

#GE ASSIGNMET 5

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
data = {'EMP ID': [1,2,3,4,5,6,7,8,9,10],
        'EMP NAME': ['satish','reeya','jay','rahul','roy','jay','vishal','serah','vishal','Prachi'],
        'Age': [21,23,40,35,26,28,29,21,29,22],
        'Salary': [50000,75000,100000,np.nan,45000,100000,np.nan,55000,np.nan,60000],
        'EMP Credits': [3.8,4,5,4.2,3.9,4.5,5,3.7,5,4.3],
        'Joining date': ["01-11-2017",np.nan,"22-09-2015","11-10-2016","08-01-2017","22-1-2018",'22-09-2015', '05-01-2016', '0
df = pd.DataFrame(data)
print(df)
print('-----')
df['Joining date'] = pd.to_datetime(df['Joining date'])
# Add new column with month names
df['month_name'] = df['Joining date'].dt.strftime('%d %B %Y')
print(df)

print('before filling\n',df)
print()
df["Salary"].fillna(50000,inplace=True)
df["Joining date"].fillna("01-01-2018",inplace=True)
print('after filling\n',df)
print(".....")
print('before duplicate\n',df)
print()
print(df.duplicated())
print(df.drop_duplicates())
print(".....")

max_credit = df['EMP Credits'].max()
max_credit_emp = df.loc[df['EMP Credits'] == max_credit, 'EMP NAME'].values.tolist()
print('\nthe max credit is',max_credit)
print('\nnames of max credit is',max_credit_emp)
```

```

max_credit = df['EMP Credits'].max()
max_credit_emp = df.loc[df['EMP Credits'] == max_credit, 'EMP NAME'].values.tolist()
print('\nthe max credit is',max_credit)
print('\nnames of max credit is',max_credit_emp)
print(".....")

df[['EMP Credits', 'Salary']].plot(kind='bar', x='EMP Credits', y='Salary', rot=0)
plt.title('Employee Credits vs Salary')
plt.xlabel('Employee Credits')
plt.ylabel('Salary')
plt.show()
df['EMP Credits'].plot(kind='pie')
plt.title('Employee Credits')
plt.show()
print(".....")
plt.figure(figsize=(8,4))
plt.subplot(1,2,1) # PLOT 1: Line Plot
df['Age'].plot.line(color='b', marker='X', x='EMP Credits',
    #xlabel = 'EMP Credits',
    #ylabel = 'Age',
    title='EMP Credits vs Age')
plt.show()
plt.subplot(1,2,2) # PLOT 2: Line Plot
df['Salary'].plot.line(color='b', marker='X', x='EMP Credits',
    #xlabel = 'EMP Credits',
    #ylabel = 'Salary',
    title='EMP Credits vs Salary')
plt.show()
print(".....")

#2..
import pandas as pd

```

	EMP ID	EMP NAME	Age	Salary	EMP Credits	Joining date
0	1	satish	21	50000.0	3.8	01-11-2017
1	2	reeya	23	75000.0	4.0	NaN
2	3	jay	40	100000.0	5.0	22-09-2015
3	4	rahul	35	NaN	4.2	11-10-2016
4	5	roy	26	45000.0	3.9	08-01-2017
5	6	jay	28	100000.0	4.5	22-1-2018
6	7	vishal	29	NaN	5.0	22-09-2015
7	8	serah	21	55000.0	3.7	05-01-2016
8	9	vishal	29	NaN	5.0	06-02-2018
9	10	Prachi	22	60000.0	4.3	05-01-2016

-----



	EMP ID	EMP NAME	Age	Salary	EMP Credits	Joining date	month_name
0	1	satish	21	50000.0	3.8	2017-01-11	11 January 2017
1	2	reeya	23	75000.0	4.0	NaT	NaN
2	3	jay	40	100000.0	5.0	2015-09-22	22 September 2015
3	4	rahul	35	NaN	4.2	2016-11-10	10 November 2016
4	5	roy	26	45000.0	3.9	2017-08-01	01 August 2017
5	6	jay	28	100000.0	4.5	2018-01-22	22 January 2018
6	7	vishal	29	NaN	5.0	2015-09-22	22 September 2015
7	8	serah	21	55000.0	3.7	2016-05-01	01 May 2016
8	9	vishal	29	NaN	5.0	2018-06-02	02 June 2018
9	10	Prachi	22	60000.0	4.3	2016-05-01	01 May 2016

before filling

	EMP ID	EMP NAME	Age	Salary	EMP Credits	Joining date	month_name
0	1	satish	21	50000.0	3.8	2017-01-11	11 January 2017
1	2	reeya	23	75000.0	4.0	NaT	NaN
2	3	jay	40	100000.0	5.0	2015-09-22	22 September 2015
3	4	rahul	35	NaN	4.2	2016-11-10	10 November 2016
4	5	roy	26	45000.0	3.9	2017-08-01	01 August 2017
5	6	jay	28	100000.0	4.5	2018-01-22	22 January 2018
6	7	vishal	29	NaN	5.0	2015-09-22	22 September 2015
7	8	serah	21	55000.0	3.7	2016-05-01	01 May 2016
8	9	vishal	29	NaN	5.0	2018-06-02	02 June 2018
9	10	Prachi	22	60000.0	4.3	2016-05-01	01 May 2016

after filling

before duplicate

	EMP ID	EMP NAME	Age	Salary	EMP Credits	Joining date	month_name
0	1	satish	21	50000.0	3.8	2017-01-11	11 January 2017
1	2	reeya	23	75000.0	4.0	2018-01-01	NaN
2	3	jay	40	100000.0	5.0	2015-09-22	22 September 2015
3	4	rahul	35	50000.0	4.2	2016-11-10	10 November 2016
4	5	roy	26	45000.0	3.9	2017-08-01	01 August 2017
5	6	jay	28	100000.0	4.5	2018-01-22	22 January 2018
6	7	vishal	29	50000.0	5.0	2015-09-22	22 September 2015
7	8	serah	21	55000.0	3.7	2016-05-01	01 May 2016
8	9	vishal	29	50000.0	5.0	2018-06-02	02 June 2018
9	10	Prachi	22	60000.0	4.3	2016-05-01	01 May 2016

0	False
1	False
2	False
3	False
4	False
5	False
6	False
7	False
8	False
9	False

..

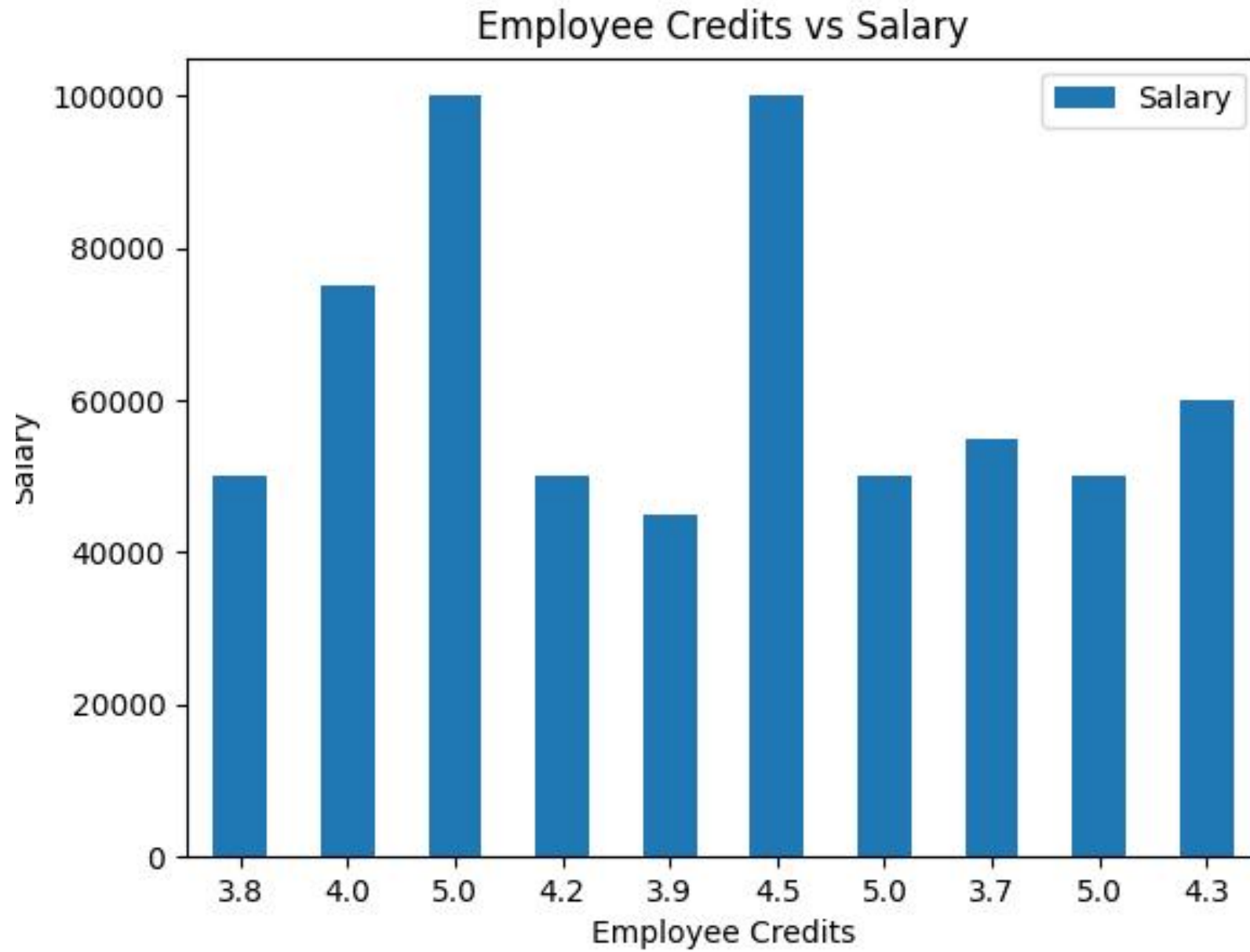


```
dtype: bool
```

	EMP ID	EMP NAME	Age	Salary	EMP Credits	Joining date	month_name
0	1	satish	21	50000.0	3.8	2017-01-11	11 January 2017
1	2	reeya	23	75000.0	4.0	2018-01-01	NaN
2	3	jay	40	100000.0	5.0	2015-09-22	22 September 2015
3	4	rahul	35	50000.0	4.2	2016-11-10	10 November 2016
4	5	roy	26	45000.0	3.9	2017-08-01	01 August 2017
5	6	jay	28	100000.0	4.5	2018-01-22	22 January 2018
6	7	vishal	29	50000.0	5.0	2015-09-22	22 September 2015
7	8	serah	21	55000.0	3.7	2016-05-01	01 May 2016
8	9	vishal	29	50000.0	5.0	2018-06-02	02 June 2018
9	10	Prachi	22	60000.0	4.3	2016-05-01	01 May 2016
.....							
.....							

```
the max credit is 5.0
```

```
names of max credit is ['jay', 'vishal', 'vishal']
```



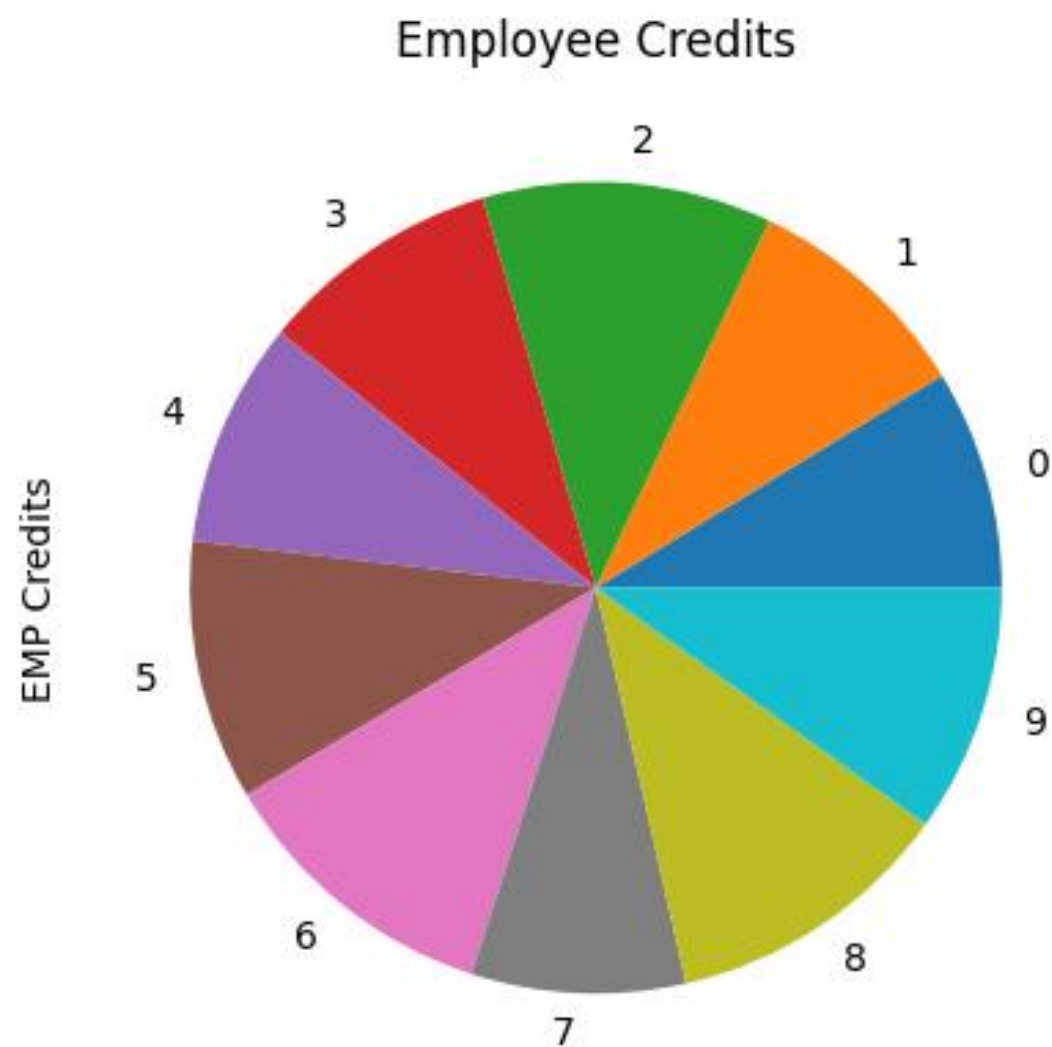
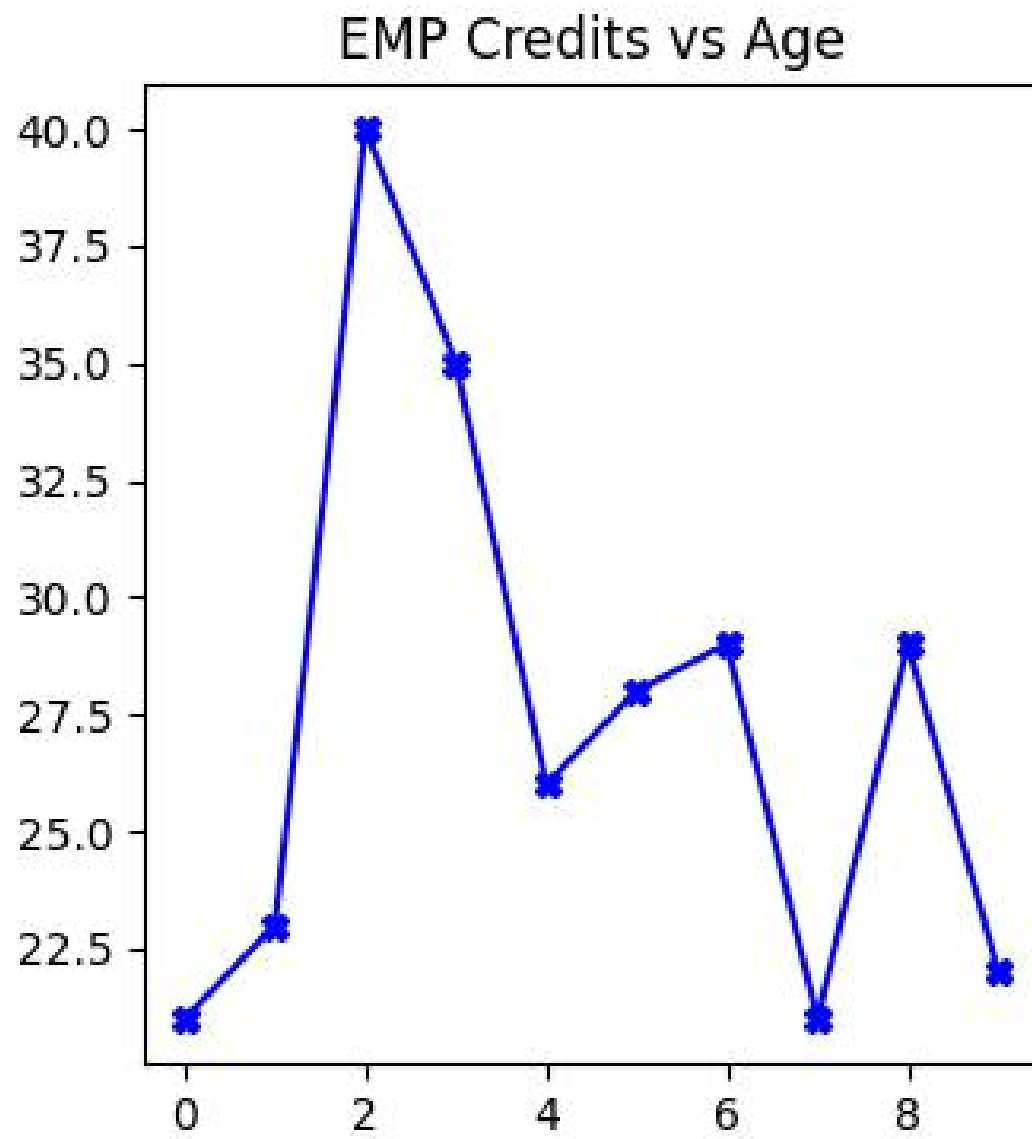
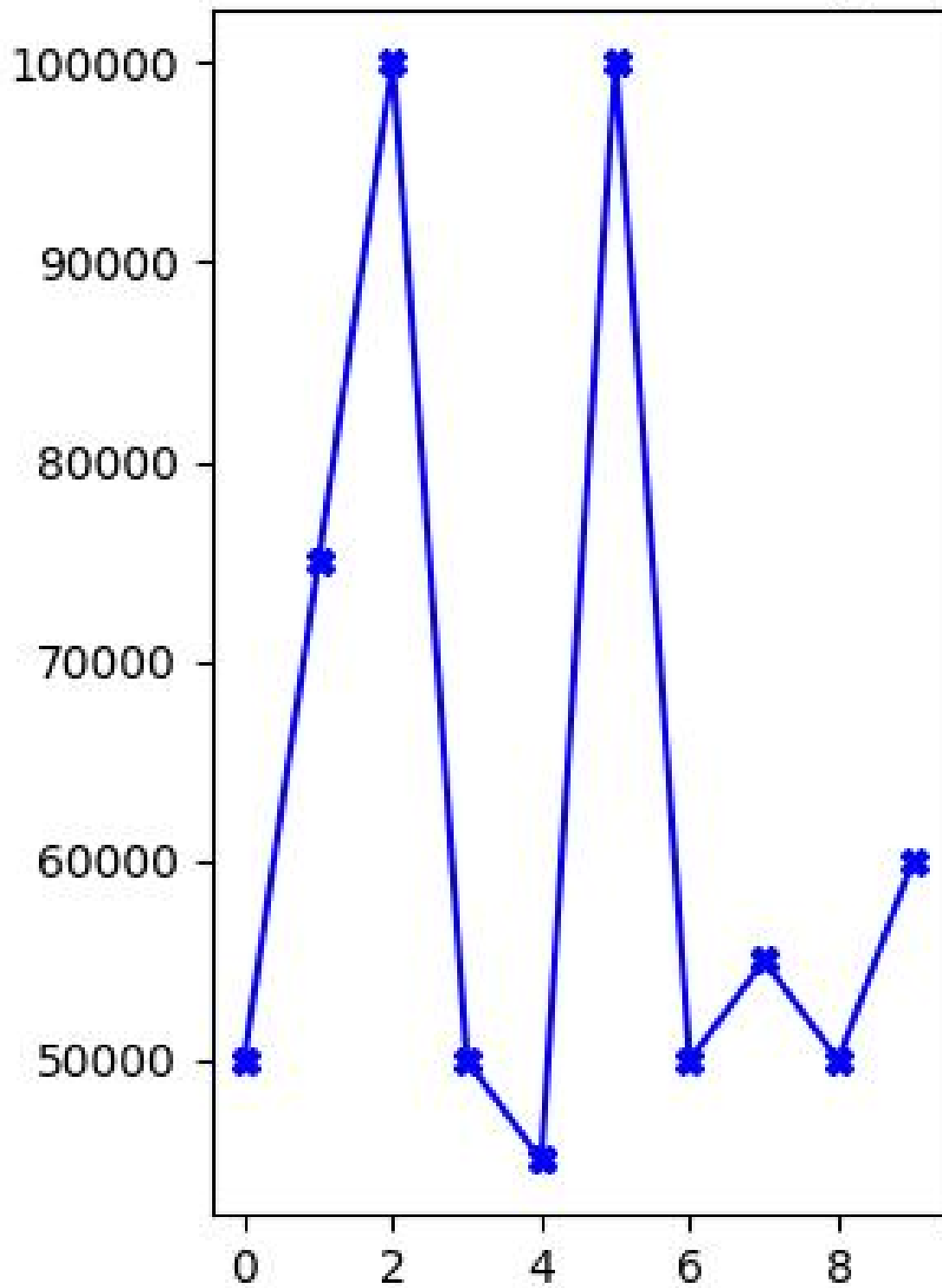




Figure 1



EMP Credits vs Salary



```
#2..
import pandas as pd
#Loading the Datas
CALORIES_DATA = pd.read_csv(r"C:\Users\Harsh\Downloads\CaloriesDataSet.csv")
print(CALORIES_DATA)
print(".....")

print('before filling\n', CALORIES_DATA)
print()
CALORIES_DATA["Calories"].fillna(300, inplace=True)
CALORIES_DATA["Date"].fillna("'2018/12/12'", inplace=True)
print('after filling\n', CALORIES_DATA)
print(".....")

print('before duplicate\n', CALORIES_DATA)
print()
print(CALORIES_DATA.duplicated())
print(CALORIES_DATA.drop_duplicates())
print(".....")

# calculate the average values for Duration, Pulse, Maxpulse, and Calories
avg_duration = CALORIES_DATA['Duration'].mean()
avg_pulse = CALORIES_DATA['Pulse'].mean()
avg_maxpulse = CALORIES_DATA['Maxpulse'].mean()
avg_calories = CALORIES_DATA['Calories'].mean()

# create a new row with the average values
avg_row = {'Duration': avg_duration, 'Pulse': avg_pulse, 'Maxpulse': avg_maxpulse, 'Calories': avg_calories}

# append the new row to the dataframe
CALORIES_DATA = CALORIES_DATA.append(avg_row, ignore_index=True)
```

```
avg_pulse = CALORIES_DATA['Pulse'].mean()
avg_maxpulse = CALORIES_DATA['Maxpulse'].mean()
avg_calories = CALORIES_DATA['Calories'].mean()

# create a new row with the average values
avg_row = {'Duration': avg_duration, 'Pulse': avg_pulse, 'Maxpulse': avg_maxpulse, 'Calories': avg_calories}

# append the new row to the dataframe
CALORIES_DATA = CALORIES_DATA.append(avg_row, ignore_index=True)

# write the updated dataframe to a new csv file
CALORIES_DATA.to_csv('updated_data.csv', index=False)
print(CALORIES_DATA)

print(".....")
CALORIES_DATA.plot(marker='X', ms=15, mec='r', mfc='r', color='r', linestyle='--', label='Makers', figsize=(5,5))
plt.xlabel('CALORIES')
plt.ylabel('Salary')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.area(color='b', label='CALORIES', figsize=(5,5))
plt.xlabel('CALORIES')
plt.ylabel('Date')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.bar(color='c', label='CALORIES', figsize=(3,3))
plt.xlabel('CALORIES')
plt.ylabel('Date')
plt.legend(loc='lower center')
plt.show()
```



```
print(".....")
CALORIES_DATA.plot(marker='X', ms=15, mec='r', mfc='r', color='r', linestyle='--', label='Makers', figsize=(5,5))
plt.xlabel('CALORIES')
plt.ylabel('Salary')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.area(color='b', label='CALORIES', figsize=(5,5))
plt.xlabel('CALORIES')
plt.ylabel('Date')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.bar(color='c', label='CALORIES', figsize=(3,3))
plt.xlabel('CALORIES')
plt.ylabel('Date')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.barh(color='c', label='CALORIES', figsize=(3,3))
plt.xlabel('CALORIES')
plt.ylabel('Date')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.box(color='m', label='CALORIES', figsize=(3,3))
plt.ylabel('Date')
plt.show()

CALORIES_DATA.plot.hist(bins=6, figsize=(3,3))
plt.ylabel('Date')
plt.show()

CALORIES_DATA.plot.pie(label='CALORIES', autopct='%1.1f%%')
```

```
plt.ylabel('Salary')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.area(color='b', label='CALORIES', figsize=(5, 5))
plt.xlabel('CALORIES')
plt.ylabel('Date')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.bar(color='c', label='CALORIES', figsize=(3, 3))
plt.xlabel('CALORIES')
plt.ylabel('Date')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.barh(color='c', label='CALORIES', figsize=(3, 3))
plt.xlabel('CALORIES')
plt.ylabel('Date')
plt.legend(loc='lower center')
plt.show()

CALORIES_DATA.plot.box(color='m', label='CALORIES', figsize=(3, 3))
plt.ylabel('Date')
plt.show()

CALORIES_DATA.plot.hist(bins=6, figsize=(3, 3))
plt.ylabel('Date')
plt.show()

CALORIES_DATA.plot.pie(label='CALORIES', autopct='%1.1f%%')
plt.ylabel('Date')
plt.show()
```

*IDLE Shell 3.11.0*					
File Edit Shell Debug Options Window Help					
	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.1
1	60	'2020/12/02'	117	145	479.0
2	60	'2020/12/03'	103	135	340.0
3	45	'2020/12/04'	109	175	282.4
4	45	'2020/12/05'	117	148	406.0
5	60	'2020/12/06'	102	127	300.0
6	60	'2020/12/07'	110	136	374.0
7	450	'2020/12/08'	104	134	253.3
8	30	'2020/12/09'	109	133	195.1
9	60	'2020/12/10'	98	124	269.0
10	60	'2020/12/11'	103	147	329.3
11	60	'2020/12/12'	100	120	250.7
13	60	'2020/12/13'	106	128	345.3
14	60	'2020/12/14'	104	132	379.3
15	60	'2020/12/15'	98	123	275.0
16	60	'2020/12/16'	98	120	215.2
17	60	'2020/12/17'	100	120	300.0
18	45	'2020/12/18'	90	112	300.0
19	60	'2020/12/19'	103	123	323.0
20	45	'2020/12/20'	97	125	243.0
21	60	'2020/12/21'	108	131	364.2
22	45	'2018/12/12'	100	119	282.0
23	60	'2020/12/23'	130	101	300.0
24	45	'2020/12/24'	105	132	246.0
25	60	'2020/12/25'	102	126	334.5
26	60	20201226	100	120	250.0
27	60	'2020/12/27'	92	118	241.0
28	60	'2020/12/28'	103	132	300.0
29	60	'2020/12/29'	100	132	280.0
30	60	'2020/12/30'	102	129	380.3
31	60	'2020/12/31'	92	115	243.0
.....					

*IDLE Shell 3.11.0*					
File Edit Shell Debug Options Window Help					
	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.1
1	60	'2020/12/02'	117	145	479.0
2	60	'2020/12/03'	103	135	340.0
3	45	'2020/12/04'	109	175	282.4
4	45	'2020/12/05'	117	148	406.0
5	60	'2020/12/06'	102	127	300.0
6	60	'2020/12/07'	110	136	374.0
7	450	'2020/12/08'	104	134	253.3
8	30	'2020/12/09'	109	133	195.1
9	60	'2020/12/10'	98	124	269.0
10	60	'2020/12/11'	103	147	329.3
11	60	'2020/12/12'	100	120	250.7
13	60	'2020/12/13'	106	128	345.3
14	60	'2020/12/14'	104	132	379.3
15	60	'2020/12/15'	98	123	275.0
16	60	'2020/12/16'	98	120	215.2
17	60	'2020/12/17'	100	120	300.0
18	45	'2020/12/18'	90	112	300.0
19	60	'2020/12/19'	103	123	323.0
20	45	'2020/12/20'	97	125	243.0
21	60	'2020/12/21'	108	131	364.2
22	45	'2018/12/12'	100	119	282.0
23	60	'2020/12/23'	130	101	300.0
24	45	'2020/12/24'	105	132	246.0
25	60	'2020/12/25'	102	126	334.5
26	60	20201226	100	120	250.0
27	60	'2020/12/27'	92	118	241.0
28	60	'2020/12/28'	103	132	300.0
29	60	'2020/12/29'	100	132	280.0
30	60	'2020/12/30'	102	129	380.3
31	60	'2020/12/31'	92	115	243.0
.....					



```
*IDLE Shell 3.11.0*
File Edit Shell Debug Options Window Help
31      60  '2020/12/31'      92      115      243.0
.....
before filling
      Duration      Date      Pulse      Maxpulse      Calories
0      60  '2020/12/01'      110      130      409.1
1      60  '2020/12/02'      117      145      479.0
2      60  '2020/12/03'      103      135      340.0
3      45  '2020/12/04'      109      175      282.4
4      45  '2020/12/05'      117      148      406.0
5      60  '2020/12/06'      102      127      300.0
6      60  '2020/12/07'      110      136      374.0
7      450  '2020/12/08'      104      134      253.3
8      30  '2020/12/09'      109      133      195.1
9      60  '2020/12/10'      98      124      269.0
10     60  '2020/12/11'      103      147      329.3
11     60  '2020/12/12'      100      120      250.7
12     60  '2020/12/12'      100      120      250.7
13     60  '2020/12/13'      106      128      345.3
14     60  '2020/12/14'      104      132      379.3
15     60  '2020/12/15'      98      123      275.0
16     60  '2020/12/16'      98      120      215.2
17     60  '2020/12/17'      100      120      300.0
18     45  '2020/12/18'      90      112      NaN
19     60  '2020/12/19'      103      123      323.0
20     45  '2020/12/20'      97      125      243.0
21     60  '2020/12/21'      108      131      364.2
22     45      NaN      100      119      282.0
23     60  '2020/12/23'      130      101      300.0
24     45  '2020/12/24'      105      132      246.0
25     60  '2020/12/25'      102      126      334.5
26     60      20201226      100      120      250.0
27     60  '2020/12/27'      92      118      241.0
28     60  '2020/12/28'      103      132      NaN
29     60  '2020/12/29'      100      132      280.0
```

*IDLE Shell 3.11.0*					
File Edit Shell Debug Options Window Help					
31	60	'2020/12/31'	92	115	243.0
after filling					
	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.1
1	60	'2020/12/02'	117	145	479.0
2	60	'2020/12/03'	103	135	340.0
3	45	'2020/12/04'	109	175	282.4
4	45	'2020/12/05'	117	148	406.0
5	60	'2020/12/06'	102	127	300.0
6	60	'2020/12/07'	110	136	374.0
7	450	'2020/12/08'	104	134	253.3
8	30	'2020/12/09'	109	133	195.1
9	60	'2020/12/10'	98	124	269.0
10	60	'2020/12/11'	103	147	329.3
11	60	'2020/12/12'	100	120	250.7
12	60	'2020/12/12'	100	120	250.7
13	60	'2020/12/13'	106	128	345.3
14	60	'2020/12/14'	104	132	379.3
15	60	'2020/12/15'	98	123	275.0
16	60	'2020/12/16'	98	120	215.2
17	60	'2020/12/17'	100	120	300.0
18	45	'2020/12/18'	90	112	300.0
19	60	'2020/12/19'	103	123	323.0
20	45	'2020/12/20'	97	125	243.0
21	60	'2020/12/21'	108	131	364.2
22	45	'2018/12/12'	100	119	282.0
23	60	'2020/12/23'	130	101	300.0
24	45	'2020/12/24'	105	132	246.0
25	60	'2020/12/25'	102	126	334.5
26	60	20201226	100	120	250.0
27	60	'2020/12/27'	92	118	241.0
28	60	'2020/12/28'	103	132	300.0
29	60	'2020/12/29'	100	132	280.0

```
*IDLE Shell 3.11.0*
File Edit Shell Debug Options Window Help
31      60  '2020/12/31'      92      115      243.0
.....
before duplicate
      Duration      Date      Pulse      Maxpulse      Calories
0      60  '2020/12/01'      110      130      409.1
1      60  '2020/12/02'      117      145      479.0
2      60  '2020/12/03'      103      135      340.0
3      45  '2020/12/04'      109      175      282.4
4      45  '2020/12/05'      117      148      406.0
5      60  '2020/12/06'      102      127      300.0
6      60  '2020/12/07'      110      136      374.0
7      450  '2020/12/08'      104      134      253.3
8      30  '2020/12/09'      109      133      195.1
9      60  '2020/12/10'      98      124      269.0
10     60  '2020/12/11'      103      147      329.3
11     60  '2020/12/12'      100      120      250.7
12     60  '2020/12/12'      100      120      250.7
13     60  '2020/12/13'      106      128      345.3
14     60  '2020/12/14'      104      132      379.3
15     60  '2020/12/15'      98      123      275.0
16     60  '2020/12/16'      98      120      215.2
17     60  '2020/12/17'      100      120      300.0
18     45  '2020/12/18'      90      112      300.0
19     60  '2020/12/19'      103      123      323.0
20     45  '2020/12/20'      97      125      243.0
21     60  '2020/12/21'      108      131      364.2
22     45  '2018/12/12'      100      119      282.0
23     60  '2020/12/23'      130      101      300.0
24     45  '2020/12/24'      105      132      246.0
25     60  '2020/12/25'      102      126      334.5
26     60      20201226      100      120      250.0
27     60  '2020/12/27'      92      118      241.0
28     60  '2020/12/28'      103      132      300.0
29     60  '2020/12/29'      100      132      280.0
```





	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.1
1	60	'2020/12/02'	117	145	479.0
2	60	'2020/12/03'	103	135	340.0
3	45	'2020/12/04'	109	175	282.4
4	45	'2020/12/05'	117	148	406.0
5	60	'2020/12/06'	102	127	300.0
6	60	'2020/12/07'	110	136	374.0
7	450	'2020/12/08'	104	134	253.3
8	30	'2020/12/09'	109	133	195.1
9	60	'2020/12/10'	98	124	269.0
10	60	'2020/12/11'	103	147	329.3
11	60	'2020/12/12'	100	120	250.7
12	60	'2020/12/12'	100	120	250.7
13	60	'2020/12/13'	106	128	345.3
14	60	'2020/12/14'	104	132	379.3
15	60	'2020/12/15'	98	123	275.0
16	60	'2020/12/16'	98	120	215.2
17	60	'2020/12/17'	100	120	300.0
18	45	'2020/12/18'	90	112	NaN
19	60	'2020/12/19'	103	123	323.0
20	45	'2020/12/20'	97	125	243.0
21	60	'2020/12/21'	108	131	364.2
22	45	NaN	100	119	282.0
23	60	'2020/12/23'	130	101	300.0
24	45	'2020/12/24'	105	132	246.0
25	60	'2020/12/25'	102	126	334.5

Figure 1

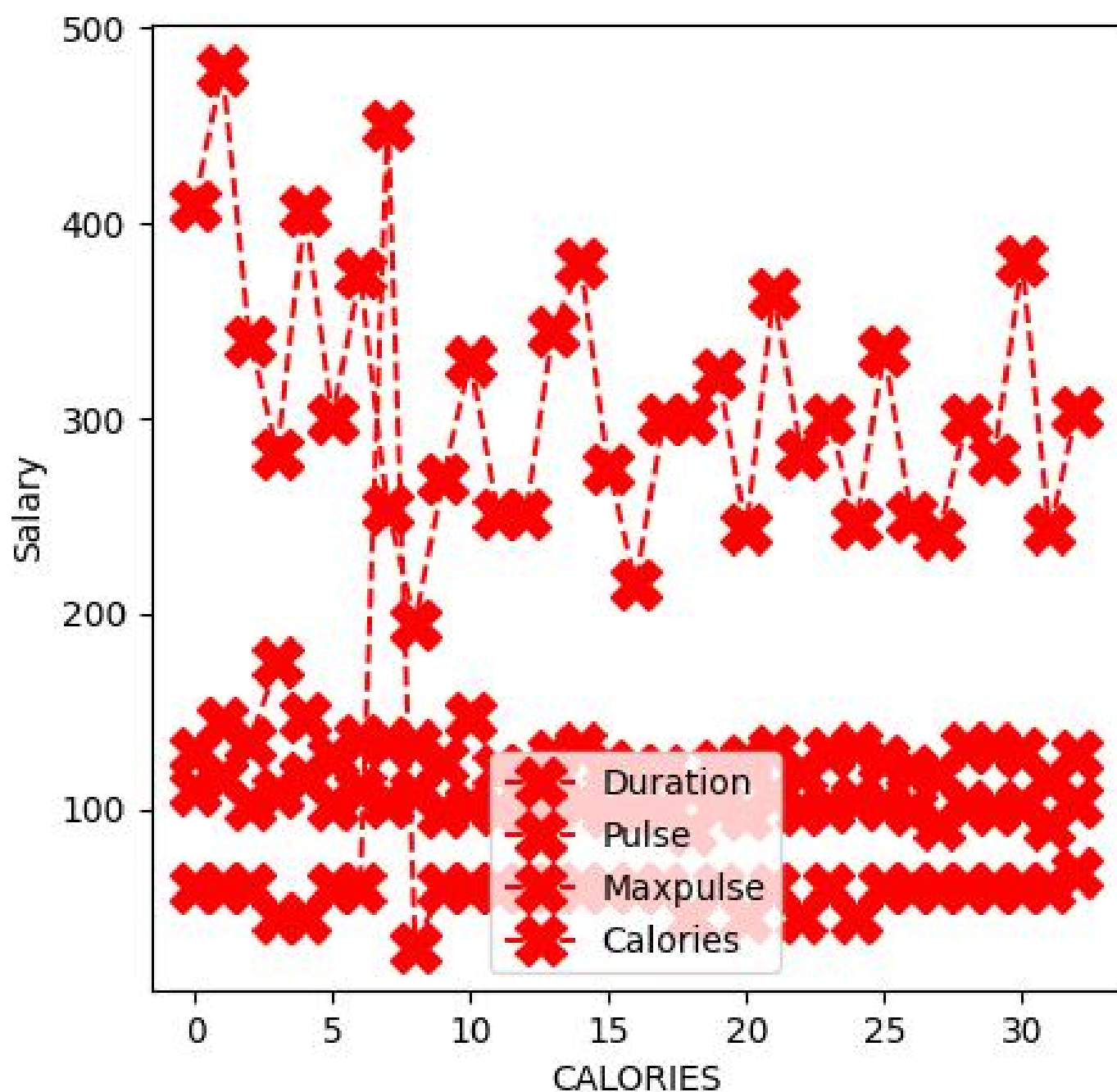


Figure 1

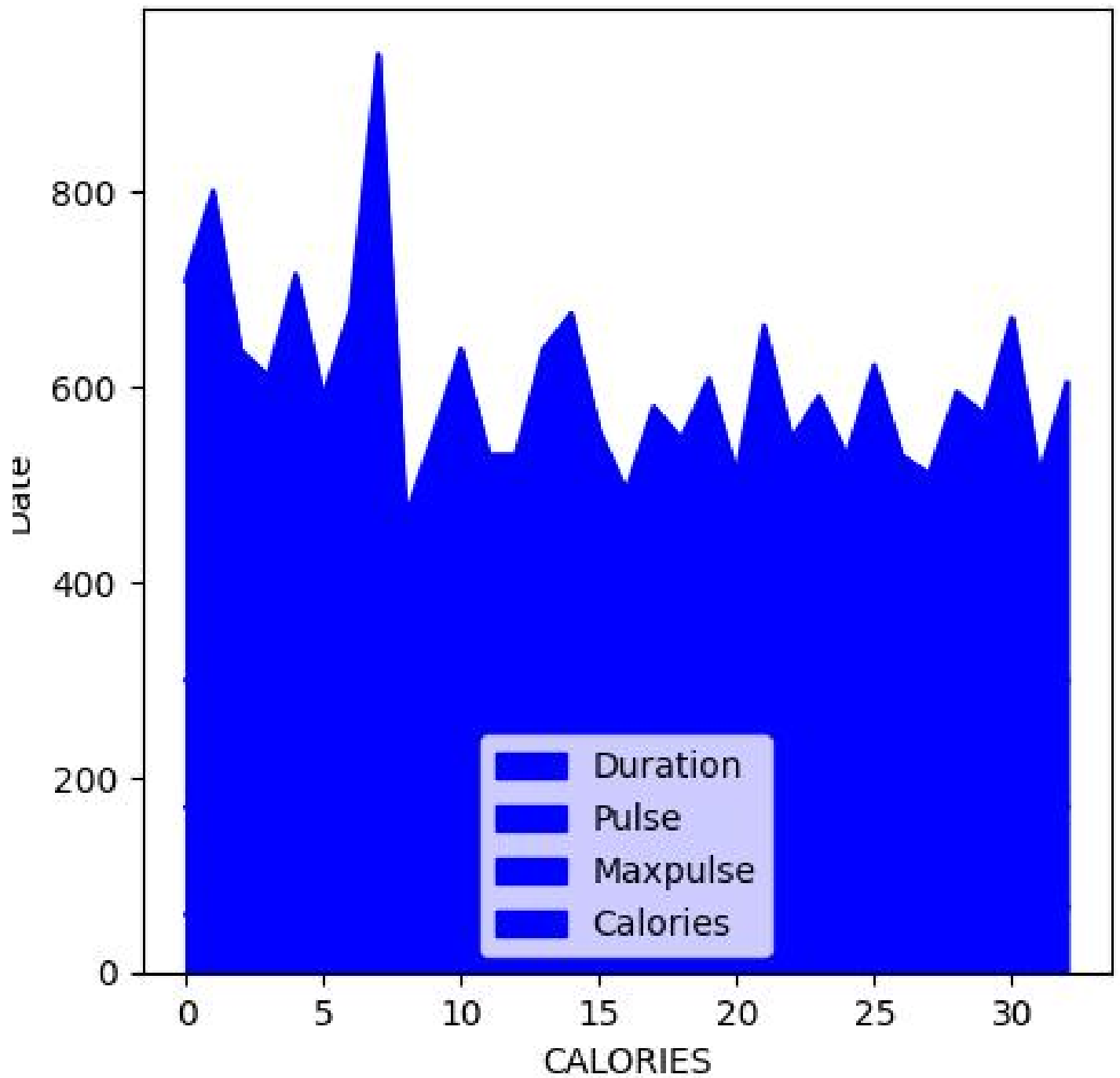






Figure 1

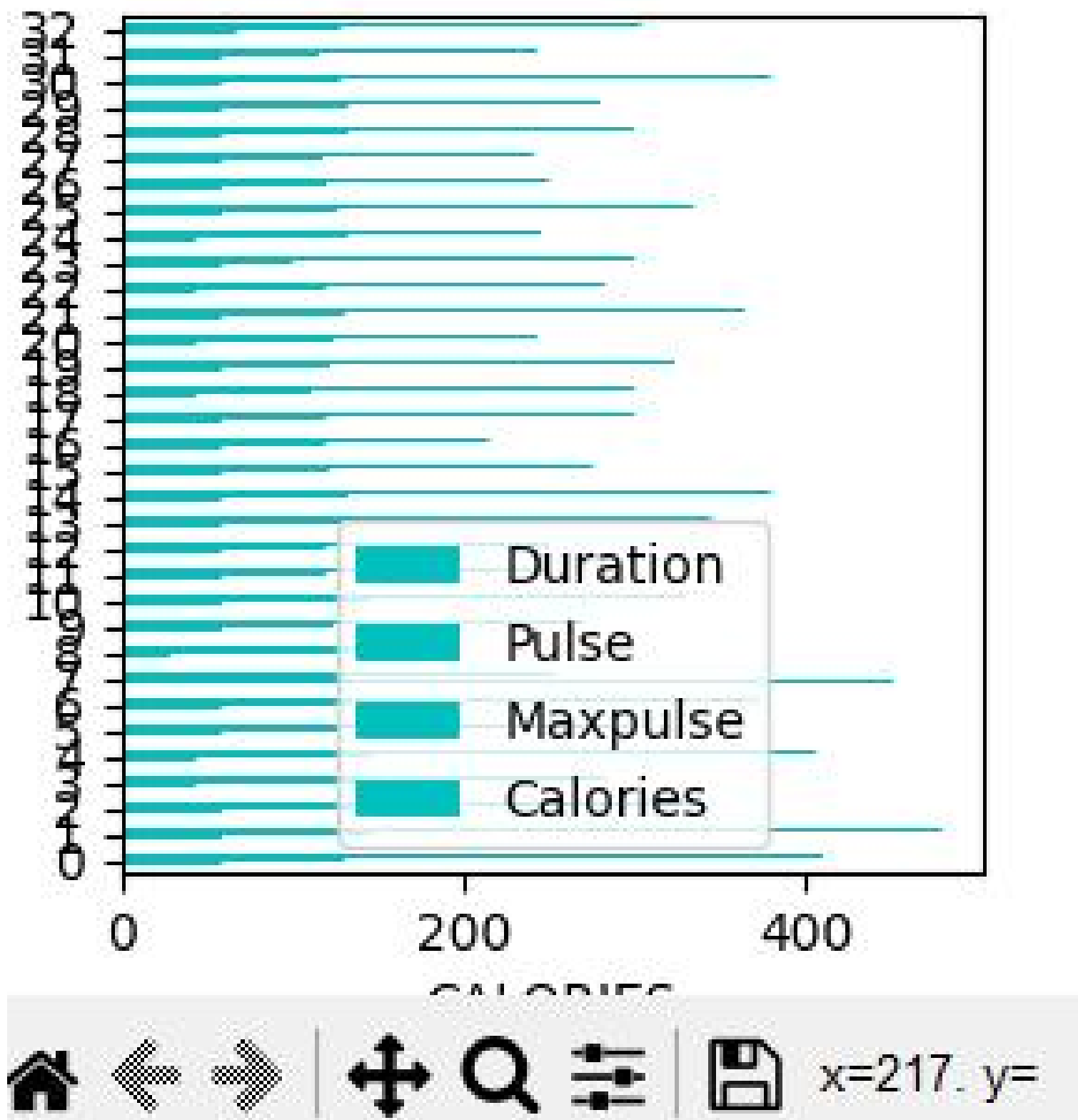


Figure 1

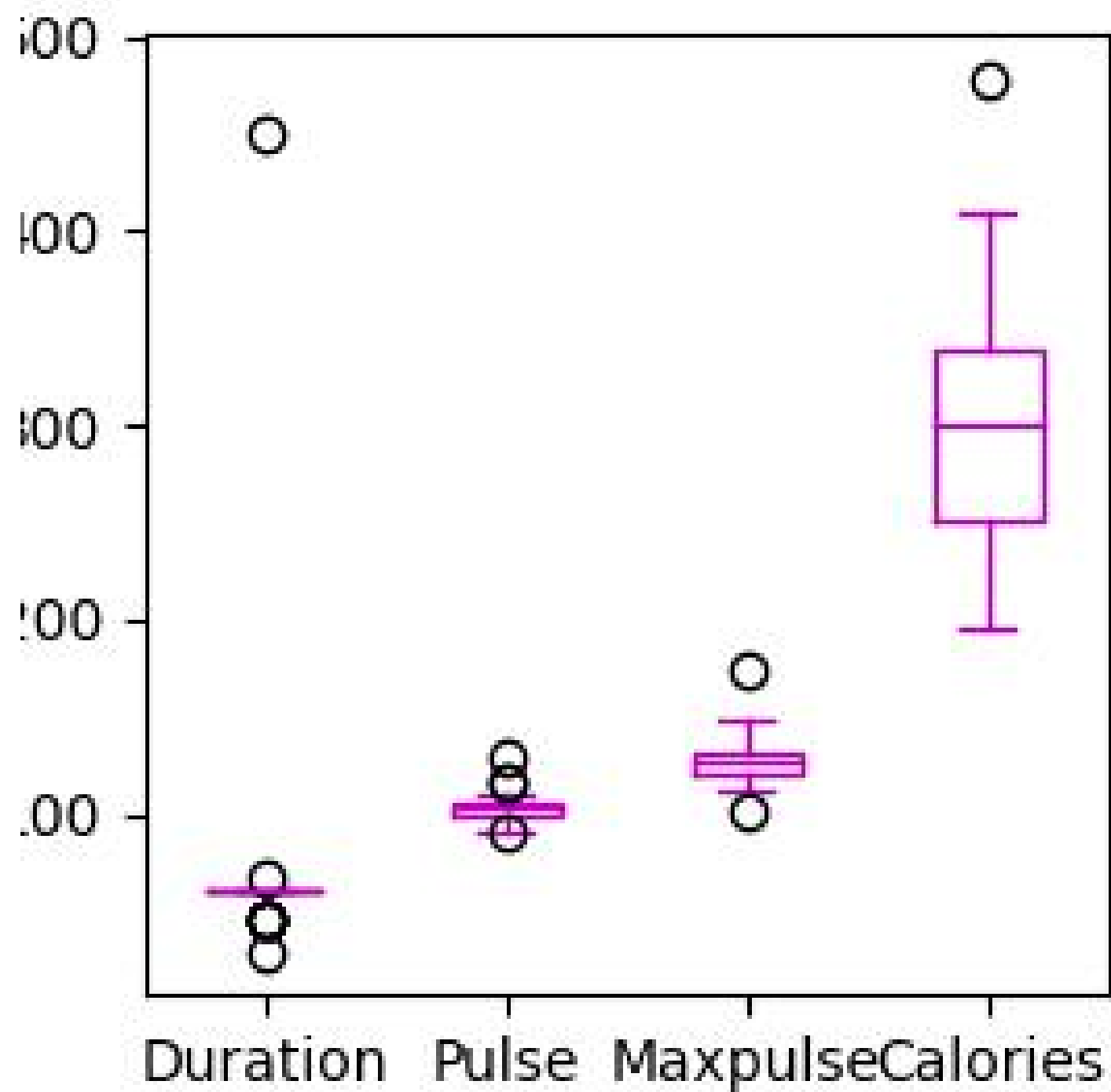


Figure 1

