



jupyter Project



Logout

☰ Menu

Trusted

Python 3



importing pandas library

In [20]:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 %matplotlib inline
4 import numpy as np
```

reading Data1 csv file

In [2]:

```
1 df1=pd.read_csv("Data1.csv")
```

reading initial rows





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Markdown



reading initial rows

In [3]:

1 df1.head()

Out[3]:

ID
0 38441
1 40403
2 3709
3 37422
4 12527

In [4]:

1 df1.shape





jupyter Project



Logout

☰ Menu

Trusted

Python 3



In [4]:

1 df1.shape

Out[4]:

(448, 1)

In [5]:

1 df1=pd.get_dummies(df1)

creating train dataset

In [6]:

1 train=df1[1:224]

creating test dataset

In [7]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Markdown



creating train dataset

In [6]:

1 train=df1[1:224]

creating test dataset

In [7]:

1 test=df1[225:]

In [8]:

1 x_train=train.drop('ID',axis=1)

In [9]:

1 x_test=test['ID']

In [10]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



C



Markdown



creating test dataset

In [7]:

```
1 test=df1[225:]
```

In [8]:

```
1 x_train=train.drop('ID',axis=1)
```

In [9]:

```
1 x_test=test['ID']
```

In [10]:

```
1 y_train=train.drop('ID',axis=1)
```

In [11]:

```
1 true_p=test['ID']
```

In [12]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Markdown



In [12]:

1 from sklearn.linear_model import

In [13]:

1 lreg=LinearRegression()

In [21]:

1 lreg.fit(x_train,x_test)

Out[21]:

LinearRegression()

creating dummies

In [15]:

1 x_train=pd.get_dummies('x_train')

In [16]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



C



Markdown



creating dummies

In [15]:

```
1 x_train=pd.get_dummies('x_train')
```

In [16]:

```
1 x_train.shape
```

Out[16]:

(1, 1)

In [17]:

```
1 x_test=pd.get_dummies('x_test')
```

In [18]:

```
1 x_train.fillna(0,inplace=True)
```

In [19]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



In [17]:

```
1 x_test=pd.get_dummies('x_test')
```

In [18]:

```
1 x_train.fillna(0,inplace=True)
```

In [19]:

```
1 x_test.fillna(0,inplace=True)
```

In [22]:

```
1 pred=lreg.predict(x_test)
```

In [23]:

```
1 pred
```

Out[23]:

```
array([[1.]])
```





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Markdown



reading Data2 csv file

In [24]:

```
1 df2=pd.read_csv("Data2.csv")
```

In [25]:

```
1 df2.head()
```

Out[25]:

	age
0	32
1	78
2	31
3	57
4	45

To [26]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



In [26]:

1 df2.tail()

Out[26]:

	age
234	27
235	57
236	31
237	33
238	46

In [27]:

1 df2.shape

Out[27]:

(239, 1)

In [28]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



In [28]:

```
1 df2['age'].mean()
```

Out[28]:

39.77824267782427

In [29]:

```
1 df2['age'].median()
```

Out[29]:

37.0

In [30]:

```
1 df2['age'].mode()
```

Out[30]:

```
0    32
dtype: int64
```





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



C



Markdown



computing variance

In [31]:

```
1 var=df2['age'].var(ddof=0)
```

In [32]:

```
1 print(var)
```

101.78764377374343

computing standard deviation

In [33]:

```
1 std=df2['age'].std(ddof=0)
```

In [34]:

```
1 print(std)
```





jupyter Project



Logout

☰ Menu

Trusted

Python 3



1 print(std)

10.088986260955231

In [35]:



1 (df2['age']<28)/len(df2['age'])

Out[35]:

0	0.000000
1	0.000000
2	0.000000
3	0.000000
4	0.000000
...	
234	0.004184
235	0.000000
236	0.000000
237	0.000000
238	0.000000

Name: age, Length: 239,
dtype: float64





jupyter Project



Logout

☰ Menu

Trusted

Python 3



In [36]:

1 df2['age']<28

Out[36]:

```
0      False
1      False
2      False
3      False
4      False
...
234     True
235    False
236    False
237    False
238    False
Name: age, Length: 239,
dtype: bool
```

drawing histogram

In [37]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



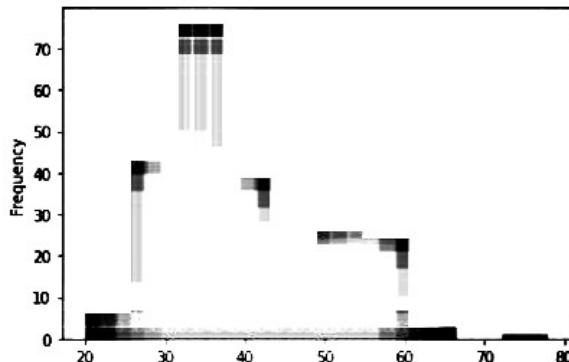
drawing histogram

In [37]:

```
1 df2['age'].plot.hist()
```

Out[37]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7eebfb0040>
```



In [38]:

```
1 df2[df2['age']<28].plot.hist()
```





jupyter Project



Logout

☰ Menu

Trusted

Python 3

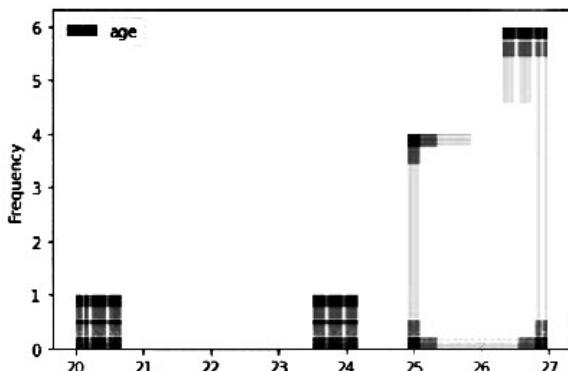


In [38]:

```
1 df2[df2['age']<28].plot.hist()
```

Out[38]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7eea  
a25d90>
```



drawing box plot

In [39]:

```
1 df2[df2['age']>28].plot.box()
```





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Markdown



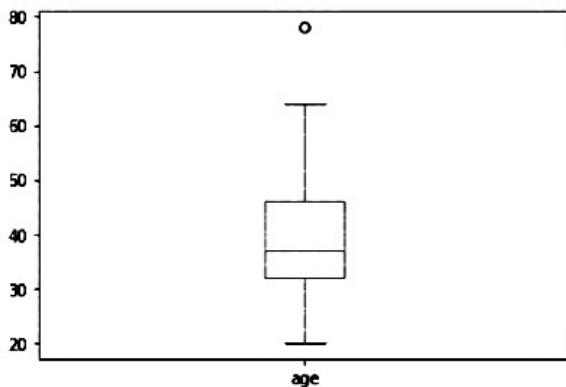
drawing box plot

In [39]:

```
1 df2['age'].plot.box()
```

Out[39]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7eec317df0>
```



In [40]:

```
1 df2[df2['age']<28].plot.box()
```





jupyter Project



Logout

☰ Menu

Trusted

Python 3

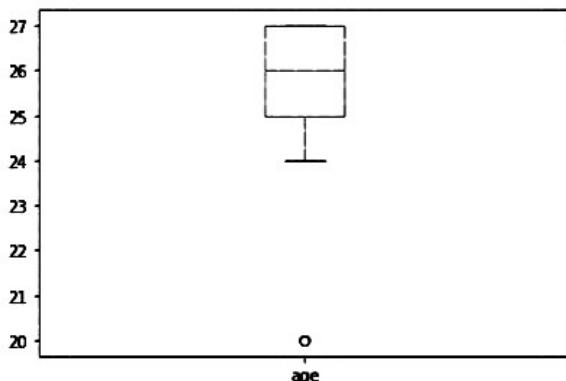


In [40]:

```
1 df2[df2['age']<28].plot.box()
```

Out[40]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7eea9afb50>
```



drawing bargraph

In [41]:

```
1 df2[df2['age']<28].plot.bar()
```



12





jupyter Project



Logout

☰ Menu

Trusted

Python 3

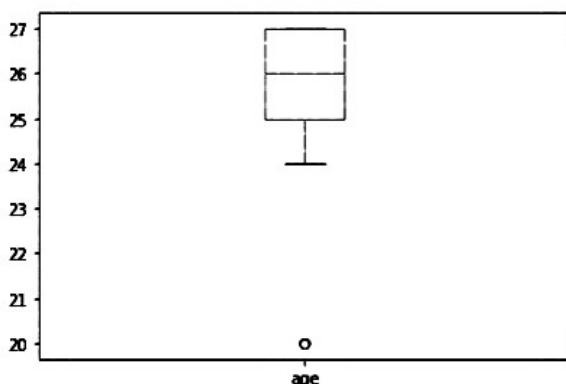


In [40]:

```
1 df2[df2['age']<28].plot.box()
```

Out[40]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7eea9afb50>
```



drawing bargraph

In [41]:

```
1 df2['age'].hist()
```



12





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Markdown



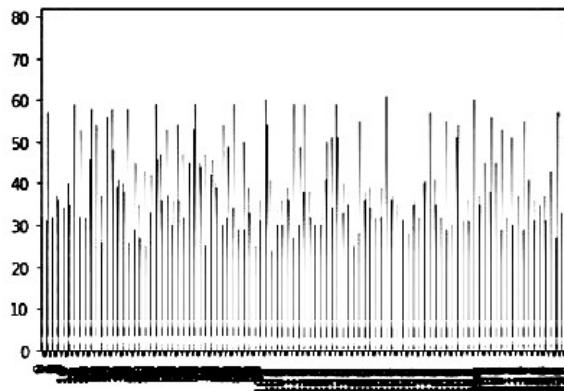
drawing bargraph

In [41]:

```
1 df2['age'].plot.bar()
```

Out[41]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7eea8f5b50>
```



In [42]:

```
1 df2[df2['age']<28].plot.bar()
```





jupyter Project



Logout

☰ Menu

Trusted

Python 3

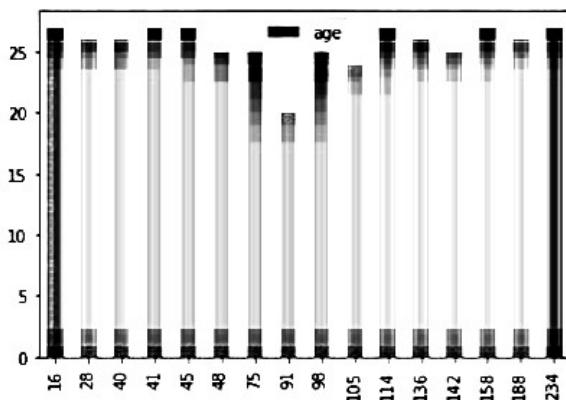


In [42]:

```
1 df2[df2['age']<28].plot.bar()
```

Out[42]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7eea00e4f0>
```



reading Data3 csv file

In [43]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



C



Markdown



reading Data3 csv file

In [43]:

```
1 df3=pd.read_csv("Data3.csv")
```

In [44]:

```
1 df3.head()
```

Out[44]:

JOB

0	services
1	retired
2	self-employed
3	services
4	blue-collar

In [45]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



In [45]:

1 df3.shape

Out[45]:

(465, 1)

In [46]:

1 df3.dtypes

Out[46]:

JOB object
dtype: object

In [47]:

1 df3.describe()

Out[47]:

JOB





jupyter Project



Logout

☰ Menu

Trusted

Python 3



In [47]:

```
1 df3.describe()
```

Out[47]:

JOB	
count	465
unique	11
top	blue-collar
freq	120

In [48]:

```
1 df3['JOB'].value_counts()
```

Out[48]:

blue-collar	120
management	95
technician	69
admin.	51





jupyter Project



Logout

☰ Menu

Trusted

Python 3



In [48]:

```
1 df3['JOB'].value_counts()
```

Out[48]:

blue-collar	120
management	95
technician	69
admin.	51
services	44
retired	20
self-employed	17
housemaid	16
entrepreneur	16
unemployed	13
student	4

Name: JOB, dtype: int64

plotting histogram

In [49]:

```
1 plt.hist(x='JOB', data=df)
```





jupyter Project



Logout

☰ Menu

Trusted

Python 3

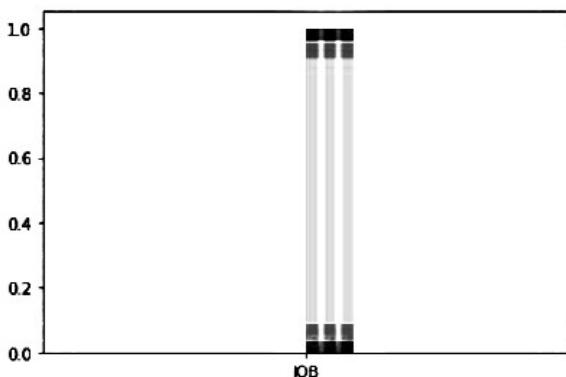


In [49]:

```
1 plt.hist(x='JOB', data='histogram')
```

Out[49]:

```
(array([0., 0., 0., 0.,
       0., 1., 0., 0., 0.,
       0.]),
 array([-0.5, -0.4, -0.
       3, -0.2, -0.1, 0. , 0.
       1, 0.2, 0.3, 0.4, 0.
       5]),
 <a list of 10 Patch obj
 ects>)
```



12





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Markdown



reading Data4 csv file

In [4]:

```
1 df4=pd.read_csv("Data4.csv")
```

In [5]:

```
1 df4.head()
```

Out[5]:

	duration
0	20
1	372
2	676
3	65
4	111

In [6]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



In [6]:

1 df4.tail()

Out[6]:

duration

179	142
180	258
181	149
182	173
183	101

In [7]:

1 df4.iloc[181]

Out[7]:

duration	149
Name:	181, dtype: int64





jupyter Project



Logout

☰ Menu

Trusted

Python 3



1 df4.iloc[5:179]

Out[8]:

duration	
5	455
6	309
7	120
8	149
9	119
...	...
174	531
175	8
176	164
177	147
178	69

174 rows × 1 columns





jupyter Project

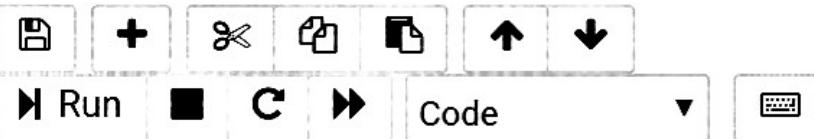


Logout

☰ Menu

Trusted

Python 3



In [9]:

```
1 df4['duration'].mean()
```

Out[9]:

280.32608695652175

In [10]:

```
1 df4['duration'].median()
```

Out[10]:

192.5

In [11]:

```
1 df4['duration'].mode()
```

Out[11]:

0	55
1	62
2	86
3	175





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



C



Code



1 df4['duration'].mode()

Out[11]:

```
0      55
1      62
2      86
3     175
4     192
dtype: int64
```

computing variance ↴

In [58]:

1 var=df4['duration'].var(ddof=0)

In [59]:

1 print(var)

76530.04584120982





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Code



computing standard deviation

In [60]:

```
1 std=df4['duration'].std(ddof=0)
```

In [61]:

```
1 print(std)
```

276.64064387072597

In [62]:

```
1 df4['duration']<120
```

Out[62]:

0	True
1	False
2	False
3	True
4	True





jupyter Project

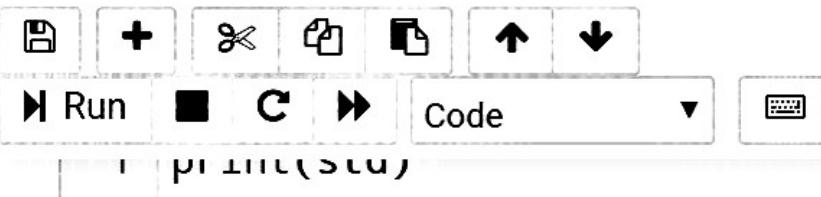


Logout

☰ Menu

Trusted

Python 3



276.64064387072597

In [62]:



1 df4['duration']<120

Out[62]:

```
0      True
1     False
2     False
3      True
4      True
...
179    False
180    False
181    False
182    False
183    True
Name: duration, Length:
184, dtype: bool
```



Project - Jupyter Notebook



jupyter Project

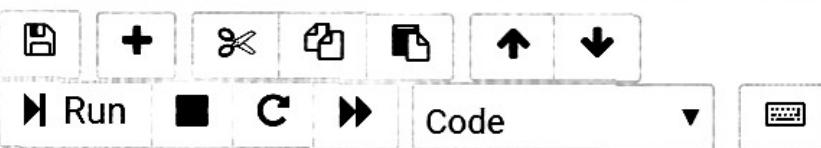


Logout

☰ Menu

Trusted

Python 3



plotting histogram

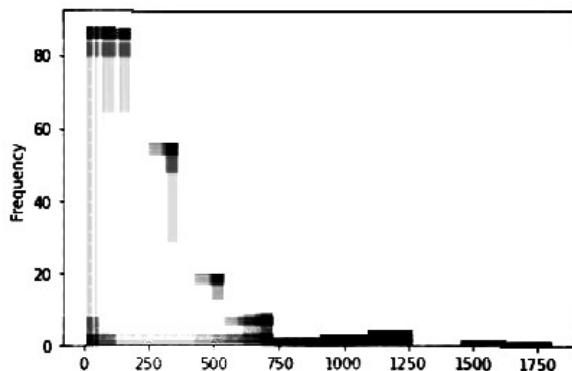
In [63]:

```
1 df4['duration'].plot.hist()
```



Out[63]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ee9d545e0>
```



plotting boxplot



12





jupyter Project

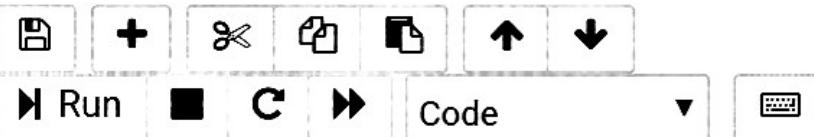


Logout

☰ Menu

Trusted

Python 3



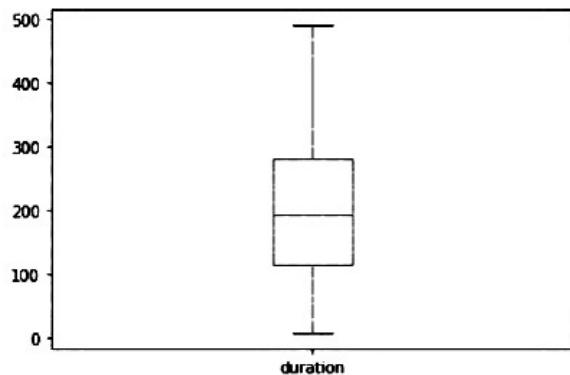
plotting boxplot

In [17]:

```
1 df4['duration'].plot.box()
```

Out[17]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x71a16659a0>
```



In [16]:

```
1 df4.loc[df4['duration']>500, 'dura
```





jupyter Project

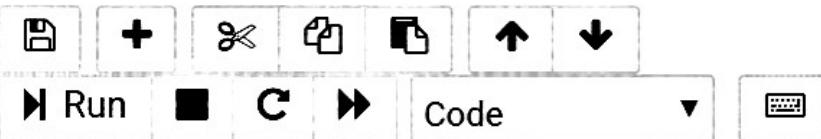


Logout

☰ Menu

Trusted

Python 3



In [16]:

1 df4.loc[df4['duration']>500, 'dura

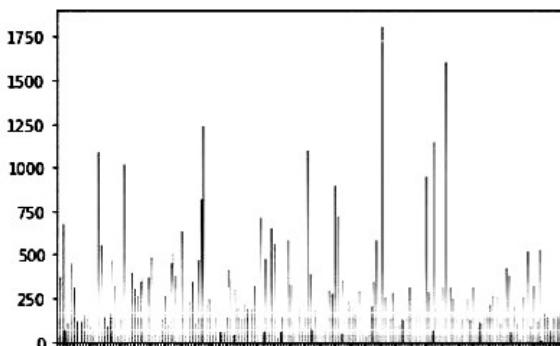
plotting bargraph

In [65]:

1 df4['duration'].plot.bar()

Out[65]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ee9cca550>
```





jupyter Project

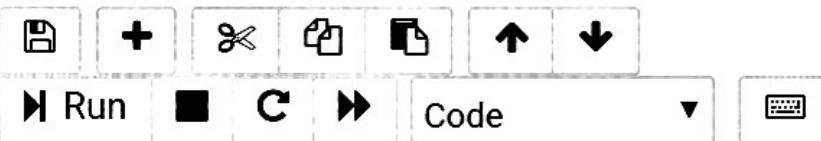


Logout

☰ Menu

Trusted

Python 3

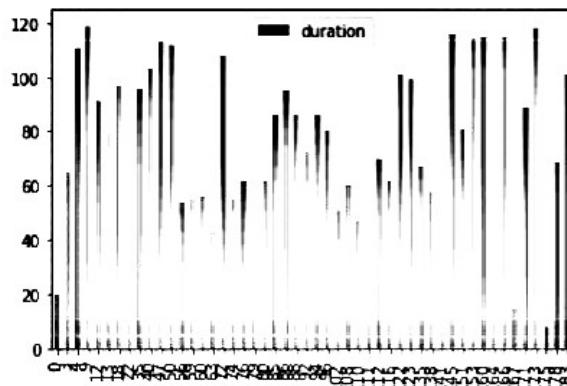


In [68]:

```
1 df4[df4['duration']<120].plot.bar
```

Out[68]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ee9446e80>
```



In [67]:

```
1 df4[df4['duration']<120].dropna()
```

Out[67]:





jupyter Project

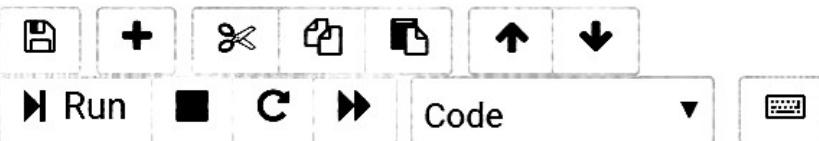


Logout

☰ Menu

Trusted

Python 3



In [67]:



```
1 df4[df4['duration']<120].dropna()
```

Out[67]:

duration	
0	20
3	65
4	111
9	119
12	91
13	79
18	97
22	30
35	96
40	103
47	113
50	112





jupyter Project

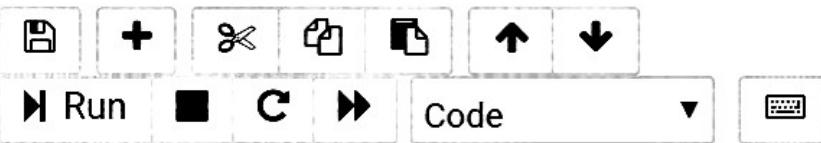


Logout

☰ Menu

Trusted

Python 3



151 81

153 114

160 115

164 55

166 115

167 15

171 89

173 118

175 8

178 69

183 101

In [69]:



1 df4.shape

Out[69]:

(184, 1)





jupyter Project

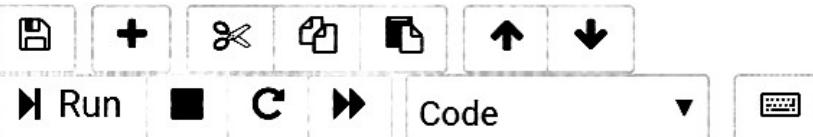


Logout

☰ Menu

Trusted

Python 3



finding IQR and other quantiles

In [70]:

```
1 q1=df4['duration'].quantile(0.25)
```

In [71]:

```
1 q2=df4['duration'].quantile(0.5)
```

In [72]:

```
1 q3=df4['duration'].quantile(0.75)
```

In [73]:

```
1 q4=df4['duration'].quantile(1)
```

In [74]:

```
1 IQR=q3-q1
```





jupyter Project

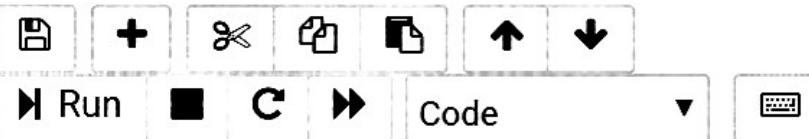


Logout

☰ Menu

Trusted

Python 3



In [74]:

1 IQR=q3-q1

In [75]:

1 print(IQR)

208.25

reading Data5 csv file

In [76]:

1 df5=pd.read_csv("Data5.csv")

In [77]:

1 df5.head()

Out[77]:

HOUSING





jupyter Project

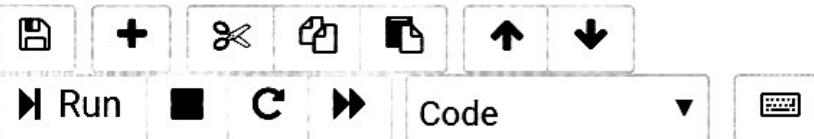


Logout

☰ Menu

Trusted

Python 3



In [78]:

1 df5.shape

Out[78]:

(97, 1)

In [79]:

1 df5.dtypes

Out[79]:

HOUSING object
dtype: object

In [80]:

1 df5.describe()

Out[80]:

HOUSING





jupyter Project

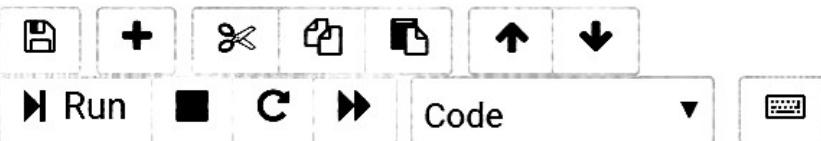


Logout

☰ Menu

Trusted

Python 3



In [80]:

1 df5.describe()

Out[80]:

HOUSING

count	97
unique	2
top	yes
freq	53

In [81]:

1 df5['HOUSING'].value_counts()

Out[81]:

yes	53
no	44
Name:	HOUSING, dtype: int64





jupyter Project

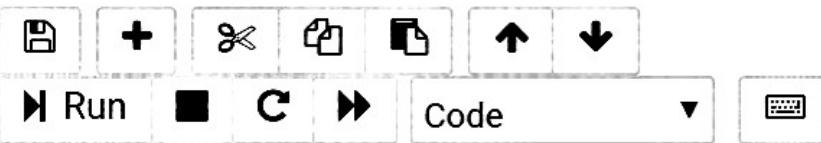


Logout

☰ Menu

Trusted

Python 3



In [82]:

```
1 df5[df5['HOUSING']=='yes']
```

Out[82]:

HOUSING

0	yes
2	yes
3	yes
5	yes
6	yes
9	yes
11	yes
12	yes
13	yes
16	yes
17	yes
18	yes





jupyter Project

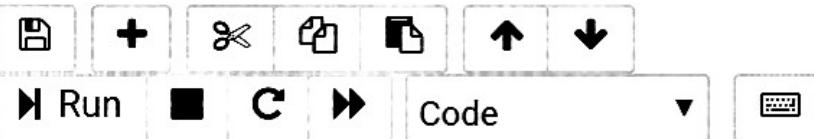


Logout

☰ Menu

Trusted

Python 3



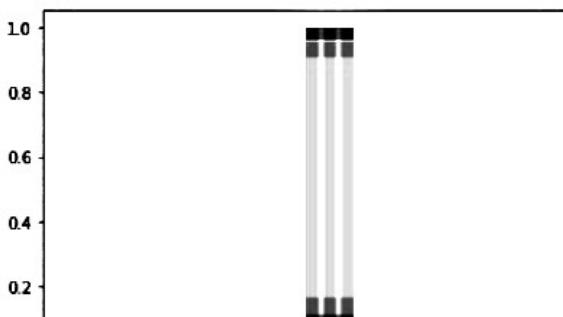
plotting histogram

In [83]:

```
1 plt.hist(x='HOUSING', data='histog
```

Out[83]:

```
(array([0., 0., 0., 0.,
       0., 1., 0., 0., 0.,
       0.]),
 array([-0.5, -0.4, -0.
       3, -0.2, -0.1, 0. , 0.
       1, 0.2, 0.3, 0.4, 0.
       5]),
 <a list of 10 Patch obj
 ects>)
```



12





jupyter Project

Logout

Menu Trusted Python 3

Run Cell Code

HOUSING

In [84]:

```
1 df5['HOUSING'].iloc[7]=='no'
```

Out[84]:

True

In [85]:

```
1 df5['HOUSING'].iloc[4:8]=='yes'
```

Out[85]:

```
4    False
5     True
6     True
7    False
Name: HOUSING, dtype: bo
ol
```





jupyter Project

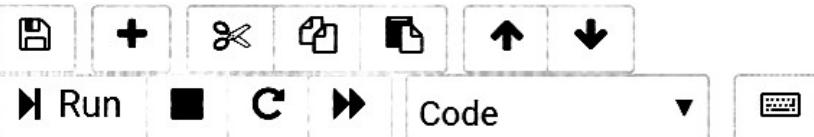


Logout

☰ Menu

Trusted

Python 3



reading Data6 csv file

In [86]:

```
1 df6=pd.read_csv("Data6.csv")
```

In [87]:

```
1 df6.head()
```

Out[87]:

MARITAL

- 0 married
- 1 divorced
- 2 single
- 3 single
- 4 divorced

In [88]:





jupyter Project

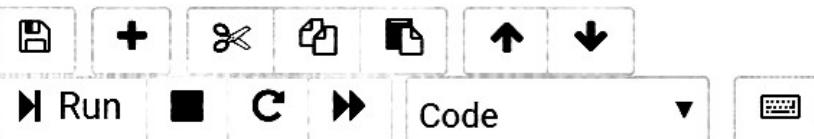


Logout

☰ Menu

Trusted

Python 3



reading Data5 csv file

In [76]:

```
1 df5=pd.read_csv("Data5.csv")
```

In [77]:

```
1 df5.head()
```

Out[77]:

HOUSING

0	yes
1	no
2	yes
3	yes
4	no

In [78]:





jupyter Project

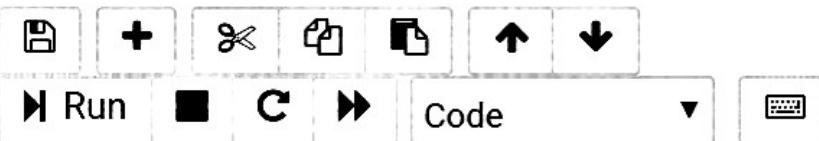


Logout

☰ Menu

Trusted

Python 3



In [88]:

1 df6.tail()

Out[88]:

MARITAL

171 married
172 single
173 married
174 divorced
175 single

In [89]:

1 df6.shape

Out[89]:

(176, 1)





jupyter Project

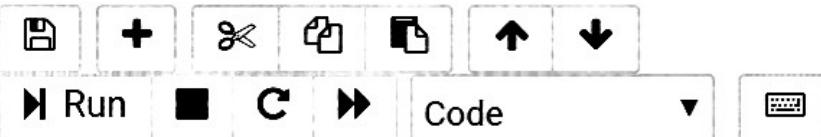


Logout

☰ Menu

Trusted

Python 3



In [90]:

1 df6.describe()

Out[90]:

MARITAL

count	176
unique	3
top	married
freq	92

In [91]:

1 df6.dtypes

Out[91]:

MARITAL	object
dtype:	object





jupyter Project

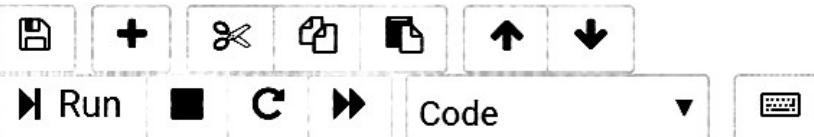


Logout

☰ Menu

Trusted

Python 3



In [92]:

```
1 df6['MARITAL'].value_counts()
```

Out[92]:

```
married      92
single       64
divorced     20
Name: MARITAL, dtype: int64
```

In [93]:

```
1 df6[df6['MARITAL']=='divorced']
```

Out[93]:

MARITAL

1	divorced
4	divorced
8	divorced





jupyter Project

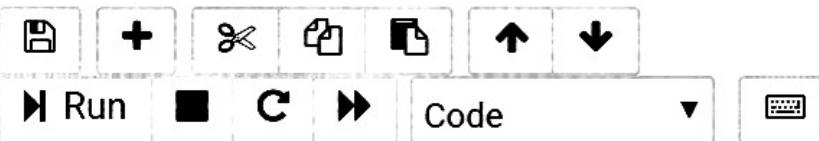


Logout

☰ Menu

Trusted

Python 3



```
1 df6[df6['MARITAL']=='divorced']
```

Out[93]:

MARITAL

- 1 divorced
- 4 divorced
- 8 divorced
- 12 divorced
- 25 divorced
- 29 divorced
- 32 divorced
- 36 divorced
- 39 divorced
- 43 divorced
- 54 divorced
- 62 divorced





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Code



In [94]:



1 df6['MARITAL'].iloc[56]=='single'

Out[94]:

False

In [95]:



1 df6['MARITAL'].iloc[56]=='married'

Out[95]:

True

In [96]:



1 plt.hist(x='MARITAL', data='histog

Out[96]:

(array([0., 0., 1., 0.,
0.]),
 array([-0.5, -0.3, -0.



jupyter Project

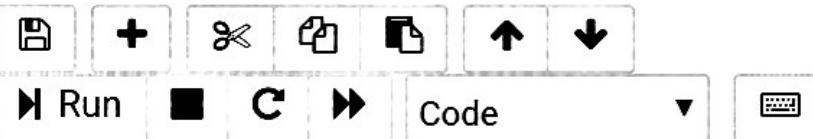


Logout

☰ Menu

Trusted

Python 3



In [99]:

1 df7.shape

Out[99]:

(119, 1)

In [100]:

1 df7.describe()

Out[100]:

CONTACT

count	119
unique	3
top	cellular
freq	67

In [101]:

1 df7.describe()





jupyter Project

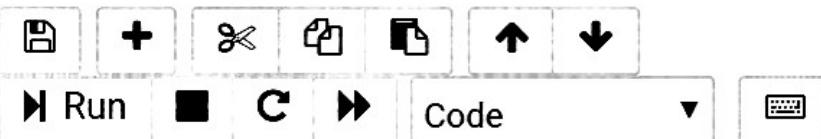


Logout

☰ Menu

Trusted

Python 3

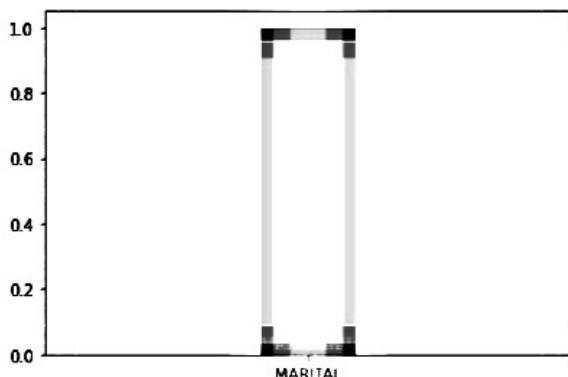


In [96]:

```
1 plt.hist(x='MARITAL', data='histog
```

Out[96]:

```
(array([0., 0., 1., 0.,
0.]),
 array([-0.5, -0.3, -0.
1, 0.1, 0.3, 0.5]),
 <a list of 5 Patch obje
cts>)
```



reading Data7 csv file





jupyter Project

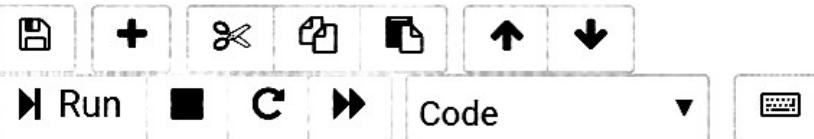


Logout

☰ Menu

Trusted

Python 3



reading Data7 csv file

In [97]:

```
1 df7=pd.read_csv("Data7.csv")
```

In [98]:

```
1 df7.head()
```

Out[98]:

CONTACT

- 0 cellular
- 1 telephone
- 2 unknown
- 3 telephone
- 4 unknown

In [99]:



12





jupyter Project

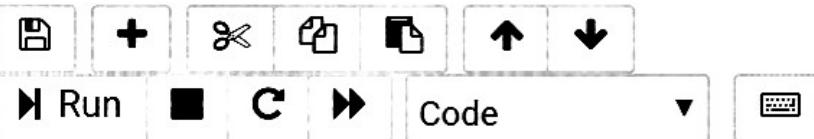


Logout

☰ Menu

Trusted

| Python 3



In [101]:

1 df7.dtypes

Out[101]:

CONTACT object
dtype: object

In [102]:

1 df7['CONTACT'].value_counts()

Out[102]:

cellular 67
unknown 38
telephone 14
Name: CONTACT, dtype: int64

In [103]:

1 df7['CONTACT']=='telephone'





jupyter Project

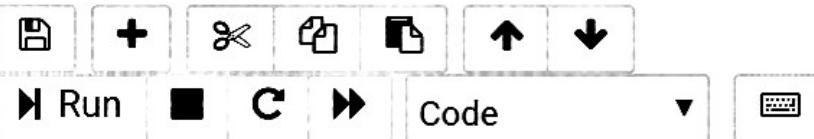


Logout

☰ Menu

Trusted

Python 3



In [103]:

```
1 df7['CONTACT']=='telephone'
```

Out[103]:

```
0      False
1      True
2     False
3      True
4     False
...
114    False
115    False
116    False
117    False
118    False
Name: CONTACT, Length: 1
19, dtype: bool
```

In [104]:

```
1 df7[df7['CONTACT']=='telephone']
```





jupyter Project

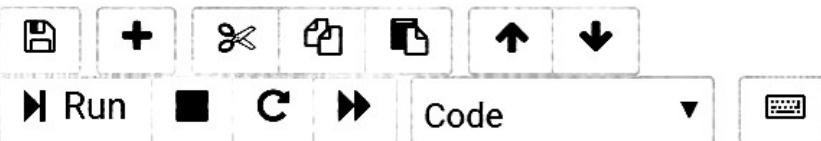


Logout

☰ Menu

Trusted

Python 3



In [104]:

1 df7[df7['CONTACT']=='telephone']

Out[104]:

CONTACT

- 1 telephone
- 3 telephone
- 22 telephone
- 30 telephone
- 32 telephone
- 38 telephone
- 43 telephone
- 57 telephone
- 71 telephone
- 88 telephone
- 100 telephone





jupyter Project

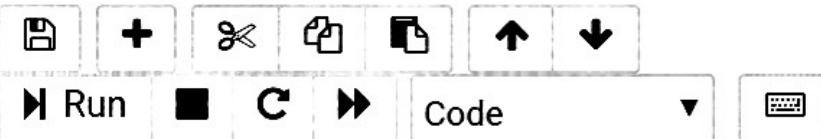


Logout

☰ Menu

Trusted

Python 3



CONTACT

- 1 telephone
- 3 telephone
- 22 telephone
- 30 telephone
- 32 telephone
- 38 telephone
- 43 telephone
- 57 telephone
- 71 telephone
- 88 telephone
- 100 telephone
- 102 telephone
- 104 telephone
- 108 telephone

In [105]:



12





jupyter Project

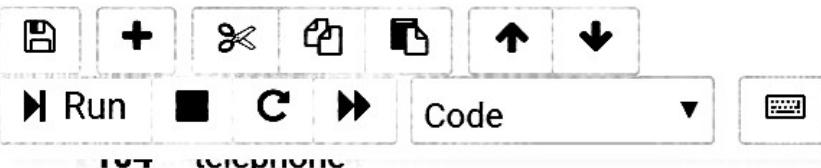


Logout

☰ Menu

Trusted

Python 3



108 telephone

In [105]:

1 df7['CONTACT'].iloc[88]=='cellula

Out[105]:

False

In [106]:

1 df7['CONTACT'].iloc[104]=='teleph

Out[106]:

True

reading Data8 csv file

In [107]:

1 df8 = pd.read_csv('Data8.csv')





jupyter Project

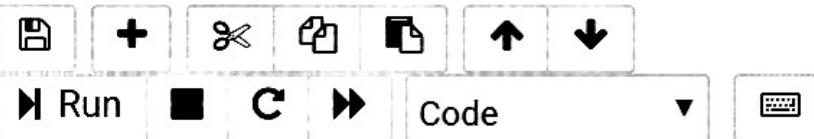


Logout

☰ Menu

Trusted

Python 3



reading Data8 csv file

In [107]:

```
1 df8=pd.read_csv("Data8.csv")
```

In [108]:

```
1 df8.head()
```

Out[108]:

LOAN

0 no

1 no

2 no

3 no

4 yes

In [109]:





jupyter Project

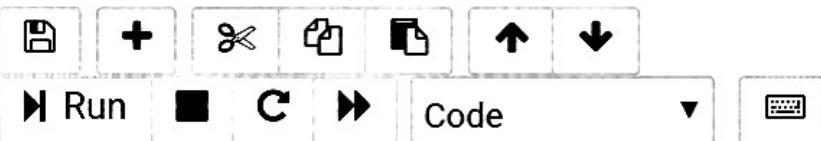


Logout

☰ Menu

Trusted

Python 3



In [109]:

1 df8.tail()

Out[109]:

LOAN

74	no
75	no
76	yes
77	yes
78	no

In [111]:

1 df8.dtypes

Out[111]:

LOAN	object
dtype:	object





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



Code



In [111]:

1 df8.dtypes

Out[111]:

```
LOAN    object
dtype: object
```

In [110]:

1 df8['LOAN'].value_counts()

Out[110]:

```
no      62
yes     17
Name: LOAN, dtype: int64
```

In [112]:

1 df8[df8['LOAN']=='no']

Out[112]:





jupyter Project

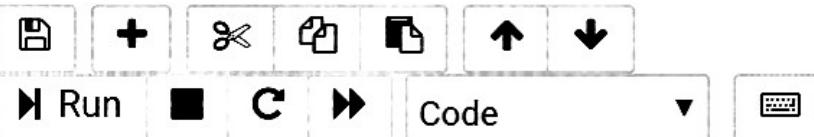


Logout

☰ Menu

Trusted

Python 3



In [110]:

```
1 df8['LOAN'].value_counts()
```

Out[110]:

```
no      62
yes     17
Name: LOAN, dtype: int64
```

In [112]:

```
1 df8[df8['LOAN']=='no']
```

Out[112]:

LOAN

0	no
1	no
2	no
3	no





jupyter Project

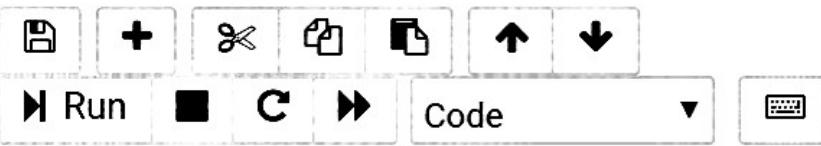


Logout

☰ Menu

Trusted

Python 3



In [111]:

Out[112]:

LOAN	
0	no
1	no
2	no
3	no
5	no
...	...
72	no
73	no
74	no
75	no
78	no

62 rows x 1 columns





jupyter Project

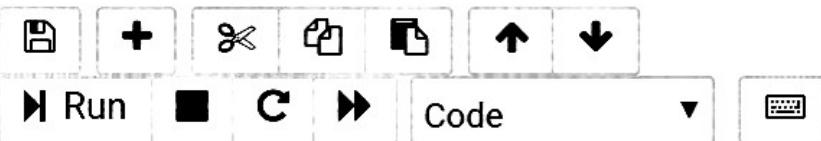


Logout

☰ Menu

Trusted

Python 3



In [114]:

```
1 df8['LOAN'].iloc[7]=='yes'
```

Out[114]:

False

In [115]:

```
1 df8['LOAN'].isnull()
```

Out[115]:

```
0    False
1    False
2    False
3    False
4    False
...
74   False
75   False
76   False
77   False
78   False
Name: LOAN, Length: 79,
```





jupyter Project

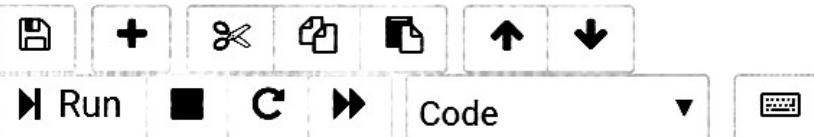


Logout

☰ Menu

Trusted

Python 3



reading Data9 csv file

In [116]:

```
1 df9=pd.read_csv("Data9.csv")
```

In [117]:

```
1 df9.head()
```

Out[117]:

campaign	
0	6
1	1
2	1
3	2
4	1

In [118]:





jupyter Project

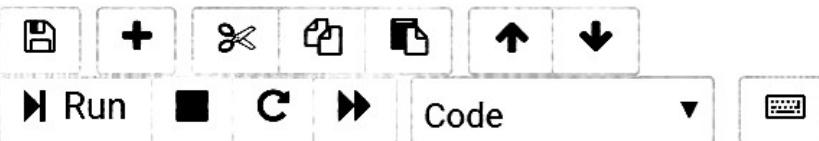


Logout

☰ Menu

Trusted

Python 3



In [118]:

1 df9.shape

Out[118]:

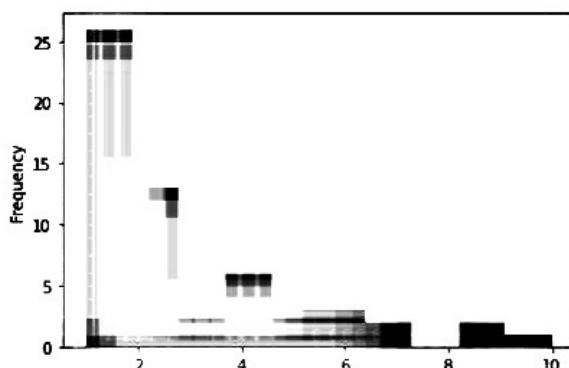
(59, 1)

In [119]:

1 df9['campaign'].plot.hist()

Out[119]:

<matplotlib.axes._subplots.AxesSubplot at 0x7ee8be51c0>





jupyter Project

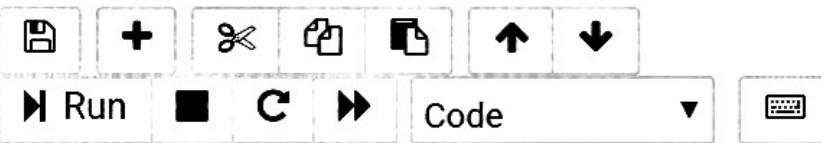


Logout

☰ Menu

Trusted

Python 3

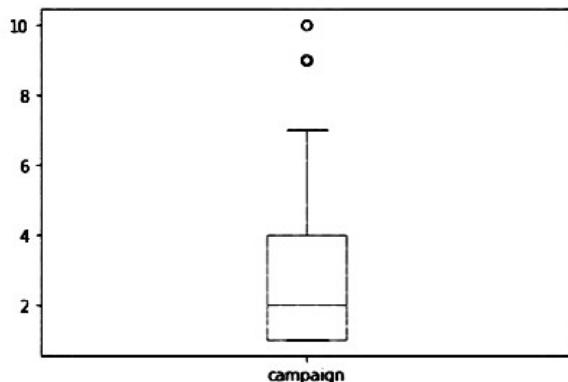


In [121]:

```
1 df9['campaign'].plot.box()
```

Out[121]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ee8a25430>
```



In [122]:

```
1 df9.mean()
```

Out[122]:





jupyter Project



Logout

☰ Menu

Trusted

Python 3



Code



In [122]:

1 df9.mean()

Out[122]:

```
campaign    2.711864
dtype: float64
```

In [123]:

1 (df9['campaign']<5).value_counts()

Out[123]:

```
True      48
False     11
Name: campaign, dtype: int64
```

In [130]:

1 q1=df9['campaign'].quantile(0.25)

In [131]:





jupyter Project

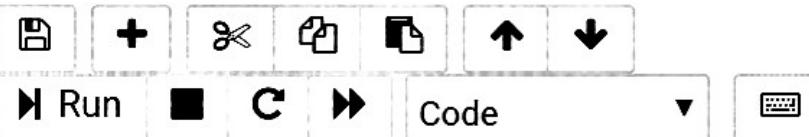


Logout

☰ Menu

Trusted

Python 3



In [130]:

```
1 q1=df9['campaign'].quantile(0.25)
```

In [131]:

```
1 q3=df9['campaign'].quantile(0.75)
```

In [132]:

```
1 IQR=q3-q1
```

In [133]:

```
1 print(IQR)
```

3.0

reading Data10 csv file

In [22]:

```
1 df10=pd.read_csv("Data10.csv")
```





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



C



Code



reading Data10 csv file

In [22]:

```
1 df10=pd.read_csv("Data10.csv")
```

In [23]:

```
1 df10.head()
```

Out[23]:

balance

0	118
---	-----

1	2787
---	------

2	144
---	-----

3	3777
---	------

4	-705
---	------

In [24]:





jupyter Project

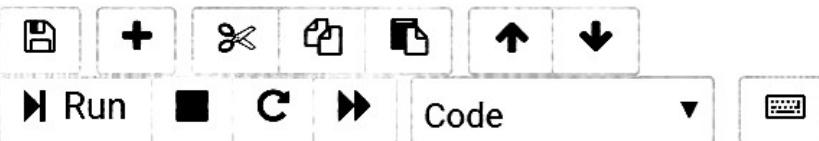


Logout

☰ Menu

Trusted

Python 3



In [24]:

1 df10.tail()

Out[24]:

balance	
13559	45
13560	2281
13561	285
13562	464
13563	2

In [25]:

1 df10.dtypes

Out[25]:

```
balance    int64
dtype: object
```





jupyter Project

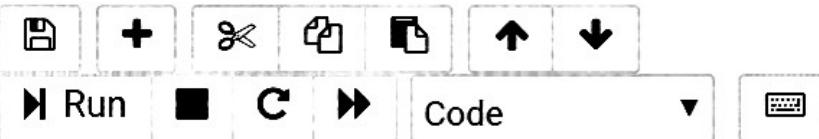


Logout

☰ Menu

Trusted

Python 3



In [25]:

1 df10.dtypes

Out[25]:

```
balance    int64
dtype: object
```

In [10]:

1 df10['balance'].mean()

Out[10]:

1358.4965349454437

In [11]:

1 df10['balance'].median()

Out[11]:

445.0





jupyter Project

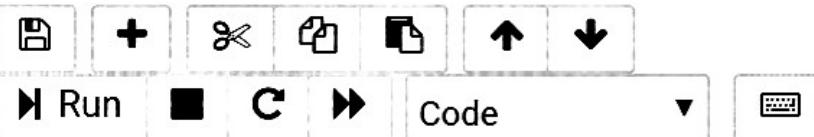


Logout

☰ Menu

Trusted

Python 3



In [12]:

```
1 df10['balance'].mode()
```

Out[12]:

```
0    0
dtype: int64
```

In [13]:

```
1 var=df10['balance'].var(ddof=0)
```

In [14]:

```
1 print(var)
```

```
9503822.167858828
```

In [15]:

```
1 std=df10['balance'].std(ddof=0)
```



12





jupyter Project



Logout

☰ Menu

Trusted

Python 3



▶ Run



C



Code



In [15]:

```
1 std=df10['balance'].std(ddof=0)
```

In [16]:

```
1 print(std)
```

3082.826976633432

In [17]:

```
1 var=9503822.16
2 std=3082.8269
3 if var==std:
4     print('equal')
5 else:
6     print('unequal')
```

unequal

In [18]:

```
1 mean=1258
```





jupyter Project

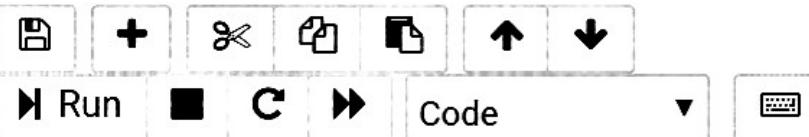


Logout

☰ Menu

Trusted

Python 3



In [18]:

```
1 mean=1358
2 median=445
3 if mean<median:
4     print('small mean')
5 else:
6     print('large mean')
```

large mean

In [19]:

```
1 df10[df10['balance']<mean].plot.h
```

Out[19]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7731
14b7f0>
```





jupyter Project

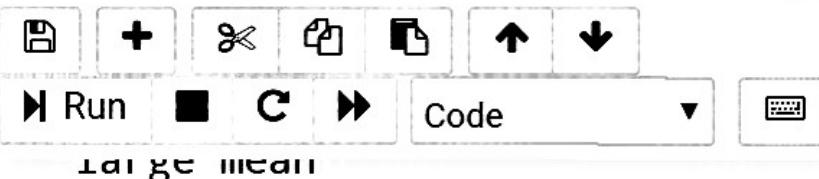


Logout

☰ Menu

Trusted

Python 3

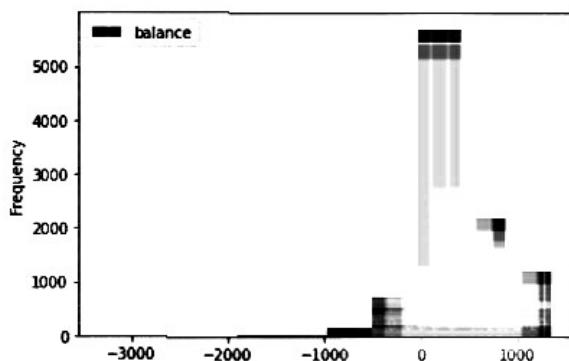


In [19]:

```
1 df10[df10['balance']<mean].plot.h
```

Out[19]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x773114b7f0>
```



In [20]:

```
1 df10[df10['balance']<median].plot
```

Out[20]:





jupyter Project

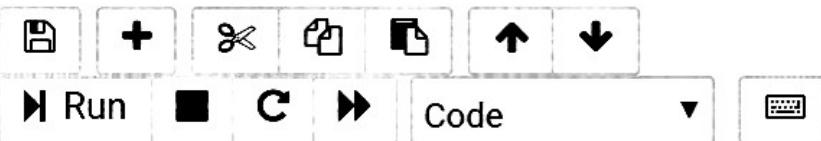


Logout

☰ Menu

Trusted

Python 3

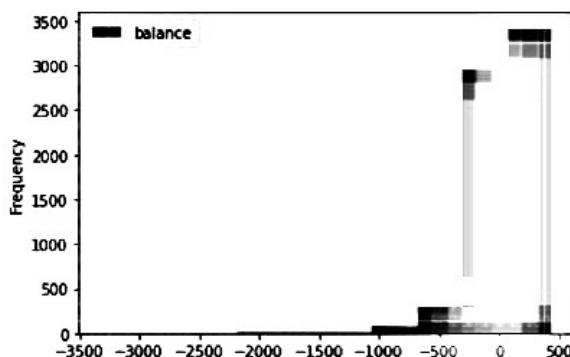


In [20]:

1 df10[df10['balance']<median].plot

Out[20]:

<matplotlib.axes._subplots.AxesSubplot at 0x772fc9460>



In [21]:

1 df10[df10['balance']<var].plot.hi

Out[21]:

<matplotlib.axes._subplots.AxesSubplot at 0x772fbf8>





jupyter Project

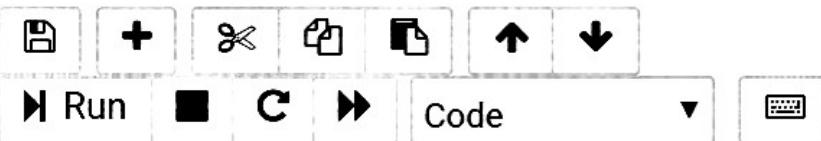


Logout

☰ Menu

Trusted

Python 3

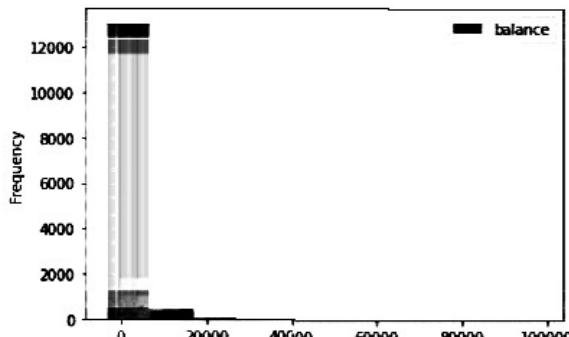


In [21]:

```
1 df10[df10['balance']<var].plot.hi
```

Out[21]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x772fb5b6d0>
```



In [22]:

```
1 df10[df10['balance']<std].plot.hi
```

Out[22]:



Project - Jupyter Notebook



jupyter Project

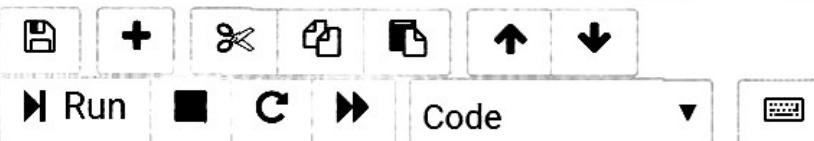


Logout

☰ Menu

Trusted

Python 3

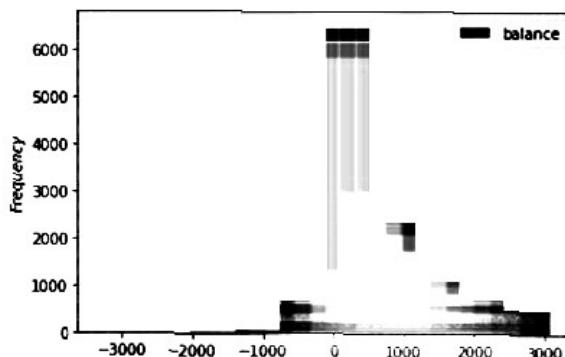


In [22]:

```
1 df10[df10['balance']<std].plot.hi
```

Out[22]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x772fb4a640>
```



In [27]:

```
1 df10['balance'].plot.box()
```

Out[27]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x772fb4a640>
```





jupyter Project

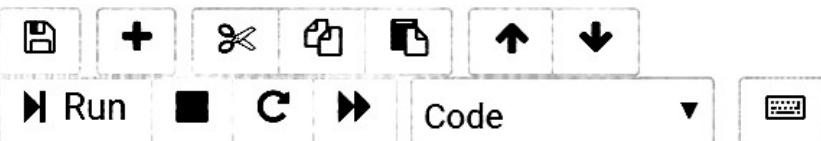


Logout

☰ Menu

Trusted

Python 3

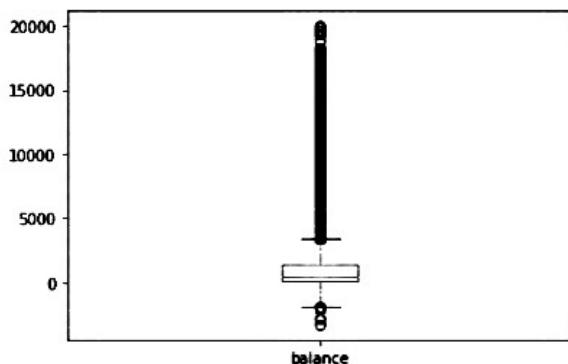


In [27]:

```
1 df10['balance'].plot.box()
```

Out[27]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x71a1645640>
```



In [26]:

```
1 df10.loc[df10['balance']>20000, 'b
```

