])
4 plt.title('문화누리 관련 검색어 별 검색 빈도수')
5
6 for i, bar in enumerate([np.sum(key_a['검색량 합계']), np.sum(key_b['검색량 합계'])]):
7     plt.text(i, bar, '{:,d} 건'.format(bar), ha='center', va='bottom', size=17)
8
9 plt.ylim(0, 1300000)
10 plt.grid(False)
11 plt.show()
```

[84114, 1174896]



#### [문화누리 키워드 검색량 데이터 분석 결과]

- 매체별로 키워드 검색량을 분석한 결과, Mobile기기를 활용한 검색량이 PC를 통한 검색량 보다 훨씬 많은 것을 알 수 있었다. 이는 문화누리카드 관련 정보에 대한 주 매개체가 Mobile 기기라는 것을 의미한다.
- 문화누리 관련 키워드 검색량의 전체 추이를 분석한 결과, 2022 01월 초와 2022년 2월 초에 검색량이 급격히 많아지는 것을 볼 수 있었다. 하지만, 모두 일시적인 상승일 뿐 이내 다시 평균 추이로 돌아오는 것을 볼 수 있다.
- 요일별 검색량을 확인해보니, 목요일에 유의미하게 가장 많은 검색량이 물리는 것을 볼 수 있었다. 이는, 주말 전에 문화누리카드의 사용처나 관련 혜택 등을 검색하는 것으로 볼 수도 있고, 혹은 다른 외부 요인이 있을 것으로 보인다.
- 문화누리 관련 키워드는 크게 '문화누리', '문화누리카드' 두 가지로 구분된다. 하지만, '문화누리카드'에 대한 검색량이 월등히 많으며 이는 문화누리카드의 발급조건, 사용처, 한도금액 등 실리적인 혜택에 대해 검색량

이 물리는 것으로 볼 수 있다.

## 2. 차상위계층 카드 가맹점 정보

[출처] : 한국문화예술위원회

[Data Description]

- 부산지역 취약계층 가맹점 데이터
- 기초생활 수급자/차상위 계층이 활용할 수 있는 문화누리카드 가맹점

[Applications]

- 제공 가맹점 대상지역 차상위 계층의 카드 사용 활성화
- 지원 카드 사용율과 제공 분야 개선

[Insight Goal]

- 제공 가맹점 매핑 및 접근성 분석

### Data Load

In [20]:

```
1 # Data Load
2 card_shop_df = pd.read_csv('data/ak_lwicc_card_mrhst_info_202103.csv')
3 card_shop_df.columns=return_col('차상위계층 카드 가맹점 정보_컬럼정의서.xls')
4 print('card_shop_df.shape : ',card_shop_df.shape)
5 card_shop_df.head(2)
```

card\_shop\_df.shape : (5027, 12)

Out[20]:

	일 련 번 호	가맹점명	가맹점위 도	가맹점경도	가맹 점시 도코 드	가 맹 점 시 도명	가맹 점시 군구 코드	가맹 점시 군구 명	가맹점 행정 코드	가맹 점행 정동 명	가맹 점구 분코 드	가맹 점구 분코 드명
0	14	GS25 R 부산동구 2점	35.136637	129.065045	21	부산 광역시	21030	동구	2103071	범일 1동	B	급식 카드 가맹 점
1	15	GS25 R 부산동구 3점	35.136637	129.065045	21	부산 광역시	21030	동구	2103071	범일 1동	B	급식 카드 가맹 점

### Data EDA & Preprocessing

In [21]:

```
1 # Data Information
2 card_shop_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5027 entries, 0 to 5026
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  --
 0   일련번호              5027 non-null   int64   
 1   가맹점명              5027 non-null   object  
 2   가맹점위도            5027 non-null   float64  
 3   가맹점경도            5027 non-null   float64  
 4   가맹점시도코드        5027 non-null   int64   
 5   가맹점시도명          5027 non-null   object  
 6   가맹점시군구코드      5027 non-null   int64   
 7   가맹점시군구명        5027 non-null   object  
 8   가맹점행정동코드      5027 non-null   int64   
 9   가맹점행정동명        5027 non-null   object  
 10  가맹점구분코드        5027 non-null   object  
 11  가맹점구분코드명      5027 non-null   object  
dtypes: float64(2), int64(4), object(6)
memory usage: 471.4+ KB
```

In [22]:

```
1 # 급식카드가맹점 : B | 문화누리가맹점 : M
2 card_shop_df['가맹점구분코드'].value_counts()
```

Out[22]:

```
B    3772
M    1255
Name: 가맹점구분코드, dtype: int64
```

In [23]:

```
1 # Feature Selection
2 card_shop_df.drop(['일련번호'],axis=1,inplace=True)
3 card_shop_df = card_shop_df[card_shop_df['가맹점구분코드명'] == '문화누리가맹점'].reset_index()
4 print('card_shop_df.shape :',card_shop_df.shape)
5 card_shop_df.head(2)
```

card\_shop\_df.shape : (1255, 12)

Out[23]:

	index	가맹점명	가맹점위도	가맹점경도	가맹점시도코드	가맹점시도명	가맹점시군구코드	가맹점시군구명	가맹점행정동코드	가맹점행정동명	가맹점구분코드	가맹점구분코드명
0	469	삼천리 자전거 동래점	35.202679	129.083916	21	부산광역시	21060	동래구	2106051	수민동	M	문화누리가맹점
1	521	행복한 스튜디오	35.152427	129.054766	21	부산광역시	21050	부산진구	2105052	부전2동	M	문화누리가맹점

In [24]:

```
1 # Subset for Mapping
2 map_subset = card_shop_df[['가맹점명','가맹점위도','가맹점경도']]
3 map_subset.columns = ['가맹점명','위도','경도']
4 print('map_subset :',map_subset.shape)
5 map_subset.head()
```

map\_subset : (1255, 3)

Out[24]:

	가맹점명	위도	경도
0	삼천리자전거동래점	35.202679	129.083916
1	행복한스튜디오	35.152427	129.054766
2	네이쳐앤티트리	35.061973	128.984193
3	신평태권도장	35.093446	128.973558
4	송무인 신금 태권도	35.252920	129.013612

## Insight Inference & Visualization

In [25]:

```
1 # 부산지역 문화누리가맹점 지도 매핑
2 busan_coord = [35.1796, 129.0756]
3 m = folium.Map(
4     location = busan_coord,
5     zoom_start=15)
6
7 for i in range(map_subset.shape[0]):
8     folium.Marker([map_subset.iloc[i,1],map_subset.iloc[i,2]],
9                   tooltip = map_subset.iloc[i,0],
10                  icon=folium.Icon('blue',icon='star')).add_to(m)
11 m
```

...

In [26]:

```
1 # Map Figure Save
2 m.save('figure/Busan_Shop_map.html')
```

[차상위계층 카드 가맹점 정보 분석 결과]

- 부산지역 내 문화누리가맹점의 분포를 확인해보았다. 생각보다, 문화누리가맹점 수가 많은 것을 알 수 있었고 접근성 면에서는 크게 문제되지 않을 것으로 예상된다. 즉, 누구나 원한다면 카드를 사용할 수 있는 인프라는 만들어져있는 상태이고, 혜택과 홍보면에서 개선점을 가져간다면 충분히 활성화 가능성이 있다.

### 3. 문화누리 결제빈도 대비 대출분석

[출처]: 국립중앙도서관

[Data Description]

- 문화누리 결제빈도 대비 대출 정보
- 지역별 문화누리 결제 빈도 대비 대출 빈도

[Applications]

- 도서관과 문화누리 카드결제 등 관련 데이터 분석

[Insight Goal]

- 문화누리 카드의 도서 대출 분야 성과분석 및 소비자 유형 분류를 통해 서비스 보완점과 개발점을 도출

## Data Load

In [32]:

```
1 # Data Load
2 frequency_flist = sorted(glob.glob('data/NL_MNURI_SETLE_FQ_VERSUS_LBRRY_LON*.csv'))
3 print(frequency_flist)
4 frequency_df = Multiple_Data_Load(frequency_flist).iloc[:, :16]
5 frequency_df.columns=return_col('문화누리 결제빈도 대비 대출분석_컬럼정의서.xls')
6 print('frequency_df.shape :', frequency_df.shape)
7 frequency_df.head()
```

```
['data\\WWWNL_MNURI_SETLE_FQ_VERSUS_LBRRY_LON_2019.csv', 'data\\WWWNL_MNURI_SETLE_FQ_VERSU
S_LBRRY_LON_2020.csv']
```

frequency\_df.shape : (4444, 16)

Out[32]:

	분석_대_상_연도	분석_대_상_월	일_차_지_역_명	이_차_지_역_명	결제_자_기준_문_화_누_리_결제_빈_도_수	가맹_점_기준_문_화_누_리_결제_빈_도_수	전체_대출_빈도_수	7_세_이하_대출_빈도_수	13_세_이하_대출_빈도_수	19_세_이하_대출_빈도_수	20_대_대출_빈도_수	30대_대출_빈도_수	40대_대출_빈도_수	50대_대출_빈도_수	60_대_대출_빈도_수	70_대_이_상_대출_빈도_수
0	2019	2	강원도	강릉시	1344	1417	10441	369.0	1416.0	741.0	543	1678	3273	1053	434	352
1	2019	3	강원도	강릉시	2577	2631	11768	540.0	1402.0	657.0	452	2303	3618	1151	619	272
2	2019	4	강원도	강릉시	1588	1625	11998	629.0	1653.0	552.0	476	2123	3519	1117	632	246
3	2019	5	강원도	강릉시	1682	1739	10576	573.0	1232.0	551.0	392	1944	3225	1008	590	323
4	2019	6	강원도	강릉시	1430	1514	12043	696.0	1741.0	662.0	478	2012	3819	1064	556	273

## Data EDA & Preprocessing



In [28]:

```
1 # Data Information
2 frequency_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4444 entries, 0 to 2221
Data columns (total 16 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   분석_대상_연도                        4444 non-null   int64
1   분석_대상_월                          4444 non-null   int64
2   일차_지역_명                          2222 non-null   object
3   이차_지역_명                          2222 non-null   object
4   결제자_기준_문화누리_결제_빈도_수    4444 non-null   int64
5   가맹점_기준_문화누리_결제_빈도_수    4444 non-null   int64
6   전체_대출_빈도_수                    4444 non-null   int64
7   7_세_이하_대출_빈도_수                2222 non-null   float64
8   13_세_이하_대출_빈도_수               2222 non-null   float64
9   19_세_이하_대출_빈도_수               2222 non-null   float64
10  20대_대출_빈도_수                     4444 non-null   int64
11  30대_대출_빈도_수                     4444 non-null   int64
12  40대_대출_빈도_수                     4444 non-null   int64
13  50대_대출_빈도_수                     4444 non-null   int64
14  60대_대출_빈도_수                     4444 non-null   int64
15  70대_이상_대출_빈도_수                4444 non-null   int64
dtypes: float64(3), int64(11), object(2)
memory usage: 590.2+ KB
```

In [ ]:

```
1 # Check MissingValue Existence
2 MissingValue_Existence(frequency_df)
```

열별 2222개의 결측치는 모두 같은 행에서 발생한 데이터이다. 해당 데이터는 배제하고 분석 지역을 좁혀 해당 지역에 집중한 인사이트를 도출하고자 한다.

In [43]:

```
1 # 일차 지역명 목록
2 frequency_df['일차_지역명'].value_counts()
```

Out[43]:

경기도	440
서울특별시	275
전라남도	220
강원도	198
부산광역시	176
충청북도	154
경상북도	132
충청남도	99
대구광역시	88
인천광역시	88
경상남도	88
전라북도	66
광주광역시	55
대전광역시	55
울산광역시	55
제주특별자치도	22
세종특별자치시	11

Name: 일차\_지역명, dtype: int64

In [45]:

```
1 # 부산광역시 문화누리카드 도서대출 결제 빈도
2 busan_freq = frequency_df[frequency_df['이차_지역_명'] == '부산광역시']
3 print('busan_freq.shape : ',busan_freq.shape)
4 busan_freq.head(2)
```

busan\_freq.shape : (176, 16)

Out[45]:

	분석_대상_연도	분석_대상_월	이차_지역_명	이차_지역_명	결제_자_기준_문화누리카드_결제_빈도_수	가맹점_기준_문화누리카드_결제_빈도_수	전체_대출_빈도_수	7_세_이하_대출_빈도_수	13_세_이하_대출_빈도_수	19_세_이하_대출_빈도_수	20_대_대출_빈도_수	30대_대출_빈도_수	40대_대출_빈도_수	50대_대출_빈도_수	60_대_대출_빈도_수
1056	2019	2	부산광역시	강서구	242	88	32944	3345.0	6539.0	888.0	722	8754	10619	1169	657
1057	2019	3	부산광역시	강서구	455	143	38155	4596.0	6969.0	753.0	822	10765	12070	1119	787

In [49]:

```
1 # 부산광역시 이차지역명
2 print(busan_freq['이차_지역_명'].value_counts())
```

```
강서구      11
금정구      11
기장군      11
남구        11
동구        11
동래구      11
부산진구    11
북구        11
사상구      11
사하구      11
서구        11
수영구      11
연제구      11
영도구      11
중구        11
해운대구    11
Name: 이차_지역_명, dtype: int64
```

# Insight Inference & Visualization

In [54]:

```
1 # 부산광역시 월별 대출 빈도수
2 busan_month_fr = busan_freq.groupby('분석_대상_월').mean()
3 print('busan_month_fr.shape :',busan_month_fr.shape)
4 busan_month_fr.head(2)
```

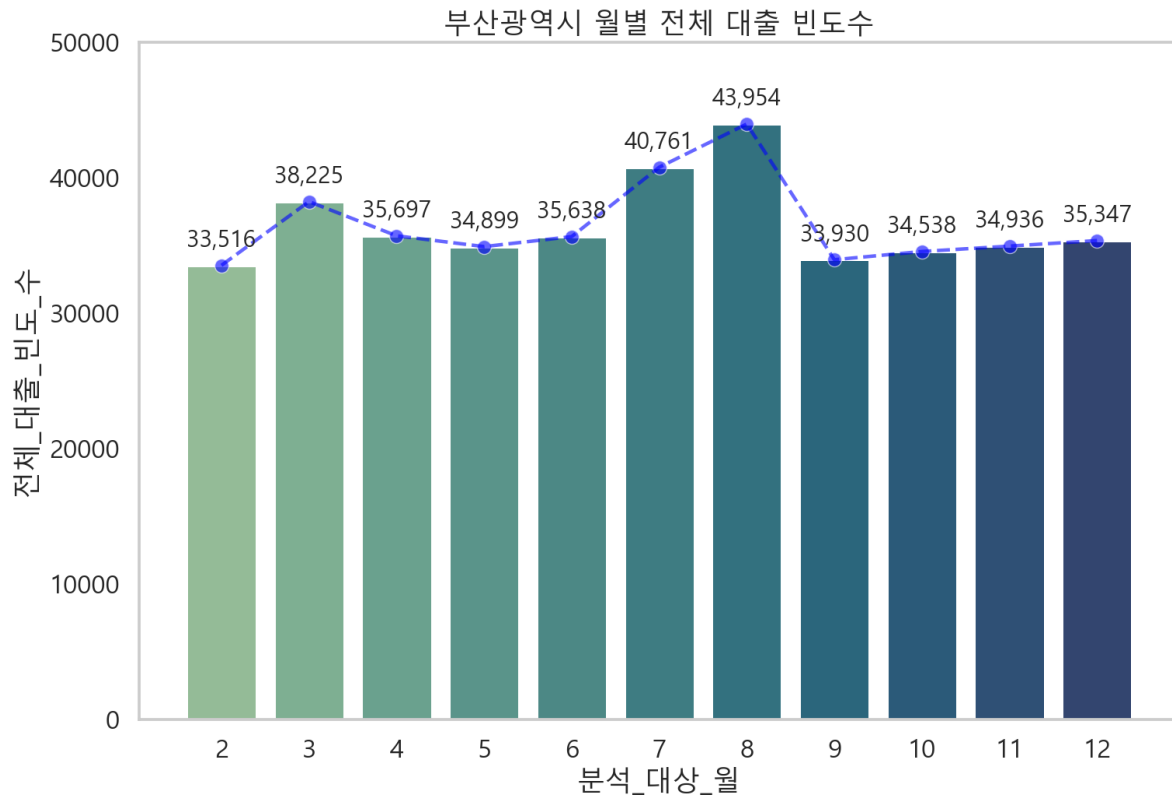
busan\_month\_fr.shape : (11, 13)

Out[54]:

	분석_대상_연도	결제자_기준_문화누리카드_결제_빈도_수	가맹점_기준_문화누리카드_결제_빈도_수	전체_대출_빈도_수	7_세_이하_대출_빈도_수	13_세_이하_대출_빈도_수	19_세_이하_대출_빈도_수	20대_대출_빈도_수	30대_대출_빈도_수
분석_대상_월									
2	2019.0	1137.500	1007.9375	33516.1250	2004.5000	5458.3125	1913.3125	1844.8750	5809.0000
3	2019.0	2351.625	2093.2500	38225.5625	2799.8125	6210.2500	1439.5625	1782.1875	7010.6250

In [88]:

```
1 # 부산광역시 월별 전체 대출 빈도수 시각화
2 plt.figure(figsize=(12,8))
3 sns.barplot(busan_month_fr.index, busan_month_fr['전체_대출_빈도수'], palette = 'crest')
4 sns.lineplot(busan_month_fr.index-2, busan_month_fr['전체_대출_빈도수'], color='b', linestyle='--')
5 sns.scatterplot(busan_month_fr.index-2, busan_month_fr['전체_대출_빈도수'], color='b', alpha=0.5)
6 plt.title('부산광역시 월별 전체 대출 빈도수')
7 for i,bar in enumerate(busan_month_fr['전체_대출_빈도수']):
8     plt.text(busan_month_fr.index[i]-2, bar+1000, '{:,.d}'.format(int(bar)),ha='center',va='bottom')
9 plt.ylim(0,50000)
10 plt.grid(False)
11 plt.show()
```



부산광역시의 월별 전체 대출 빈도수를 보면, 1월 같은 경우 카드 신청 기간이기 때문에 데이터가 없고, 8월에서 가장 많은 대출 빈도를 보이는 것을 알 수 있다.

In [114]:

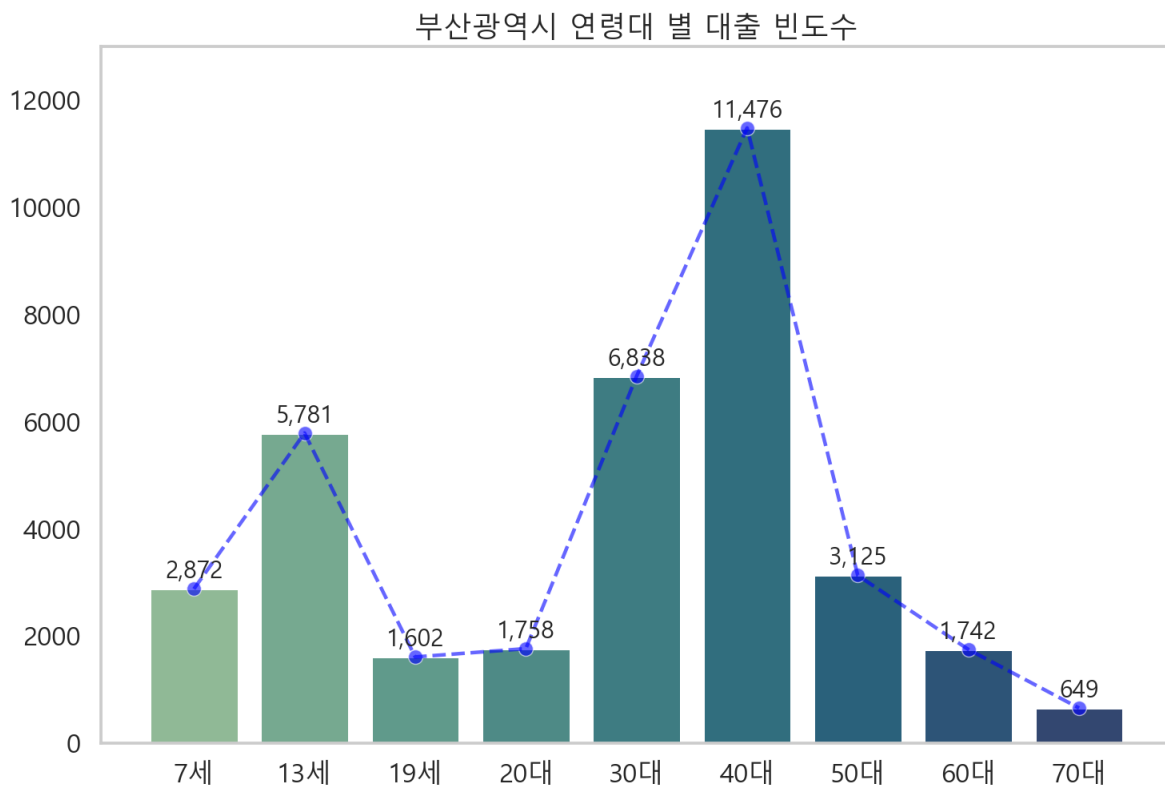
```
1 # 부산광역시 연령대 별 대출 빈도수
2 busan_age = np.mean(busan_freq.iloc[:,7:])
3 busan_age.index = busan_age.index.str[:4].str.replace('_', '')
4 busan_age
```

Out[114]:

```
7세      2872.767045
13세     5781.977273
19세     1602.840909
20대     1758.034091
30대     6838.500000
40대    11476.744318
50대     3125.500000
60대     1742.187500
70대      649.159091
dtype: float64
```

In [117]:

```
1 # 부산광역시 연령대 별 대출 빈도수 시각화
2 plt.figure(figsize=(12,8))
3 sns.barplot(busan_age.index, busan_age.values, palette = 'crest')
4 sns.lineplot(busan_age.index, busan_age.values, color='b', linestyle='--',alpha=0.6)
5 sns.scatterplot(busan_age.index, busan_age.values, color='b', alpha=0.6)
6 plt.title('부산광역시 연령대 별 대출 빈도수')
7 for i in range(len(busan_age)):
8     plt.text(busan_age.index[i], busan_age.values[i]+100, '{:.d}'.format(int(busan_age.values[i])))
9 plt.ylim(0,13000)
10 plt.grid(False)
11 plt.show()
```



부산광역시 내 문화누리사용자의 연령대 별 대출 빈도수를 확인한 결과, 40대가 가장 많은 도서 대출 빈도 수를 보였다. 19세와 20대의 경우, 상당히 낮은 대출 빈도를 보였는데, 이는 19세 ~ 20대가 도서에 관심이 없는 것으로

볼 수 있고, 60~70대의 경우 도서 대출을 하는 경위 자체가 부담인 것으로 볼 수 있다.

## 4. 문화, 문화누리 키워드 추출 정보

[출처] : 한국문화예술위원회

[Data Description]

- 문화, 문화누리 키워드 추출 정보

[Applications]

- 인터넷 뉴스, 블로그에 대한 내용을 머신러닝 기반의 키워드 추출 정보
- 문화 키워드 분석을 통해 최근 이슈가 되거나 트렌드가 되는 키워드를 분석하여 관련 이슈 현황 분석

[Insight Goal]

- 문화, 문화누리 관련 키워드에 대한 빈도수와 내용을 파악함으로써 분석 시사점 도출

## Data Load

In [149]:

```
1 # Data Load : 문화및문화누리 뉴스키워드 추출정보
2 news_flist = sorted(glob.glob('data/ak_cltur_nd_mnuri_news*.csv'))
3 print('news_flist :', len(news_flist))
4 print(news_flist)
5 news_df = Multiple_Data_Load(news_flist).iloc[:, :8]
6 news_df.columns=return_col('문화 문화누리 키워드 추출 정보_컬럼정의서.xls')[::2]
7 news_df.dropna(axis=0, inplace=True)
8 news_df.reset_index(inplace=True, drop=True)
9 print('news_df.shape :', news_df.shape)
10 news_df.head(2)
```

```
news_flist : 2
['data\\Wak_cltur_nd_mnuri_news_kwr_d_extrc_info_2020.csv', 'data\\Wak_cltur_nd_mnuri_n
ews_kwr_d_extrc_info_2021.csv']
news_df.shape : (12405, 8)
```

Out[149]:

태그 구분 코드	추출본문내용	등록일시	추출 제목 명	출 처 명	출처URL	추출 키워드 내용
0 1.0 문화	[서울경제] 서울 경제 문화레저 부 기자들이 지 난 한 주간의 문 화예술계 주요 이슈들을 ...	2.020080e+13	[문화 +]문 화계 를 구 해줘!	서울 경제	https://news.naver.com/main/read.nhn? mode=LSD&...	배 우, 대 표, 서울 경제, 기 금, 콘서 트, 뮤지 컬, 영화, 공연, 마련, 강 철 비



일련번호	태그구분코드	추출본문내용	등록일시	추출제목명	출처명	출처URL	추출키워드내용	
1	2.0	문화	어미 멧돼지와 세끼 멧돼지. ? 게티이미지뱅크“멧돼지가 다 녀갔네. 그저께 조금 늦게 ...	2.020080e+13	[삶과 문화] 멧돼지 가족의 나들이	한국일보	https://news.naver.com/main/read.nhn?mode=LSD&...	멧돼지, 애초, 가족, 고구마, 사람, 다시, 수확, 이웃, 뿌리, 텃밭

In [150]:

```
1 # Data Load : 문화및문화누리 뉴스키워드 추출정보
2 blog_flist = sorted(glob.glob('data/ak_cltur_nd_mnuri_blog*.csv'))
3 print('blog_flist :',len(blog_flist))
4 print(blog_flist)
5 blog_df = Multiple_Data_Load(blog_flist).iloc[:, :8]
6 blog_df.columns=return_col('문화 문화누리 키워드 추출 정보_컬럼정의서.xls')[::2]
7 blog_df.dropna(axis=0, inplace=True)
8 blog_df.reset_index(inplace=True, drop=True)
9 print('blog_df.shape :',blog_df.shape)
10 blog_df.head(2)
```

```
blog_flist : 2
['data\\Wak_cltur_nd_mnuri_blog_kwr_d_extrc_info_2020.csv', 'data\\Wak_cltur_nd_mnuri_b
log_kwr_d_extrc_info_2021.csv']
blog_df.shape : (6536, 8)
```

Out[150]:

태그 구분 코드	출처URL	출처 키워드 내용	출처 제목 명	등록일시	추출 본문 내용	일련 번호
0 1.0 문화	https://blog.naver.com/iloveydp/222001087949	영등포구청, 체험, 인터넷, 문화체험단, 소재	영등포구	2.020062e+13	2020년 신나는 여름방학을 위한\n자매도시 어린이 문화체험단 모집\n\n모집기간...	0

일련번호	태그분코드	추출본문내용	등록일시	추출제목명	출처명	출처URL	추출키워드내용
1	2.0	문화 엄마들 사이에서 입소문(?)난 곳\n청양 목재 문화 체험관\n입구에 식당이 있길래,...	2.020070e+13	청양 목재 문화 체험관 : 아이와 가볼 만한 곳	해피냥냥	<a href="https://blog.naver.com/hj8506/222020019837">https://blog.naver.com/hj8506/222020019837</a>	청양, 오늘, 하늘, 목재, 문화, 제법, 체험관, 이후, 입장, 불가

In [151]:

```
1 # 뉴스 및 블로그 키워드 추출 정보
2 total_df = pd.concat([news_df, blog_df], axis=0).reset_index(drop=True)
3 print('total_df.shape :', total_df.shape)
4 total_df.head(2)
```

total\_df.shape : (18941, 8)

Out[151]:

태그 구분 코드	추출본문내용	등록일시	추출 제목 명	출 처 명	출처URL	추출 키워드 내용
0 1.0 문화	[서울경제] 서울 경제 문화레저 부 기자들이 지 난 한 주간의 문 화예술계 주요 이슈들을 ...	2.020080e+13	[문화 +]문 화계 를 구 해줘!	서울 경제	https://news.naver.com/main/read.nhn? mode=LSD&...	배 우, 대 표, 서울 경제, 기 금, 콘서 트, 뮤지 컬, 영화, 공연, 마 련, 강 철 비

일련 번호	태그 구분 코드	추출본문내용	등록일시	추출 제목 명	출 처 명	출처URL	추출 키워드 내용
1	2.0	문화 어미 멧돼지와 세끼 멧돼지. ? 게티이미지뱅크 “멧돼지가 다 녀갔네. 그저께 조금 늦게 ...	2.020080e+13	[삶과 문화] 멧돼 지 가 족의 나들 이	한 국 일 보	https://news.naver.com/main/read.nhn? mode=LSD&...	멧 돼 지, 애 초, 가 족, 고 구 마, 사 람, 다 시, 수 확, 이 웃, 뿌 리, 텃 밭

## Data EDA & Preprocessing

In [152]:

```
1 # 전체 뉴스 및 블로그 키워드 정보
2 total_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18941 entries, 0 to 18940
Data columns (total 8 columns):
#   Column      Non-Null Count  Dtype
---  -
0   일련번호    18941 non-null float64
1   태그구분코드 18941 non-null object
2   추출본문내용 18941 non-null object
3   등록일시    18941 non-null float64
4   추출제목명   18941 non-null object
5   출처명      18941 non-null object
6   출처URL     18941 non-null object
7   추출키워드내용 18941 non-null object
dtypes: float64(2), object(6)
memory usage: 1.2+ MB
```

In [158]:

```
1 # WordCloud 전처리 by '추출키워드내용'
2 content = total_df['추출키워드내용'].apply(lambda x : x.replace(',',' '))
3 content = ' '.join(content)
4 content[:100]
```

Out [ 158 ] :

'배우 대표 서울경제 기금 콘서트 뮤지컬 영화 공연 마련 강철비 멧돼지 애초 가족 고구마  
사람 다시 수확 이웃 뿌리 텃밭 일본 한국외대 태평양 교수 융합일본지역학부 통해 전쟁  
리뷰 '

## Insight Inference & Visualization

In [157]:

```
1 # content를 위에서 만든 함수인 display_word_cloud로 워드클라우드를 시각화 해봅니다.
2 display_word_cloud(content, width=1200, height=500)
```

