Experiment-3

3.Consider titanic dataset and plot different graphs to visualize,Trends,Proportions,Categorical data,Uncertainties & distribution

Matplotlib

What is matplotlib?

Matplotlib is a low level graph plotting library in python that serves as a visualization utility.

Matplotlib was created by John D. Hunter.

Installation of Matplotlib

!pip install matplotlib

1.Consider titanic dataset and plot different graphs to visualize,Trends,Proportions,Categorical data,Uncertainties & distribution

A)Read the Titanic Dataset

In [1]:
```
import matplotlib.pyplot as plt
import pandas as pd
df=pd.read_csv('C:/Users/Guru Kiran/All CSV files/titanic.csv')
df
```

Out[1]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who
0	0	3	male	22.0	1	0	7.2500	S	Third	man
1	1	1	female	38.0	1	0	71.2833	C	First	woman
2	1	3	female	26.0	0	0	7.9250	S	Third	woman
3	1	1	female	35.0	1	0	53.1000	S	First	woman
4	0	3	male	35.0	0	0	8.0500	S	Third	man
...	...	...	...	...	...	...	...	...	...	...
886	0	2	male	27.0	0	0	13.0000	S	Second	man
887	1	1	female	19.0	0	0	30.0000	S	First	woman
888	0	3	female	NaN	1	2	23.4500	S	Third	woman
889	1	1	male	26.0	0	0	30.0000	C	First	man
890	0	3	male	32.0	0	0	7.7500	Q	Third	man

891 rows × 15 columns

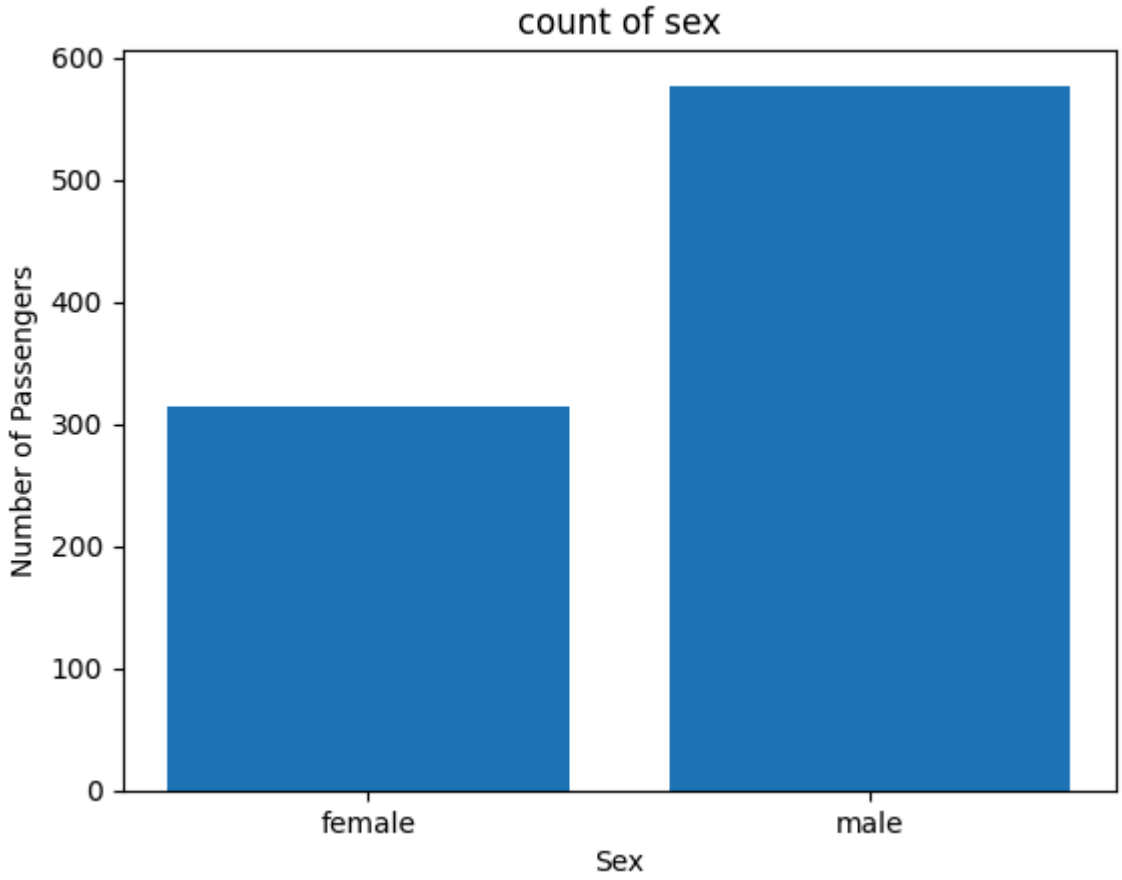
Examples:

1.Bar graph

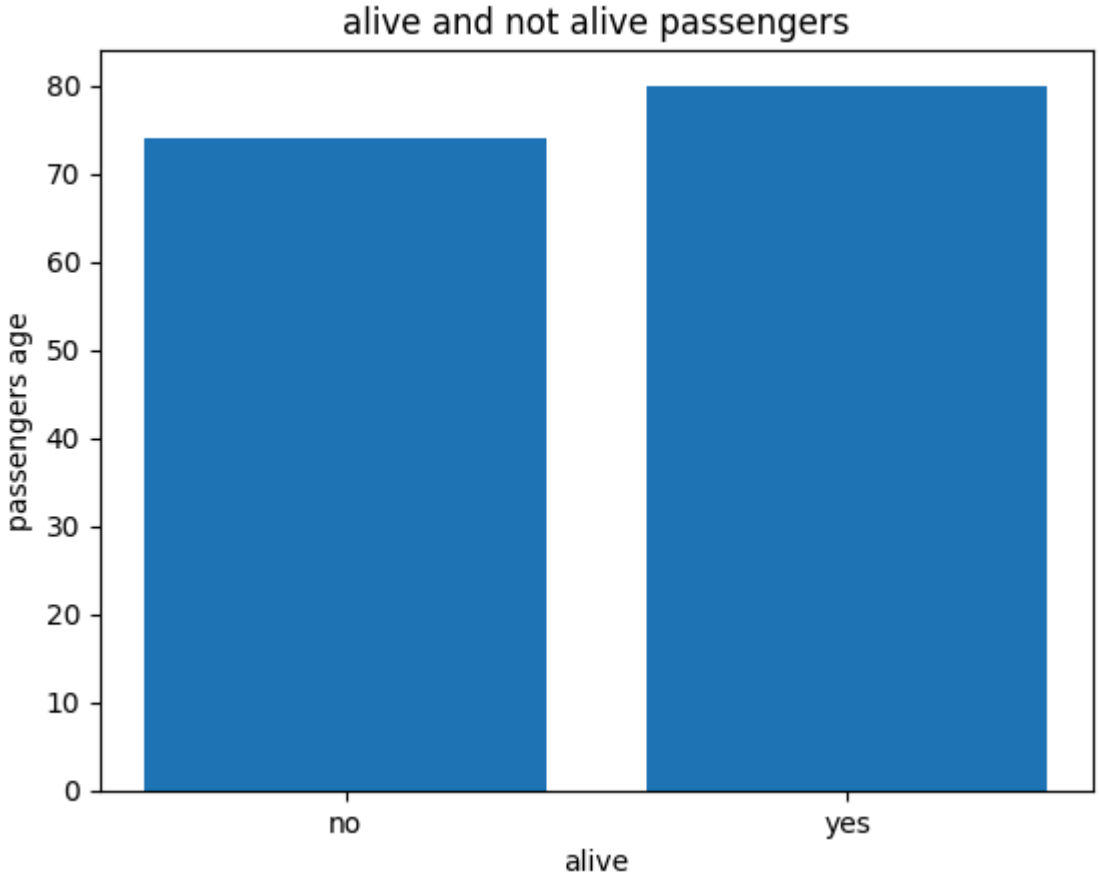
A bar graph is a visual representation of data using rectangular bars to show the values of different categories. The length or height of each bar corresponds to the data value it represents.

B)plot the graph for Categarical Data

In [2]:
```
#bor graph
a= df['sex'].value_counts().sort_index()
plt.bar(a.index, a.values)
plt.xlabel('Sex')
plt.ylabel('Number of Passengers')
plt.title('count of sex')
plt.show()
```



In [3]:
```
#bar graph
plt.bar(df['alive'],df['age'])
plt.title("alive and not alive passengers")
plt.xlabel("alive")
plt.ylabel("passengers age")
plt.show()
```

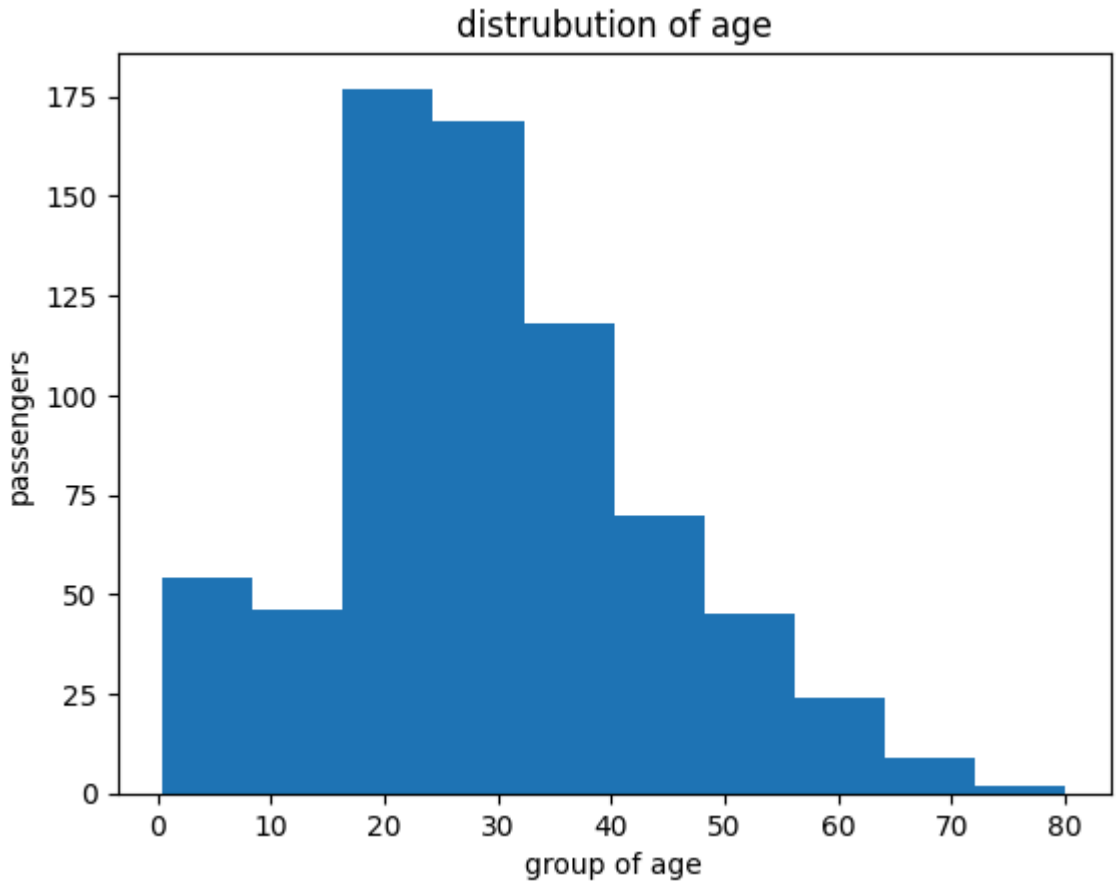


2.Histogram

A histogram is a graphical representation that shows the distribution of a numerical dataset using bars. Each bar represents the frequency of data points within a specific range (bin).

C)distribution of the age

In [4]:
```
plt.hist(df['age'],bins=10)
plt.title("distrubution of age")
plt.xlabel("group of age")
plt.ylabel("passengers")
plt.show()
```



3.line graph

A line graph is a chart that connects data points with a continuous line to show trends over time or ordered categories. It's useful for visualizing changes and patterns in data.

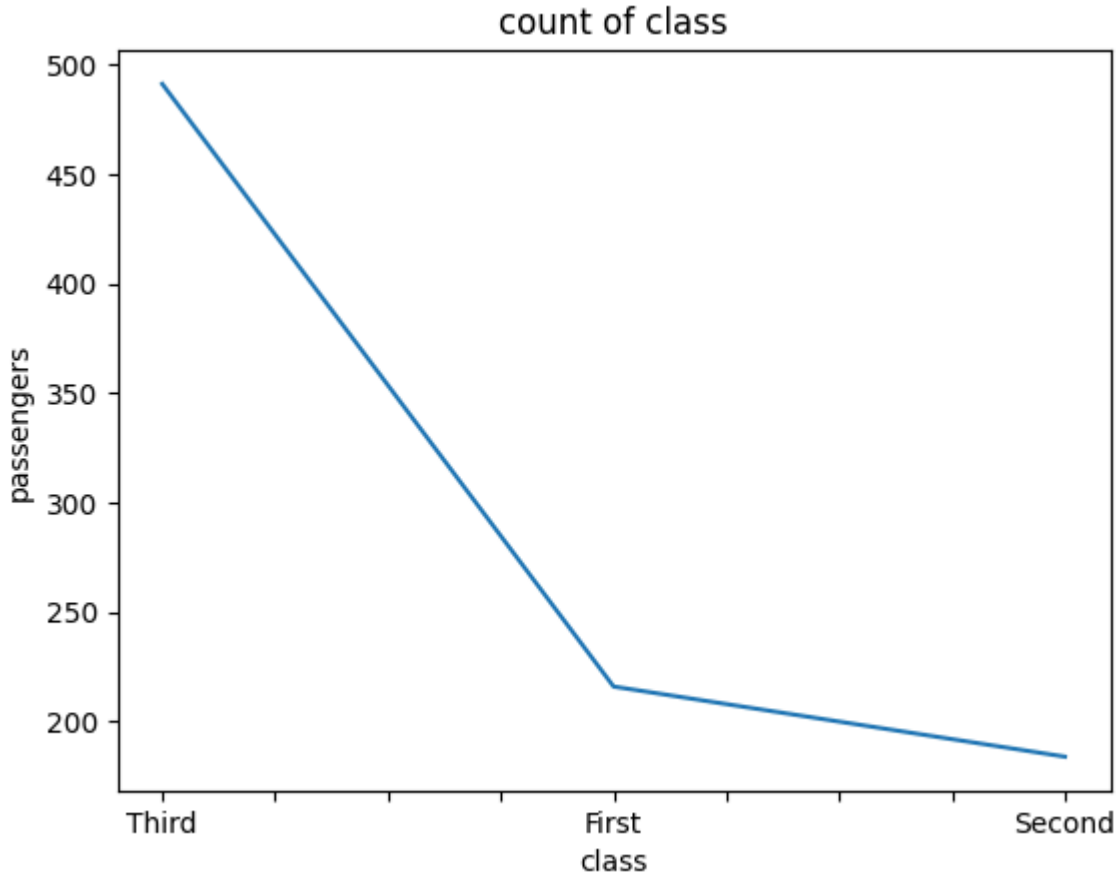
In [5]:
```
a=df['class'].value_counts()
a
```

Out[5]:
```
class
Third     491
First     216
Second    184
Name: count, dtype: int64
```

D)Plot the graph to show the Trends

In [6]:
```
a=df['class'].value_counts()
plt.title("count of class")
plt.xlabel("class")
plt.ylabel("passengers")
a.plot(kind='line')
```

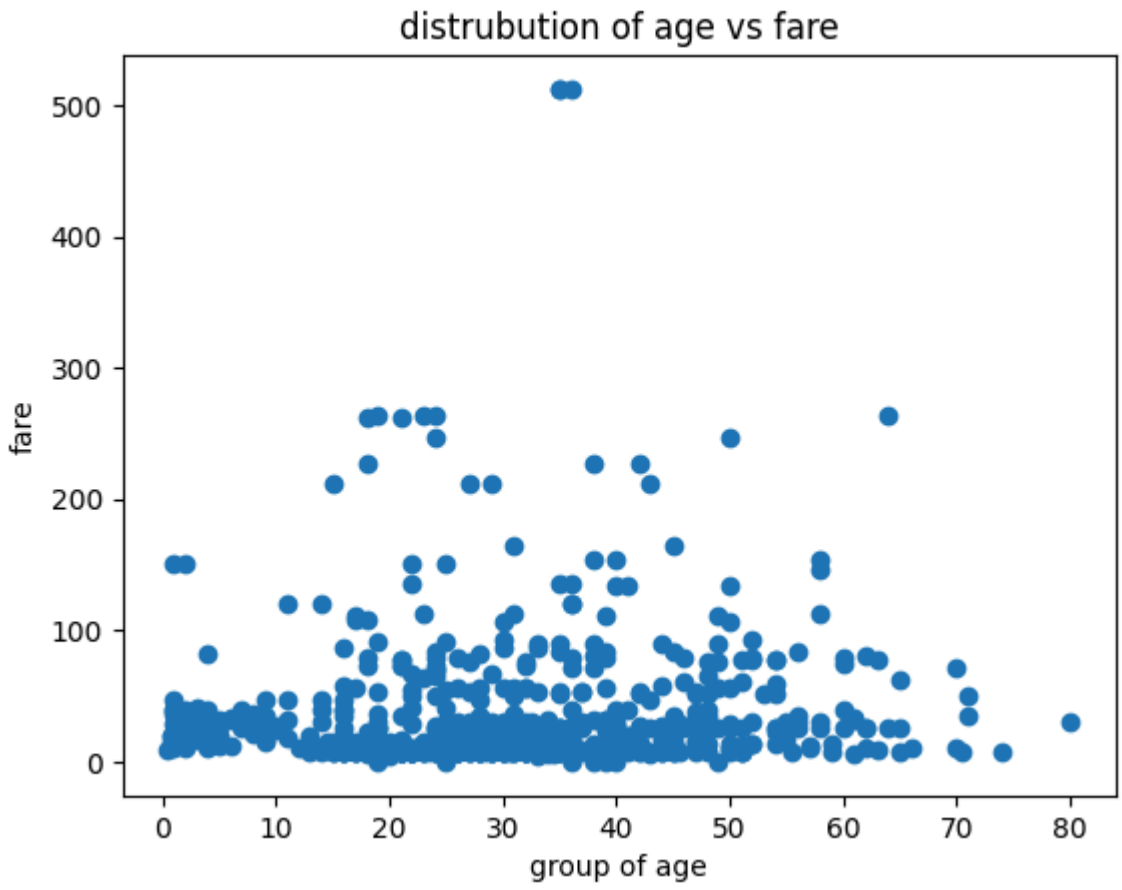
Out[6]:
```
<Axes: title={'center': 'count of class'}, xlabel='class', ylabel='passengers'>
```



4.Scatter plot

A scatter plot is a graph that displays points to show the relationship between two different variables. Each point represents an observation with values on the x- and y-axes.

In [7]:
```
plt.scatter(df['age'],df['fare'])
plt.title("distrubution of age vs fare")
plt.xlabel("group of age")
plt.ylabel("fare")
plt.show()
```

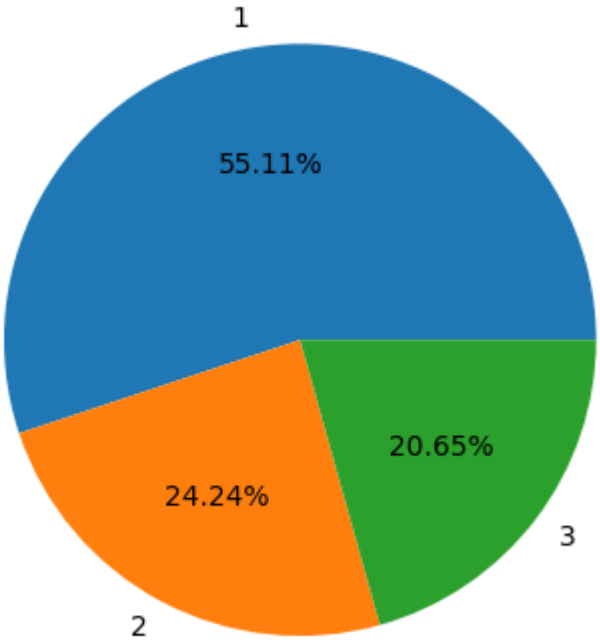


5.Piechart

A pie chart is a circular graph divided into slices to illustrate numerical proportions. Each slice represents a category's contribution to the whole.

E)Plot a pie chart showing the proportion of class column

In [8]:
```
plt.pie(df['class'].value_counts(),labels=[1,2,3],autopct='%1.2f%%')
plt.show()
```

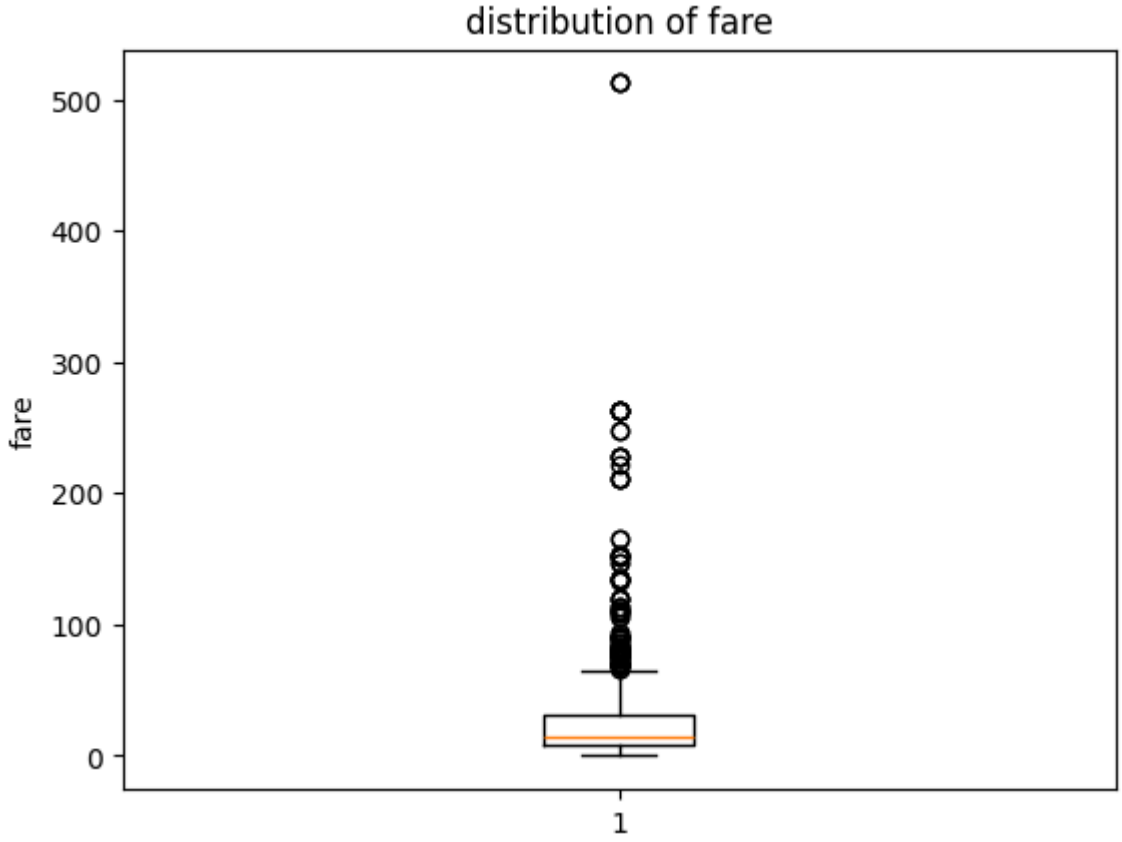


6.Boxplot

A boxplot is a graphical display that shows the distribution of a dataset through its quartiles. It highlights the median, spread, and potential outliers in the data.

F)Categorical Data,Plot the boxplot for Identify the outerliers of fare column

In [9]:
```
plt.boxplot(df['fare'])
plt.title("distribution of fare")
plt.ylabel("fare")
plt.show()
```


distribution of fare
fare
500
400
300
200
100
0
1