In [1]:
```
# 데이터 분석 라이브러리
import numpy as np
import pandas as pd

# 시각화 라이브러리
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns

# 경고 제거
import warnings
warnings.filterwarnings(action='ignore')
```

In [2]:
```
# 한글 폰트 해결
from matplotlib import font_manager, rc
font_name = font_manager.FontProperties(fname="c:/Windows/Fonts/malgun.ttf").get_name
rc('font', family=font_name)
```

유동인구 데이터 (내국인, 장기체류 외국인, 단기체류 외국인)

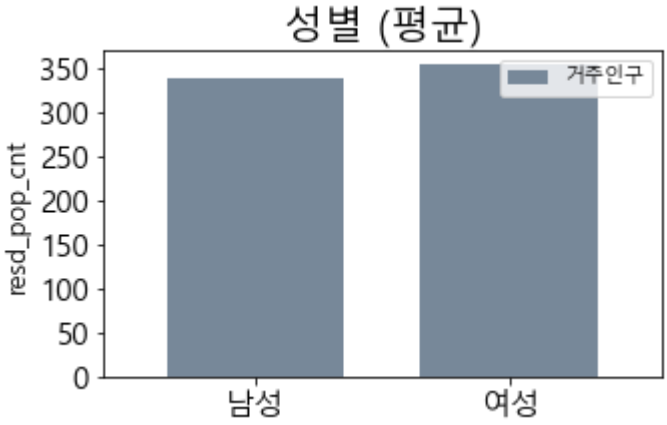
내국인 유동인구 데이터

In [62]:
```
# 데이터 불러오기
korea = pd.read_csv("KOREAN.CSV", encoding = "cp949")
korea.tail(3)
```

Out[62]:

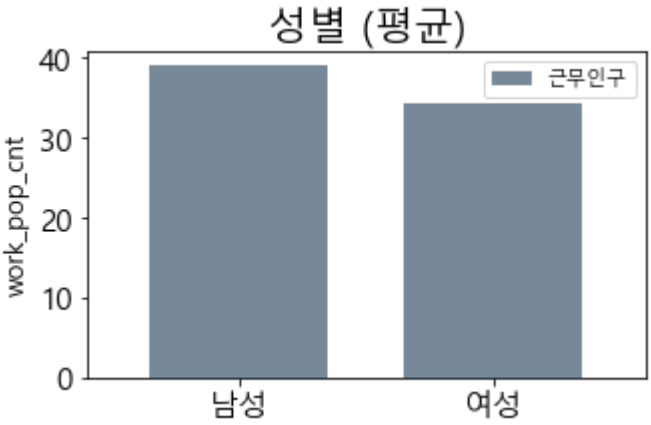
	base_date	time	city	emd_cd	emd_nm	sex	age	resd_pop_cnt	work_pop_cnt	visit_p
47250193	2021-06-30	24	서귀포시	50130620	예래동	남성	10	121.8467	0.0	
47250194	2021-06-30	24	서귀포시	50130620	예래동	남성	80	122.2190	0.0	
47250195	2021-06-30	24	서귀포시	50130620	예래동	남성	80	0.0000	0.0	

In [50]:
```
# 거주인구 성별 확인
korea.groupby('sex')['resd_pop_cnt'].mean().plot(kind = "bar", figsize = (5,3), fonts

plt.xlabel('', fontsize = 13)
plt.xticks(rotation = 0)
plt.ylabel('resd_pop_cnt', fontsize = 13)
plt.legend(["거주인구"])
plt.title("성별 (평균)", fontsize = 20)
plt.show()
```



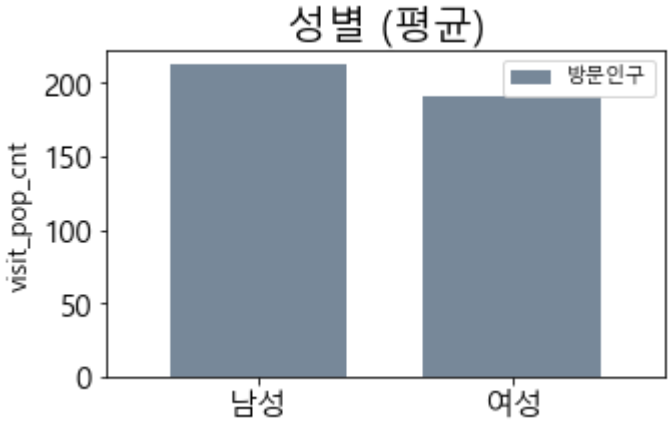
In [51]:

```
# 근무인구 성별 확인
korea.groupby('sex')['work_pop_cnt'].mean().plot(kind = "bar", figsize = (5,3), fonts

plt.xlabel('', fontsize = 13)
plt.xticks(rotation = 0)
plt.ylabel('work_pop_cnt', fontsize = 13)
plt.legend(["근무인구"])
plt.title("성별 (평균)", fontsize = 20)
plt.show()
```



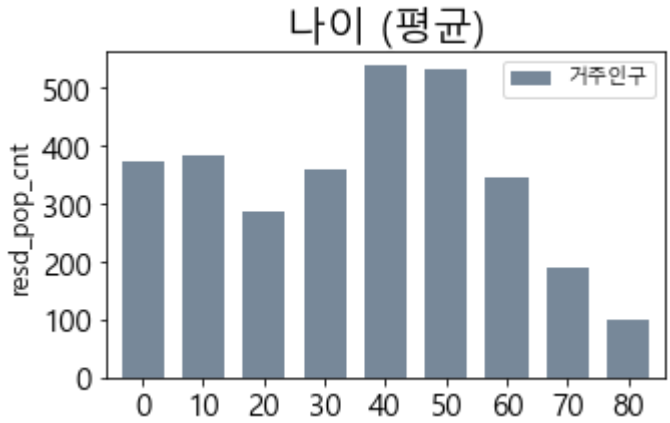
In [52]:

```
# 방문인구 성별 확인
korea.groupby('sex')['visit_pop_cnt'].mean().plot(kind = "bar", figsize = (5,3), font

plt.xlabel('', fontsize = 13)
plt.xticks(rotation = 0)
plt.ylabel('visit_pop_cnt', fontsize = 13)
plt.legend(["방문인구"])
plt.title("성별 (평균)", fontsize = 20)
plt.show()
```



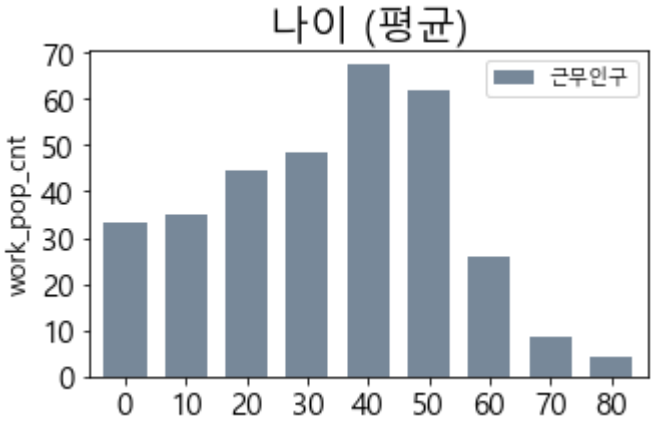
In [56]:

```
# 거주인구 나이 확인
korea.groupby('age')['resd_pop_cnt'].mean().plot(kind = "bar", figsize = (5,3), fonts

plt.xlabel('', fontsize = 13)
plt.xticks(rotation = 0)
plt.ylabel('resd_pop_cnt', fontsize = 13)
plt.legend(["거주인구"])
plt.title("나이 (평균)", fontsize = 20)
plt.show()
```



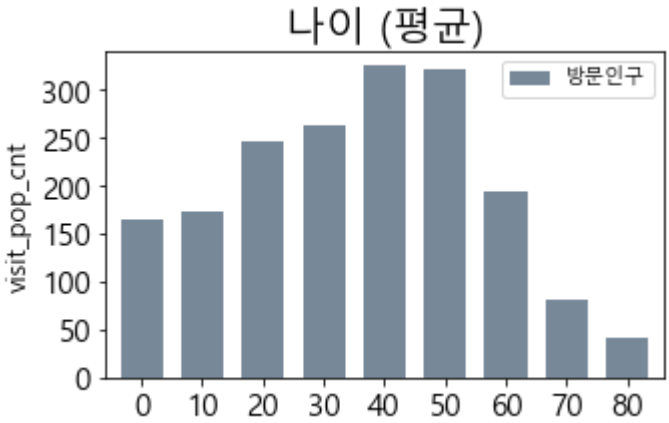
In [54]:

```
# 근무인구 나이 확인
korea.groupby('age')['work_pop_cnt'].mean().plot(kind = "bar", figsize = (5,3), fonts

plt.xlabel('', fontsize = 13)
plt.xticks(rotation = 0)
plt.ylabel('work_pop_cnt', fontsize = 13)
plt.legend(["근무인구"])
plt.title("나이 (평균)", fontsize = 20)
plt.show()
```



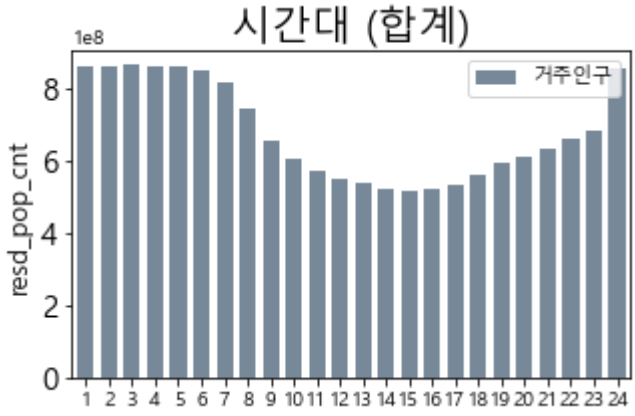
In [55]:

```
# 방문인구 나이 확인
korea.groupby('age')['visit_pop_cnt'].mean().plot(kind = "bar", figsize = (5,3), font

plt.xlabel('', fontsize = 13)
plt.xticks(rotation = 0)
plt.ylabel('visit_pop_cnt', fontsize = 13)
plt.legend(["방문인구"])
plt.title("나이 (평균)", fontsize = 20)
plt.show()
```

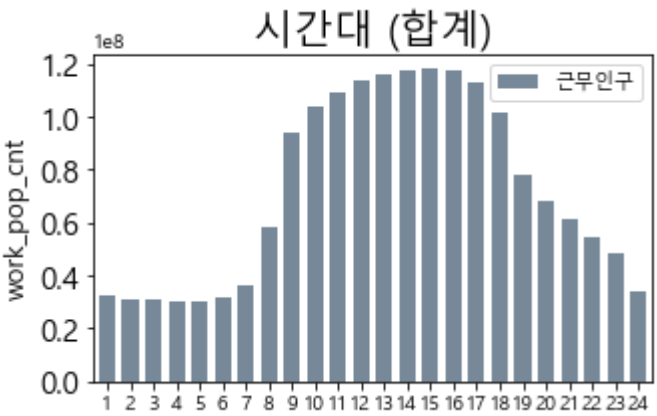


In [35]:

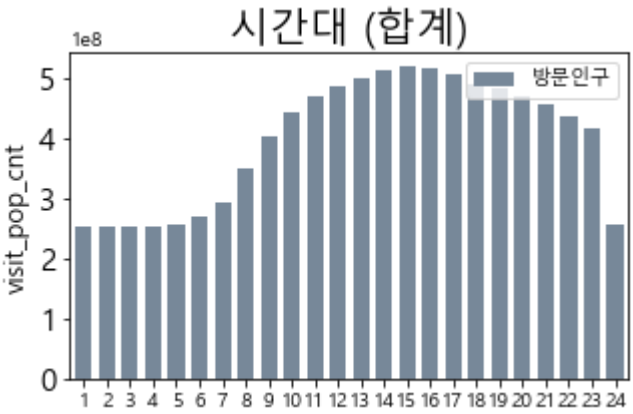
```
# 시간대별 거주인구 확인
korea.groupby('time')['resd_pop_cnt'].sum().plot(kind = "bar", figsize = (5,3), fonts

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 10, rotation = 0)
plt.ylabel('resd_pop_cnt', fontsize = 13)
plt.legend(["거주인구"],loc="upper right")
plt.title("시간대 (합계)", fontsize = 20)
plt.show()
```



In [36]:
```
# 시간대별 근무인구 확인
korea.groupby('time')['work_pop_cnt'].sum().plot(kind = "bar", figsize = (5,3), fonts

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 10, rotation = 0)
plt.ylabel('work_pop_cnt', fontsize = 13)
plt.legend(["근무인구"])
plt.title("시간대 (합계)", fontsize = 20)
plt.show()
```



In [37]:
```
# 시간대별 방문인구 확인
korea.groupby('time')['visit_pop_cnt'].sum().plot(kind = "bar", figsize = (5,3), font

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 10, rotation = 0)
plt.ylabel('visit_pop_cnt', fontsize = 13)
plt.legend(["방문인구"],loc="upper right")
plt.title("시간대 (합계)", fontsize = 20)
plt.show()
```



장기체류 유동인구 데이터

In [13]:
```
# 장기체류 외국인 데이터 불러오기
LONG_TERM_FRGN = pd.read_csv("LONG_TERM_FRGN.CSV", encoding = "cp949")
LONG_TERM_FRGN.tail()
```

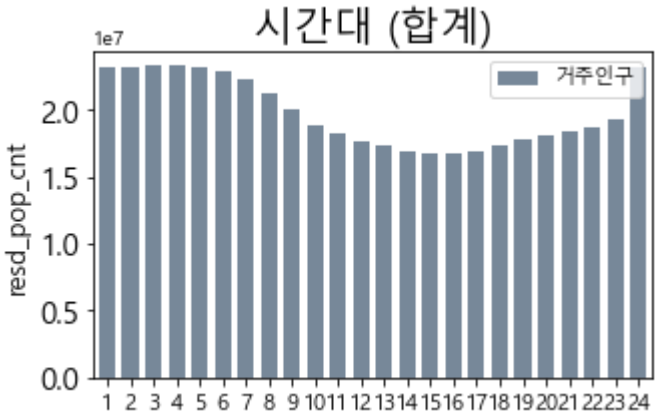
Out[13]:

	base_date	time	nationality	city	emd_cd	emd_nm	resd_pop_cnt	work_pop_cnt	visit
23837312	2021-06-30	24	THA	서귀포시	50130620	예래동	0.0000	0.0	

	base_date	time	nationality	city	emd_cd	emd_nm	resd_pop_cnt	work_pop_cnt	visit
23837313	2021-06-30	24	TWN	서귀포시	50130620	예래동	0.0000	0.0	
23837314	2021-06-30	24	USA	서귀포시	50130620	예래동	22.0624	0.0	
23837315	2021-06-30	24	USA	서귀포시	50130620	예래동	0.0000	0.0	
23837316	2021-06-30	24	VNM	서귀포시	50130620	예래동	3.7754	0.0	

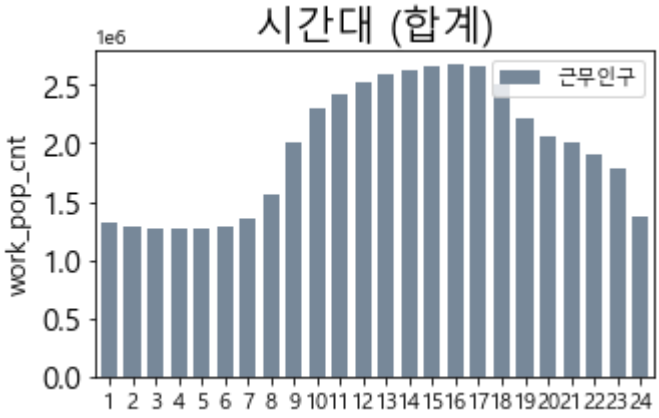
In [38]:

```
# 거주인구 시간 확인
LONG_TERM_FRGN.groupby('time')['resd_pop_cnt'].sum().plot(kind = "bar", figsize = (5,

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 11, rotation = 0)
plt.ylabel('resd_pop_cnt', fontsize = 13)
plt.legend(["거주인구"],loc="upper right")
plt.title("시간대 (합계)", fontsize = 20)
plt.show()
```



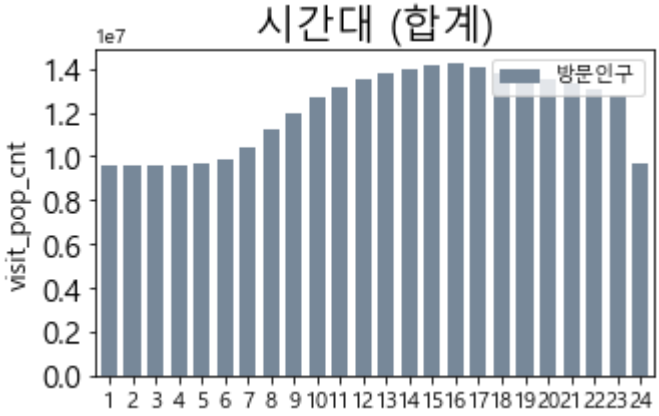
In [39]:

```
# 시간대별 근무인구 확인
LONG_TERM_FRGN.groupby('time')['work_pop_cnt'].sum().plot(kind = "bar", figsize = (5,

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 11, rotation = 0)
plt.ylabel('work_pop_cnt', fontsize = 13)
plt.legend(["근무인구"],loc="upper right")
plt.title("시간대 (합계)", fontsize = 20)
plt.show()
```



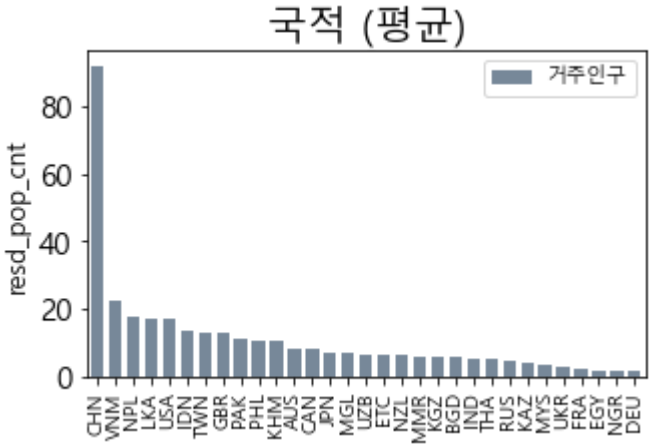
In [41]:

```
# 시간대별 방문인구 확인
LONG_TERM_FRGN.groupby('time')['visit_pop_cnt'].sum().plot(kind = "bar", figsize = (5

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 11, rotation = 0)
plt.ylabel('visit_pop_cnt', fontsize = 13)
plt.legend(["방문인구"],loc="upper right")
plt.title("시간대 (합계)", fontsize = 20)
plt.show()
```



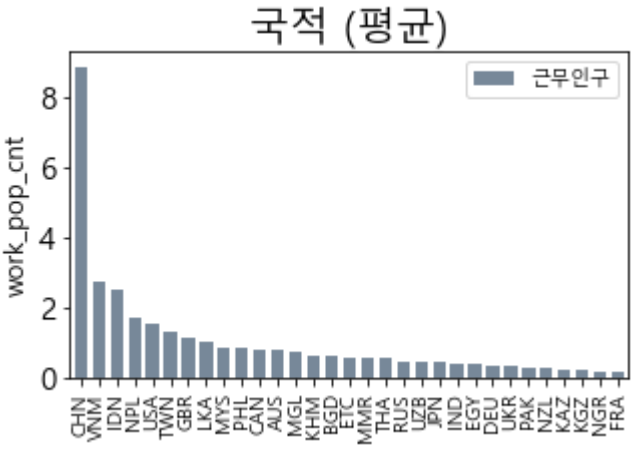
In [60]:

```
# 국적별 거주인구 확인
LONG_TERM_FRGN.groupby('nationality')['resd_pop_cnt'].mean().sort_values("index",asce

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 10, rotation = 90)
plt.ylabel('resd_pop_cnt', fontsize = 13)
plt.legend(["거주인구"],loc="upper right")
plt.title("국적 (평균)", fontsize = 20)
plt.show()
```



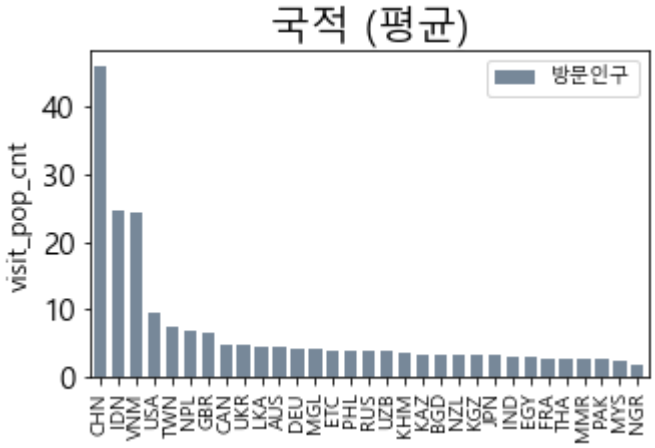
In [58]:

```
# 국적별 거주인구 확인
LONG_TERM_FRGN.groupby('nationality')['work_pop_cnt'].mean().sort_values("index",asce

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 10, rotation = 90)
plt.ylabel('work_pop_cnt', fontsize = 13)
plt.legend(["근무인구"],loc="upper right")
plt.title("국적 (평균)", fontsize = 20)
plt.show()
```



In [59]:

```
# 국적별 방문인구 확인
LONG_TERM_FRGN.groupby('nationality')['visit_pop_cnt'].mean().sort_values("index",asc

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 10, rotation = 90)
plt.ylabel('visit_pop_cnt', fontsize = 13)
plt.legend(["방문인구"],loc="upper right")
plt.title("국적 (평균)", fontsize = 20)
plt.show()
```



In [20]:
```
# 년-월만 추출
em_cnt_M = LONG_TERM_FRGN["base_date"]

c3_list = []
for i in em_cnt_M:
    i = str(i)  # 숫자 -> 문자로 변환
    i = i[0:7]  # month 부분만 추출
    c3_list.append(i)  # list 로 저장
#print(c2_list)

LONG_TERM_FRGN.insert(1,"base_date_ym",c3_list)
```

In [21]:
```
LONG_TERM_FRGN["people"] = LONG_TERM_FRGN["resd_pop_cnt"] + LONG_TERM_FRGN["work_pop_c
LONG_TERM_FRGN.head(3)
```

Out[21]:

	base_date	base_date_ym	time	nationality	city	emd_cd	emd_nm	resd_pop_cnt	work_pop_cn
0	2018-01-01	2018-01	1	CAN	제주시	50110250	한림읍	6.3381	0.000
1	2018-01-01	2018-01	1	CHN	제주시	50110250	한림읍	442.1091	0.000
2	2018-01-01	2018-01	1	CHN	제주시	50110250	한림읍	0.0000	7.869

단기체류 외국인 데이터

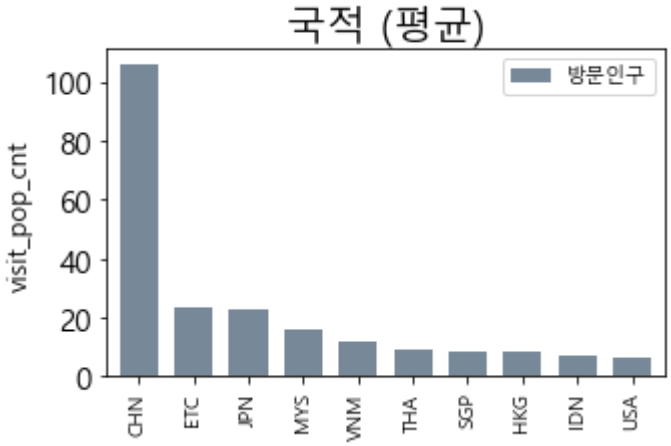
In [22]:
```
SHORT_TERM_FRGN = pd.read_csv("SHORT_TERM_FRGN.CSV", encoding = "cp949")
SHORT_TERM_FRGN.tail(3)
```

Out[22]:

	base_date	time	nationality	city	emd_cd	emd_nm	visit_pop_cnt
6059041	2021-06-30	24	ETC	서귀포시	50130610	중문동	11.5473
6059042	2021-06-30	24	CHN	서귀포시	50130620	예래동	0.2013
6059043	2021-06-30	24	HKG	서귀포시	50130620	예래동	0.0166

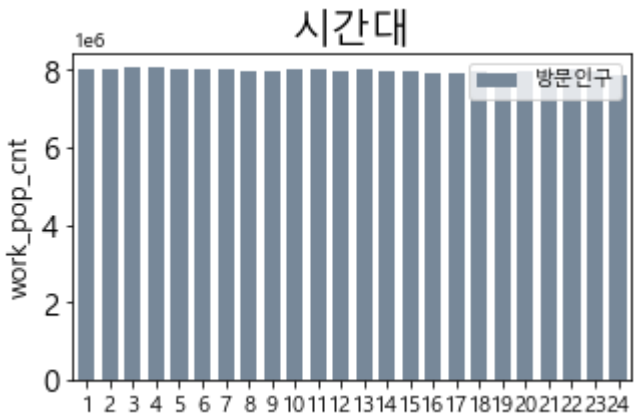
In [61]:
```
# 국적별 방문인구 확인
```

```
SHORT_TERM_FRGN.groupby('nationality')['visit_pop_cnt'].mean().sort_values("index",as

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 10, rotation = 90)
plt.ylabel('visit_pop_cnt', fontsize = 13)
plt.legend(["방문인구"],loc="upper right")
plt.title("국적 (평균)", fontsize = 20)
plt.show()
```



In [24]:

```
# 시간대별 방문인구 확인
SHORT_TERM_FRGN.groupby('time')['visit_pop_cnt'].sum().plot(kind = "bar", figsize = (

plt.xlabel('', fontsize = 13)
plt.xticks(fontsize = 11, rotation = 0)
plt.ylabel('work_pop_cnt', fontsize = 13)
plt.legend(["방문인구"],loc="upper right")
plt.title("시간대", fontsize = 20)
plt.show()
```



In [25]:

```
# 시간대별로 중복이 있으므로 일별 평균 후 -> 월별 합

LONG_TERM_FRGN_ = LONG_TERM_FRGN["mean_visit"] = LONG_TERM_FRGN.groupby(["base_date"]

SHORT_TERM_FRGN_ = SHORT_TERM_FRGN["mean_visit"] = SHORT_TERM_FRGN.groupby(["base_dat
LONG_TERM_FRGN_ = LONG_TERM_FRGN_.reset_index()
SHORT_TERM_FRGN_ = SHORT_TERM_FRGN_.reset_index()
```

In [26]:

```
# 년-월만 추출
em_cnt_M = LONG_TERM_FRGN_["base_date"]

c3_list = []
for i in em_cnt_M:
```

```
    i = str(i)  # 숫자 -> 문자로 변환
    i = i[0:7]  # month 부분만 추출
    c3_list.append(i)  # list 로 저장
#print(c2_list)

LONG_TERM_FRGN_.insert(1,"base_date_ym",c3_list)

em_cnt_M = SHORT_TERM_FRGN_["base_date"]

c3_list = []
for i in em_cnt_M:
    i = str(i)  # 숫자 -> 문자로 변환
    i = i[0:7]  # month 부분만 추출
    c3_list.append(i)  # list 로 저장
#print(c2_list)

SHORT_TERM_FRGN_.insert(1,"base_date_ym",c3_list)
```

In [27]:

```
LONG_TERM_FRGN_.head(3)
```

Out[27]:

	base_date	base_date_ym	visit_pop_cnt
0	2018-01-01	2018-01	9.989425
1	2018-01-02	2018-01	9.441093
2	2018-01-03	2018-01	9.596244

In [46]:

```
# 단기&장기체류 외국인 방문인구 그래프 확인
plt.figure(figsize=(15,4))  # size 지정

plt.plot(LONG_TERM_FRGN_["base_date_ym"].unique(), LONG_TERM_FRGN_.groupby('base_date_
plt.plot(LONG_TERM_FRGN_["base_date_ym"].unique(), SHORT_TERM_FRGN_.groupby('base_dat

plt.xticks(fontsize = 11, rotation = 45)
plt.ylabel("", fontsize = 13)
plt.legend(["장기체류 외국인 방문인구","단기체류 외국인 방문인구"], loc = "upper righ
plt.title("일평균 유동인구 (합계)", fontsize = 18)
plt.show()
```