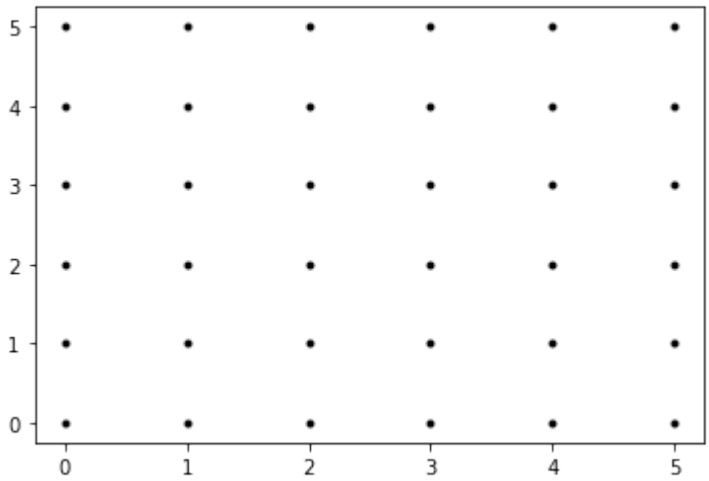
In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [4]:
```
xvalues = list(range(0,6,1))
yvalues = list(range(0,6,1))

x, y = np.meshgrid(xvalues, yvalues)

plt.plot(x,y, marker='.', color='k', linestyle='none')
```

Out[4]:
```
[<matplotlib.lines.Line2D at 0x2ddad5e4220>,
 <matplotlib.lines.Line2D at 0x2ddad5e4250>,
 <matplotlib.lines.Line2D at 0x2ddad5e42b0>,
 <matplotlib.lines.Line2D at 0x2ddad5e4370>,
 <matplotlib.lines.Line2D at 0x2ddad5e4430>,
 <matplotlib.lines.Line2D at 0x2ddad5e44f0>]
```



In [29]:
```
# 학번 생성 (1 ~ 100000)
student_ids = np.arange(1, 1000001)

# 점수 생성 (0 ~ 100 사이의 랜덤 값)
scores = np.random.uniform(0, 101, size=1000000)

# 데이터프레임 생성
df = pd.DataFrame({'학번': student_ids, '점수': scores})
```

In [30]:
```
df.to_csv('export_dataframe.csv', index=False, header=True, encoding='utf-8')
```

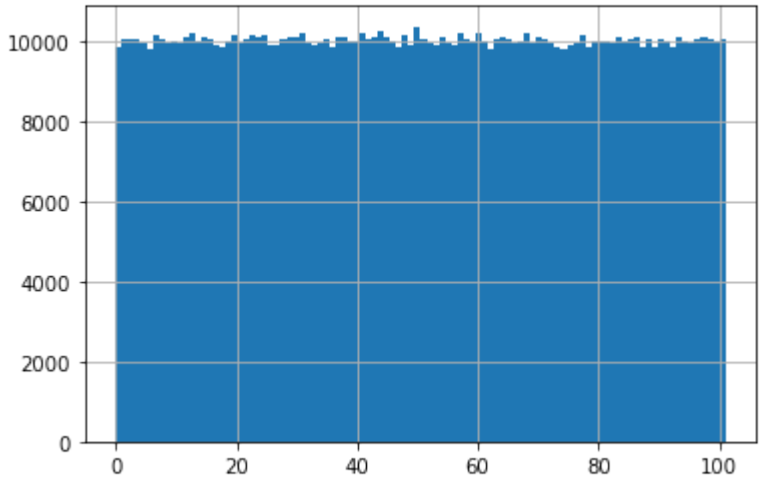
In [31]:
```
df.head()
```

Out[31]:

	학번	점수
0	1	70.054444
1	2	61.750503
2	3	57.876944
3	4	36.554949
4	5	62.517230

In [32]:
```
df['점수'].hist(bins=100)
```

Out[32]: <AxesSubplot:>



In [33]:
```
df2 = pd.read_csv("export_dataframe.csv")
```

In [34]:
```
df3 = df2.sort_values(by='점수', ascending=False)
df3
```

Out[34]:

	학번	점수
496382	496383	100.999997
197242	197243	100.999960
98356	98357	100.999952
323834	323835	100.999906
639625	639626	100.999869
...	...	...
634657	634658	0.000452
32016	32017	0.000444
511308	511309	0.000444
11949	11950	0.000134
136574	136575	0.000012

1000000 rows × 2 columns

In [37]:
```
print("평균 = ", df3['점수'].mean())
print("분산 = ", df3['점수'].var())
```

```
평균 =  50.4795261613508
분산 =  849.0094843848324
```

In [38]:
```
top = int(1000000 * 0.05)
bottom = int(1000000 * 0.95)

df4 = df3.iloc[top:bottom]
```

```
print("평균 = ", df4['점수'].mean())
print("분산 = ", df4['점수'].var())
```

```
평균 =  50.47635059110327
분산 =  687.4190374756876
```

In [45]:

```
def getGrade(x,gradeList):
    if(x['점수'] >= gradeList[0]):
        return "1";
    elif(x['점수'] >= gradeList[1]):
        return "2";
    elif(x['점수'] >= gradeList[2]):
        return "3";
    elif(x['점수'] >= gradeList[3]):
        return "4";
    elif(x['점수']>= gradeList[4]):
        return "5";

first_cut = df3['점수'].iloc[int(1000000*0.05)]
second_cut = df3['점수'].iloc[int(1000000*0.13)]
third_cut = df3['점수'].iloc[int(1000000*0.24)]
forth_cut = df3['점수'].iloc[int(1000000*0.72)]
fifth_cut = df3['점수'].iloc[999999]

gradecut_list = np.array([first_cut,second_cut,third_cut,forth_cut,fifth_cut])
print(gradecut_list)
print(df3.iloc[0])
df3['등급'] = df3.apply(lambda x: getGrade(x,gradecut_list),axis='columns')
print(df3)
```

```
[9.59639850e+01 8.78504647e+01 7.67253388e+01 2.82933437e+01
 1.18326300e-05]
학번         496383
점수     100.999997
등급              1
Name: 496382, dtype: object
            학번          점수 등급
496382  496383  100.999997  1
197242  197243  100.999960  1
98356    98357  100.999952  1
323834  323835  100.999906  1
639625  639626  100.999869  1
...        ...         ... ..
634657  634658    0.000452  5
32016    32017    0.000444  5
511308  511309    0.000444  5
11949    11950    0.000134  5
136574  136575    0.000012  5

[1000000 rows x 3 columns]
```

In [47]:

```
df5 = df3.groupby('등급')
print("₩nmean₩n",df5['점수'].mean().sort_values(ascending=False))
print("₩nvar₩n",df5['점수'].var().sort_values(ascending=False))
print("₩nstd₩n",df5['점수'].std().sort_values(ascending=False))
print("₩nstd₩n",df5['점수'].size().sort_values(ascending=False))
df5['점수'].size().sort_values(0,ascending=True).plot()
plt.xlabel('grade')
```

```
mean
 등급
1    98.478915
2    91.906222
3    82.286565
```

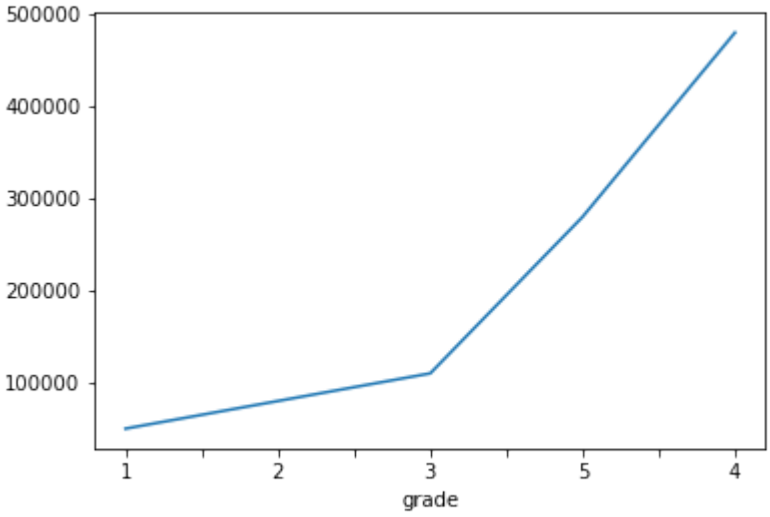
```
4    52.473166
5    14.158417
Name: 점수, dtype: float64

var
 등급
4    195.032780
5     66.597046
3     10.306307
2      5.489875
1      2.110987
Name: 점수, dtype: float64

std
 등급
4    13.965414
5     8.160701
3     3.210344
2     2.343048
1     1.452924
Name: 점수, dtype: float64

std
 등급
4    480000
5    279999
3    110000
2     80000
1     50001
Name: 점수, dtype: int64
```

Out[47]: Text(0.5, 0, 'grade')



In [ ]: