

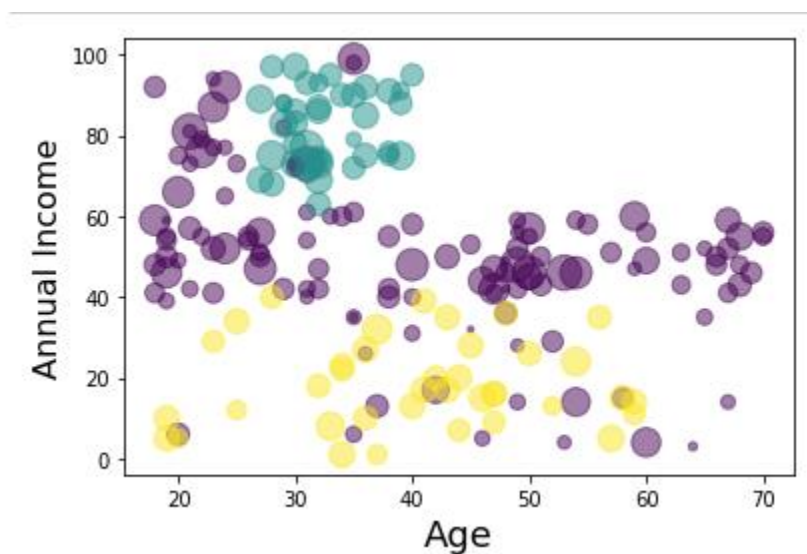
مقدمه

در این پروژه از دو الگوریتم k-means و DBSCAN برای خوشه بندی داده های مربوط به مشتریان یک فروشگاه استفاده شده است.

برای کوتاه شدن گزارش جزییات دیتاست و کدنویسی در گزارش ذکر نشده است.

الگوریتم K-means

داده ها برای خوشه بندی ابتدا استاندارد سازی شدند تا قابل استفاده برای الگوریتم ها باشند. پس از این کار یک بار با تعداد خوشه 3 با استفاده از الگوریتم k-means خوشه بندی شدند که شکل خروجی خوشه ها بصورت شکل می باشد.

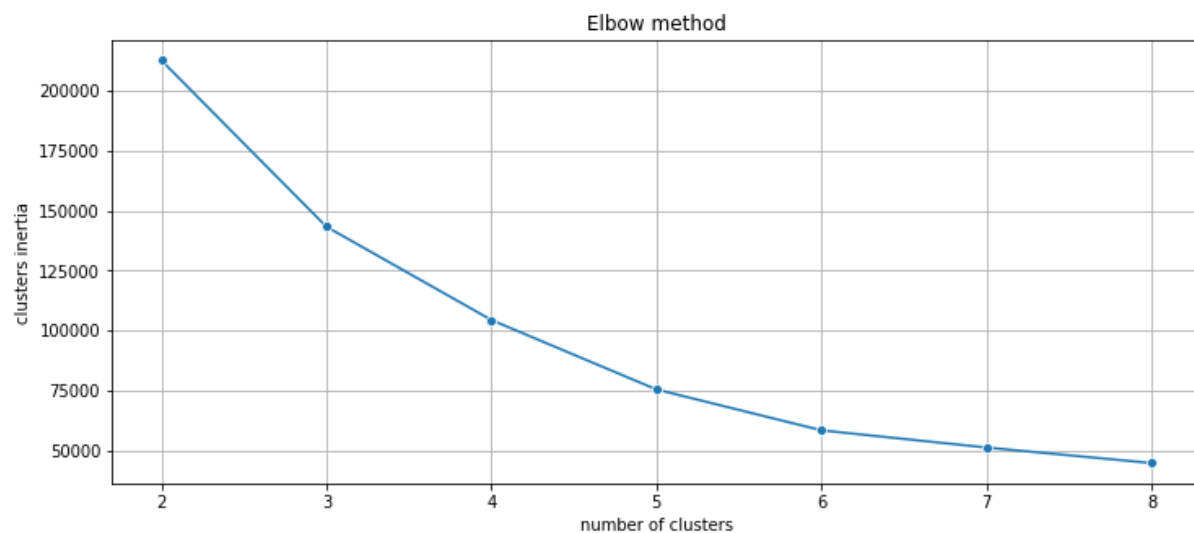


شکل 1

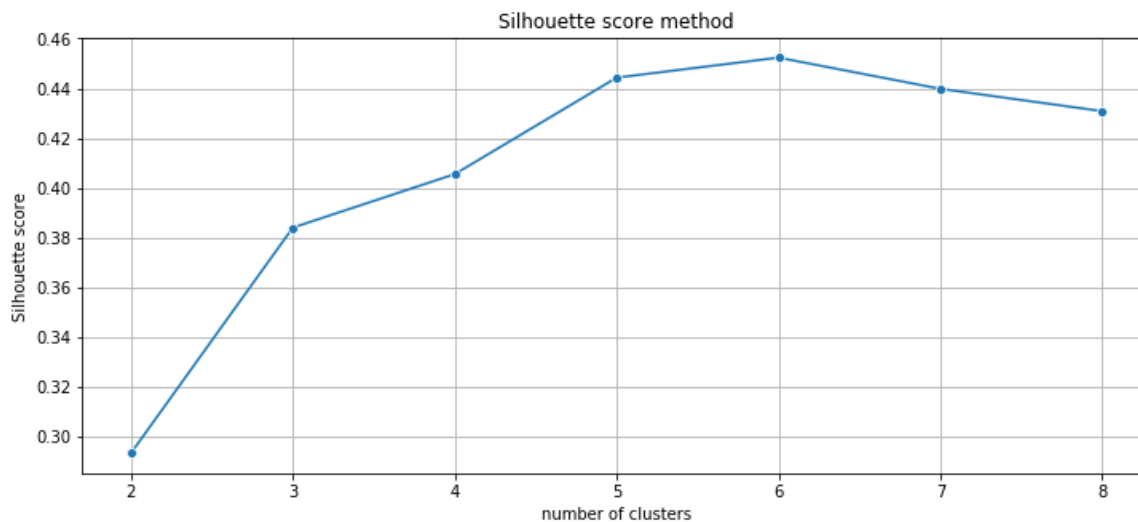
پس از این کار با استفاده از معیار silhouette و همچنین معیار Elbow برای بدست آوردن بهترین k برای این الگوریتم استفاده شد که مقادیر خروجی و نمودارها بهترین مقدار برای هریک از معیارها نشان می دادند که ما از معیار silhouette استفاده کرده و مقدار 6 را برای خوشه بندی جدید انتخاب کردیم.

```
silhouette: [0.293166070535953, 0.3839349967742105, 0.40546302077733304, 0.44428597560893024, 0.4523443947724053, 0.43978902692  
261157, 0.4307991110543737]  
Elbow: [212840.16982097185, 143342.751571706, 104366.15145556198, 75350.77917248776, 58300.44332159069, 51116.986944634315, 446  
37.39564017064]
```

شکل 2



شکل 3

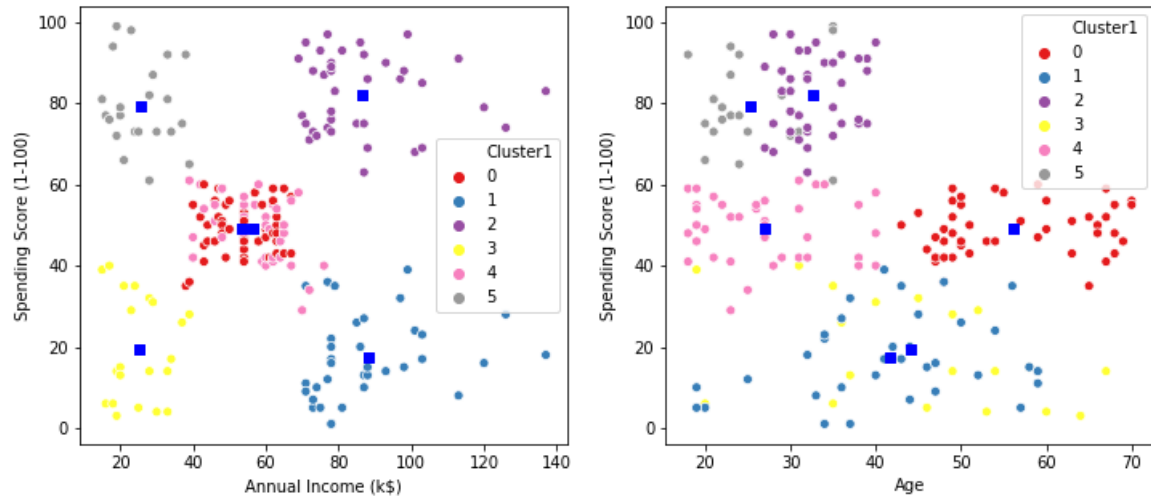


شکل 4

پس از انتخاب 6 به عنوان تعداد خوشه های الگوریتم، الگوریتم k-means یک بار دیگر با این تعداد خوشه انجام شد و با اضافه کردن یک ستون برای برچسب گذاری داده های خوشه بندی شده پراکندگی داده ها بصورت شکل 5 بدست آمد. همچنین پراکندگی داده های بصورت نمودارهای Age-Spending Score و Annual Income-Spending Score که در آنها خوشه ها نیز مشخص شده اند مانند شکل شماره 6 بدست آمد.

Cluster	
0	45
1	38
2	35
3	22
4	39
5	21

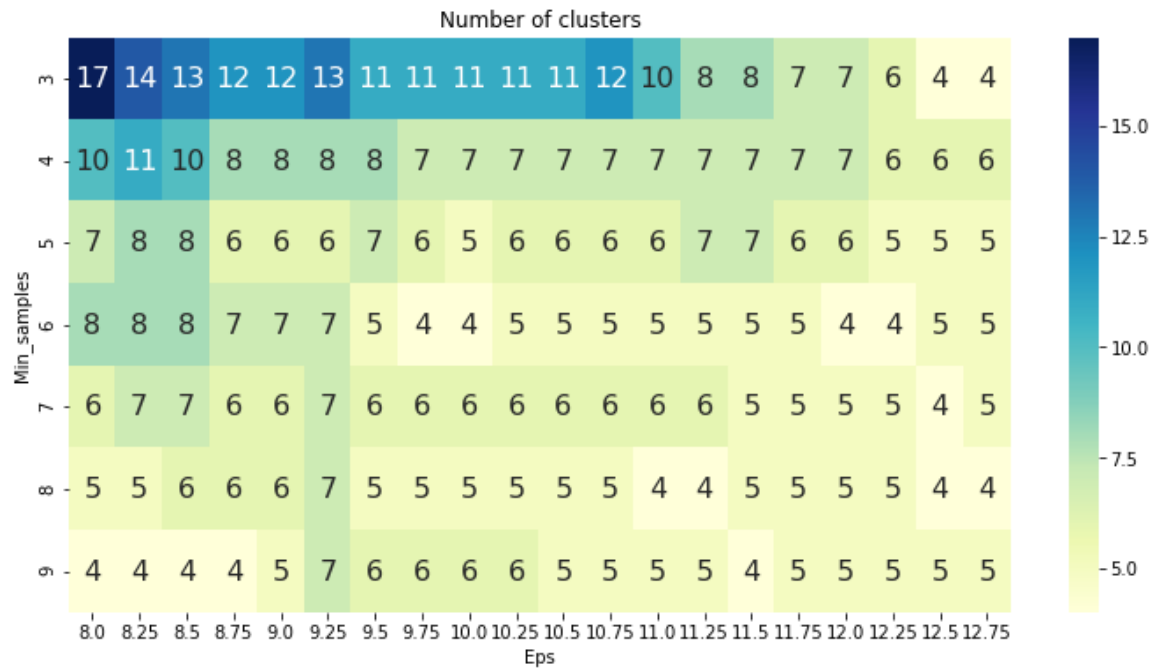
شکل 5



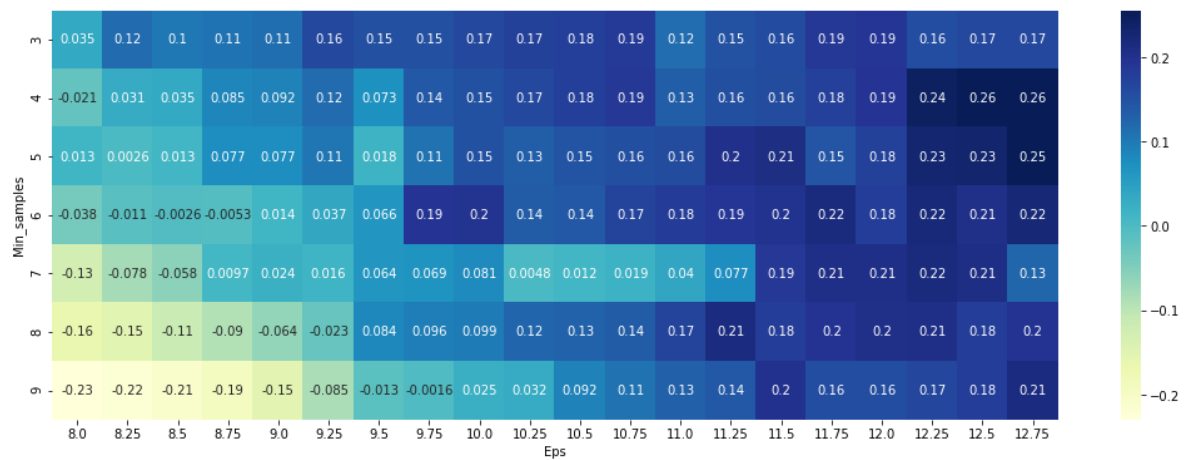
شکل 6

الگوریتم DBSCAN

برای الگوریتم DBSCAN در همان ابتدا یک آرایه از مقادیر $\min\text{-sup}$ و ϵ تعریف شد تا مقدار بهینه آنها بدست آید و با استفاده از آن خوشه بندی انجام گیرد. همانطور که در شکل 7 مشخص است برای تعریف شده ما می توانیم از 5 تا 17 خوشه در نظر بگیریم.



شکل 7



شکل 8

در شکل 8 مقادیر min-sup و eps با توجه به معیار silhouette رسم شده است. رنگ آبی پر رنگ نشان دهنده مقدار بزرگتری برای معیار مورد نظر است. که در اینجا مقدار 0.26 برای این معیار در eps=12.5 و min-sup=4 اتفاق می افتد. حال با توجه به شکل 7 تعداد خوشه ها با توجه به این مقادیر برابر 6 می باشد.

[0.03523183319491217, -0.02147416705080717, 0.01311730537912744, -0.037856968288658935, -0.12691137408529196, -0.16320109865956545, -0.22952472168827678, 0.11794025651864473, 0.03091945987798986, 0.0026120067709137795, -0.010657568508570952, -0.07843325081254296, -0.15076821608044708, -0.2193812749149287, 0.10218339749091918, 0.03525515433758843, 0.012783230221530086, -0.0026411846829178433, -0.05764928433699911, -0.11442367185842997, -0.20829335215639208, 0.10794176025451761, 0.08512060045848535, 0.07687636405239015, -0.005314265607171622, 0.009673519062197957, -0.08983717779622592, -0.19084530198917327, 0.11395369330062792, 0.09200586725651252, 0.07687636405239015, 0.01373675522866092, 0.023655146382314446, -0.06361003273116211, -0.1530431679636375, 0.16483355817918394, 0.11846615029723534, 0.10554053872016375, 0.0372788643854904, 0.015950137387842282, -0.023239937942889494, -0.08462004420998577, 0.15499294818931003, 0.07254893597582747, 0.018201407008572045, 0.06567255669809864, 0.06444348602118771, 0.08399598179227934, -0.0130309632955094, 0.15499294818931003, 0.14261641906699515, 0.11054539328361283, 0.19341987538058883, 0.06855506549725476, 0.0955975365497244, -0.001584748732399255, 0.16560157362154815, 0.15189248068906694, 0.15212635871472988, 0.19707521745228654, 0.0812837289733531, 0.09898520185653577, 0.02468558647609917, 0.17425323061017678, 0.16610511388044952, 0.1336309533528798, 0.1366332763483743, 0.004843955463492065, 0.12337401106706146, 0.03156783060937614, 0.17821547946937702, 0.178572631571689, 0.14780003400502686, 0.14193935975919467, 0.011774728028104718, 0.12879537285067205, 0.09189298227515615, 0.185851182732256, 0.18853553072957918, 0.15572208551296238, 0.17358818660372244, 0.01936798866335313, 0.13693358011171713, 0.11000267286520568, 0.11717691378269142, 0.1328534443680357, 0.1596939430611063, 0.1823977185500609, 0.04028490861311685, 0.1709530532622651, 0.13149700111467832, 0.15025800529738428, 0.1553787598552803, 0.2042311686234455, 0.19115951187565833, 0.07740817973401846, 0.21467738602928685, 0.1428008181012056, 0.1610286491755767, 0.15914281104161046, 0.21156364147287732, 0.1998732041485828, 0.1876826925144944, 0.18129578168387675, 0.2009501086531089, 0.1883306352794154, 0.17646363387231617, 0.145784128646343, 0.21690226569266902, 0.20675480670600443, 0.19581362491608117, 0.15824383382786478, 0.19188770917280448, 0.19385774986665416, 0.1759177615505368, 0.1806368944323813, 0.20675480670600443, 0.20491356452653534, 0.1586288701453676, 0.1574262862826218, 0.24462445365183694, 0.23170713241201668, 0.2229439727113055, 0.21690226569266902, 0.20675480670600443, 0.1681167265586275, 0.16571403162446946, 0.2555628510249617, 0.23413283117329386, 0.20574046006288735, 0.20825444946885996, 0.18164715046828095, 0.1800826997804269, 0.16571403162446946, 0.25525407621469237, 0.25403926785302217, 0.22429511667319135, 0.1250767619866061, 0.19585367512365615, 0.21397863336174366]

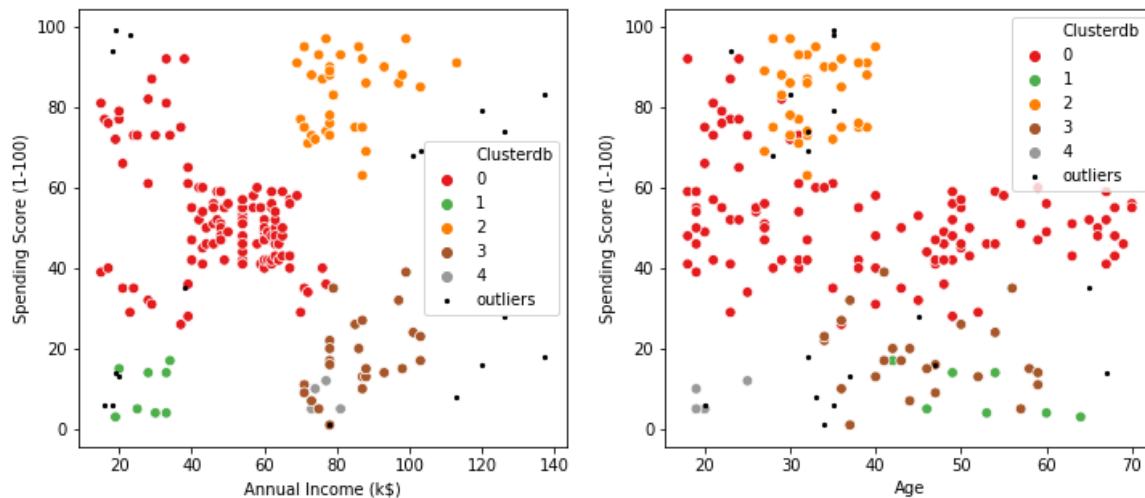
شکل 9

مشخص است که این تعداد در روش خوشه بندی قبلی نیز بدست آمده بود. حال با توجه به این مقادیر خوشه بندی را انجام دادیم.

Clusterdb	
-1	18
0	112
1	8
2	34
3	24
4	4

شکل 10

با توجه به شکل 10 پراکندگی داده ها مشخص است. همانطور که مشخص است خوشه 1- مربوط به داده های اوتلایر می باشد که در اینجا تعداد این داده ها برابر 18 عدد می باشد که این مورد در روش k-means قابل استخراج نبود. نحوه خوشه بنده با محورهای همانند شکل 6 بصورت شکل 11 بدست آمد. همانطور که در شکل مشخص است تعداد خوشه ها برابر 6 می باشد که البته یک مورد اوتلایر ها بوده و تعداد خوشه های اصلی 5 می باشد.



شکل 11

نتیجه گیری

با توجه به کارهای انجام شده و روشهای پیاده سازی شده برای این پروژه و مقایسه معیار Silhouette خواهیم تصمیم گیری کنیم به دلیل بالا بودن این معیار در روش k-means که برابر 0.45 بود و این معیار در روش DBSCAN برابر 0.26 بود روش خوشه بندی k-means نتیجه بهتری دارد ولی چنانچه خواهیم با توجه به قدرت تشخیص اوتلایر نگاه کنیم روش DBSCAN بهتر می باشد. و این بستگی به پروژه ما دارد. چون در این پروژه هدف خوشه بندی بوده است روش k-means نتیجه بهتری داشته و این روش پیشنهاد می شود.