

In [1]:

```
%matplotlib notebook
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:

```
#전세계 각 대륙별 서울시 외국인 관광객수
#코로나 바이러스로 인해서 외국인 관광객이 급감 1년사이 급감하는 것을 알 수 있다.
```

In [5]:

```
city=pd.read_csv('seoul_visitor_utf8.csv',engine='python')
```

In [6]:

```
city=city.dropna() #결측값 제거
#seoul_visitor.csv파일은 dropna()만으로 각 칼럼의 결측값만 제거시 올바른 데이터 파악 가능한 csv
파일
city=city.set_index('대륙')
# 대륙별 전체 관광객수/남성관광객수/여성관광객수/승무원관광객수 파악하기 위해 index를 '대륙'으로
설정
city
```

Out[6]:

	국적	인원	전년동기	성장률	구성비	남성인원	남성인원성장률	여성인원	여성인원성장률	승무원인원	승무원인원성장률
대륙											
아시아주	중국	332,474	535,536	-37.9	29.3	129,908	-31.3	191,236	-40.4	11,330	-55.3
아시아주	일본	193,705	197,209	-1.8	17.1	75,992	-4.7	115,397	0.1	2,316	8.4
아시아주	대만	75,738	64,488	17.4	6.7	22,794	12.1	52,063	20.1	881	11.9
아시아주	홍콩	58,761	63,998	-8.2	5.2	20,931	-12.5	37,018	-5.7	812	-1.5
아시아주	필리핀	37,019	45,529	-18.7	3.3	12,278	14.2	13,297	37.0	11,444	-54.4
...
아프리카	아프리카 기타	2,578	2,589	-0.4	0.2	1,796	27.8	476	20.2	306	-61.2
아프리카	아프리카 소계	3,312	3,575	-7.4	0.3	2,145	20.9	794	10.6	373	-65.6
기타	국적미상	54	59	-8.5	0.0	35	66.7	7	-46.2	12	-52
교포	교포	22,416	22,765	-1.5	2.0	9,738	-0.8	12,678	-2.1	0	-
교포	교포 소계	22,416	22,765	-1.5	2.0	9,738	-0.8	12,678	-2.1	0	-

65 rows × 11 columns

In [7]:

city.head()

Out[7]:

	국 적	인원	전년동 기	성장 률	구성 비	남성인 원	남성인 원성장 률	여성인 원	여성인 원성장 률	승무원 인원	승무 원인성 장률
대륙											
아시아주	중국	332,474	535,536	-37.9	29.3	129,908	-31.3	191,236	-40.4	11,330	-55.3
아시아주	일본	193,705	197,209	-1.8	17.1	75,992	-4.7	115,397	0.1	2,316	8.4
아시아주	대만	75,738	64,488	17.4	6.7	22,794	12.1	52,063	20.1	881	11.9
아시아주	홍콩	58,761	63,998	-8.2	5.2	20,931	-12.5	37,018	-5.7	812	-1.5
아시아주	필리핀	37,019	45,529	-18.7	3.3	12,278	14.2	13,297	37.0	11,444	-54.4

In [8]:

```
asia_models=city.loc['아시아주']['인원'].values
europe_models=city.loc['구주']['인원'].values
america_models=city.loc['미주']['인원'].values
pacific_models=city.loc['대양주']['인원'].values
africa_models=city.loc['아프리카']['인원'].values
```

In [9]:

```
asia_people=asia_models[len(asia_models)-1]
europe_people=europe_models[len(europe_models)-1]
america_people=america_models[len(america_models)-1]
pacific_people=pacific_models[len(pacific_models)-1]
africa_people=africa_models[len(africa_models)-1]
```

In [10]:

```
results=[asia_people,europe_people,america_people,pacific_people,africa_people]
```

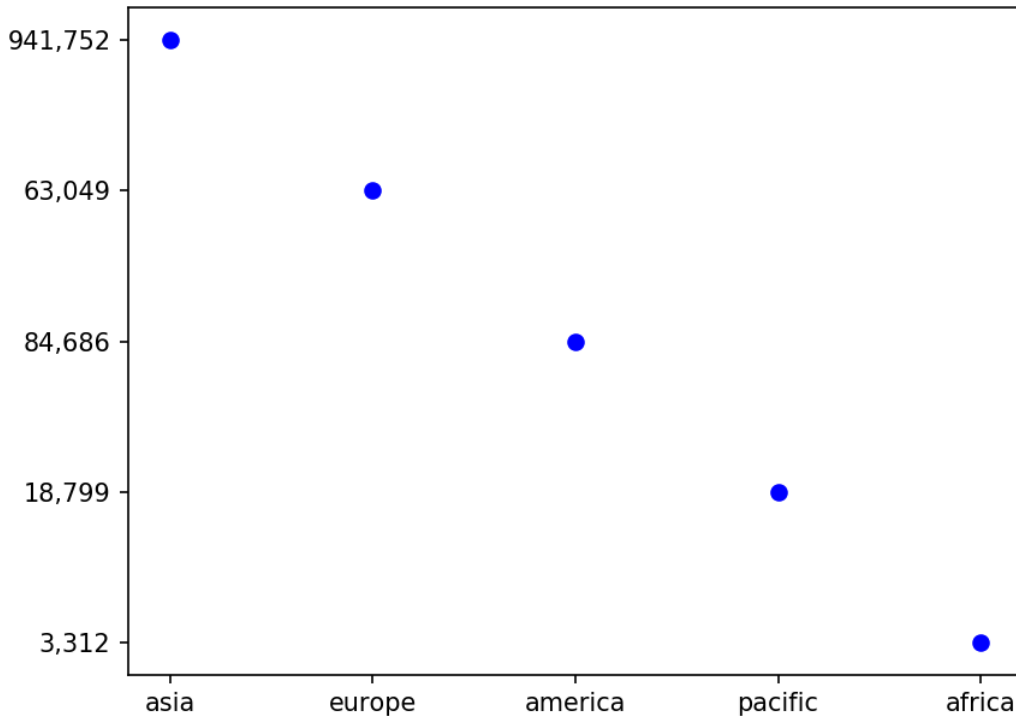
In [11]:

```
names=['asia','europe','america','pacific','africa']
```

In [12]:

```
#5개의 대륙별 전체 관광객수를 시각화
fig=plt.figure()
fig.suptitle('Seoul Visitor Area Comparison')
plt.gca().invert_yaxis()
plt.scatter(names,results,color='blue')
plt.show()
```

Seoul Visitor Area Comparison



In [13]:

```
man_results=[]
woman_results=[]
crew_results=[]
```

In [14]:

```
#대륙별 남자 관광객수 표현하기
```

In [15]:

```
asia_man_models=city.loc['아시아주']['남성인원'].values
america_man_models=city.loc['미주']['남성인원'].values
europe_man_models=city.loc['구주']['남성인원'].values
pacific_man_models=city.loc['대양주']['남성인원'].values
africa_man_models=city.loc['아프리카']['남성인원'].values
```

In [16]:

```
asia_men=asia_man_models[len(asia_man_models)-1]
europe_men=europe_man_models[len(europe_man_models)-1]
america_men=america_man_models[len(america_man_models)-1]
pacific_men=pacific_man_models[len(pacific_man_models)-1]
africa_men=africa_man_models[len(africa_man_models)-1]
```

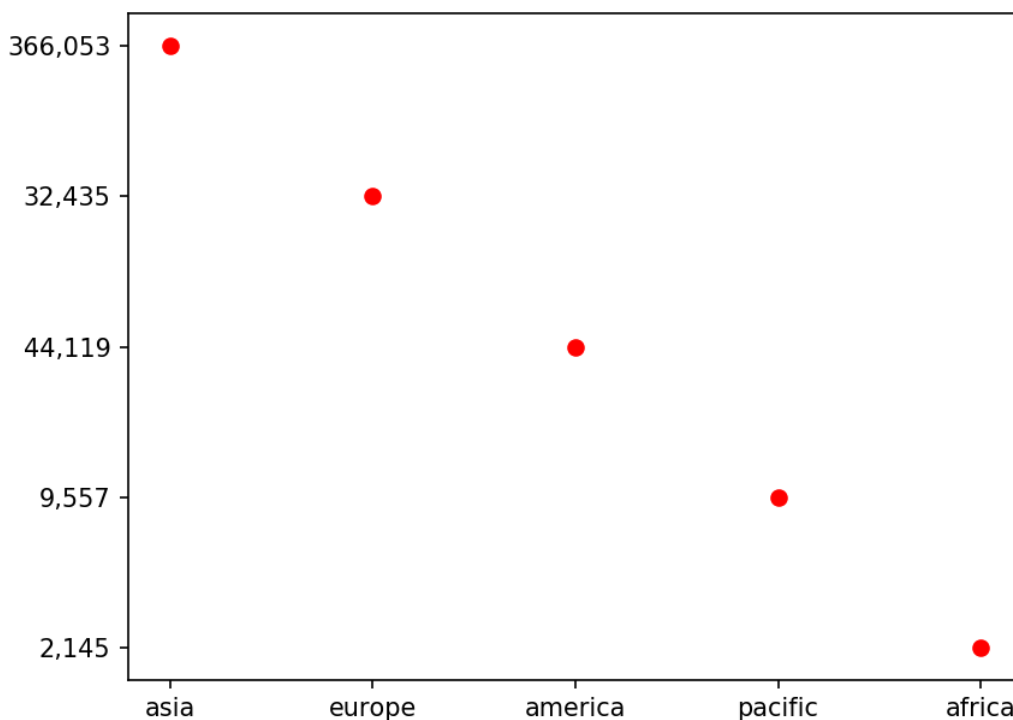
In [17]:

```
man_results=[asia_men,europe_men,america_men,pacific_men,africa_men]
```

In [18]:

```
#대륙별 남자 관광객수 시각화
fig=plt.figure()
fig.suptitle('Seoul Men Visitor Area Comparison')
plt.gca().invert_yaxis()
plt.scatter(names,man_results,color='red')
plt.show()
```

Seoul Men Visitor Area Comparison



In [19]:

```
#대륙별 여자 관광객수 표현하기
```

In [20]:

```
asia_woman_models=city.loc['아시아주']['여성인원'].values
america_woman_models=city.loc['미주']['여성인원'].values
europe_woman_models=city.loc['구주']['여성인원'].values
pacific_woman_models=city.loc['대양주']['여성인원'].values
africa_woman_models=city.loc['아프리카']['여성인원'].values
```

In [21]:

```
asia_women=asia_woman_models[len(asia_woman_models)-1]
europe_women=europe_woman_models[len(europe_woman_models)-1]
america_women=america_woman_models[len(america_woman_models)-1]
pacific_women=pacific_woman_models[len(pacific_woman_models)-1]
africa_women=africa_woman_models[len(africa_woman_models)-1]
```

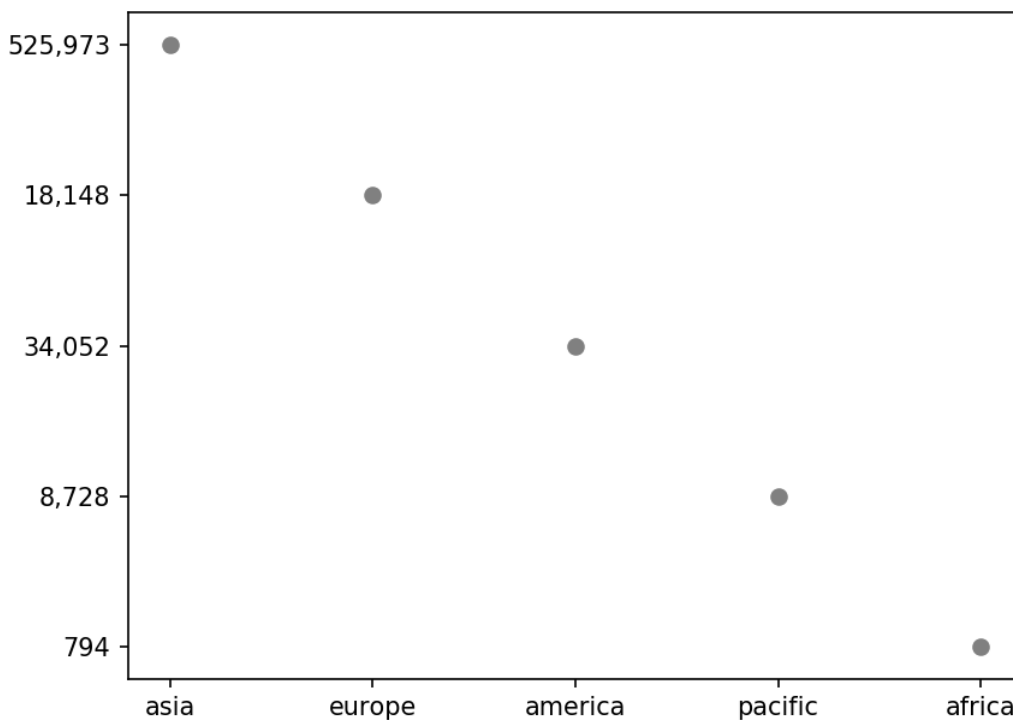
In [22]:

```
woman_results=[asia_women,europe_women,america_women,pacific_women,africa_women]
```

In [23]:

```
#대륙별 여자 관광객수 시각화
fig=plt.figure()
fig.suptitle('Seoul Women Visitor Area Comparison')
plt.gca().invert_yaxis()
plt.scatter(names,woman_results,color='gray')
plt.show()
```

Seoul Women Visitor Area Comparison



In [24]:

```
#대륙별 승무원관광객수 표현하기
```

In [25]:

```
asia_crew_models=city.loc['아시아주']['승무원인원'].values
america_crew_models=city.loc['미주']['승무원인원'].values
europe_crew_models=city.loc['구주']['승무원인원'].values
pacific_crew_models=city.loc['대양주']['승무원인원'].values
africa_crew_models=city.loc['아프리카']['승무원인원'].values
```

In [26]:

```
asia_crew=asia_crew_models[len(asia_crew_models)-1]
europe_crew=europe_crew_models[len(europe_crew_models)-1]
america_crew=america_crew_models[len(america_crew_models)-1]
pacific_crew=pacific_crew_models[len(pacific_crew_models)-1]
africa_crew=africa_crew_models[len(africa_crew_models)-1]
```

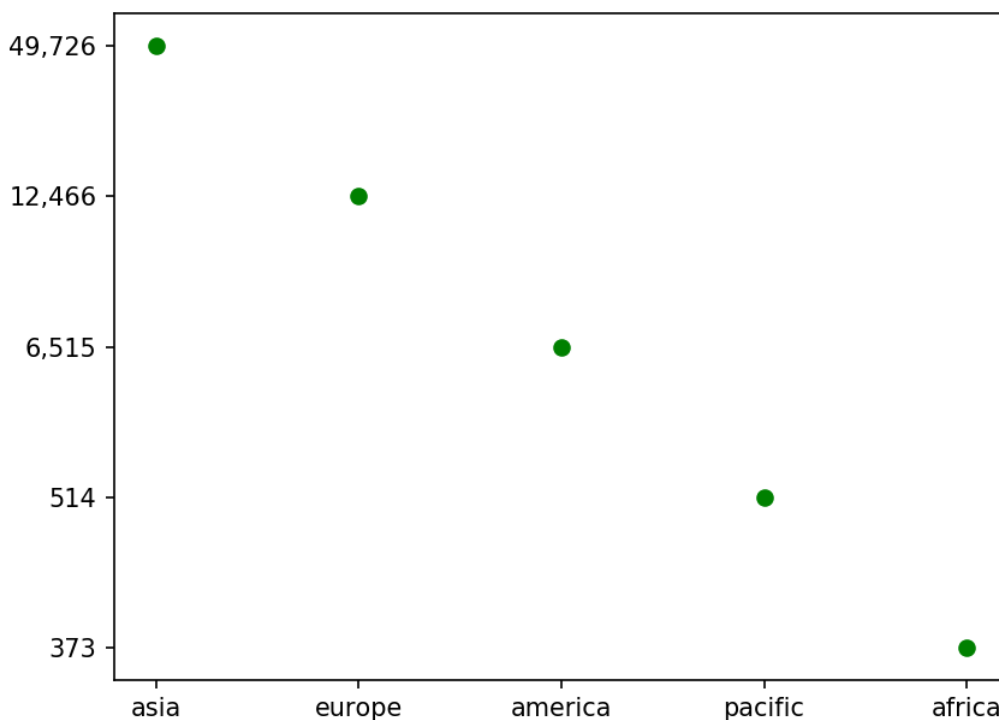
In [27]:

```
crew_results=[asia_crew,europe_crew,america_crew,pacific_crew,africa_crew]
```

In [28]:

```
#대륙별 승무원관광객수 시각화
fig=plt.figure()
fig.suptitle('Seoul Crew Visitor Area Comparison')
plt.gca().invert_yaxis()
plt.scatter(names,crew_results,color='green')
plt.show()
```

Seoul Crew Visitor Area Comparison



In [29]:

```
#전체 data의 데이터 타입을 string->float로 바꾸기 위해서 sep=', '을 추가
```

In [30]:

```
from sklearn.model_selection import train_test_split
visitors=pd.read_table('seoul_visitors_utf8.csv',engine='python',sep=',',thousands = ',')
```

In [31]:

```
visitors.head()  
visitors=visitors.dropna()  
visitors
```


Out[31]:

	label	대 륙	국 적	인원	전년동기	성장 률	구성 비	남성인원	남성 인원 성장 률	여성인원	여성 인원 성장 률	승무원 인원
3	1.0	아 시 아 주	중 국	332474.0	535536.0	-37.9	29.3	129908.0	-31.3	191236.0	-40.4	11330.0
4	1.0	아 시 아 주	일 본	193705.0	197209.0	-1.8	17.1	75992.0	-4.7	115397.0	0.1	2316.0
5	1.0	아 시 아 주	대 만	75738.0	64488.0	17.4	6.7	22794.0	12.1	52063.0	20.1	881.0
6	1.0	아 시 아 주	홍 콩	58761.0	63998.0	-8.2	5.2	20931.0	-12.5	37018.0	-5.7	812.0
7	1.0	아 시 아 주	필 리 핀	37019.0	45529.0	-18.7	3.3	12278.0	14.2	13297.0	37.0	11444.0
...
63	5.0	아 프 리 카	아 프 리 카 기 타	2578.0	2589.0	-0.4	0.2	1796.0	27.8	476.0	20.2	306.0
64	5.0	아 프 리 카	아 프 리 카 소 계	3312.0	3575.0	-7.4	0.3	2145.0	20.9	794.0	10.6	373.0
65	6.0	기 타	국 적 미 상	54.0	59.0	-8.5	0.0	35.0	66.7	7.0	-46.2	12.0
67	7.0	교 포	교 포	22416.0	22765.0	-1.5	2.0	9738.0	-0.8	12678.0	-2.1	0.0
68	7.0	교 포	교 포 소 계	22416.0	22765.0	-1.5	2.0	9738.0	-0.8	12678.0	-2.1	0.0

65 rows × 13 columns

In [32]:

```
#knn 머신러닝 알고리즘과 남성인원, 여성인원, 승무원인원을 가지고 해당 대륙을 추측
```

In [33]:

```
from matplotlib import cm
#feature
X=visitors[['남성인원', '여성인원', '승무원인원']]
Y=visitors['label']
X_train, X_test, Y_train, Y_test=train_test_split(X,Y,random_state=0)
```

In [34]:

```
from sklearn.neighbors import KNeighborsClassifier
#parameter change
k_range=range(1,10)
counts=[]

for k in k_range:
    knn=KNeighborsClassifier(n_neighbors=k)
    knn.fit(X_train,Y_train)
    counts.append(knn.score(X_test,Y_test))
```

In [35]:

```
counts
```

Out[35]:

```
[0.4117647058823529,
 0.35294117647058826,
 0.4117647058823529,
 0.35294117647058826,
 0.4117647058823529,
 0.4117647058823529,
 0.4117647058823529,
 0.4117647058823529,
 0.4117647058823529,
 0.4117647058823529]
```

In [36]:

```
lookup_area_name=dict(zip(visitors.label.unique(), visitors.대륙.unique()))
```

In [37]:

```
area_prediction=knn.predict([[7524,4483,908]])
lookup_area_name[area_prediction[0]]
```

Out[37]:

```
'아시아주'
```

In []: