PANDAS PART - 5 BY VIRAT TIWARI

October 16, 2023

[16]:
```
import pandas as pd
```

[17]:
```
data={"A":[1,2,3,4],
      "B":[5,6,4,9],
      "Name":["Virat","Yash","Rohit","Parth"]}
```

[18]:
```
df=pd.DataFrame(data)
```

[19]:
```
df
```

[19]:

	A	B	Name
0	1	5	Virat
1	2	6	Yash
2	3	4	Rohit
3	4	9	Parth

[20]:
```
df.set_index("A",inplace=True)
```

[21]:
```
df
```

[21]:

A	B	Name
1	5	Virat
2	6	Yash
3	4	Rohit
4	9	Parth

[22]:
```
df.reset_index()
```

[22]:

	A	B	Name
0	1	5	Virat
1	2	6	Yash
2	3	4	Rohit
3	4	9	Parth

[23]:
```
df
```

[23]:
```
   B   Name
A
1  5  Virat
2  6   Yash
3  4  Rohit
4  9  Parth
```

[24]:
```
df.reset_index(inplace=True)
```

[25]:
```
df
```

[25]:
```
   A  B   Name
0  1  5  Virat
1  2  6   Yash
2  3  4  Rohit
3  4  9  Parth
```

[26]:
```
data={"A":[1,2,3,4],
      "B":[5,6,4,9],
      "Name":["Virat","Yash","Rohit","Parth"]}
df1=pd.DataFrame(data,index=["a","b","c","d"])
```

[27]:
```
df1
```

[27]:
```
   A  B   Name
a  1  5  Virat
b  2  6   Yash
c  3  4  Rohit
d  4  9  Parth
```

[28]:
```
df1.reindex(["b","c","d","a"])
```

[28]:
```
   A  B   Name
b  2  6   Yash
c  3  4  Rohit
d  4  9  Parth
a  1  5  Virat
```

[29]:
```
df1
```

[29]:
```
   A  B   Name
a  1  5  Virat
b  2  6   Yash
c  3  4  Rohit
d  4  9  Parth
```

[30]:
```
df2=df1.reindex(["b","c","d","a"])
```

[31]:
```
df2
```

[31]:
```
   A  B   Name
b  2  6   Yash
c  3  4  Rohit
d  4  9  Parth
a  1  5  Virat
```

[32]:
```
df1
```

[32]:
```
   A  B   Name
a  1  5  Virat
b  2  6   Yash
c  3  4  Rohit
d  4  9  Parth
```

[33]:
```
for i ,j in df1.iterrows():
    print(i,j)
```

```
a A           1
B           5
Name    Virat
Name: a, dtype: object
b A           2
B           6
Name    Yash
Name: b, dtype: object
c A           3
B           4
Name    Rohit
Name: c, dtype: object
d A           4
B           9
Name    Parth
Name: d, dtype: object
```

[34]:
```
df1
```

[34]:
```
   A  B   Name
a  1  5  Virat
b  2  6   Yash
c  3  4  Rohit
d  4  9  Parth
```

[35]:
```
for i  in df1.iteritems():
    print(i)
```

```
('A', a    1
```

```
b    2
c    3
d    4
Name: A, dtype: int64)
('B', a    5
b    6
c    4
d    9
Name: B, dtype: int64)
('Name', a    Virat
b     Yash
c    Rohit
d    Parth
Name: Name, dtype: object)

/tmp/ipykernel_1244/520842960.py:1: FutureWarning: iteritems is deprecated and
will be removed in a future version. Use .items instead.
  for i  in df1.iteritems():
```

[36]:
```
df1
```

[36]:

	A	B	Name
a	1	5	Virat
b	2	6	Yash
c	3	4	Rohit
d	4	9	Parth

[37]:
```
[i for i in df["A"]]
```

[37]:
```
[1, 2, 3, 4]
```

[38]:
```
df1
```

[38]:

	A	B	Name
a	1	5	Virat
b	2	6	Yash
c	3	4	Rohit
d	4	9	Parth

[39]:
```
def test(x):
    return x.sum()

df1.apply(test)
```

[39]:
```
A                        10
B                        24
Name    ViratYashRohitParth
dtype: object
```

[40]:
```
df1
```

[40]:

	A	B	Name
a	1	5	Virat
b	2	6	Yash
c	3	4	Rohit
d	4	9	Parth

[41]:
```
df3=df1[["A","B"]]
```

[42]:
```
df3
```

[42]:

	A	B
a	1	5
b	2	6
c	3	4
d	4	9

[43]:
```
# This is how we get the square

df3.applymap(lambda x:x**2)
```

[43]:

	A	B
a	1	25
b	4	36
c	9	16
d	16	81

[44]:
```
df1
```

[44]:

	A	B	Name
a	1	5	Virat
b	2	6	Yash
c	3	4	Rohit
d	4	9	Parth

[45]:
```
df
```

[45]:

	A	B	Name
0	1	5	Virat
1	2	6	Yash
2	3	4	Rohit
3	4	9	Parth

[46]:
```
df.sort_values("Name",inplace=True)
```

[47]:
```
df
```

[47]:
```
   A  B   Name
3  4  9  Parth
2  3  4  Rohit
0  1  5  Virat
1  2  6   Yash
```

[48]:
```
df
```

[48]:
```
   A  B   Name
3  4  9  Parth
2  3  4  Rohit
0  1  5  Virat
1  2  6   Yash
```

[49]:
```
# This is how we sor the indexes in ascending and descending orders

df.sort_index(ascending=True)
```

[49]:
```
   A  B   Name
0  1  5  Virat
1  2  6   Yash
2  3  4  Rohit
3  4  9  Parth
```

[50]:
```
df.sort_index(ascending=True,inplace = True)
```

[51]:
```
df
```

[51]:
```
   A  B   Name
0  1  5  Virat
1  2  6   Yash
2  3  4  Rohit
3  4  9  Parth
```

[52]:
```
df
```

[52]:
```
   A  B   Name
0  1  5  Virat
1  2  6   Yash
2  3  4  Rohit
3  4  9  Parth
```

[53]:
```
df4=pd.DataFrame({"Desc":["Hey ! This is vIrat Tiwari Aspiring Data Scientist .␣
 ↪Currently I enrolled in Data Science Masters course by PW Skills and I␣
 ↪already finised Python , Statistics and Machine Learning as well . It was␣
 ↪great experience with the pw team and their course is amazing."]})
```

[54]:
```
df4
```

[54]:
```
                                                Desc
0  Hey ! This is vIrat Tiwari Aspiring Data Scien…
```

[55]:
```
pd.set_option("display.max_colwidth",1000)

df4=pd.DataFrame({"Desc":["Hey ! This is vIrat Tiwari Aspiring Data Scientist .␣
 ↪Currently I enrolled in Data Science Masters course by PW Skills and I␣
 ↪already finised Python , Statistics and Machine Learning as well . It was␣
 ↪great experience with the pw team and their course is amazing."]})
```

[56]:
```
df4
```

[56]:
```
                    Desc
0  Hey ! This is vIrat Tiwari Aspiring Data Scientist . Currently I enrolled in
Data Science Masters course by PW Skills and I already finised Python ,
Statistics and Machine Learning as well . It was great experience with the pw
team and their course is amazing.
```

[57]:
```
df4
```

[57]:
```
                    Desc
0  Hey ! This is vIrat Tiwari Aspiring Data Scientist . Currently I enrolled in
Data Science Masters course by PW Skills and I already finised Python ,
Statistics and Machine Learning as well . It was great experience with the pw
team and their course is amazing.
```

[58]:
```
pd.set_option("display.max_colwidth",1000)

df4=pd.DataFrame({"Desc":["Hey ! This is vIrat Tiwari Aspiring Data Scientist .␣
 ↪Currently I enrolled in Data Science Masters course by PW Skills and I␣
 ↪already finised Python , Statistics and Machine Learning as well . It was␣
 ↪great experience with the pw team and their course is amazing.","I hope you␣
 ↪all cover the same concepts and learn many more technologies that helpful in␣
 ↪data scinece"]})
```

[59]:
```
df4
```

[59]:
```
                    Desc
0  Hey ! This is vIrat Tiwari Aspiring Data Scientist . Currently I enrolled in
Data Science Masters course by PW Skills and I already finised Python ,
Statistics and Machine Learning as well . It was great experience with the pw
team and their course is amazing.
1
I hope you all cover the same concepts and learn many more technologies that
helpful in data scinece
```

[60]:
```
# This is how we get lenth of the characters

df4["len"]=df4["Desc"].apply(len)
```

[61]:
```
df4
```

[61]:
```
                    Desc  \
0  Hey ! This is vIrat Tiwari Aspiring Data Scientist . Currently I enrolled in
Data Science Masters course by PW Skills and I already finised Python ,
Statistics and Machine Learning as well . It was great experience with the pw
team and their course is amazing.
1
I hope you all cover the same concepts and learn many more technologies that
helpful in data scinece

   len
0  260
1  100
```

[62]:
```
t="This is virat tiwari"
len(t.split())
```

[62]:
```
4
```

[63]:
```
df
```

[63]:
```
   A  B   Name
0  1  5  Virat
1  2  6   Yash
2  3  4  Rohit
3  4  9  Parth
```

[64]:
```
df["A"][0]
```

[64]:
```
1
```

[65]:
```
df
```

[65]:
```
   A  B   Name
0  1  5  Virat
1  2  6   Yash
2  3  4  Rohit
3  4  9  Parth
```

1 STATISTICAL FUNCTIONS

[66]:
```
df
```

[66]:
```
   A  B   Name
0  1  5  Virat
1  2  6   Yash
2  3  4  Rohit
3  4  9  Parth
```

[67]:
```
df["A"].mean()
```

[67]:
```
2.5
```

[68]:
```
df["B"].median()
```

[68]:
```
5.5
```

[69]:
```
df["Name"].mode()
```

[69]:
```
0    Parth
1    Rohit
2    Virat
3     Yash
Name: Name, dtype: object
```

[70]:
```
df["A"].mode()
```

[70]:
```
0    1
1    2
2    3
3    4
Name: A, dtype: int64
```

[71]:
```
df["B"].mode()
```

[71]:
```
0    4
1    5
2    6
3    9
Name: B, dtype: int64
```

[72]:
```
df["A"].std()
```

[72]:
```
1.2909944487358056
```

[73]:
```
df["A"].sum()
```

[73]:
```
10
```

[74]:
```
df["B"].sum()
```

[74]:
```
24
```

[75]:
```
df["A"].min()
```

[75]:
```
1
```

[76]:
```
df["B"].min()
```

[76]:
```
4
```

[77]:
```
df["A"].max()
```

[77]:
```
4
```

[78]:
```
df["B"].max()
```

[78]:
```
9
```

[79]:
```
df["A"].var()
```

[79]:
```
1.6666666666666667
```

[80]:
```
df["B"].var()
```

[80]:
```
4.666666666666667
```

[81]:
```
df5=pd.DataFrame({"a":[1,2,3,4,5,6,7,8,9]})
```

[82]:
```
df5
```

[82]:

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

[83]:
```
df5["a"].rolling(window=1).mean()
```

[83]:
```
0    1.0
1    2.0
2    3.0
3    4.0
4    5.0
5    6.0
6    7.0
7    8.0
8    9.0
Name: a, dtype: float64
```

[84]:
```
df5["a"].rolling(window=2).mean()
```

[84]:
```
0    NaN
1    1.5
2    2.5
3    3.5
4    4.5
5    5.5
6    6.5
7    7.5
8    8.5
Name: a, dtype: float64
```

[85]:
```
df5["a"].rolling(window=3).mean()
```

[85]:
```
0    NaN
1    NaN
2    2.0
3    3.0
4    4.0
5    5.0
6    6.0
7    7.0
8    8.0
Name: a, dtype: float64
```

[86]:
```
# Date Functionality
```

[87]:
```
import pandas as pd

datee=pd.date_range(start="2023-04-23" , end="2022-06-23")
```

[88]:
```
datee
```

[88]:
```
DatetimeIndex([], dtype='datetime64[ns]', freq='D')
```

[89]:
```
df_datee=pd.DataFrame({"datee":datee})
```

[90]:
```
df_datee
```

[90]:
```
Empty DataFrame
Columns: [datee]
Index: []
```

[94]:
```
pd.Timedelta(days=1,hours=5,minutes=45)
```

[94]:
```
Timedelta('1 days 05:45:00')
```

[95]:
```
dt=pd.to_datetime("2023-05-11")
```

[97]:
```
td=pd.Timedelta(days=1)
```

[98]:
```
dt+td
```

[98]:
```
Timestamp('2023-05-12 00:00:00')
```

2 Categorical Data

[99]:
```
data=["Virat","Yash","Rohit","Parth"]
```

[101]:
```
cat=pd.Categorical(data)
```

[102]:
```
cat
```

[102]:
```
['Virat', 'Yash', 'Rohit', 'Parth']
Categories (4, object): ['Parth', 'Rohit', 'Virat', 'Yash']
```

[103]:
```
cat.value_counts()
```

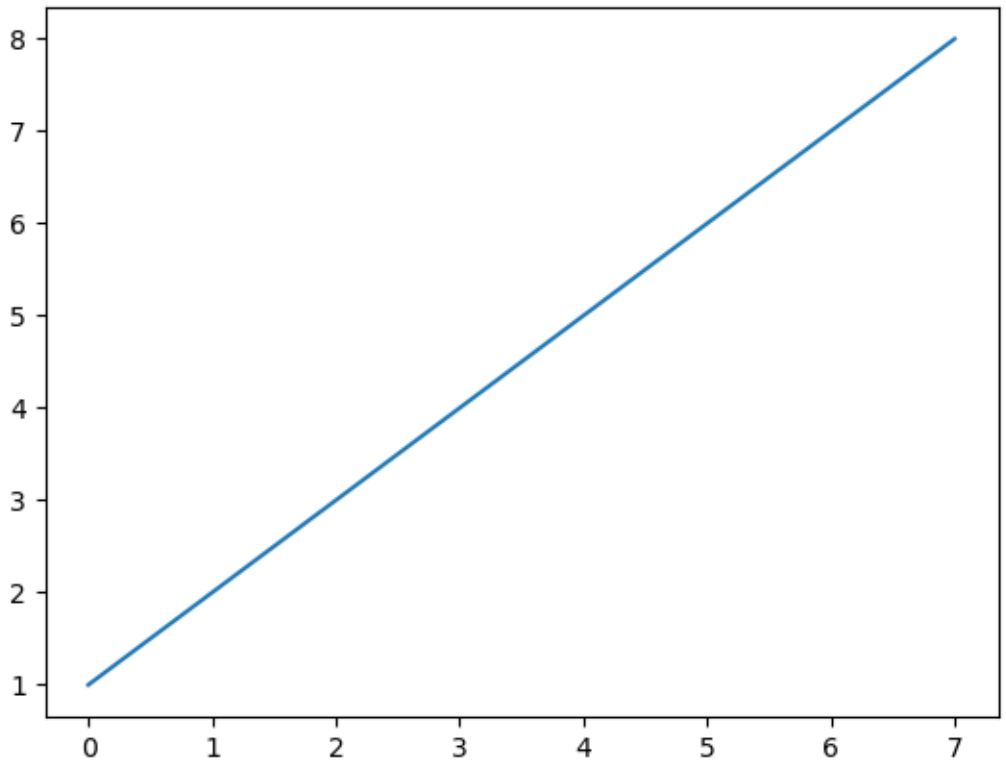
[103]:
```
Parth    1
Rohit    1
Virat    1
Yash     1
dtype: int64
```

3 Data Visualization in Pandas

[107]:
```
d=pd.Series([1,2,3,4,5,6,7,8])
```

[109]:
```
d.plot()
d
```

[109]:
```
0    1
1    2
2    3
3    4
4    5
5    6
6    7
7    8
dtype: int64
```



[110]:
```
df=pd.DataFrame({"a":[1,2,3,4,5],
                 "b":[6,7,8,9,7]})
```

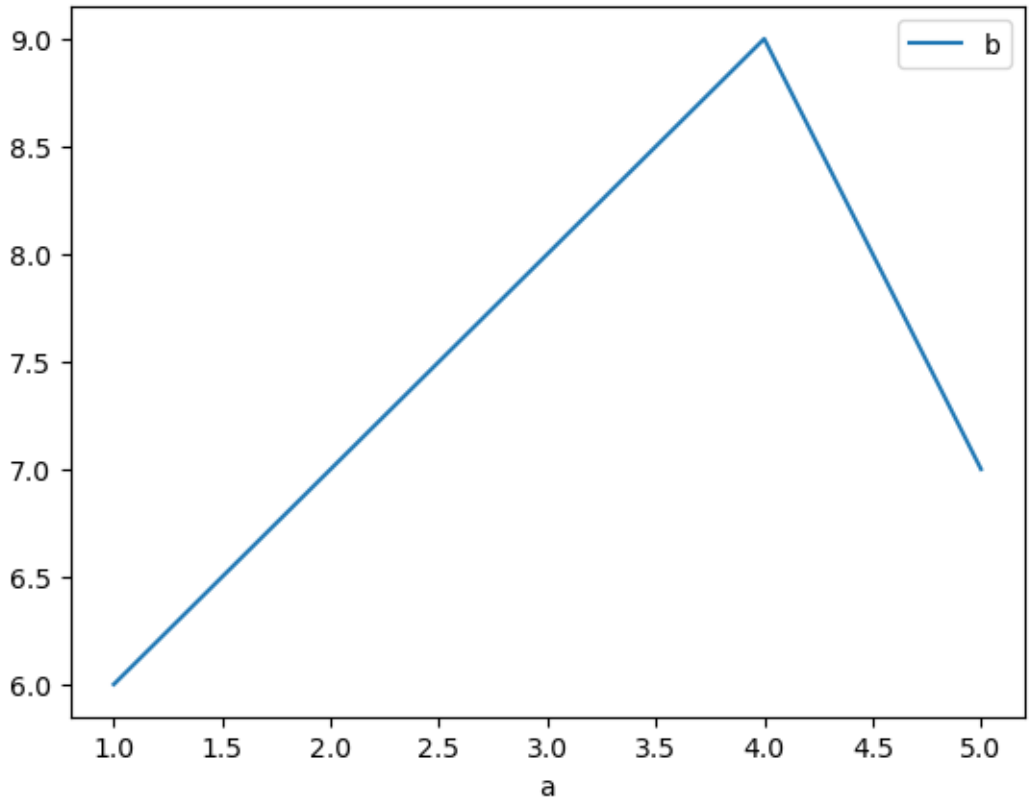
[111]:
```
df
```

[111]:

	a	b
0	1	6
1	2	7
2	3	8
3	4	9
4	5	7

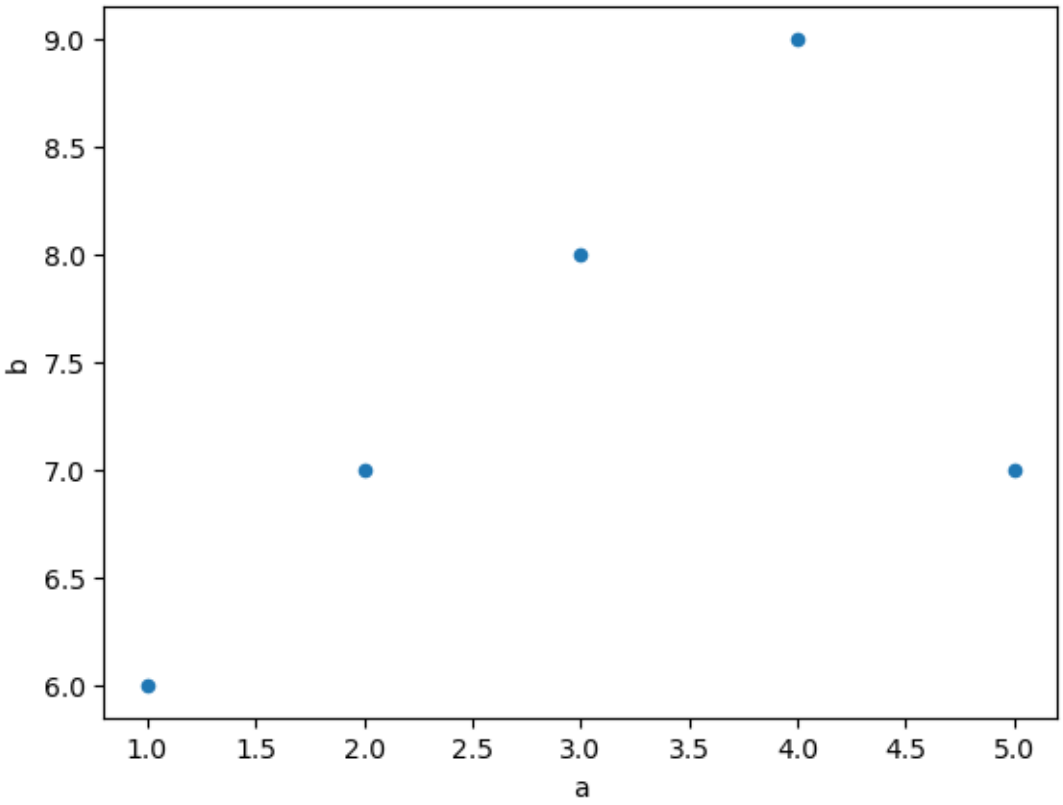
[113]:
```
df.plot(x="a",y="b")
```

[113]: <AxesSubplot: xlabel='a'>



[114]:
```
df.plot.scatter(x="a",y="b")
```

[114]: <AxesSubplot: xlabel='a', ylabel='b'>



[117]:
```
df
```

[117]:

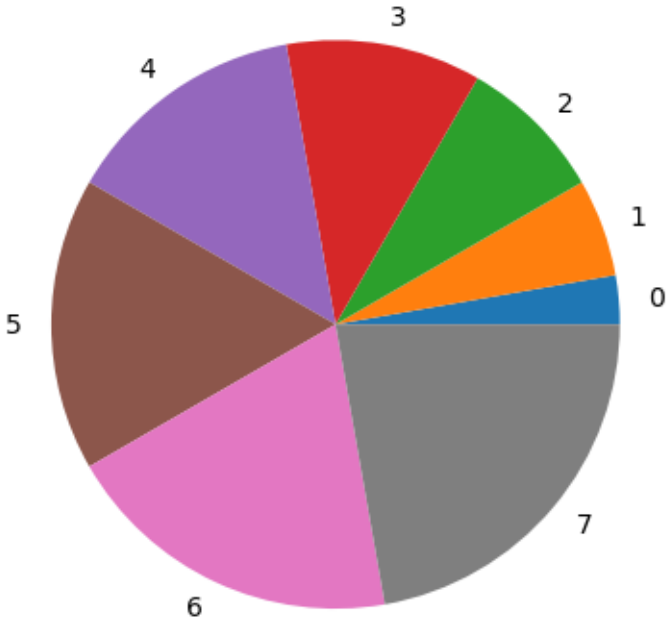
	a	b
0	1	6
1	2	7
2	3	8
3	4	9
4	5	7

[118]:
```
# Graph is used for finding the TREND of Data
```

[119]:
```
d=pd.Series([1,2,3,4,5,6,7,8])
```

[120]:
```
d.plot.pie()
```

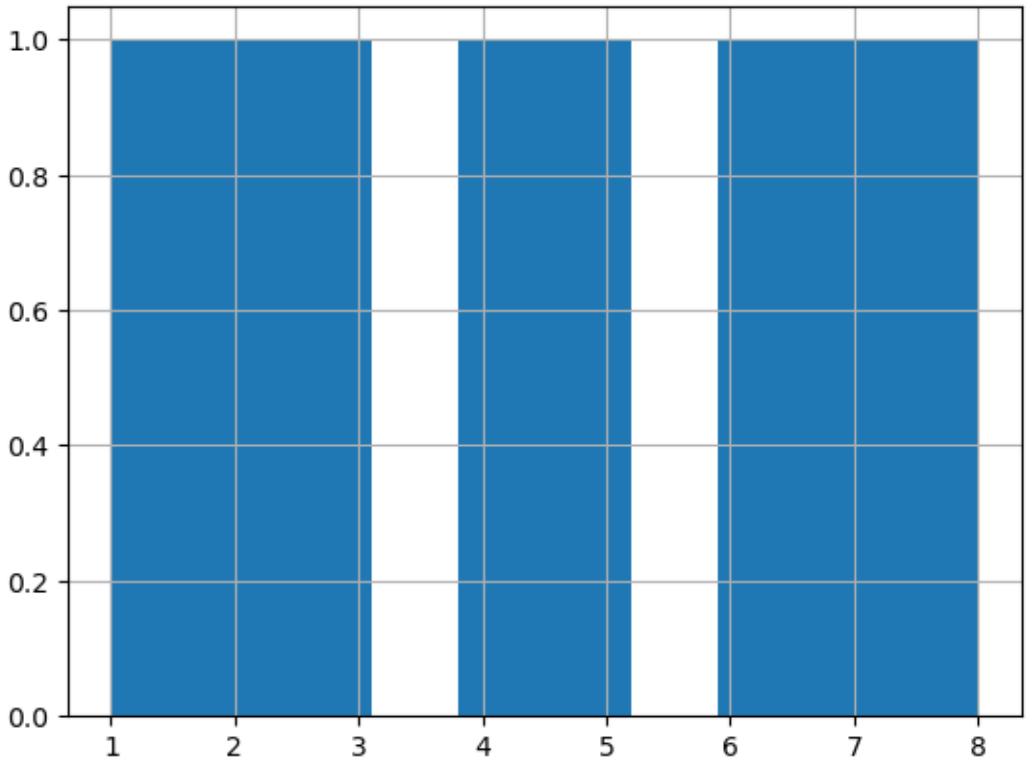
[120]:
```
<AxesSubplot: >
```



[121]:
```
d=pd.Series([1,2,3,4,5,6,7,8])
```

[122]:
```
d.hist()
```

[122]: <AxesSubplot: >



4 THANK YOU SO MUCH !!

5 YOURS VIRAT TIWARI :)