

Lifestyle Clustering using K-Means

MONTH YEAR

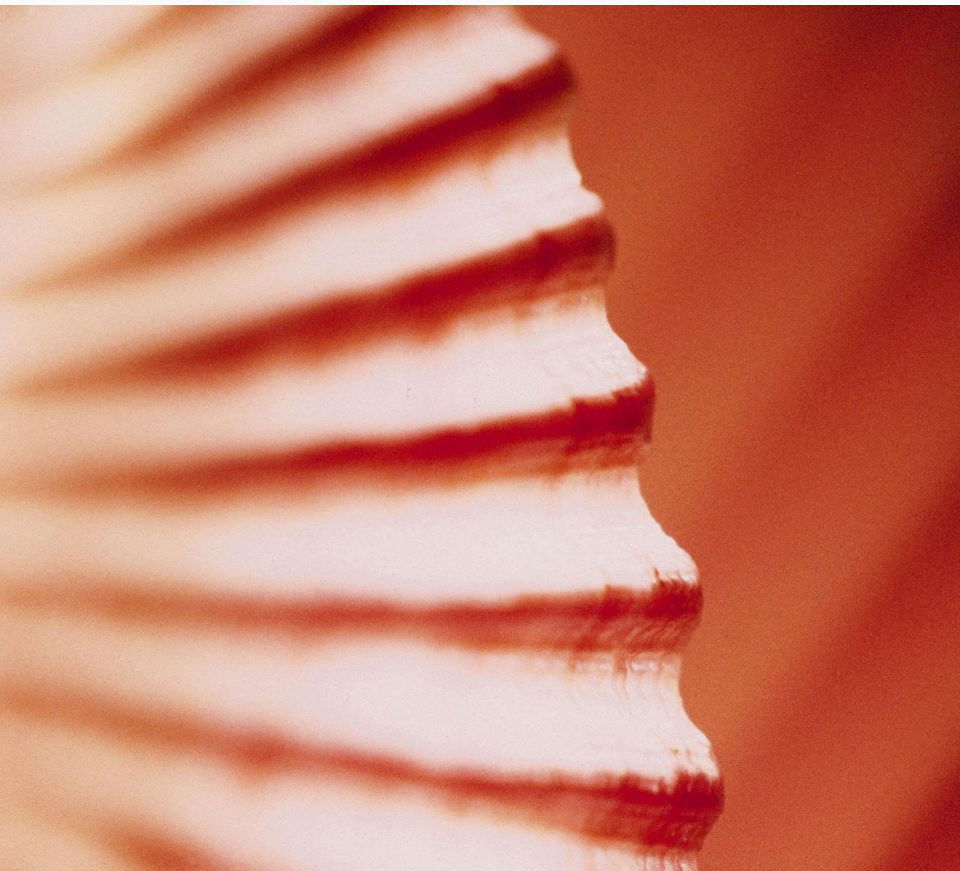
Lingwei Si

tools:Python, pandas,
sklearn,matplotlib,seaborn



Agenda

MONTH YEAR



- 01 Motivation and the choose of the methodology
- 02 Exploratory data analysis
- 03 The choose of K
- 04 Clustering result
- 05 Interpretation

What variables?

We want to figure out how could people in the college become a friend. Our assumption is that students in the college with similar lifestyles will become a good friend.

- 1.study_hours_per_day
- 2.Extracurricular_Hours_Per_Day
- 3.Sleep_Hours_Per_Day
- 4.Social_Hours_Per_Day
- 5.Physical_Activity_Hours_Per_Da
y
- 6.GPA
- 7.Stress_Level

WHY K-means?

-
- There are many continuous variables in our datasets.
 - The dataset shows meaningful natural group segregation.

Exploratory Data Analysis

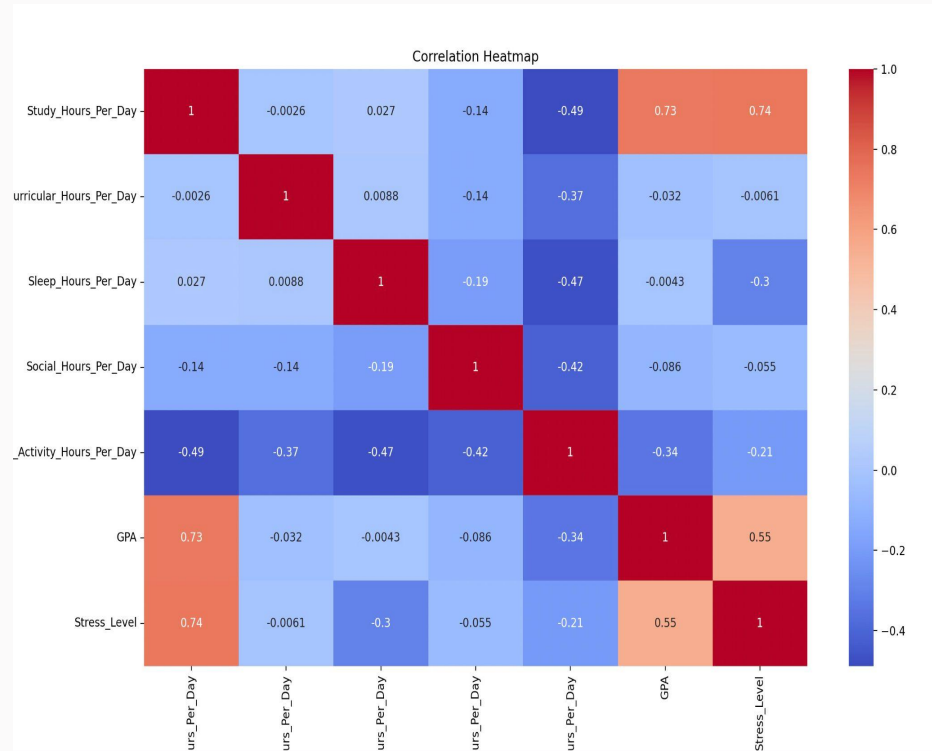
MONTH YEAR

The correlation heatmap shows the correlation coefficient between each two variables.

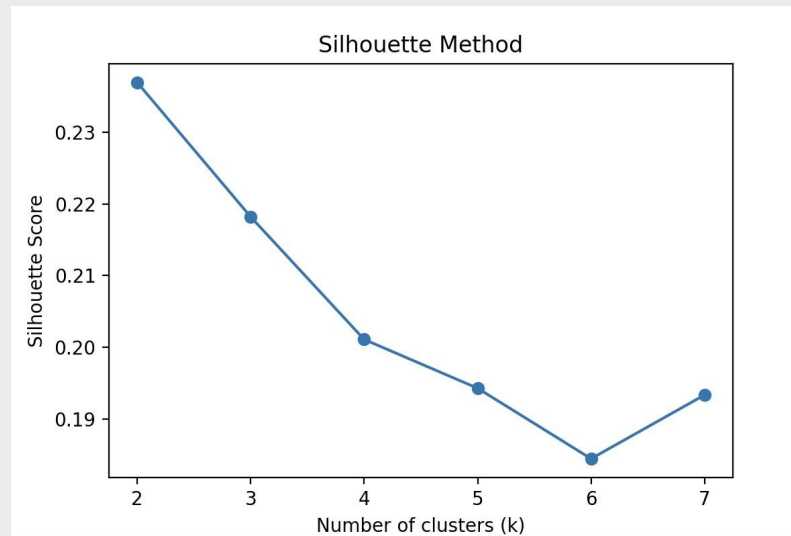
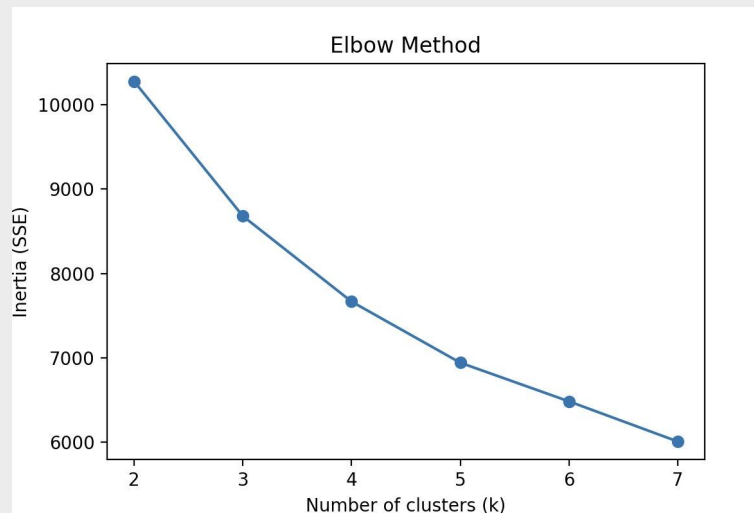
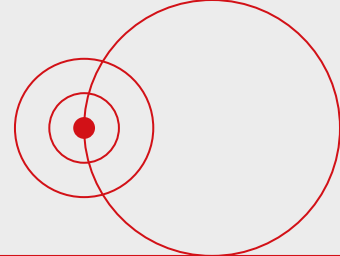
From this graph, I found that the correlations between each two variables are low, which satisfies our expectations.

Also, we check all the histogram and found all groups are distributed in a reasonable way. There are no outliers.

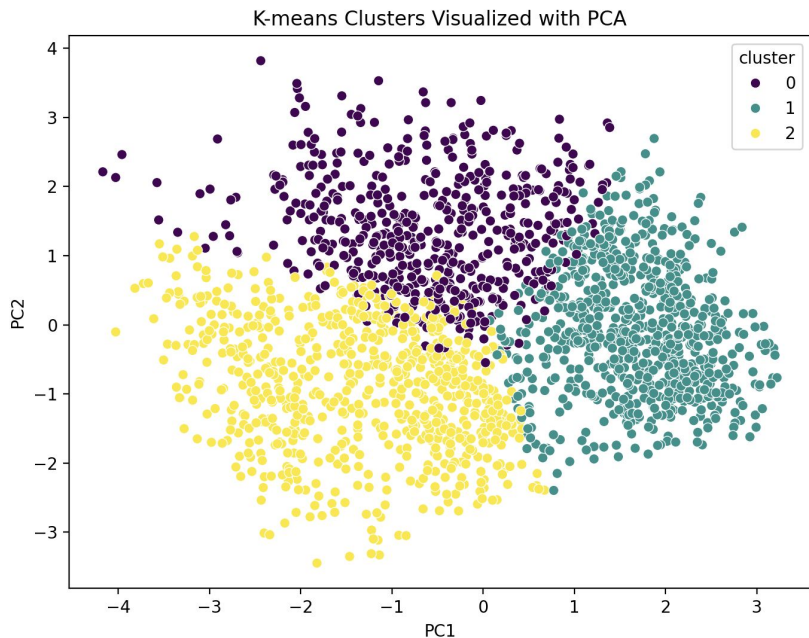
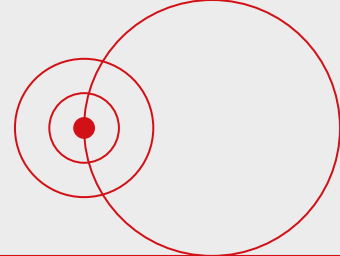
All in all, since the number of the variables are not so many, we decide to keep all the variables to construct our model



HOW to choose K?



Visualizations_graph

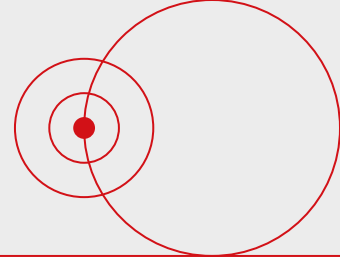


The PCA visualization shows three well-separated groups.

cluster	Study_Hours_Per_Day	Extracurricular_Hours_Per_Day	Sleep_Hours_Per_Day	Social_Hours_Per_Day	Physical_Activity_Hours_Per_Day	GPA	Stress_Level
0	-0.350977	-0.254293	-1.004464	-0.169963	1.013533	-0.231768	0.306753
1	1.021497	0.104274	0.184863	-0.080137	-0.680076	0.841192	0.769677
2	-0.820448	0.079791	0.555910	0.212211	-0.037578	-0.718494	-1.045794

1etheriander@peter@lucelienchenyingsdeMacBook Air dataset 8: CDA

Results and Interpretations



ALL IN ALL, we get the conclusion that there are three kinds of college students

Cluster 0 **Moderate-study but High-Stress Group**

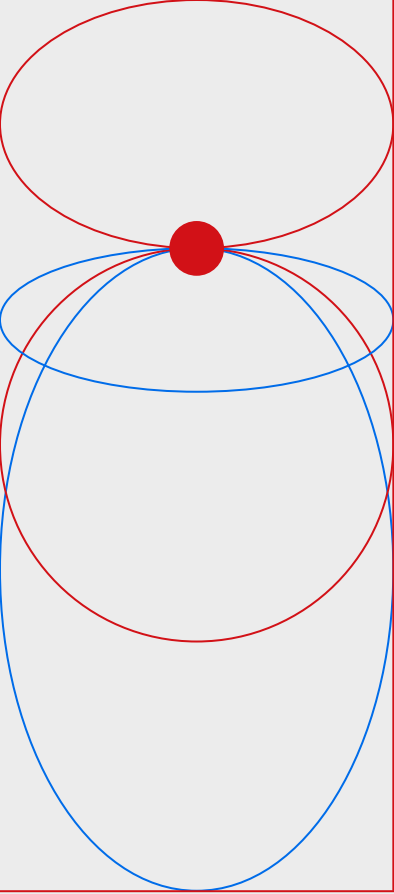
Not so many study hours
Stress is relatively high
Low sleep
The highest physical
activity

Cluster 1 **High Achieving and High-Stress Group**

High Study hours
High Stress and High
GPA
Low physical activity

Cluster 2 **Relaxed and Social Group**

Lowest study hours
and GPA
Lowest stress
High sleep and social
time



*Thank
you*