

گزارش تمرین دوم داده‌کاوی

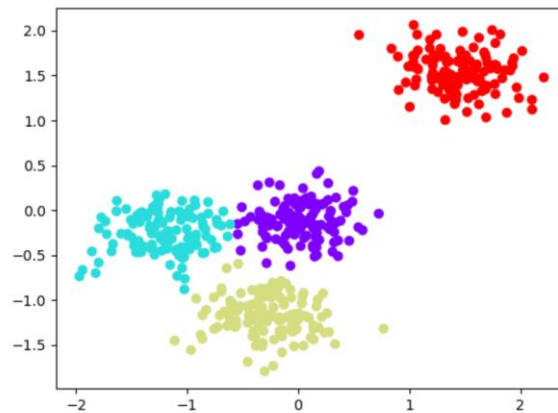
مهدیس صفری ۹۵۳۱۰۵۱

پیداهسازی اول

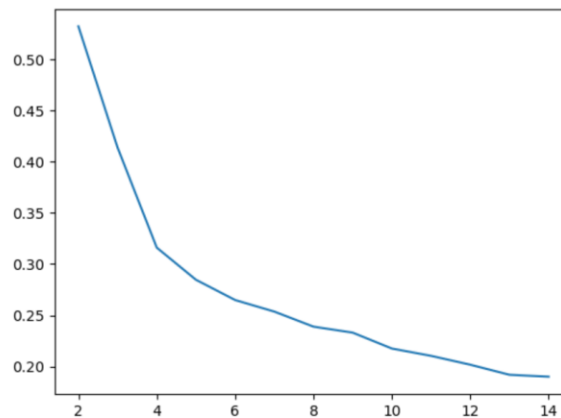
به ازای $k = 4$ و $\text{min_iteration} = 15$

for $k=4$ the average distance between the cluster center and the data points in that cluster :

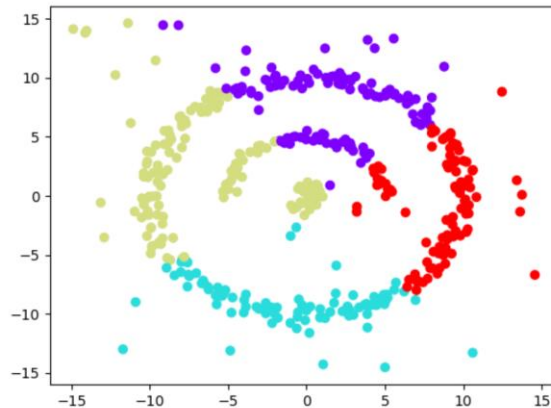
[۰/۳۱۶۰۰۸۸۱۹۵۷۷۳۴۷۴۴ , ۰/۳۳۶۱۲۳۳۲۰۹۴۰۰۴۵۰۵ , ۰/۳۲۳۸۳۶۸۹۴۰۵۴۱۰۰۹ , ۰/۲۸۸۶۵۹۷۳۹۷۵۹۵۵۶۴]



for $k=4$ cluster error : 0.34779997588214495



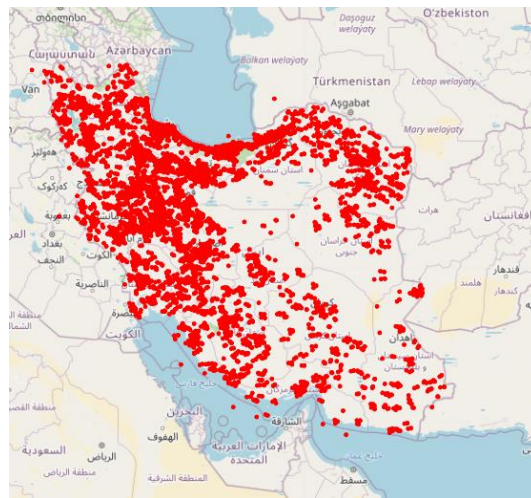
optimum $K = 4$



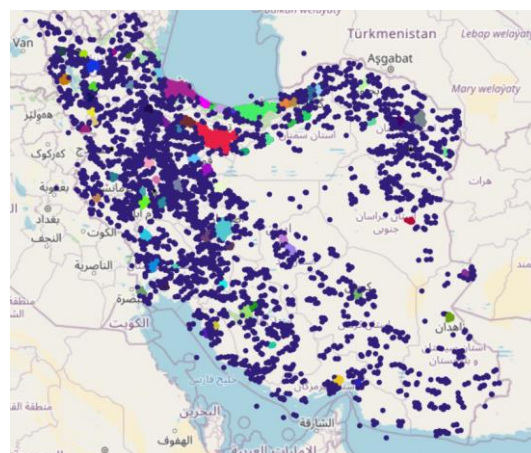
زیرا kmean به صورت خطی خوشه‌بندی می‌کند. برای خوشه‌بندی های مبتنی بر همسایگی و تراکم DBSCAN پیشنهاد می‌شود.

لازم به ذکر است بخش دوم و سوم پیاده سازی در [google collaboration](#) انجام شده اند.

پیاده‌سازی دوم

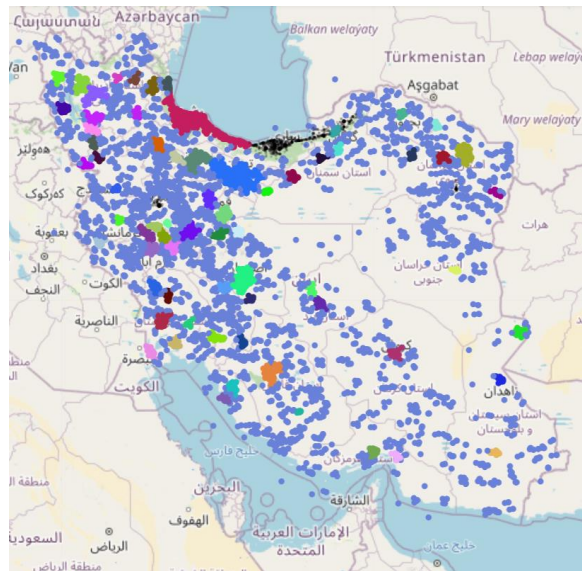


به ازای $\text{eps}=0.1$, $\text{min_samples}=30$:



بهترین مقادیر حدود : $\text{eps}=0.1-0.15$, $\text{min_samples}=30-40$

: $\text{eps}=0.15$, $\text{min_samples}=30$



پیاپی سازی سوم

سایز عکس به ازای $k=256$ دو برابر $k=16$ می باشد.

