In [1]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```
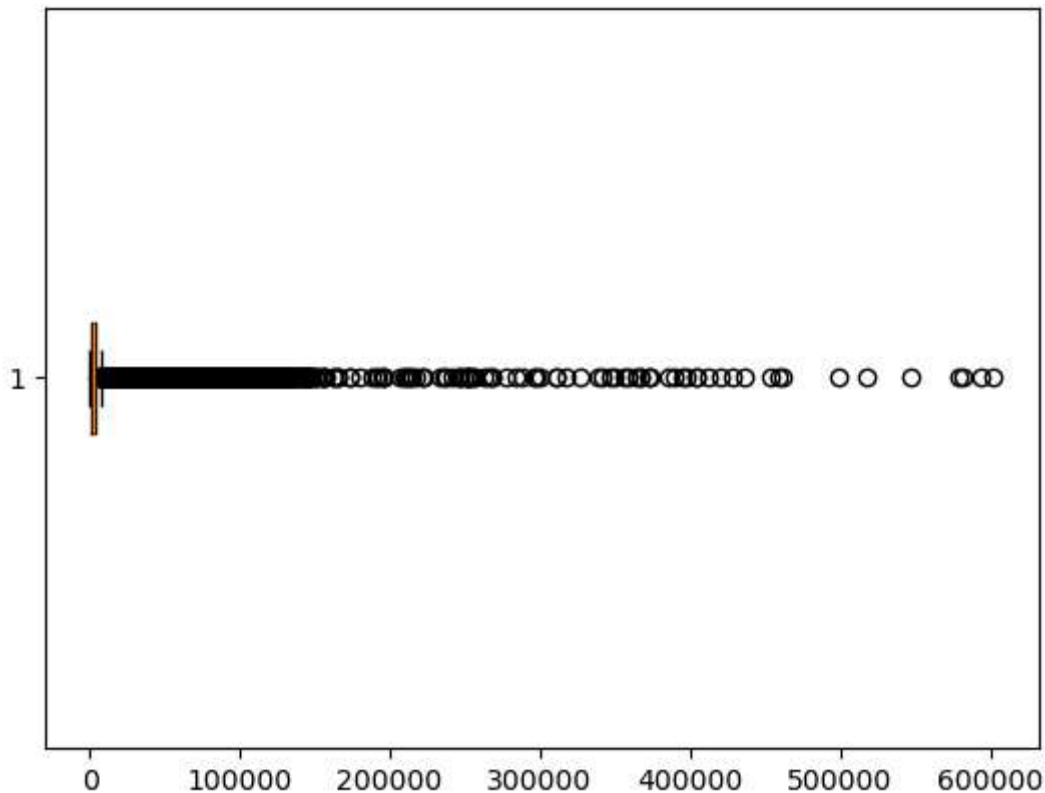
In [2]:
```python
visa_df=pd.read_csv(r"C:\Users\kolli\Documents\Naresh IT\data files\\Visadataset.cs
visa_df.head(2)
```
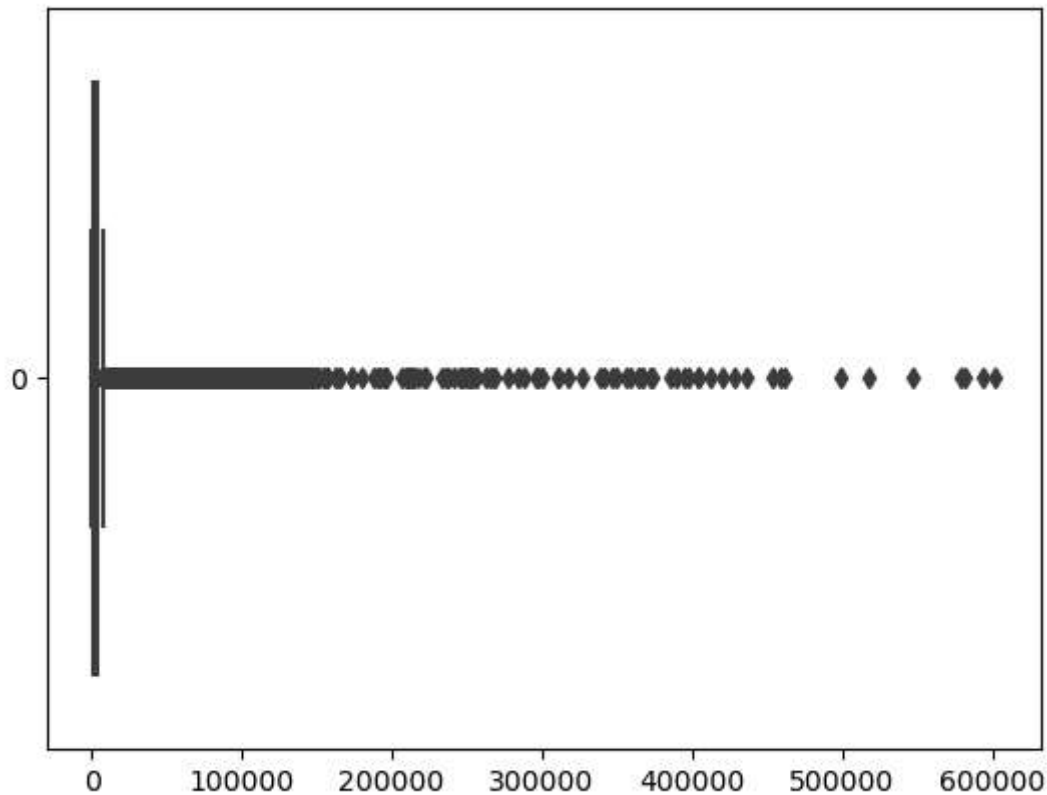
Out[2]:

|	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	

In [3]:
```python
plt.boxplot(visa_df['no_of_employees'],vert=False)
plt.show()
```


In [6]:
```python
sns.boxplot(visa_df['no_of_employees'],orient='h')
plt.show()
```
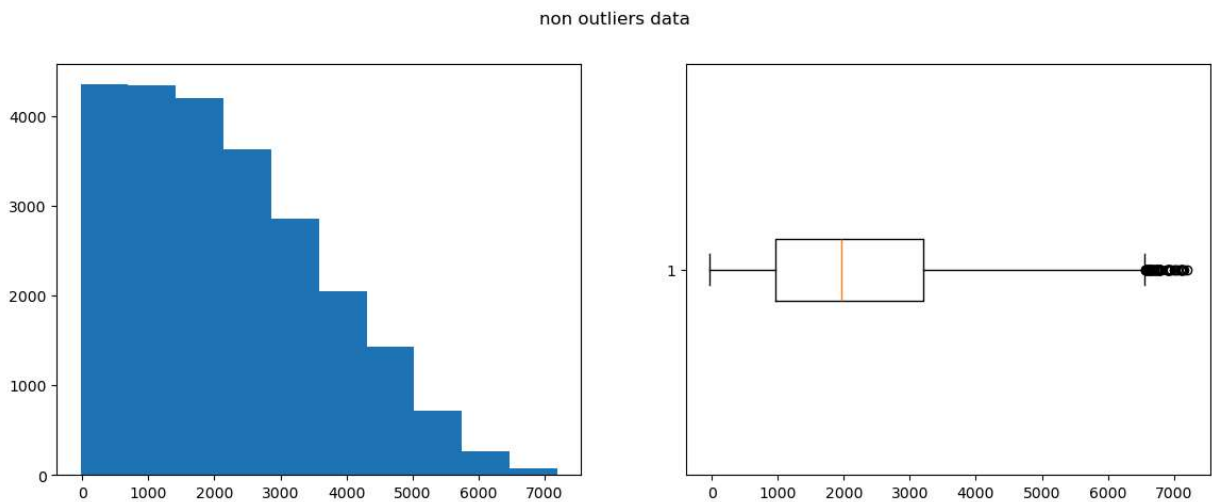
In [7]:
```python
plt.figure(figsize=(14,5))
plt.subplot(1,2,1).hist(visa_df['no_of_employees'])
plt.subplot(1,2,2).boxplot(visa_df['no_of_employees'],vert=False)
plt.show()
```
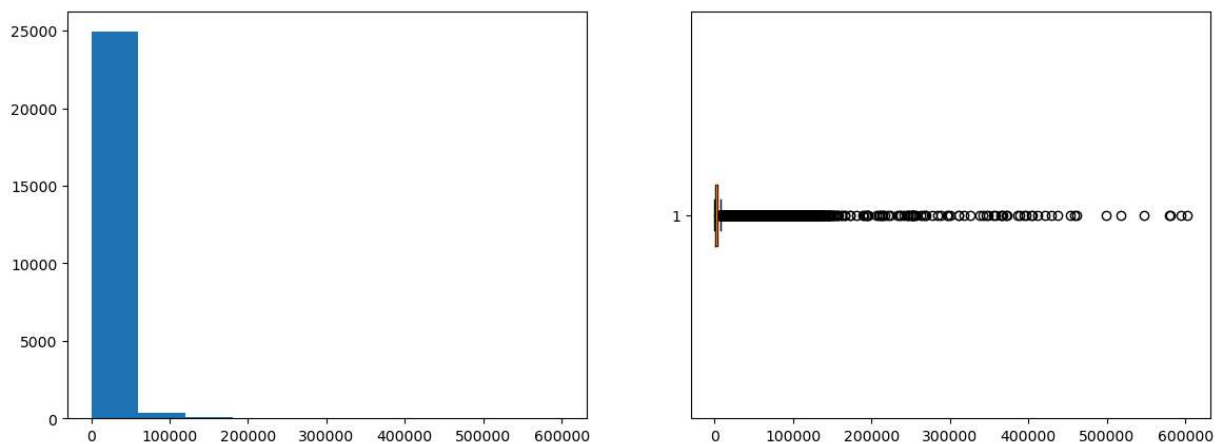

In [9]:
```python
Q1=np.percentile(visa_df['no_of_employees'],25)
Q3=np.percentile(visa_df['no_of_employees'],75)
IQR=Q3-Q1
LB=Q1-1.5*IQR
UB=Q3+1.5*IQR
con1=visa_df['no_of_employees']>LB
con2=visa_df['no_of_employees']<UB
non_outliers_data=visa_df[con1&con2]
len(non_outliers_data)
```
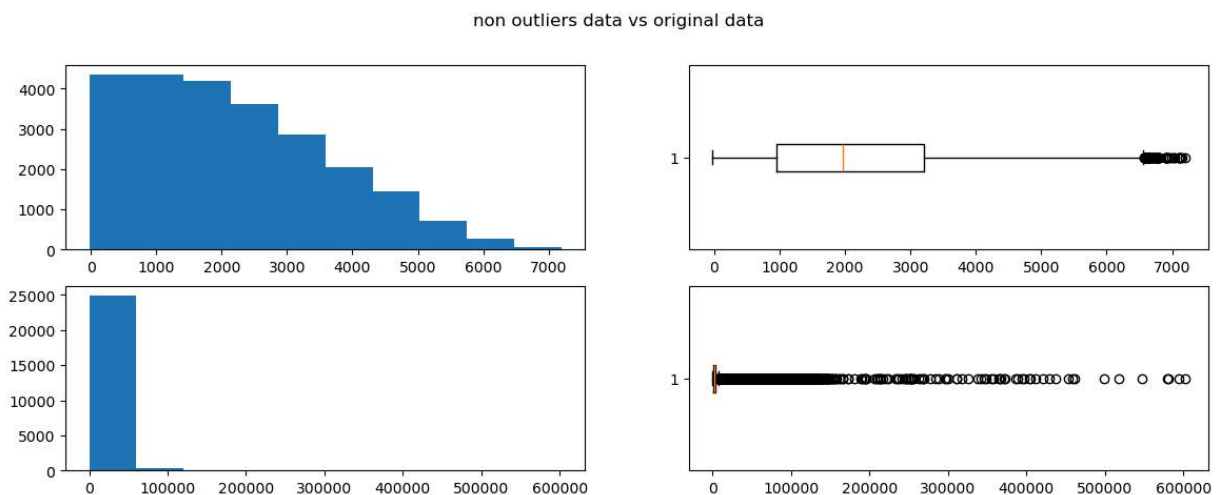
Out[9]: 23924

In [10]:
```python
plt.figure(figsize=(14,5))
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'],vert=False)
plt.show()
```


In [11]:
```python
plt.figure(figsize=(14,5))
plt.suptitle("non outliers data vs original data")
plt.subplot(2,2,1).hist(non_outliers_data['no_of_employees'])
plt.subplot(2,2,2).boxplot(non_outliers_data['no_of_employees'],vert=False)

plt.subplot(2,2,3).hist(visa_df['no_of_employees'])
plt.subplot(2,2,4).boxplot(visa_df['no_of_employees'],vert=False)
plt.show()
```


In [13]:
```python
plt.figure(figsize=(14,5))
plt.suptitle("non outliers data vs original data")
plt.subplot(2,2,1).hist(non_outliers_data['no_of_employees'])
plt.subplot(2,2,1).hist(visa_df['no_of_employees'])

plt.subplot(2,2,2).boxplot(non_outliers_data['no_of_employees'],vert=False)
plt.show()
```

non outliers data vs original data
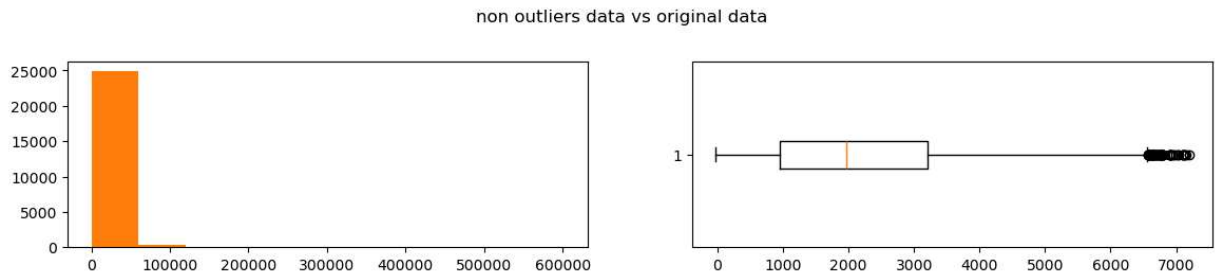
In [14]:
```python
import numpy as np
import matplotlib.pyplot as plt


# Conditions for outliers
con1 = visa_df['no_of_employees'] < LB
con2 = visa_df['no_of_employees'] > UB
con = con1 | con2

# Calculate the median
median = visa_df['no_of_employees'].median()

# Replace outliers with the median
new_data = np.where(con, median, visa_df['no_of_employees'])

# Ensure visa_df_copy is a copy of visa_df
visa_df_copy = visa_df.copy()
visa_df_copy['no_of_employees'] = new_data

# Plot the boxplots
plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)
plt.boxplot(visa_df['no_of_employees'])
plt.title('Original Data')

plt.subplot(1, 2, 2)
plt.boxplot(new_data)
plt.title('Data with Outliers Replaced by Median')

plt.show()
```
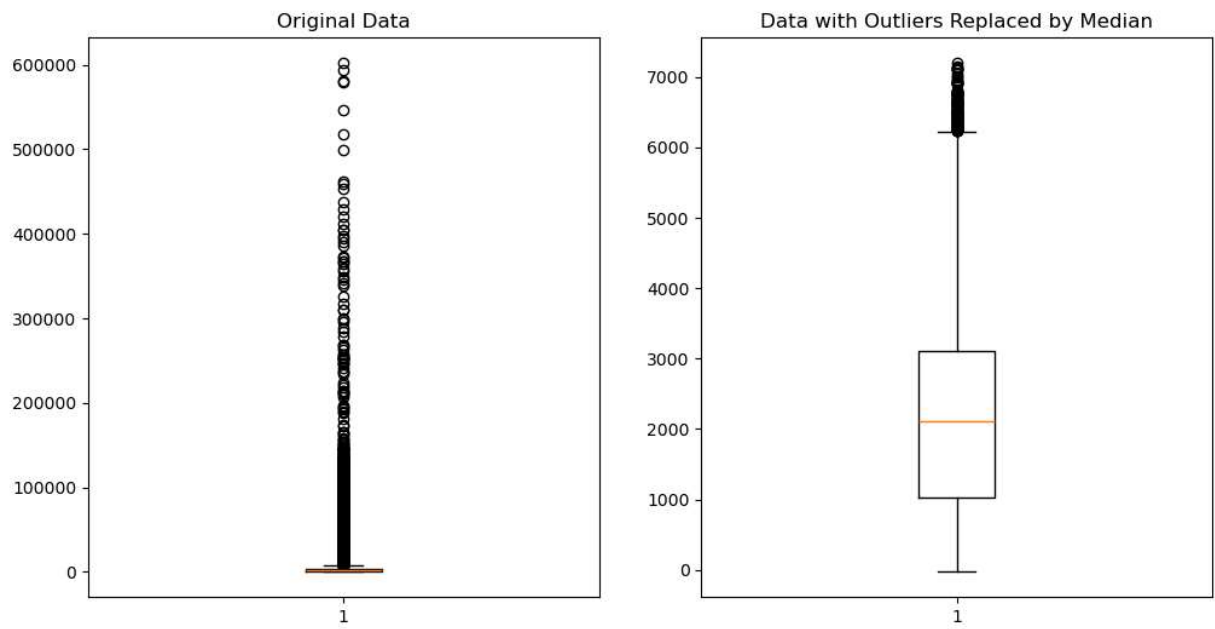
In [ ]: