ASSIGNMENT 2

Name: Aakanksha Bhondve

Grno: 21810939

Rollno: 322004

Comp B1

Aim –

Perform the following operations using Python on the data sets.
Compute and display summary statistics for each feature available in the dataset. (eg. minimum value, maximum value, mean, range, standard deviation, variance and percentiles. Data Visualization-Create a histogram for each feature in the dataset to illustrate the feature distributions.

Theory –

Statistics –

1) Measures of Central Tendency

Mean : Sum of all entries divided by number of entries

Median: Middle value of sorted Data.

Mode: Entry with highest frequency

2) Measures of Dispersion

Standard Deviation : To know how close the entries are to the mean

Variance: Square of Standard Deviation.

3) Minimum and maximum value

Min: Least value among all the entries.

Max: Highest value among all the entries.

4) Percentile

A score compared to other scores in the same set.

Visualization – Visualization is used to represent the statistics using various types of Graphs like Histogram, Barchart, Scatter chart etc.

Code and Output –

In [3]:
```
import pandas as pd
import numpy as np
```

In [4]:
```
file = pd.read_csv('telecom.csv')
file.head()
```

Out[4]:

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Total night minutes	Total night calls	Total night charge	Total intl minutes	Total intl calls	Total intl charge	Cus s
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99	16.78	244.7	91	11.01	10.0	3	2.70	
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103	16.62	254.4	103	11.45	13.7	3	3.70	
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110	10.30	162.6	104	7.32	12.2	5	3.29	
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88	5.26	196.9	89	8.86	6.6	7	1.78	
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122	12.61	186.9	121	8.41	10.1	3	2.73	

In [5]:
```
length = file['Account length']
length.head()
```

Out[5]:
```
0    128
1    107
2    137
3     84
4     75
Name: Account length, dtype: int64
```

Maximum value

In [6]:
```
np.max(length)
```

Out[6]: 243

Minimum value

In [7]:
```
np.min(length)
```

Out[7]: 1

Mean value

In [8]:
```
np.mean(length)
```

Out[8]: 101.06480648064806

Standard deviation

In [9]:
```
np.std(length)
```

Out[9]: 39.81613156715945

Variance

In [10]:
```
length.var()
```

Out[10]:
```
1585.8001205882892
```

Data Visualization

Data Visualization is the presentation of data in graphical format. It helps people understand the significance of data by summarizing and presenting huge amount of data in a simple and easy-to-understand format and helps communicate information clearly and effectively.

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [3]:
```
file = pd.read_csv('telecom.csv')
file.head()
```

Out[3]:

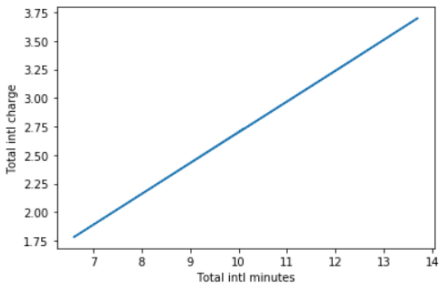
	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Total night minutes	Total night calls	Total night charge	Total intl minutes	Total intl calls	Total intl charge	Cus s
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99	16.78	244.7	91	11.01	10.0	3	2.70	
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103	16.62	254.4	103	11.45	13.7	3	3.70	
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110	10.30	162.6	104	7.32	12.2	5	3.29	
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88	5.26	196.9	89	8.86	6.6	7	1.78	
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122	12.61	186.9	121	8.41	10.1	3	2.73	

Line Plot

A line chart or line graph is a type of chart which displays information as a series of data points called 'markers' connected by straight line segments.

In [5]:
```
x = file['Total intl minutes'].head()
y = file['Total intl charge'].head()
plt.xlabel('Total intl minutes')
plt.ylabel('Total intl charge')
plt.plot(x,y)
```

Out[5]:
```
[<matplotlib.lines.Line2D at 0x1b4f0899308>]
```

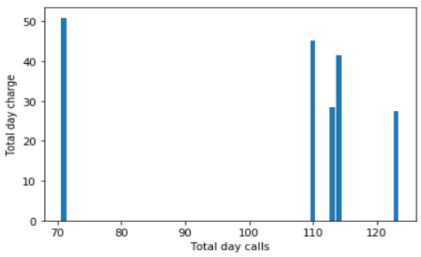


Bar Plot

A bar chart is used to show a comparison among different attributes, or it can show a comparison of items over time.

In [14]:
```
x = file['Total day calls'].head()
y = file['Total day charge'].head()
plt.xlabel('Total day calls')
plt.ylabel('Total day charge')
plt.bar(x,y)
```

Out[14]:
```
<BarContainer object of 5 artists>
```

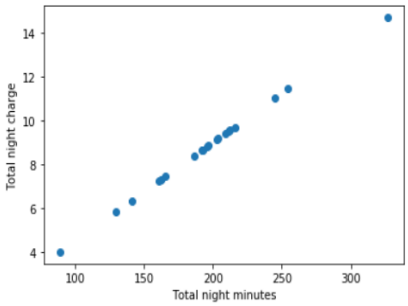


Scatter Plot

A scatter chart shows the relationship between two different variables and it can reveal the distribution trends. It should be used when there are many different data points, and you want to highlight similarities in the data set. This is useful when looking for outliers and for understanding the distribution of your data.

In [15]:
```
x = file['Total night minutes'].head(20)
y = file['Total night charge'].head(20)
plt.xlabel('Total night minutes')
plt.ylabel('Total night charge')
plt.scatter(x,y)
```

Out[15]:
```
<matplotlib.collections.PathCollection at 0x210cba84848>
```

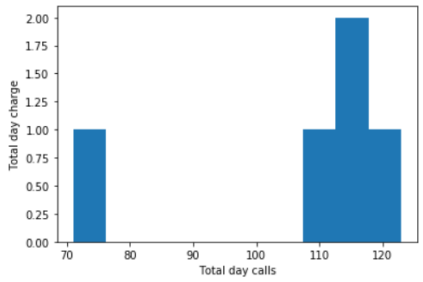


Histogram

The histogram represents the frequency of occurrence of specific phenomena which lie within a specific range of values and arranged in consecutive and fixed intervals.

In [16]:
```
x = file['Total day calls'].head()
y = file['Total day charge'].head()
plt.xlabel('Total day calls')
plt.ylabel('Total day charge')
plt.hist(x)
```

Out[16]:
```
(array([1., 0., 0., 0., 0., 0., 0., 1., 2., 1.]),
 array([ 71. ,  76.2,  81.4,  86.6,  91.8,  97. , 102.2, 107.4, 112.6,
        117.8, 123. ]),
 <a list of 10 Patch objects>)
```



Pie Chart

A pie chart shows a static number and how categories represent part of a whole the composition of something. A pie chart represents numbers in percentages, and the total sum of all segments needs to equal 100%.

In [17]:
```
x = file['Total day calls'].head()
y = file['Total day charge'].head()
plt.xlabel('Total day calls')
plt.ylabel('Total day charge')
plt.pie(x)
```

Out[17]:
```
([<matplotlib.patches.Wedge at 0x210cbb7b648>,
  <matplotlib.patches.Wedge at 0x210cbb7bb88>,
  <matplotlib.patches.Wedge at 0x210cbb83548>,
  <matplotlib.patches.Wedge at 0x210cbb83648>,
  <matplotlib.patches.Wedge at 0x210cbb88ec8>],
 [Text(0.8751588285671863, 0.6664060509786135, ''),
  Text(-0.4868823658088613, 0.9863800291289186, ''),
  Text(-1.054099023170609, -0.31444435016512545, ''),
  Text(-0.20381753045456996, -1.080952549504094, ''),
  Text(0.8631935549684825, -0.6818334742889011, '')])
```

Total day calls