Data Science Concepts:

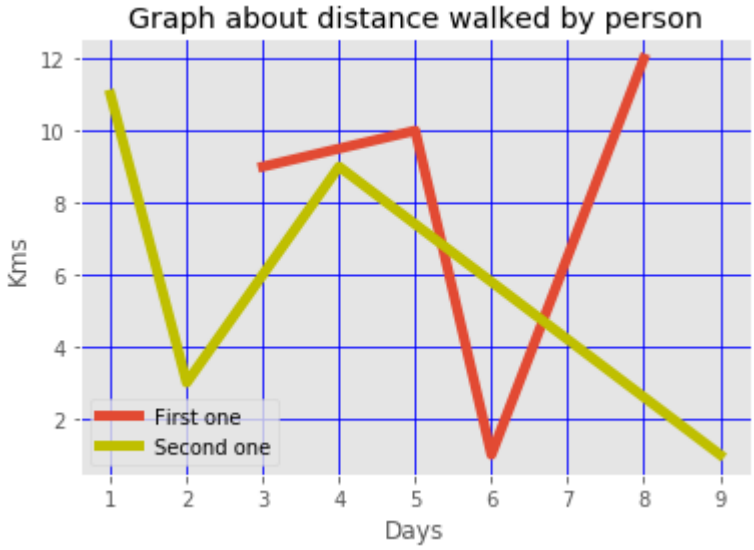
- Numpy
- Matplotlib
- Pandas

#Matplotlib: import matplotlib

1. Bar graph
2. Histogram
3. Scatter graph
4. Area graph
5. Pie chart

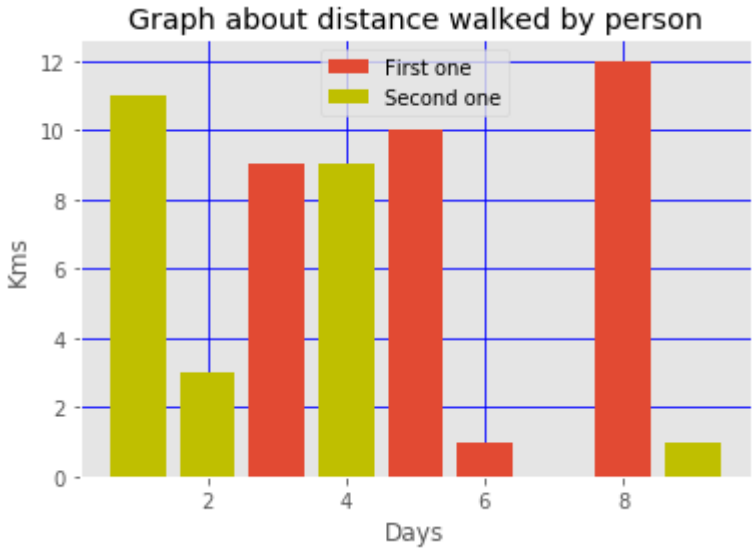
In [14]:

```
from matplotlib import pyplot as plt
from matplotlib import style
style.use("ggplot")
x=[3,5,6,8]
y=[9,10,1,12]
x1=[1,2,4,9]
y1=[11,3,9,1]
plt.title("Graph about distance walked by person")
plt.xlabel("Days")
plt.ylabel("Kms")
plt.plot(x,y,label="First one",linewidth=5)
plt.plot(x1,y1,color="y",label="Second one",linewidth=5)
plt.legend()
plt.grid(True,color="b")
plt.show()
```



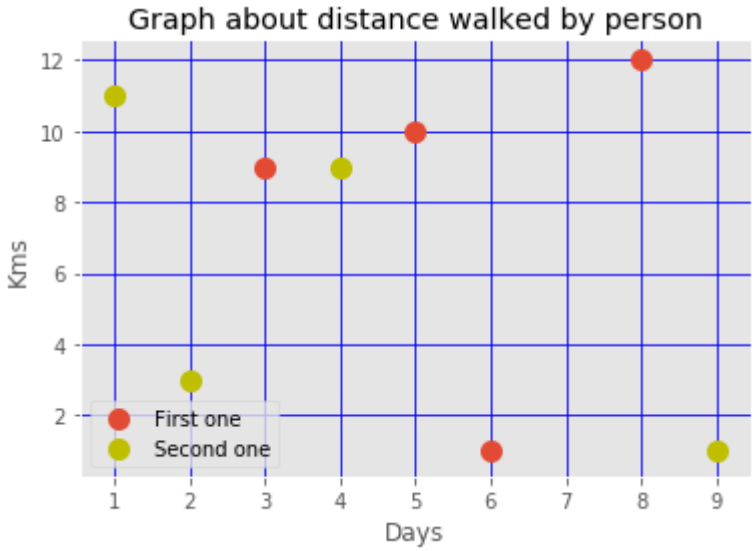
In [15]:

```
from matplotlib import pyplot as plt
from matplotlib import style
style.use("ggplot")
x=[3,5,6,8]
y=[9,10,1,12]
x1=[1,2,4,9]
y1=[11,3,9,1]
plt.title("Graph about distance walked by person")
plt.xlabel("Days")
plt.ylabel("Kms")
plt.bar(x,y,label="First one",linewidth=5)
plt.bar(x1,y1,color="y",label="Second one",linewidth=5)
plt.legend()
plt.grid(True,color="b")
plt.show()
```



In [16]:

```
from matplotlib import pyplot as plt
from matplotlib import style
style.use("ggplot")
x=[3,5,6,8]
y=[9,10,1,12]
x1=[1,2,4,9]
y1=[11,3,9,1]
plt.title("Graph about distance walked by person")
plt.xlabel("Days")
plt.ylabel("Kms")
plt.scatter(x,y,label="First one",linewidth=5)
plt.scatter(x1,y1,color="y",label="Second one",linewidth=5)
plt.legend()
plt.grid(True,color="b")
plt.show()
```



Pandas

- It is an open source library that produces high performance data manipulation and preparation.
- Using pandas we can accomplish five typical steps in processing and analysis of data regardless of original data.
  - Data loading
  - Prepare
  - Manipulate
  - Model
  - Analyse

Pandas data structures

- Pandas series.
- Pandas data frames.

Pandas Series

- It is 1-dimensional array.
- These are mostly used in analysis.
- There are various way of creating data frames.They are
  - Using dictionaries
  - Using numpy arrays

- Using .csv files
- Using .txt files

In [1]:

```
import numpy
import pandas as pd
pd.__version__
```

Out[1]:

```
'0.25.1'
```

In [21]:

```
#Creation of series using pandas
a=pd.Series ([1,2,3,4])
a
```

Out[21]:

```
0    1
1    2
2    3
3    4
dtype: int64
```

In [23]:

```
a=pd.Series((1,2,3,4))
a
```

Out[23]:

```
0    1
1    2
2    3
3    4
dtype: int64
```

In [24]:

```
b=a=pd.Series([1,2,3,4],index=['a','b','c','d'])
b
```

Out[24]:

```
a    1
b    2
c    3
d    4
dtype: int64
```

In [26]:

```
print(b[0],b[2],b[1],b[3])
print(b[:])
print(b['a'])
print(b['a':'c'])
```

```
1 3 2 4
a    1
b    2
c    3
d    4
dtype: int64
1
a    1
b    2
c    3
dtype: int64
```

In [27]:

```
s={'name':'a','marks':93,'CGPA':9}
s
```

Out[27]:

```
{'name': 'a', 'marks': 93, 'CGPA': 9}
```

In [28]:

```
pd.Series(s)
```

Out[28]:

```
name      a
marks    93
CGPA      9
dtype: object
```

In [30]:

```
pd.date_range('20-12-2019','30-12-2019')
```

Out[30]:

```
DatetimeIndex(['2019-12-20', '2019-12-21', '2019-12-22', '2019-12-23',
               '2019-12-24', '2019-12-25', '2019-12-26', '2019-12-27',
               '2019-12-28', '2019-12-29', '2019-12-30'],
              dtype='datetime64[ns]', freq='D')
```

In [31]:

```
pd.date_range('20-12-2019','30-12-2019',periods=4)
```

Out[31]:

```
DatetimeIndex(['2019-12-20 00:00:00', '2019-12-23 08:00:00',
               '2019-12-26 16:00:00', '2019-12-30 00:00:00'],
              dtype='datetime64[ns]', freq=None)
```

In [2]:

```
t=pd.Series([1,2,3,4],date)
t
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-2-fc8de526686c> in <module>
----> 1 t=pd.Series([1,2,3,4],date)
      2 t

NameError: name 'date' is not defined
```

In [ ]:

```

```