

سوال ۳)

در این سوال مشاهده می‌کنیم که با دیتاستی به شکل زیر رو به رو هستیم:

	39.37962	44.39011
0	39.33063	44.14934
1	39.20938	44.76798
2	39.27789	44.54155
3	39.29137	44.46958
4	39.28080	44.46657
...
18187	33.70264	59.26388
18188	32.59978	59.40304
18189	33.54605	56.94985
18190	34.18181	58.02116
18191	25.55712	57.12496

18192 rows × 2 columns

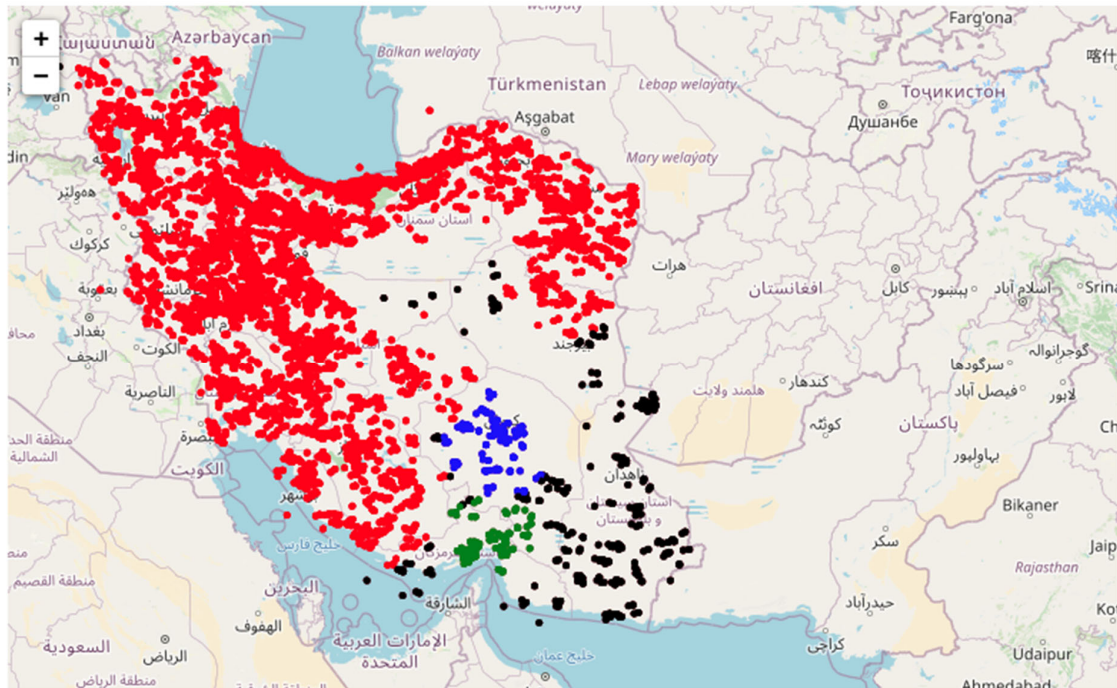
سپس اگر موقعیت مکانی بیماران را مصور کنیم، نقشه ایران به صورت زیر خواهد شد:



حالا اگر شعاع را برابر ۱ و حداقل تعداد نقاط را برابر ۱۵۰ در نظر بگیریم داریم:

```
start = calc_DBSCAN(1, 150)
for index, row in start.iterrows():
    color = ''
    if row[2] != -1:
        color = colors[int(row[2])]
    else:
        color = "black"
    folium.Circle(location=[row[0], row[1]], radius=1, color=color, fill=True).add_to(m)
```

Clusters: 0 -1 1 2

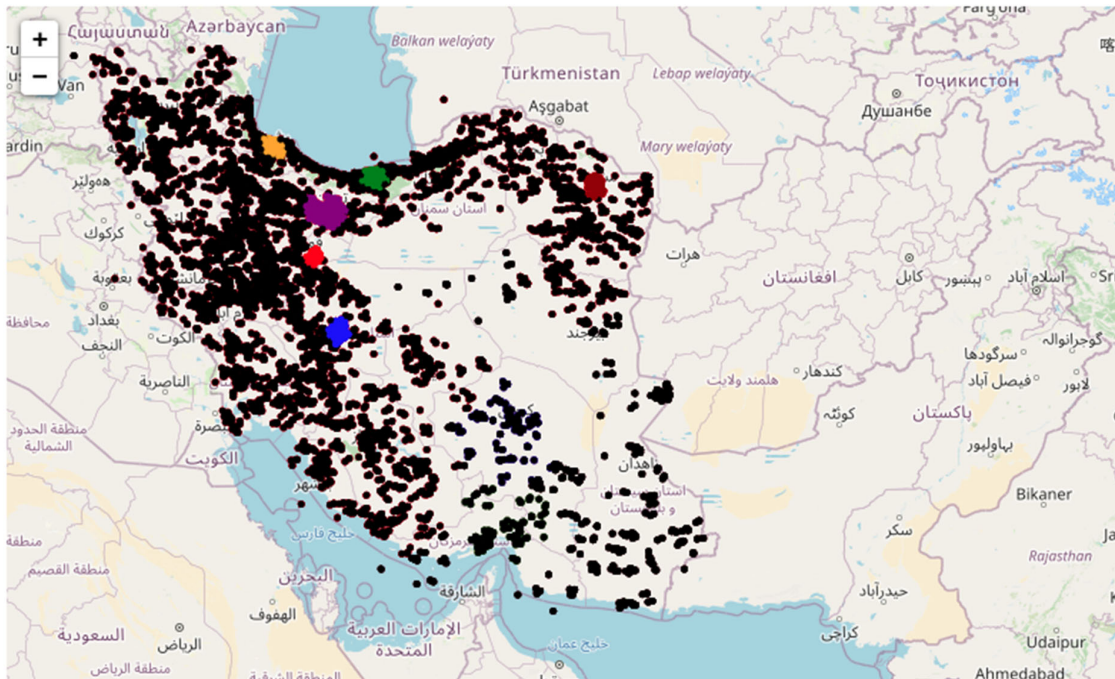


می‌بینیم که تعداد زیادی از شهرها را در یک خوشه قرار داده که مناسب نیست.

حالا اگر شعاع را ۰.۲ و حداقل نقاط را ۳۰۰ در نظر بگیریم، داریم:

```
final_test = calc_DBSCAN(0.2, 300)
for index, row in final_test.iterrows():
    color = ''
    if row[2] != -1:
        color = colors[int(row[2])]
    else:
        color = "black"
    folium.Circle(location=[row[0], row[1]], radius=1, color=color, fill=True).add_to(m)
```

Clusters: -1 4 0 1 2 3 5



و همان‌طور که قابل مشاهده است، شهرهای تهران، اصفهان، مشهد، ساری، رشت و قم به عنوان خوشه‌های ما که متراکم‌تر هستند، مشخص شده‌اند.

سوال ۴)

در این سوال هم ابتدا من یک زیر مجموعه‌ی ۴۰ تایی را استخراج کردم و با کمک PCA، ابعاد آن را کاهش دادم و به صورت ۲ بعدی در آورده ام و سپس مصور شده‌ی این دادگان به صورت زیر قابل مشاهده هستند:

