

بدلیل آشنایی با پکیج ها و ساختار زبان برنامه نویسی R ، کد های مربوطه در زبان R پیاده سازی شده است.
در ابتدا حدود ۱۰۰ خط کد برای هر عدد دیتا بیس به صورت مستقل نوشته شده بود که بعد در انتهای کار به این نتیجه رسیدم که برای خلاصه کردن کار، تمامی فرایند به صورت Functionalized در آمده و بر روی هر دیتا ست تابع `do_my_homework_Q1` فراخوانی شود.

در خطوط ابتدایی کتابخانه های مربوطه و دیتا ست ها پس از دانلود شدن و قرار گرفتن در دایرکتوری فایل `۰۲.r` لود شدند.

کتابخانه `fossil` برای استفاده از تابع `rand.index` آن و کتابخانه `factoextra` برای نمایش یکسری از کلاسترینگ ها به کار می آید. `Ggplot2` برای کارهای گرافیکی ساده و `pheatmap` ، همانطور که از اسمش مشخص است `heatmap` ترسیم میکند.

تابع محاسبه `purity index` در ادامه آمده است. به وسیله تهیه نوعی `confusion matrix` و انجام محاسبات مربوطه تابع تعریف میشود. (در مورد صحیح بودن نحوه محاسبه شک دارم و از `stackoverflow` کمک گرفتم. ولی اطمینان ندارم)

تابع اصلی `do_my_homework_Q1` چهار پارامتر ورودی میگیرد. داده ها، مقدار `k` مطابق وبسایتی که داده ها دانلود شده ،اینکه دسته بندی دقیق را میدانیم یا خیر(که داده Boolean هست) و یک رشته `tag` برای قراردادن در بالای `plot` ها.

به طور خلاصه، یک دیتا فریم کمکی برای محاسبه `rand index` از همان ابتدا تهیه میشود که مقادیرش ارام ارام تکمیل میگردد. متد های `kmean` و سه نوع `clustering` درخواست شده به وسیله فاصله `Euclidean` محاسبه میشوند و تعداد مختلفی `plot` تولید میشود که میتوانید نتایج را در انتهای همین فایل ببینید.

در انتهای کار، پس از پیاده سازی متدهای `clustering`، نوعی هیت مپ بوسیله دیتا فریم کمکی که در ابتدای کد تولید کردیم ، رسم میکنیم. هیت مپ بوسیله مقادیر `rand.index` رسم شده است

هیچ نوع کتابخانه ای برای `spectral clustering` پیدا نکردم. فلذا هیت مپ ۴در۴ هست.

پس از پیاده سازی تابع `do_my_homework_Q1` حال ان را برای هر کدام از داده ها فراخوانی میکنیم. و نتیجه ها را در ادامه قرار داده ام. کد پس از اجرا شدن، نتایج را در یک فایل `pdf` تحت عنوان `graphical_results.pdf`

ذخیره میکند. همچنین خودم این کد را در محیط **Rstudio** اجرا کردم و نتیجه را تحت عنوان **Code_Text_Outputs.txt** مشاهده میکنید