

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [3]:

```
df=pd.read_csv('E:\\headbrain.csv')
df
```

Out[3]:

	Gender	Age Range	Head Size(cm^3)	Brain Weight(grams)
0	1	1	4512	1530
1	1	1	3738	1297
2	1	1	4261	1335
3	1	1	3777	1282
4	1	1	4177	1590
...	...	...	...	...
232	2	2	3214	1110
233	2	2	3394	1215
234	2	2	3233	1104
235	2	2	3352	1170
236	2	2	3391	1120

237 rows × 4 columns

In [5]:

```
df.isnull().sum()
```

Out[5]:

```
Gender          0
Age Range       0
Head Size(cm^3) 0
Brain Weight(grams) 0
dtype: int64
```

In [7]:

```
df.shape
```

Out[7]:

```
(237, 4)
```

In [10]:

```
df.head()
```

Out[10]:

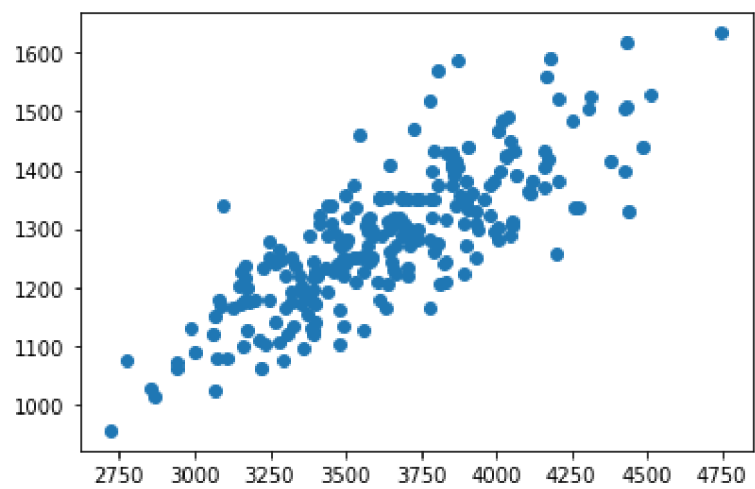
	Gender	Age Range	Head Size(cm^3)	Brain Weight(grams)
0	1	1	4512	1530
1	1	1	3738	1297
2	1	1	4261	1335
3	1	1	3777	1282
4	1	1	4177	1590

In [13]:

```
plt.scatter(df['Head Size(cm^3)'],df['Brain Weight(grams)'])
```

Out[13]:

<matplotlib.collections.PathCollection at 0x1736a30>

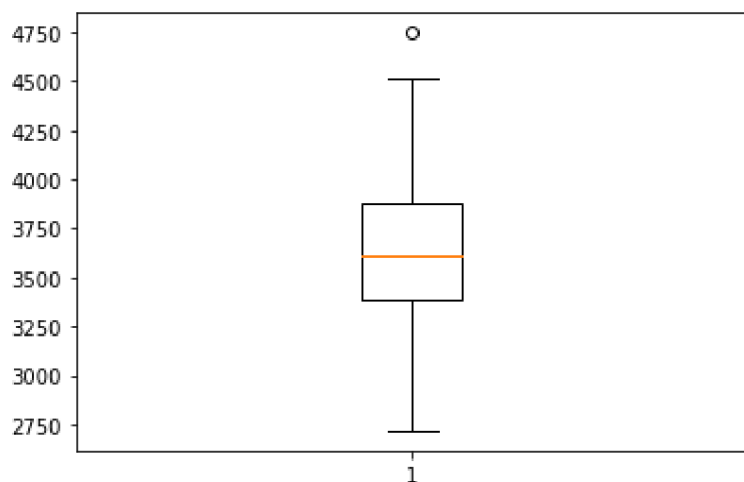


In [15]:

```
plt.boxplot(df['Head Size(cm^3)'])
```

Out[15]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x55a7aa8>,  
<matplotlib.lines.Line2D at 0x55a7c58>],  
'caps': [<matplotlib.lines.Line2D at 0x55a7e08>,  
<matplotlib.lines.Line2D at 0x55a7fb8>],  
'boxes': [<matplotlib.lines.Line2D at 0x55a78f8>],  
'medians': [<matplotlib.lines.Line2D at 0x55b7190>],  
'fliers': [<matplotlib.lines.Line2D at 0x55b7340>],  
'means': []}
```



In [25]:

```
sum_x=0  
for i in df['Head Size(cm^3)']:  
    sum_x=sum_x+df['Head Size(cm^3)']  
print(sum_x)
```

```
0      1069344  
1       885906  
2      1009857  
3       895149  
4       989949
```

...

```
232     761718  
233     804378  
234     766221  
235     794424  
236     803667
```

Name: Head Size(cm^3), Length: 237, dtype: int64

In [73]:

```
x=df['Head Size(cm^3)']
y=df['Brain Weight(grams)']
n=np.size(x)
x_mean=x.mean()
y_mean=y.mean()
m=(np.sum(x*y)- n*x_mean*y_mean)/(np.sum(x*x)-n*x_mean*x_mean)
c=y_mean-m*x_mean
y_pred=m*x+c
```

Out[73]:

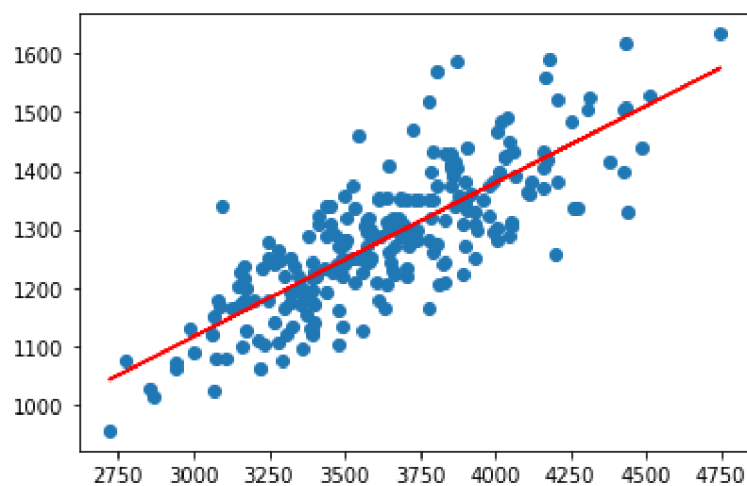
```
0      1514.166601
1      1310.272292
2      1448.045837
3      1320.546036
4      1425.917772
...
232     1172.235318
233     1219.652599
234     1177.240476
235     1208.588567
236     1218.862311
Name: Head Size(cm^3), Length: 237, dtype: float64
```

In [75]:

```
plt.scatter(x,y)
plt.plot(x,y_pred,color='red')
```

Out[75]:

[<matplotlib.lines.Line2D at 0xa6cc838>]



In [77]:

```
error = y - y_pred
se = np.sum(error**2)
print('squared error is', se)

mse = se/n
print('mean squared error is', mse)
```

```
squared error is 1232728.014636552
mean squared error is 5201.384028002329
```