Step-1: Import the pacakges

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Step-2:Read visa datasets

In [5]:
```
file_path=r'C:\Users\omkar\OneDrive\Documents\Gen_AI\Data_files\Visadataset.csv'
visa_df=pd.read_csv(file_path)
```

Step-3: Divide into categorical and Numerical columns

In [14]:
```
types=visa_df.dtypes.to_dict()
cat=[i for i,j in types.items() if j=='object']
num=[i for i,j in types.items() if j!='object']
```

In [20]:
```
len(cat),len(num)
```

Out[20]: (9, 3)

In [22]:
```
cat=visa_df.select_dtypes(include='object').columns
num=visa_df.select_dtypes(exclude='object').columns
```

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

file_path=r'C:\Users\omkar\OneDrive\Documents\Gen_AI\Data_files\Visadataset.csv'
visa_df=pd.read_csv(file_path)

cat=visa_df.select_dtypes(include='object').columns
num=visa_df.select_dtypes(exclude='object').columns
```

Step-4: How to read a column

In [35]:
```
visa_df['continent']
visa_df[['continent']]
cols=['continent','case_status']
visa_df[cols]
```

Out[35]:

	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 [39]:
```
visa_df['continent']
visa_df[['continent']]
visa_df.loc[:,'continent']
visa_df.loc[:,['continent']]
Id=cat.to_list().index('continent')
visa_df.iloc[:,Id]
visa_df.iloc[:,[Id]]
```

Out[39]:

	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

In [37]:
```
cat
```

Out[37]:
```
Index(['case_id', 'continent', 'education_of_employee', 'has_job_experience',
       'requires_job_training', 'region_of_employment', 'unit_of_wage',
       'full_time_position', 'case_status'],
      dtype='object')
```

Continent Columnan analysis

In [44]:
```
visa_df[['continent']]  # Dataframe
visa_df['continent']  # series
visa_df['continent'].values # Values 25480
```

Out[44]:
```
array(['Asia', 'Asia', 'Asia', ..., 'Asia', 'Asia', 'Asia'], dtype=object)
```

In [46]:
```
len(visa_df['continent'].values)
```

Out[46]:
```
25480
```

Unique

In [51]:
```
visa_df['continent'].unique()
```

Out[51]:
```
array(['Asia', 'Africa', 'North America', 'Europe', 'South America',
       'Oceania'], dtype=object)
```

In [53]:
```
visa_df['continent'].nunique()
```

Out[53]:
```
6
```

TASK:

- we want to know how many asia applicants are from total data 25480

In [70]:
```
visa_df  # Dataframe
visa_df['continent'] # Column
con=visa_df['continent']=='Asia' # Condition
len(visa_df[con]) # selection
```

Out[70]:
```
16861
```

In [72]:
```
con=visa_df['continent']=='Asia'
con=visa_df['continent']=='Asia'
con=visa_df['continent']=='Asia'
con=visa_df['continent']=='Asia'
con=visa_df['continent']=='Asia'
con=visa_df['continent']=='Asia'

# take a empty list
# Apply for loop
con=visa_df['continent']==i
len(visa_df[con])
```

In [74]:
```
lables=visa_df['continent'].unique()
for i in lables:
    print(i)
```

```
Asia
Africa
North America
Europe
South America
Oceania
```

In [78]:
```
l=[]
lables=visa_df['continent'].unique()
for i in lables:
    con=visa_df['continent']==i
    count=len(visa_df[con])
    l.append(count)
```

In [80]:
```
l
```

Out[80]:
```
[16861, 551, 3292, 3732, 852, 192]
```

In [82]:
```
lables
```

Out[82]:
```
array(['Asia', 'Africa', 'North America', 'Europe', 'South America',
       'Oceania'], dtype=object)
```

In [90]:
```
continet_df=pd.DataFrame(zip(lables,l),
                         columns=['Continent','Number of Applicants'])
continet_df
```

Out[90]:

	Continent	Number of Applicants
0	Asia	16861
1	Africa	551
2	North America	3292
3	Europe	3732
4	South America	852
5	Oceania	192

In [92]:
```
continet_df.to_csv('Continent_data',index=False)
```

Value counts

In [94]:
```
visa_df['continent'].value_counts()
```

Out[94]:
```
continent
Asia             16861
Europe            3732
North America     3292
South America      852
Africa             551
Oceania            192
Name: count, dtype: int64
```

In [97]:
```
visa_df['continent'].value_counts().keys()
```

Out[97]:
```
Index(['Asia', 'Europe', 'North America', 'South America', 'Africa',
       'Oceania'],
      dtype='object', name='continent')
```

In [105…
```
visa_df['continent'].value_counts().index
```

Out[105…
```
Index(['Asia', 'Europe', 'North America', 'South America', 'Africa',
       'Oceania'],
      dtype='object', name='continent')
```

In [99]:
```
visa_df['continent'].value_counts().values
```

Out[99]:
```
array([16861,  3732,  3292,   852,   551,   192], dtype=int64)
```

In [107…
```
visa_df['continent'].value_counts()
keys=visa_df['continent'].value_counts().keys()
values=visa_df['continent'].value_counts().values
pd.DataFrame(zip(keys,values),
             columns=['Continent','Number of Applicants'])
```

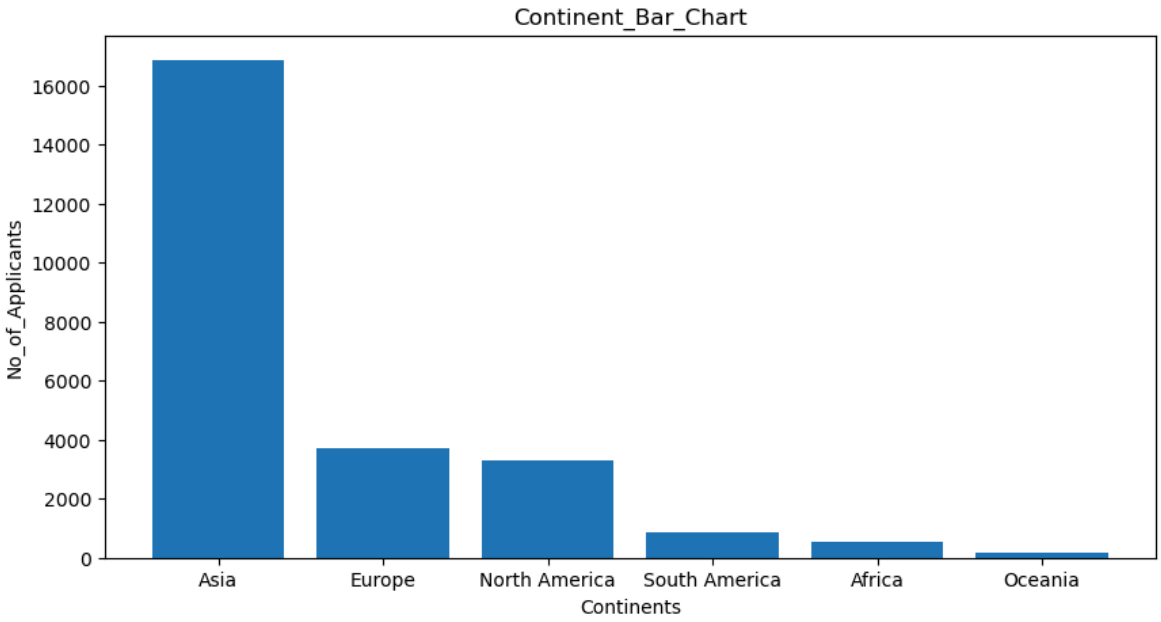
Out[107…

	Continent	Number of Applicants
0	Asia	16861
1	Europe	3732
2	North America	3292
3	South America	852
4	Africa	551
5	Oceania	192

Dont work on CASE ID:25480

Bar chart

In [123…
```
plt.figure(figsize=(10,5))
plt.bar(keys,values)
plt.xlabel('Continents')
plt.ylabel('No_of_Applicants')
plt.title('Continent_Bar_Chart')
plt.savefig('continent_bar_chart.jpg')
plt.show()
```


Continent_Bar_Chart
No_of_Applicants
16000
14000
12000
10000
8000
6000
4000
2000
0
Asia
Europe
North America
South America
Africa
Oceania
Continents

In [ ]: