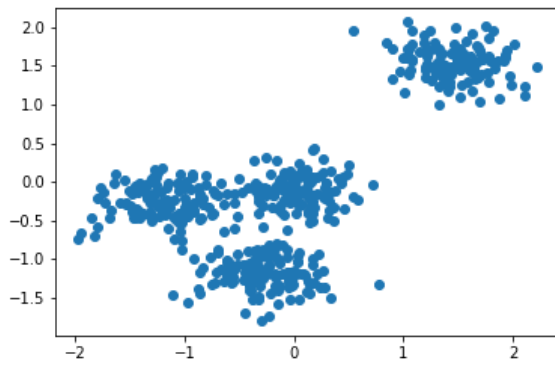


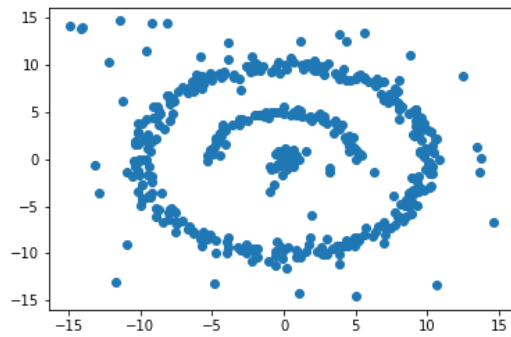
یک.

(الف)

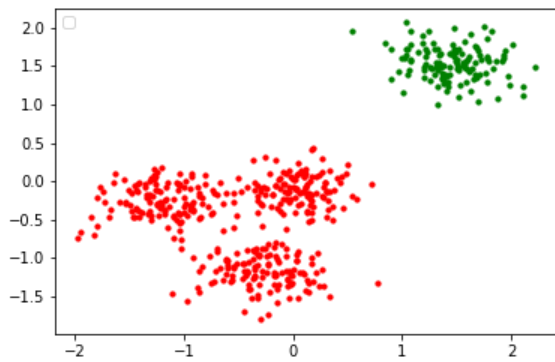
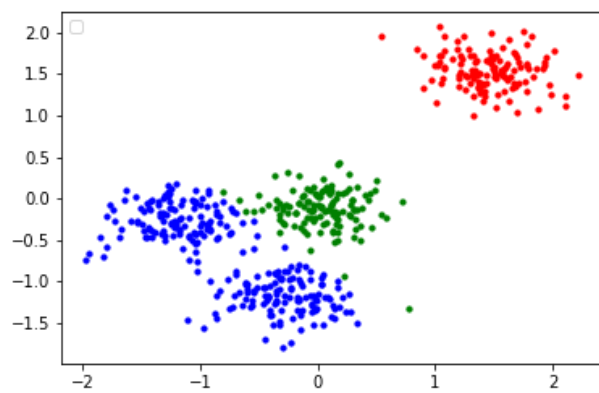
دیتاست ۱:

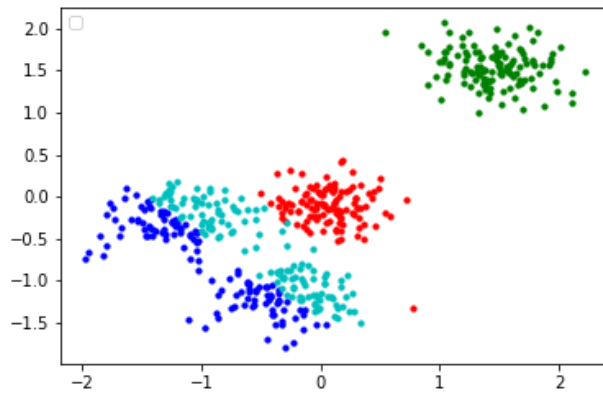


دیتاست ۲:



(ب)

برای $k = 2$:برای $k = 3$:

برای $k=4$:

(پ)

برای $k=2$ خطاهای خوشه:

```
[0.3185998331070036, 0.7493196570175052]
```

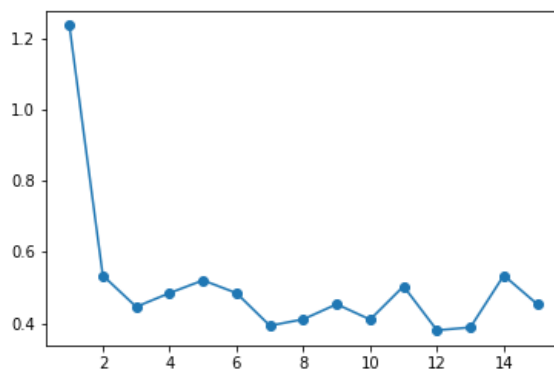
برای $k=3$ خطاهای خوشه:

```
[0.6991856737562374, 0.3185998331070036, 0.3200719062826248]
```

برای $k=4$ خطاهای خوشه:

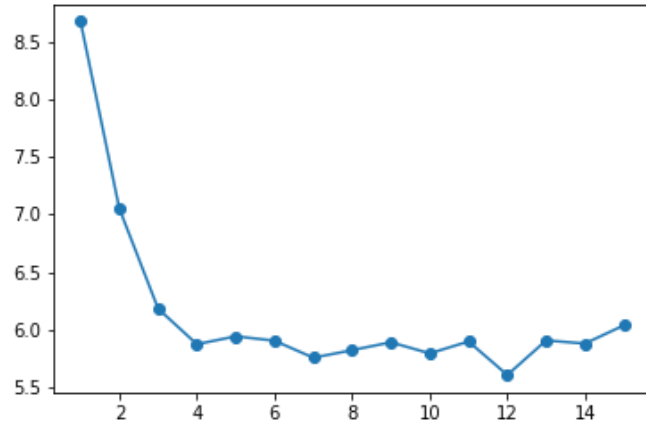
```
[0.2988268220910847,
0.3185998331070036,
0.6656815591710529,
0.6558541721601958]
```

(ت)

برای $k=2$ خطای خوشه بندی: 0.5339597450622544برای $k=3$ خطای خوشه بندی: 0.44595247104862196برای $k=4$ خطای خوشه بندی: 0.4847405966323342(ث) نمودار خطاهای خوشه‌بندی برای k های مختلف:

ج) از روی شکل می‌توان فهمید که $k=12$ کم‌ترین مقدار خطا را دارد و بهینه‌ترین مقدار است.

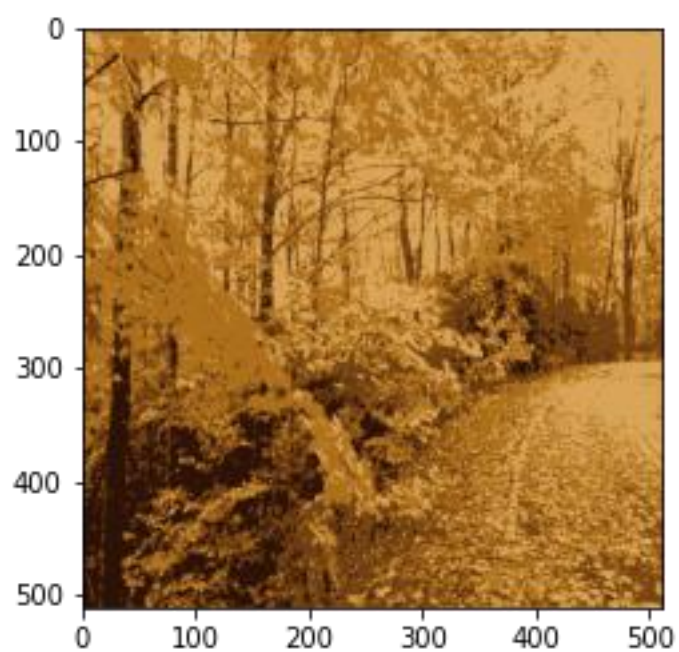
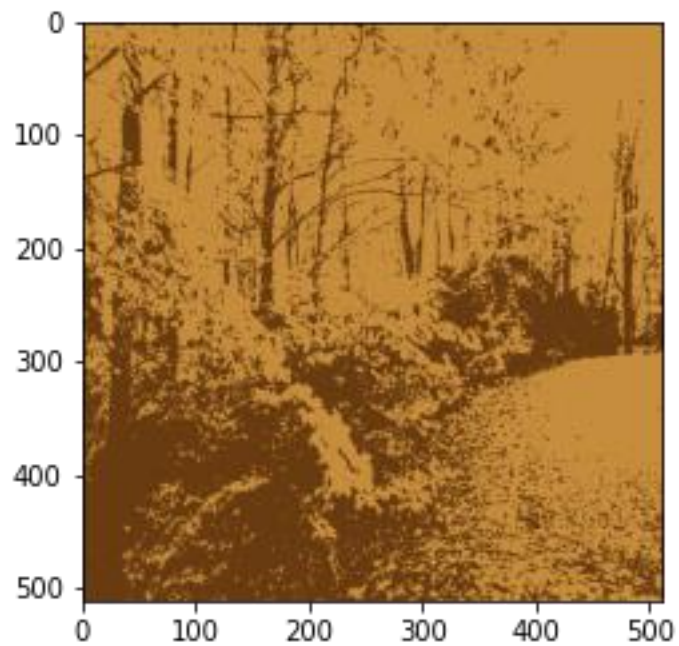
چ) نکته‌ی جالب این است که باز بهترین مقدار ۱۲ خواهد بود!

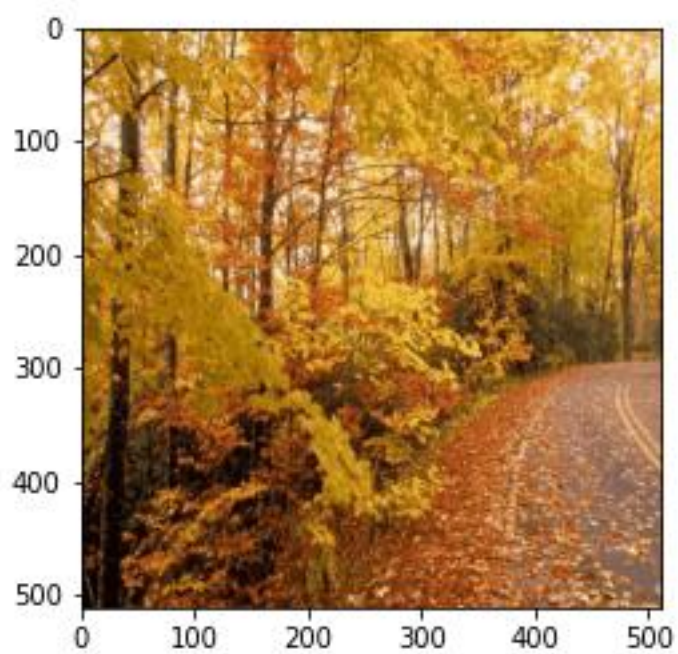
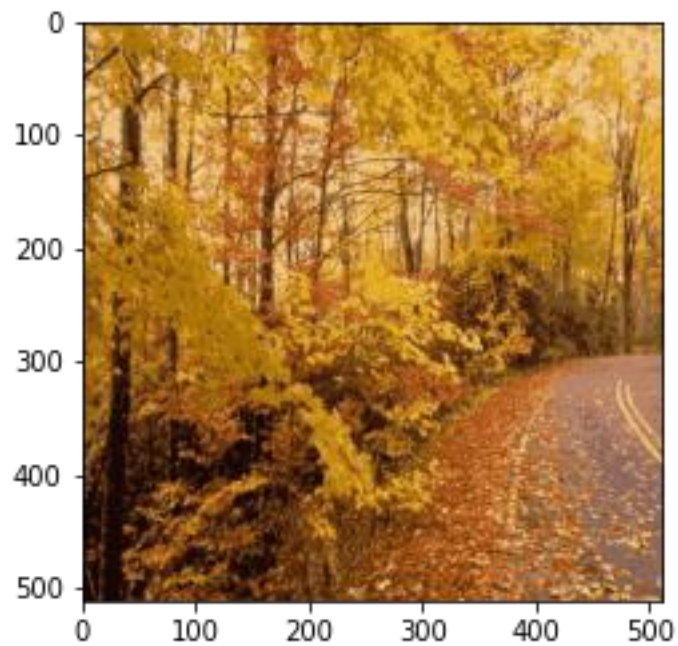


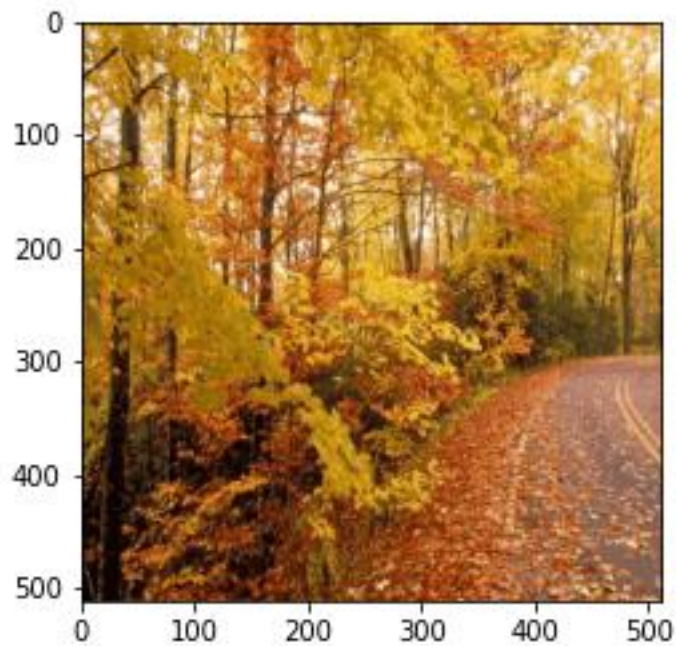
دلیل خوب نبودن این روش (ارور ۱۵) این است که این دیتا در فضای ۲ بعدی مرزهای مشخصی ندارد و باید با متدهایی همچون SVM آن را به فضاهای با بعد بیش‌تر برد تا بتوان آن را درست دسته بندی کرد.

دو.

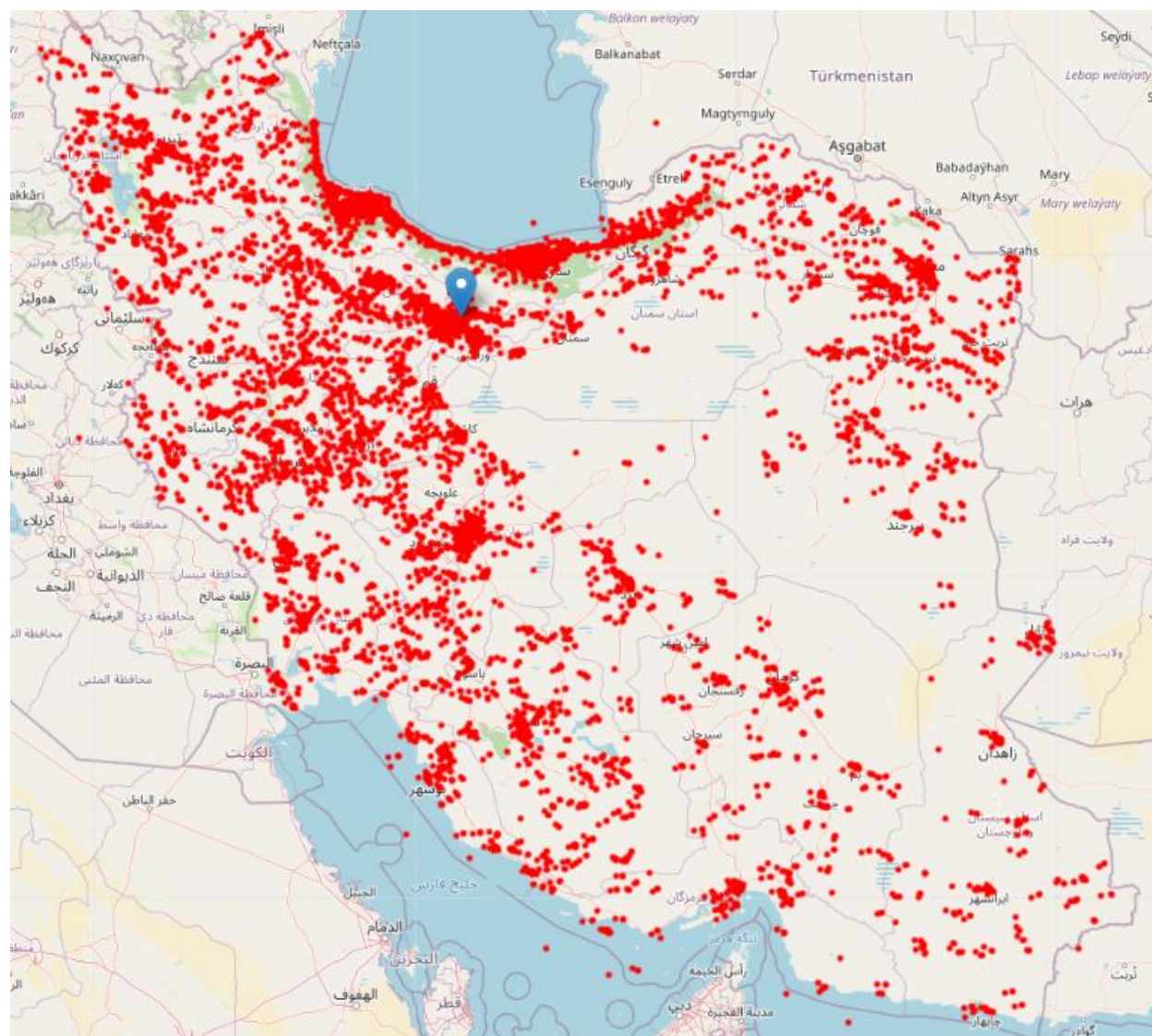
به ترتیب از بالا به پایین برای $k=2, k=4, k=16, k=32, k=64$







(الف)



(ب) جواب با پارامترهای $\text{eps}=0.5$, $\text{min_sample}=100$

```
Estimated number of clusters: 12
Estimated number of noise points: 2124
```

(پ) با پارامترهای $\text{eps}=0.3$, $\text{min_sample}=320$

