In [ ]:

```
- how to create a column

- how to drop a column

- shape-size

- head-tail

- take-loc-iloc

- info- len

- is null
```

�������� − 1
:

Import required packages

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

�������� − 2

Read the data

In [2]: In [3]:

```
file_path="C:\\Users\\omkar\\OneDrive\\Do

cuments\\Data science\\Naresh IT\\

visa_df=pd.read_csv(file_path)


visa_df.head()
```

```
- read the data

- how to create a data frame
```

Out[3]: **case_id continent education_of_employee has_job_experience requires_job_training no_** **0** EZYV01

Asia High School N N **1** EZYV02 Asia Master's Y N **2** EZYV03 Asia Bachelor's N Y **3** EZYV04 Asia Bachelor's N N **4**

EZYV05 Africa Master's Y N

�������� − 3

Read a column

In [4]:

```
visa_df.col
umns
```

Out[4]:

```
Index(['case_id', 'continent', 'education_of_employee', 'has_job_experienc
e',
   'requires_job_training',   'no_of_employees',   'yr_of_estab',
 'region_of_employment',    'prevailing_wage',    'unit_of_wage',
 'full_time_position', 'case_status'],
  dtype='object')
```

In [5]:

```
visa_df['conti
nent']
```

Out[5]:

```
0 Asia
1 Asia
2 Asia
3 Asia
4 Africa
  ...
25475 Asia
25476 Asia
25477 Asia
25478 Asia
25479 Asia
Name: continent, Length: 25480, dtype: object
```

In [8]:

```
type(visa_df['con
tinent'])
```

Out[8]:

```
pandas.core.series.Series
```

In [7]:

```
cols=['continent']
visa_df[cols]


#visa_df['continent']=======                    ===== series
                                                #visa_df[['continent']]
                                                ======== dataframe(table)
```

Out[7]:

	continent
0	Asia
1	Asia
2	Asia
3	Asia
4	Africa
...	...
25475	Asia
25476	Asia
25477	Asia
25478	Asia
25479	Asia

25480 rows × 1 columns

```
              type(visa_df[
In [9]:       cols])

Out[9]: pandas.core.frame.DataFrame
             visa_df.conti
In [10]:     nent

Out[10]: 0 Asia
          1 Asia
          2 Asia
          3 Asia
          4 Africa
            ...
          25475 Asia
          25476 Asia
          25477 Asia
          25478 Asia
          25479 Asia
          Name: continent, Length: 25480, dtype: object
                        ']]
                        #visa_df.continent
In [ ]: In [11]:


                        # Two columns at
                        time
#visa_df['continent'    cols=['continent','c
]                       ase_status']
#visa_df[['continent    visa_df[cols]
```

Out[11]: **continent case_status** **0** Asia

Denied

1 Asia Certified

2 Asia Denied

3 Asia Denied

4 Africa Certified

... ...

25475 Asia Certified

25476 Asia Certified

25477 Asia Certified

25478 Asia Certified

25479 Asia Certified

25480 rows × 2 columns

������������

```
                                for dataframe : [[]] # apply
In [13]:                        only for series :[]
# first read the column
# the apply unique              visa_df['continent'].unique()

# dont apply unique operation

Out[13]: array(['Asia', 'Africa', 'North America', 'Europe', 'South America',
          'Oceania'], dtype=object)
                          len(visa_df['continen
                          t'].unique())
In [14]:

Out[14]: 6
                      len(visa_df['cont
                      inent'])
In [15]:

Out[15]: 25480
```

���

���

���

���

��

```
In [16]:                    number of unique
visa_df['continent'
                            lables
].nunique() #

Out[16]: 6
                          visa_df[['continent','
                          case_status']]
In [17]:

Out[17]: continent case_status 0 Asia

            Denied

            1 Asia Certified

            2 Asia Denied

            3 Asia Denied

            4 Africa Certified

            ... ... ...

         25475 Asia Certified

         25476 Asia Certified

         25477 Asia Certified

         25478 Asia Certified

         25479 Asia Certified
```

25480 rows × 2 columns

In [ ]: In [23]:

```
# we read continent column
# we understood there 6 unique
lables are there # these 6 unique
lables repeating and toatl 25480
observations # how many are asia
are there
# how many are africa are there


visa_df['continent']=='Asia'

# do you want to know how many True
# how many rows are satisfying
condition
# how many observations are having
continent as asia
```

Out[23]:

```
0 True
1 True
2 True
3 True
4 False
  ...
25475 True
25476 True
25477 True
25478 True
25479 True
Name: continent, Length: 25480, dtype: bool
```

In [25]:

```
visa_df[visa_df['contin
ent']=='Asia']
```

Out[25]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_traini
0	EZYV01	Asia	High School	N	
1	EZYV02	Asia	Master's	Y	
2	EZYV03	Asia	Bachelor's	N	
3	EZYV04	Asia	Bachelor's	N	
5	EZYV06	Asia	Master's	Y	
...	...	...	...	...	
25475	EZYV25476	Asia	Bachelor's	Y	
25476	EZYV25477	Asia	High School	Y	
25477	EZYV25478	Asia	Master's	Y	
25478	EZYV25479	Asia	Master's	Y	
25479	EZYV25480	Asia	Bachelor's	Y	

16861 rows × 12 columns

```
len(visa_df[visa_df['cont
inent']=='Asia'])
```

In [26]:

Out[26]: 16861

In [ ]: In [28]:

In [30]: In [32]:

```
visa_df # complete df
visa_df['continent'] # column
visa_df['continent']=='Asia' #
one label
visa_df[visa_df['continent']=='As
ia'] # df
len(visa_df[visa_df['continent']=
='Asia']) # len


############### BAD
WAY###############################
#
print(len(visa_df[visa_df['contin
ent']=='Asia']))
print(len(visa_df[visa_df['contin
ent']=='Africa']))
print(len(visa_df[visa_df['contin
ent']=='North America']))
print(len(visa_df[visa_df['contin
ent']=='Europe']))
print(len(visa_df[visa_df['contin
ent']=='South America']))
print(len(visa_df[visa_df['contin
ent']=='Oceania']))
```

```
continents=visa_df['continent'].u
nique()
# for loop
print(len(visa_df[visa_df['contin
ent']==i]))
```

```
16861
551
3292
3732
852
192
```

```
continents=visa_df['continent'].u
nique()
for i in continents:

count=len(visa_df[visa_df['contin
ent']==i])  print(i,':',count)
```

```
Asia : 16861
Africa : 551
North America : 3292
Europe : 3732
South America : 852
Oceania : 192
```

```
count=[]
continents=visa_df['continent'].u
nique()
for i in continents:

c=len(visa_df[visa_df['continent'
]==i])
 count.append(c)
```

```
count
```

```
Out[32]: [16861, 551, 3292, 3732, 852, 192]
```

```
In [37]:
continents=visa_df['continent'].uniqu
e()
count=[len(visa_df[visa_df['continent
']==i]) for i in continents]

continents_df=pd.DataFrame(zip(contin
ents,count),
columns=['Continents','Count'])

continents_df.to_csv('continetns_info
.csv',index=False)
```

```
In [ ]: In [45]:
```

```
visa_df # complete df
visa_df['continent'] # column
visa_df['continent']=='Asia' # one label
visa_df[visa_df['continent']=='Asia'] # df
len(visa_df[visa_df['continent']=='Asia'])
# len

############################################
##################################
print(len(visa_df[visa_df['continent']=='A
sia']))
print(len(visa_df[visa_df['continent']=='A
frica']))
print(len(visa_df[visa_df['continent']=='N
orth America']))
print(len(visa_df[visa_df['continent']=='E
urope']))
print(len(visa_df[visa_df['continent']=='S
outh America']))
print(len(visa_df[visa_df['continent']=='O
ceania']))

############################################
##################################
continents=visa_df['continent'].unique()
for i in continents:

count=len(visa_df[visa_df['continent']==i]
)
 print(i,':',count)

############################################
##################################

continents=visa_df['continent'].unique()
count=[len(visa_df[visa_df['continent']==i
]) for i in continents]

continents_df=pd.DataFrame(zip(continents,
count),
 columns=['Continents','Count'])

continents_df.to_csv('continetns_info.csv'

,index=False)


��
��������-�����������
�

pd.Series(count,index=continents)
```

```
Out[45]: Asia 16861 Africa 551
          North America 3292
          Europe 3732
          South America 852
          Oceania 192
          dtype: int64
```

```
In [39]:                visa_df['continent'].
                        value_counts()
```

```
Out[39]: continent
          Asia 16861
          Europe 3732
          North America 3292
          South America 852
          Africa 551
          Oceania 192
          Name: count, dtype: int64
```

```
In [ ]: In [46]:           visa_df['continent'].value_cou
                           nts()

                           # in order to creaate a
# Always learn how a method a   dataframe
giving answer # will im able   # we need two list
to write same answer with out  # or one dictionary
method
                           # from value counts create two
                           lists
# How to create a dataframe    # values
using value counts # or using  # keys
series
```

```
Out[46]: continent
          Asia 16861
          Europe 3732
          North America 3292
          South America 852
          Africa 551
          Oceania 192
          Name: count, dtype: int64
```

```
                                    count=visa_df['continent'].value_coun
                                    ts().values
In [50]:                            pd.DataFrame(zip(continents,count),co
# Method-1:                         lumns=['continetns','count'])
continents=visa_df['continent'].value
_counts().keys()
```

Out[50]: **continetns count** **0** Asia

16861

1 Europe 3732

2 North America 3292

3 South America 852

4 Africa 551

5 Oceania 192

In [57]:

```
# Method-2
dict1=dict(visa_df['continent'].value_cou
nts())
print(dict1)
# 16861 is a scalar value, it is not in a
list
pd.DataFrame(dict1,index=['count'])
```

```
{'Asia': 16861, 'Europe': 3732, 'North
America': 3292, 'South America': 85 2,
'Africa': 551, 'Oceania': 192}
```

Out[57]: **Asia Europe North America South America Africa Oceania count** 16861 3732

3292 852 551 192

In [62]:

```
# Method-3
dict1=dict(visa_df['continent'].va
lue_counts()) keys=dict1.keys()
values=dict1.values()

pd.DataFrame(zip(keys,values),colu
mns=['Continent','Count'])
```

Out[62]: **Continent Count** **0** Asia

16861

1 Europe 3732

2 North America 3292

3 South America 852

4 Africa 551

5 Oceania 192

In [68]:

```
dict1={}
keys=visa_df['continent'].va
lue_counts().keys()
dict1['continents']=keys
dict1
```

```
Out[68]: {'continents': Index(['Asia', 'Europe', 'North America', 'South America',
'Africa',
  'Oceania'],
  dtype='object', name='continent')}
```

In [69]:

```
continents_df

# one column=
categorical :
Continents # collumn=
numerical : Count
```

Out[69]: **Continents Count**

0 Asia 16861

1 Africa 551

2 North America 3292

3 Europe 3732

4 South America 852

5 Oceania 192

barplot
pie chart

������-���������

barplot
x-axis: categorical column
y-axis: numerical column
where you are taking the data: continents_df

In [71]:
```
continents_df      # we are creating
                   from scratch
```

Out[71]: **Continents Count** **0** Asia

16861

1 Africa 551

2 North America 3292

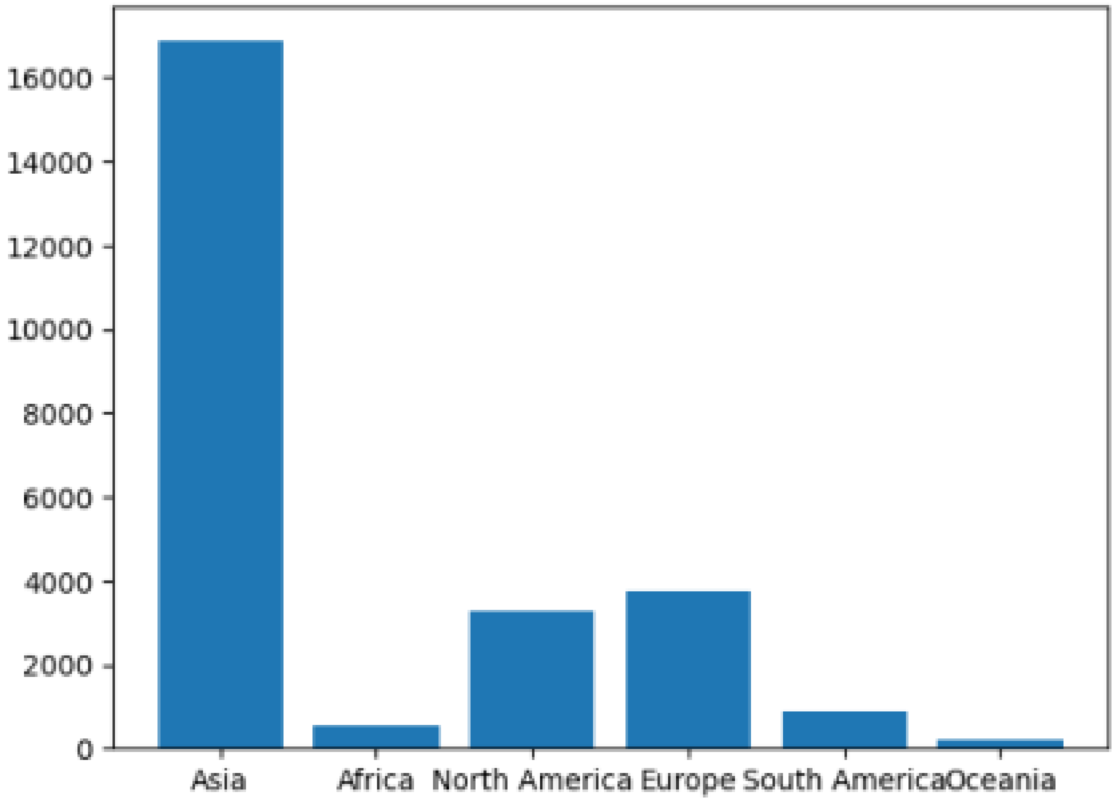
3 Europe 3732

4 South America 852

5 Oceania 192

In [70]:
```
plt.bar('Continent       data=continents_df
s',  'Count',        )
```

Out[70]:
```
<BarContainer object of 6 artists>
```



In [ ]:

In [74]:

```
# Method-1:
visa_df['continent'].value_counts()
continents=visa_df['continent'].value_counts().keys()
count=visa_df['continent'].value_counts().values
contint_data=pd.DataFrame(zip(continents,count),
columns=['continetns','count'])
contint_data
```

Out[74]:

	continetns	count
0	Asia	16861
1	Europe	3732
2	North America	3292
3	South America	852
4	Africa	551
5	Oceania	192

In [84]:

```
plt.figure(figsize=(10,5))
# 10= horizontal x
# 5= vertical y
plt.bar('continetns','count',data=contint_data)
plt.title("Bar chart")
plt.xlabel("Continents")
plt.ylabel("Count")
plt.savefig("continents_bar.jpg")
plt.show()
```

Reading a cat column

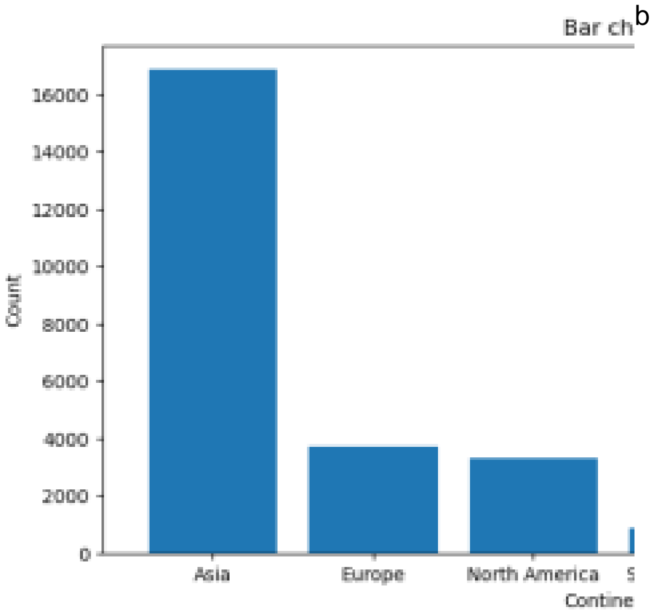
unique

nunique

value counts

frequency table (dataframe)

barplot



In [ ]:

```
# read the data
```

In [ ]:

```
# Date: 12-12-2023
```

In [ ]: In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:

```
# Whenever you open notebook as fresh
# you need to run
# packages
```

```
file_path="C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\

visa_df=pd.read_csv(file_path)
visa_df.head(2)
```

Out[3]: **case_id continent education_of_employee has_job_experience requires_job_training no_** **0** EZYV01 Asia High School N N **1** EZYV02 Asia Master's Y N

bar plot using seaborn

```
In [ ]: In [ ]:
```

```
# in order to draw a bar chart we
required frequency table # continent
column
# we created one more data frame
```

```
# haveing each lable frequncy
# asia 16k
# africa

# the above things are required, if you

want to draw bar chart using matplo #

saeborn will take directly the original

column from original data frame
```

In [ ]: In [4]:

- matplotlib bar chart requires 3 values

- x axis : categorical column

- y axis : numerical column

- data name

- from original data frame visadf ,

the original column is continent - we

created another table which has labels

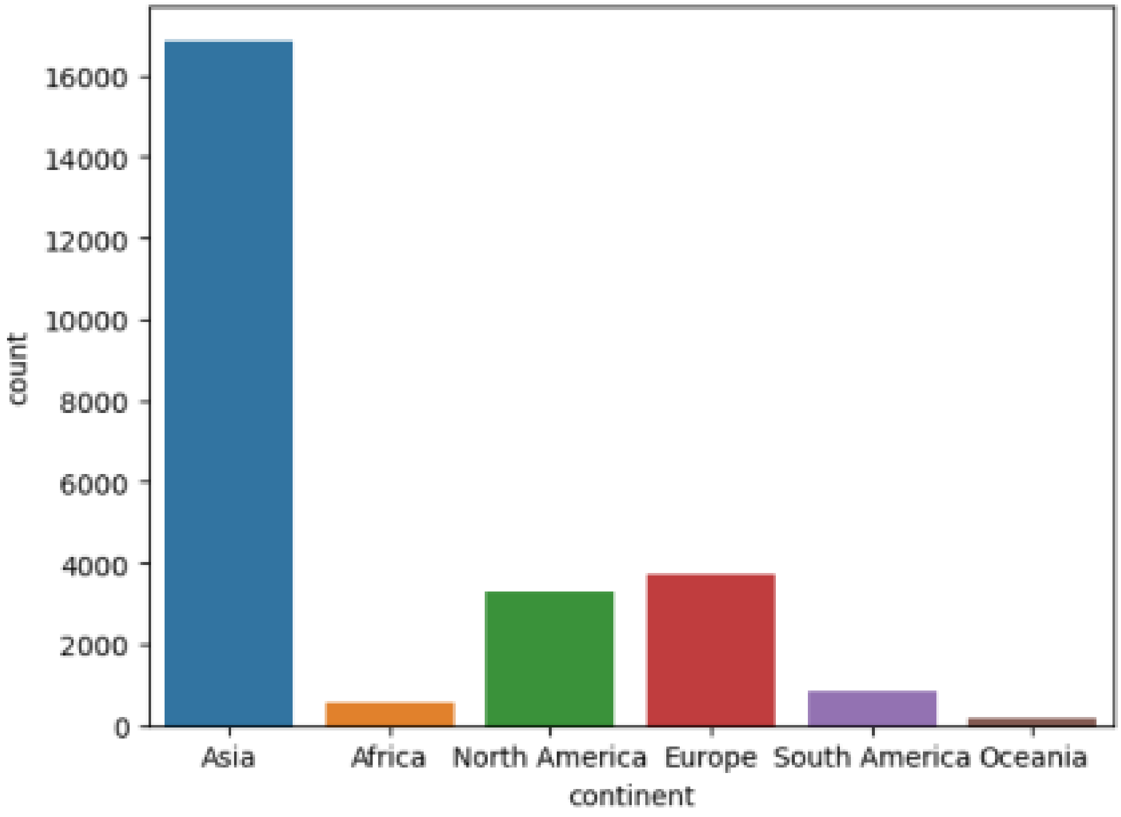
and its count

|continetns| count|
0 Asia 16861
1 Europe 3732
2 North America 3292
3 South America 852
4 Africa 551
5 Oceania 192

seaborn requires 2 values
data: original dataframe : visa_df
x : original column name : 'continent'

```
import seaborn as sns
sns.countplot(data=visa_df,x='continen
t')
```

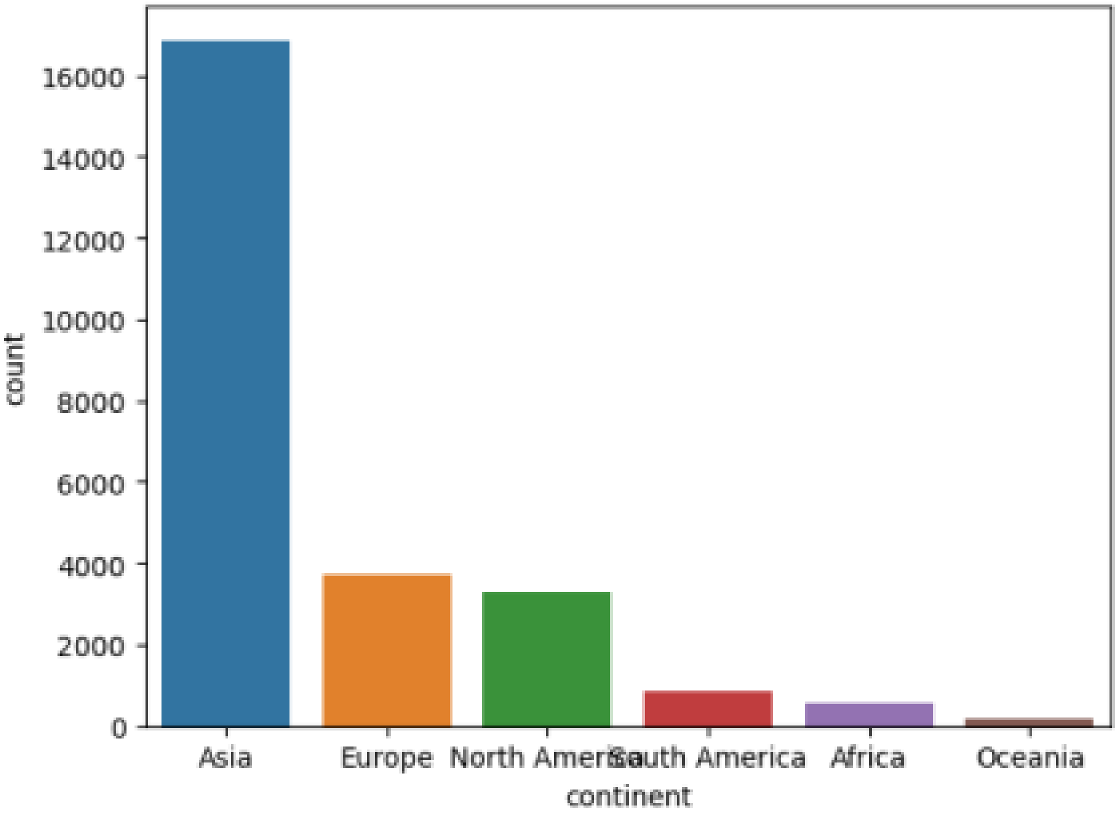
```
Out[4]: <Axes: xlabel='continent', ylabel='count'>
```



In [5]:

```
import seaborn as sns
labels=['Asia','Europe','
North America',  'South   f,
America','Africa','Oceani  x='continent',
a']                        order=labels)
sns.countplot(data=visa_d
```

```
Out[5]: <Axes: xlabel='continent', ylabel='count'>
```

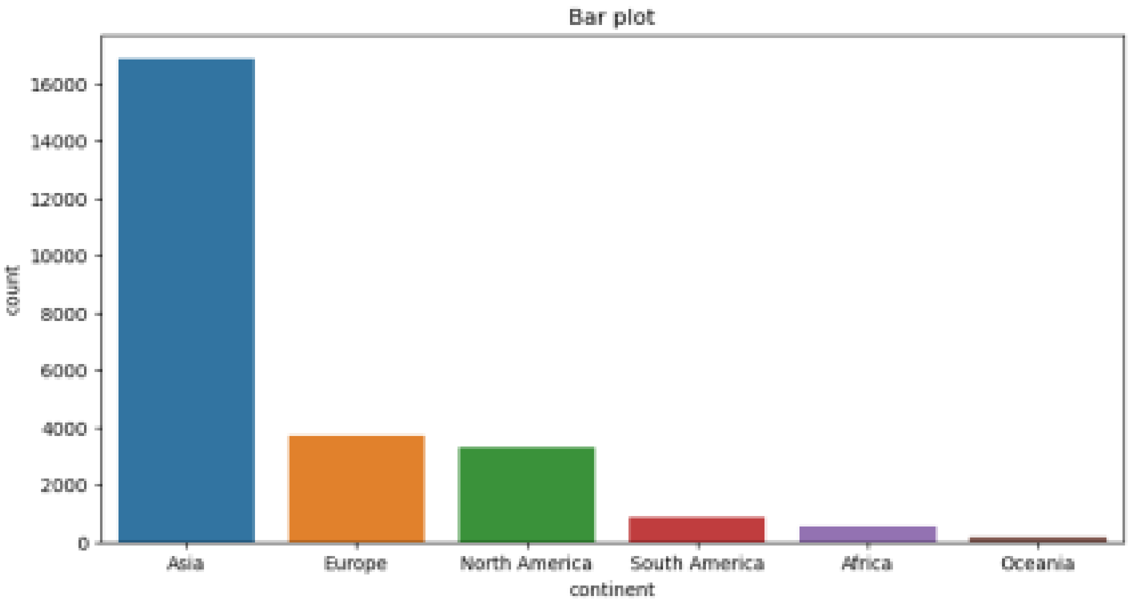


```
In [9]:                    e_counts().keys()
visa_df['continent'].valu
```

```
Out[9]: Index(['Asia', 'Europe', 'North America', 'South America', 'Africa',
          'Oceania'],
          dtype='object', name='continent')
```

```
In [11]: import seaborn as sns
```

```
labels=visa_df['continent'].value_counts().keys()
plt.figure(figsize=(10,5))
sns.countplot(data=visa_df,
 x='continent',
 order=labels)
plt.title("Bar plot")
plt.savefig("Continent_bar_seaborn")
plt.show()
```

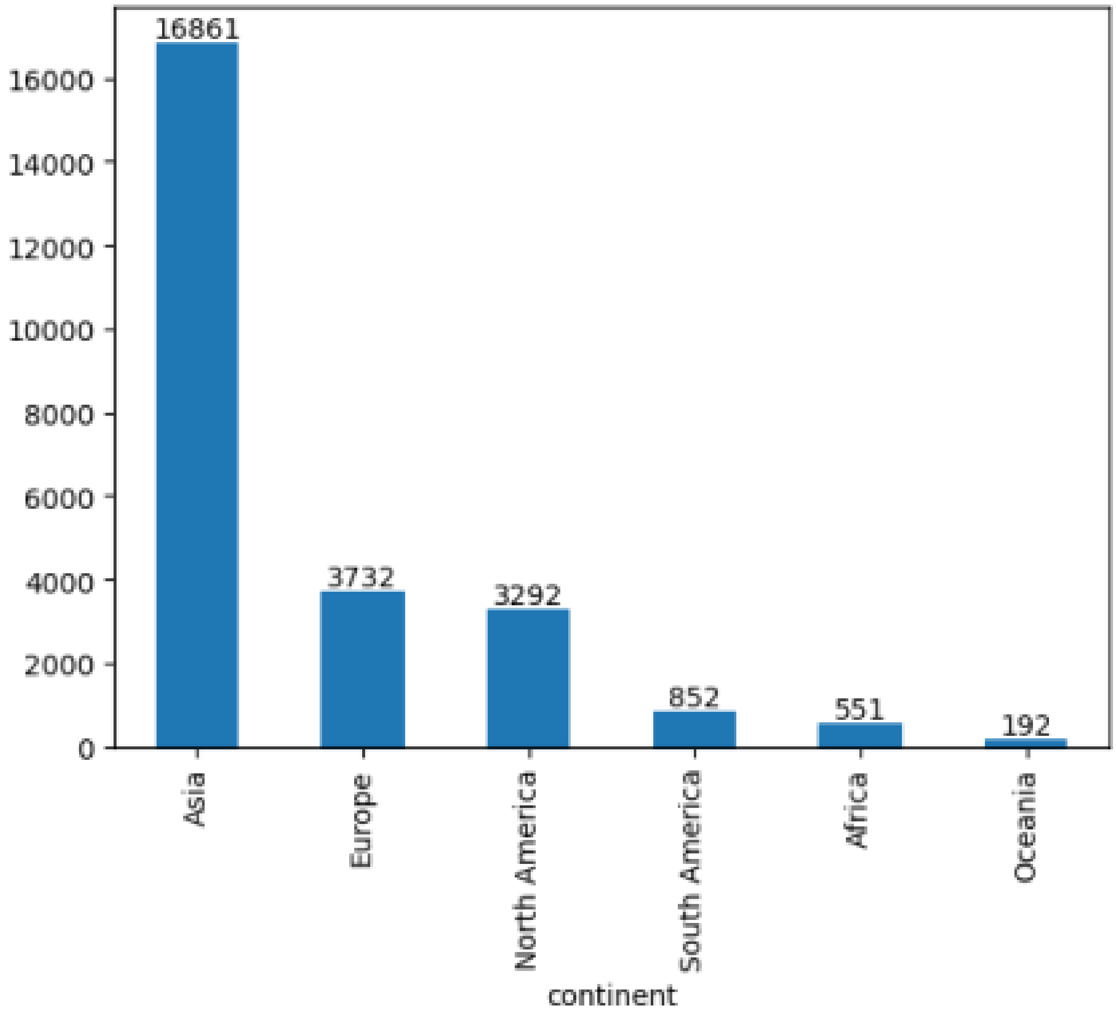


In [ ]: In [ ]:

```
#Method-1: using matplotlib

###################################
Reading the data#########################
file_path="C:\\Users\\omkar\\OneDrive\\Do
cuments\\Data science\\Naresh IT\\

visa_df=pd.read_csv(file_path)
visa_df.head(2)


############################## Create a
frequency table ####################

visa_df['continent'].value_counts()
continents=visa_df['continent'].value_counts().keys()
count=visa_df['continent'].value_counts()
.values
contint_data=pd.DataFrame(zip(continents,
count),
 columns=['continetns','count'])

contint_data

######################################plot
###################################
plt.figure(figsize=(10,5))
# 10= horizontal x
# 5= vertical y
plt.bar('continetns','count',data=contint
_data)
plt.title("Bar chart")
plt.xlabel("Continents")
plt.ylabel("Count")
plt.savefig("continents_bar.jpg")
plt.show()


# Method-2: Seaborn

###################################
Reading the data#########################
file_path="C:\\Users\\omkar\\OneDrive\\Do
cuments\\Data science\\Naresh IT\\

visa_df=pd.read_csv(file_path)
visa_df.head(2)

#########################################
plot############################## import
seaborn as sns
labels=visa_df['continent'].value_counts(
).keys()
plt.figure(figsize=(10,5))
sns.countplot(data=visa_df,
 x='continent',
 order=labels)
plt.title("Bar plot")
plt.savefig("Continent_bar_seaborn")
plt.show()
```

In [19]: In [29]:

```
# Method-3: using value counts
count=visa_df['continent'].value_counts()
ax=count.plot(kind='bar') # ax= axes
ax.bar_label(ax.containers[0])
plt.show()
```



������-��ℎ�������

```
keys=visa_df['continent'].value_counts(normalize=True).keys()
values=visa_df['continent'].value_counts(normalize=True).values
pd.DataFrame(zip(keys,values),
 columns=['Continent','Relative frequency'])
```

Out[29]: **Continent Relative frequency** **0** Asia

0.661735

1 Europe 0.146468

2 North America 0.129199

3 South America 0.033438

4 Africa 0.021625

5 Oceania 0.007535

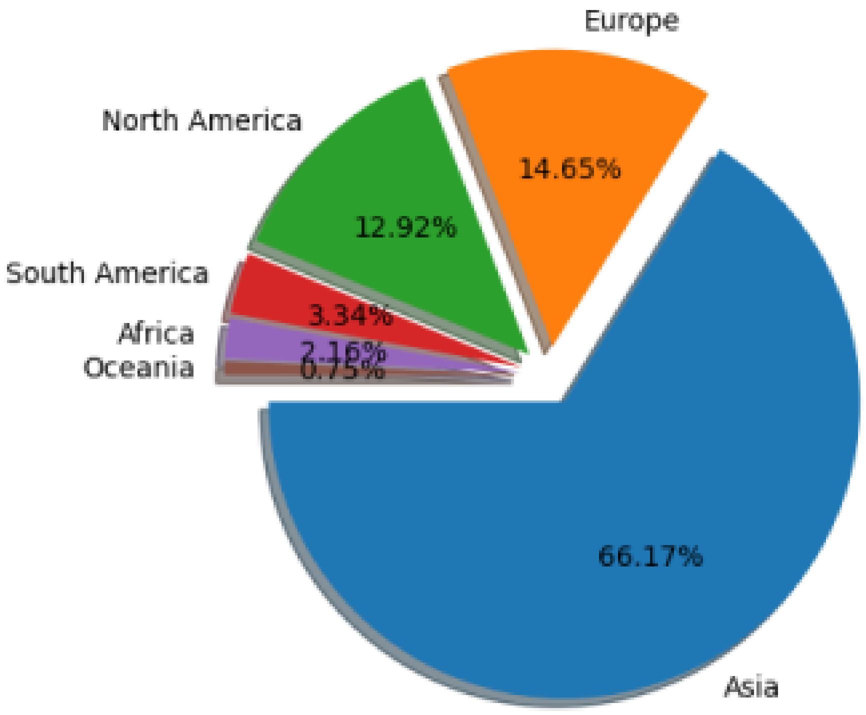
In [47]:

```
keys
```

```
Out[47]: Index(['Asia', 'Europe', 'North America', 'South America', 'Africa',
           'Oceania'],
            dtype='object', name='continent')
```

In [48]: In [52]:

```
plt.pie(x=values,
 labels=keys,
 autopct="%0.2f%%",
 shadow=True,
 startangle=180,
 radius=1,
 explode=[0.1,0.1,0.1,0.1,0.1,0.1]) # 66% 66.76 plt.show()
```



```
data_types=dict(visa_df.dtypes)
cat=[i for i in data_types if data_types[i]=='O'] cat
```

```
Out[52]: ['case_id',
           'continent',
           'education_of_employee',
           'has_job_experience',
           'requires_job_training',
           'region_of_employment',
           'unit_of_wage',
           'full_time_position',
           'case_status']
```

```
In [58]:
```

```
# this will save dataframes where python file existed
```

create a folder

take the entire path

add double slash at the end

conactenate with your file name

```
file_path='C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\'
```

```
file_path+'{}.csv'.format(i)

# This will save the data frames in a seperate folder
```

In [60]: In [62]:

```
data_types=dict(visa_df.dtypes)
cat=[i for i in data_types if data_types[i]=='O']

for i in cat[1:]:
 visa_df[i].value_counts()
 value1=visa_df[i].value_counts().keys()
 value2=visa_df[i].value_counts().values
 data=pd.DataFrame(zip(value1,value2),
 columns=[i,'count'])

 data.to_csv('{}.csv'.format(i))
```

```
data_types=dict(visa_df.dtypes)
cat=[i for i in data_types if data_types[i]=='O']

for i in cat[1:]:
 visa_df[i].value_counts()
 value1=visa_df[i].value_counts().keys()
 value2=visa_df[i].value_counts().values
 data=pd.DataFrame(zip(value1,value2),
 columns=[i,'count'])

 data.to_csv(file_path+'{}.csv'.format(i))
```

In [49]:

```
visa_df['continent'].value_counts()
continents=visa_df['continent'].value_counts().keys()
count=visa_df['continent'].value_counts().values
contint_data=pd.DataFrame(zip(continents,count),
columns=['continetns','count'])   v")

contint_data.to_csv("continent.cs
```

Out[49]: **continetns count 0** Asia

16861

1 Europe 3732

2 North America 3292

3 South America 852

4 Africa 551

5 Oceania 192

```
In [50]:
visa_df['case_status'].value_coun
ts()
continents=visa_df['case_status']
.value_counts().keys()
count=visa_df['case_status'].valu
```

```
e_counts().values
contint_data=pd.DataFrame(zip(con
tinents,count),
columns=['case_status','count'])

contint_data
```

Out[50]: **continetns count 0**

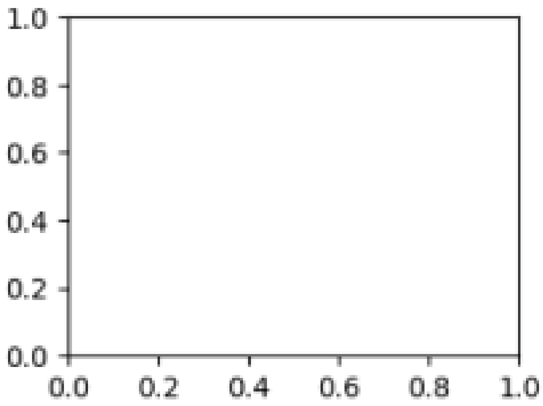
Certified 17018

1 Denied 8462

```
In [ ]:

In [65]:
```

```
plt.subplot(2,
2, 2)
```

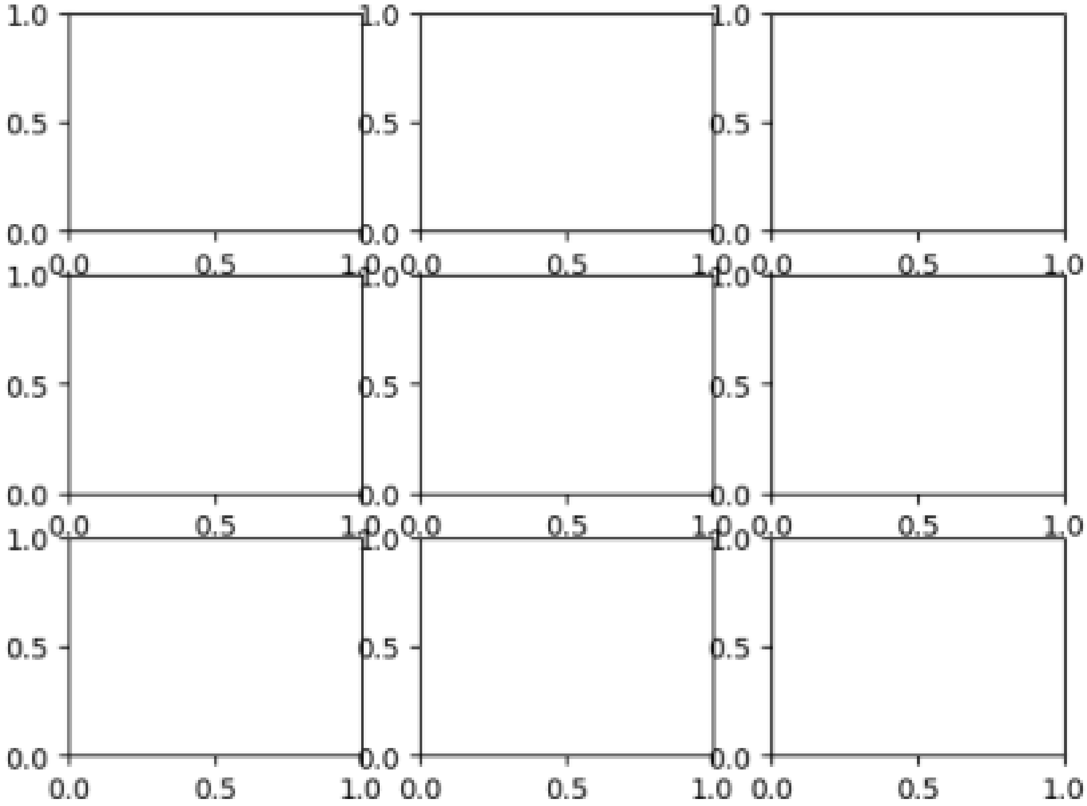
```
Out[65]: <Axes: >
```



```
In [70]:
plt.subplot(3,3,1
)
########
code############
plt.subplot(3,3,2
```

```
)
#########
code######
plt.subplot(3,3,3
)
plt.subplot(3,3,4
)                 )
plt.subplot(3,3,5plt.subplot(3,3,8
)                 )
plt.subplot(3,3,6plt.subplot(3,3,9
)                 )
plt.subplot(3,3,7
```

```
Out[70]: <Axes: >
```



```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```