In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
visa_df=pd.read_csv(r"C:\Users\abhis\OneDrive\Documents\Narash it\data files folder
visa_df
```

Out[2]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_trair
0	EZYV01	Asia	High School	N	
1	EZYV02	Asia	Master's	Y	
2	EZYV03	Asia	Bachelor's	N	
3	EZYV04	Asia	Bachelor's	N	
4	EZYV05	Africa	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	

25480 rows × 12 columns

In [3]:
```
visa_df.head(6)
```

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	
5	EZYV06	Asia	Master's	Y	N	

In [4]:
```
visa_df['no_of_employees']
```

Out[4]:
```
0        14513
1         2412
2        44444
3           98
4         1082
         ...
25475     2601
25476     3274
25477     1121
25478     1918
25479     3195
Name: no_of_employees, Length: 25480, dtype: int64
```

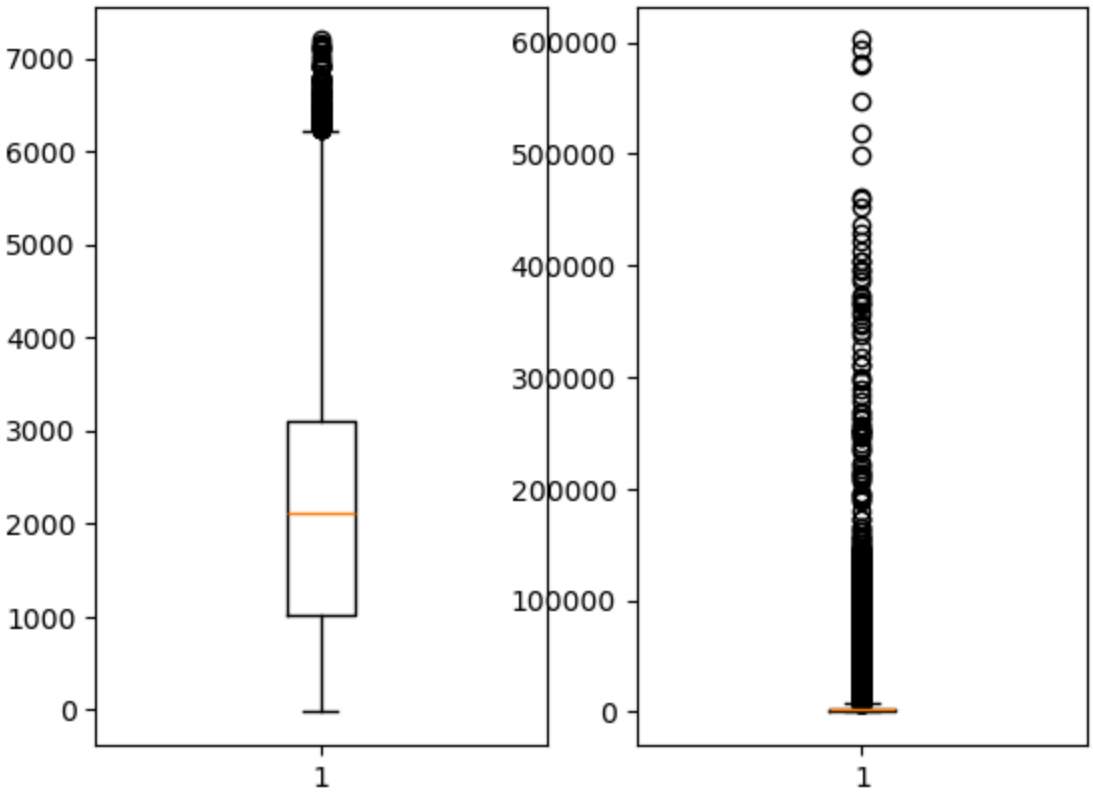
In [5]:
```
len(visa_df['no_of_employees'])
```

Out[5]:
```
25480
```

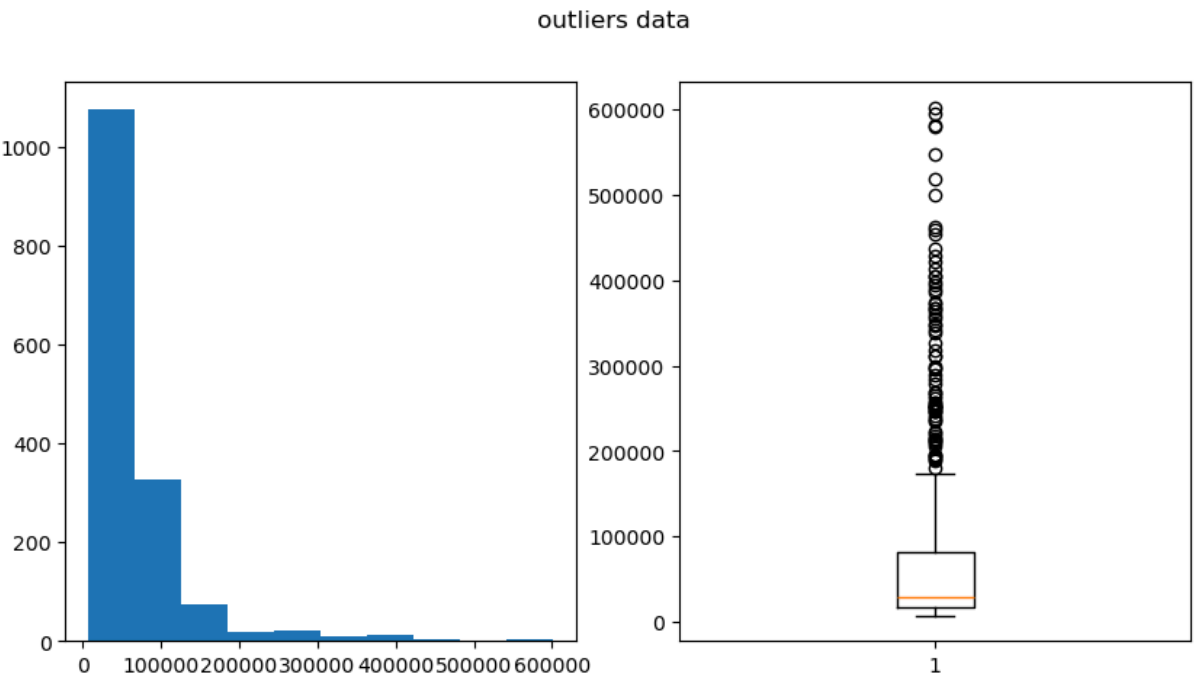
In [6]:
```
Q1=round(np.percentile(visa_df['no_of_employees'],25),2)
Q3=round(np.percentile(visa_df['no_of_employees'],75),2)
IQR=Q3-Q1
LB=Q1-1.5*IQR
UB=Q3+1.5*IQR
con1=visa_df['no_of_employees']<LB
con2=visa_df['no_of_employees']>UB
outliers_data=visa_df[con1 | con2]
len(outliers_data)
```

Out[6]:
```
1556
```

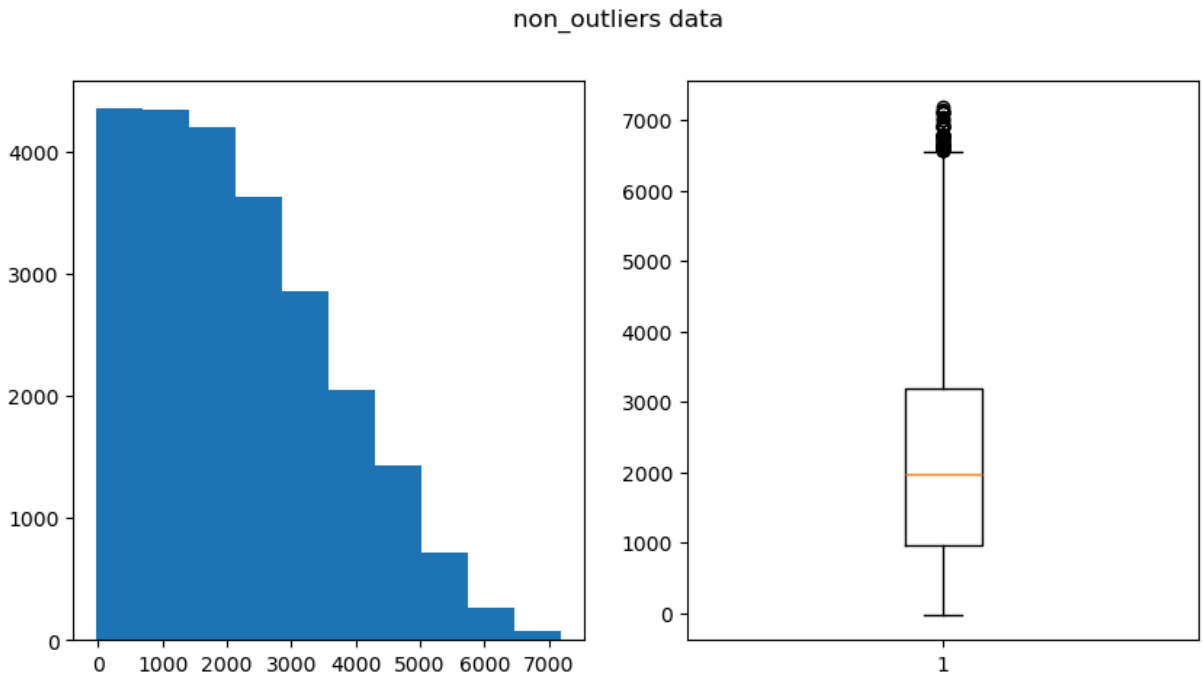
In [8]:
```
l1=[]
median = visa_df['no_of_employees'].median()
for value in visa_df['no_of_employees'].values:
 if value < LB or value > UB:
     l1.append(median)
 else:
     l1.append(value)
visa_df_copy=visa_df.copy()
visa_df_copy['no_of_employees']=l1
plt.subplot(1,2,1).boxplot(visa_df_copy['no_of_employees'])
plt.subplot(1,2,2).boxplot(visa_df['no_of_employees'])
plt.show()
```



In [9]:
```
con1=visa_df['no_of_employees']<LB
con2=visa_df['no_of_employees']>UB
outliers_data=visa_df[con1 | con2]
len(outliers_data)
plt.figure(figsize=(10,5))
plt.suptitle(' outliers data')
plt.subplot(1,2,1).hist(outliers_data['no_of_employees'])
plt.subplot(1,2,2).boxplot(outliers_data['no_of_employees'])
plt.show()
```



In [10]:
```
con3=visa_df['no_of_employees']>LB
con4=visa_df['no_of_employees']<UB
non_outliers_data=visa_df[con3 & con4]
len(non_outliers_data)
plt.figure(figsize=(10,5))
plt.suptitle('non_outliers data')
plt.subplot(1,2,1).hist(non_outliers_data['no_of_employees'])
plt.subplot(1,2,2).boxplot(non_outliers_data['no_of_employees'])
plt.show()
```



In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: