Clustering the feedback from Turkiye-Student-Evaluation Dataset.

- An Unsupervised Learning Problem.

- Problem - Clustering Technique to deal with given data.

  3 clustering analysis methods:

  K-Means   -   Agglomerative Clustering   -   DBSCAN

  on "Turkiye Student Evaluation" dataset and compared their performances.

Description :

- instr: Instructor's identifier; values taken from {1,2,3}

- class: Course code (descriptor); values taken from {1-13}

- repeat: Number of times the student is taking this course; values taken from {0,1,2,3,…}

- attendance: Code of the level of attendance; values from {0, 1, 2, 3, 4}

- difficulty: Level of difficulty of the course as perceived by the student; values taken
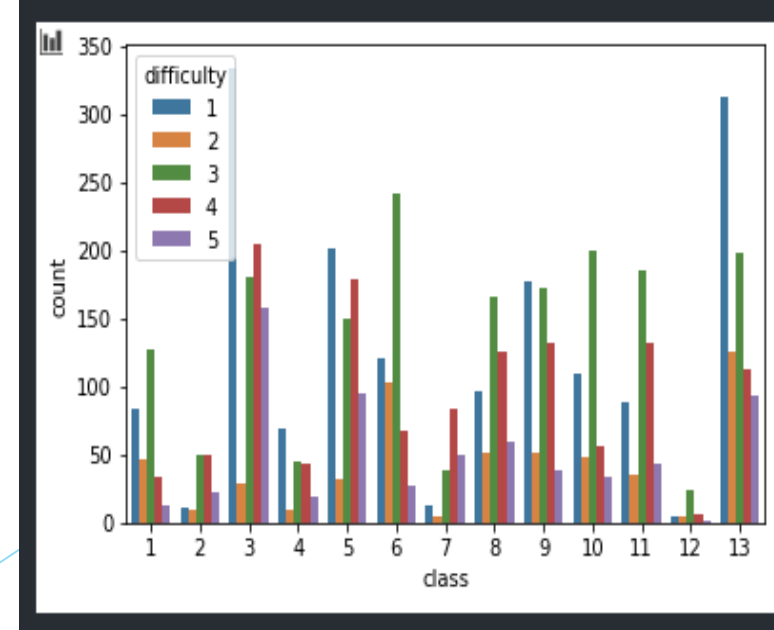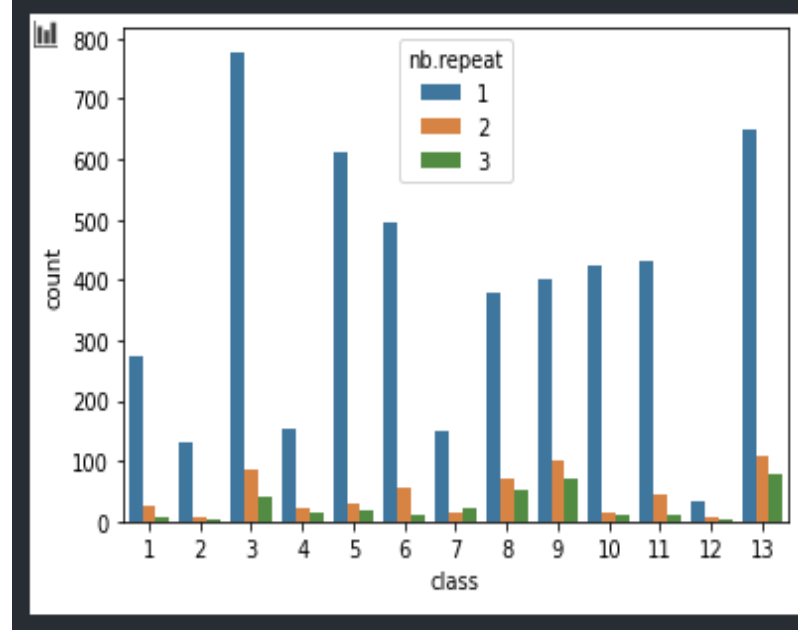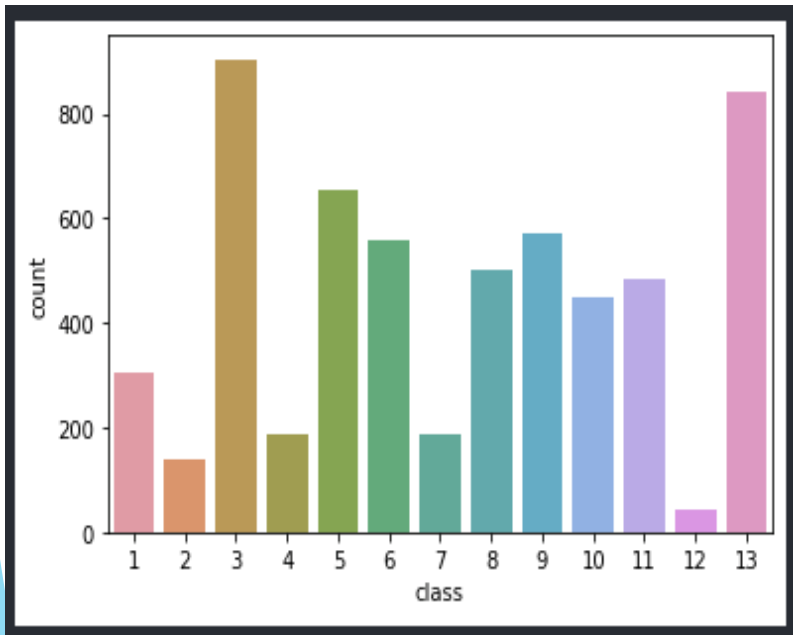
Attribute Information

And 28 different questions: Q1-Q28 values are taken from {1,2,3,4,5}

Some question examples:

Q1: The semester course content, teaching method and evaluation system were provided at the start.
Q2: The course aims and objectives were clearly stated at the beginning of the period.
Q3: The course was worth the amount of credit assigned to it.
…

Q14: The Instructor came prepared for classes.
Q15: The Instructor taught in accordance with the announced lesson plan.
Q16: The Instructor was committed to the course and was understandable.

…

Q27: The Instructor provided solutions to exams and discussed them with students.
Q28: The Instructor treated all students in a right and objective manner

Data Pre-Processing

- ▶ Our dataset contains 5820 rows * 33 columns

- ▶ As we can see there is no missing values in our dataset.

- ▶ That's why we did not do any cleaning operation

```
RangeIndex: 5820 entries, 0 to 5819
Data columns (total 33 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   instr       5820 non-null   int64
 1   class       5820 non-null   int64
 2   nb.repeat   5820 non-null   int64
 3   attendance  5820 non-null   int64
 4   difficulty  5820 non-null   int64
 5   Q1          5820 non-null   int64
 6   Q2          5820 non-null   int64
 7   Q3          5820 non-null   int64
 8   Q4          5820 non-null   int64
 9   Q5          5820 non-null   int64
 10  Q6          5820 non-null   int64
 11  Q7          5820 non-null   int64
 12  Q8          5820 non-null   int64
 13  Q9          5820 non-null   int64
 14  Q10         5820 non-null   int64
 15  Q11         5820 non-null   int64
 16  Q12         5820 non-null   int64
 17  Q13         5820 non-null   int64
 18  Q14         5820 non-null   int64
 19  Q15         5820 non-null   int64
 20  Q16         5820 non-null   int64
 21  Q17         5820 non-null   int64
 22  Q18         5820 non-null   int64
 23  Q19         5820 non-null   int64
 24  Q20         5820 non-null   int64
 25  Q21         5820 non-null   int64
 26  Q22         5820 non-null   int64
 27  Q23         5820 non-null   int64
 28  Q24         5820 non-null   int64
 29  Q25         5820 non-null   int64
 30  Q26         5820 non-null   int64
 31  Q27         5820 non-null   int64
 32  Q28         5820 non-null   int64
dtypes: int64(33)
memory usage: 1.5 MB
```
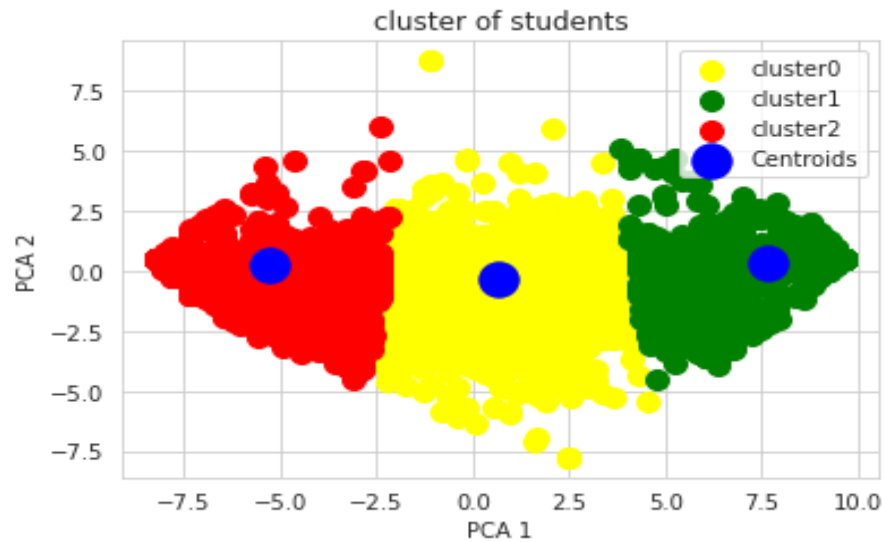
K-MEANS

- We continued with first 7 questions also on the clustering algorithms.

- To analyze the cluster size for the k-means algorithm, at first we created this elbow graph. And saw that 3 may be our number of clusters.
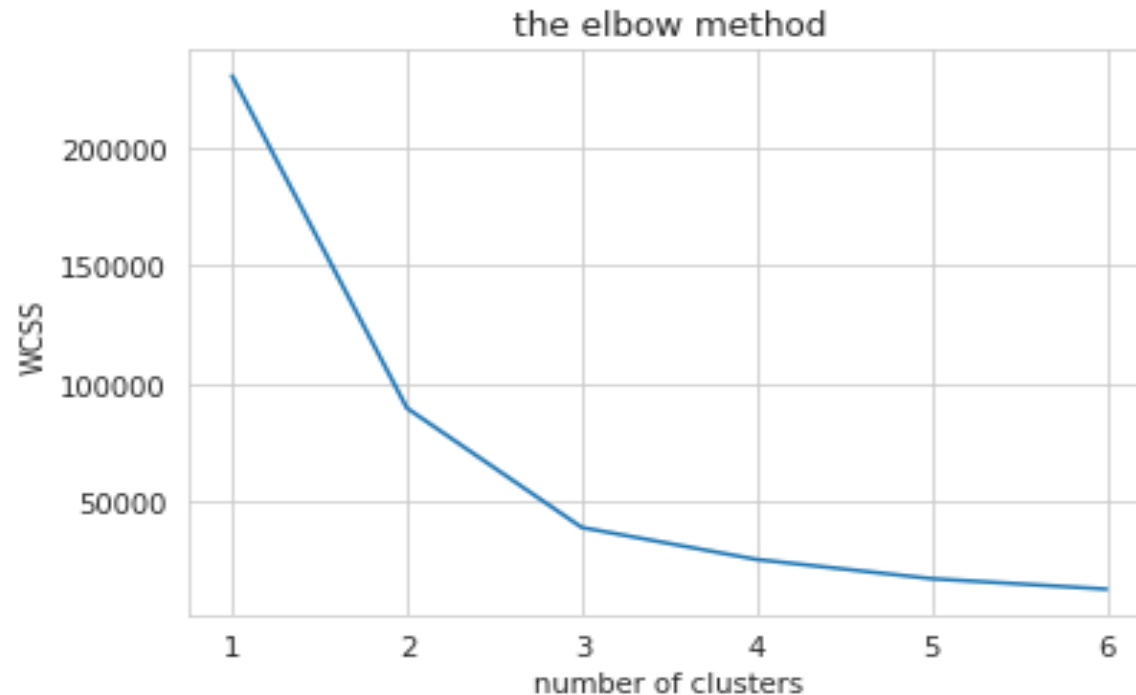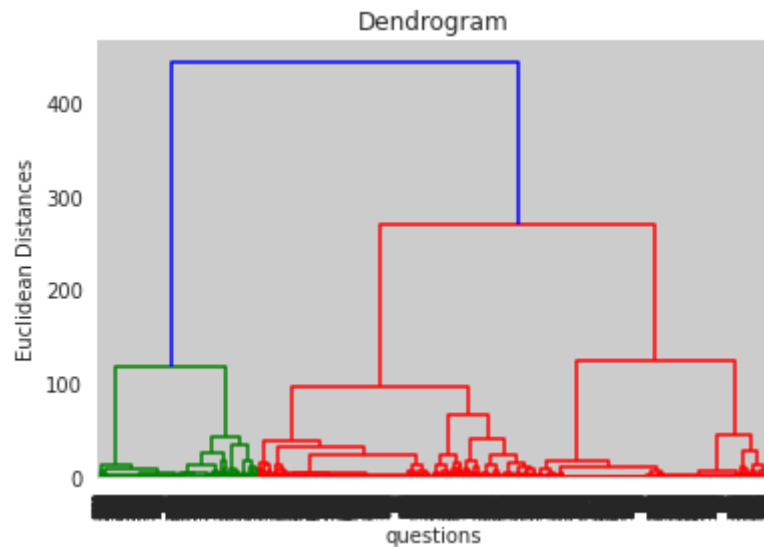
K-MEANS

▶ Then with sklearn.cluster's KMeans library, we clustered our data. And we plotted the results:

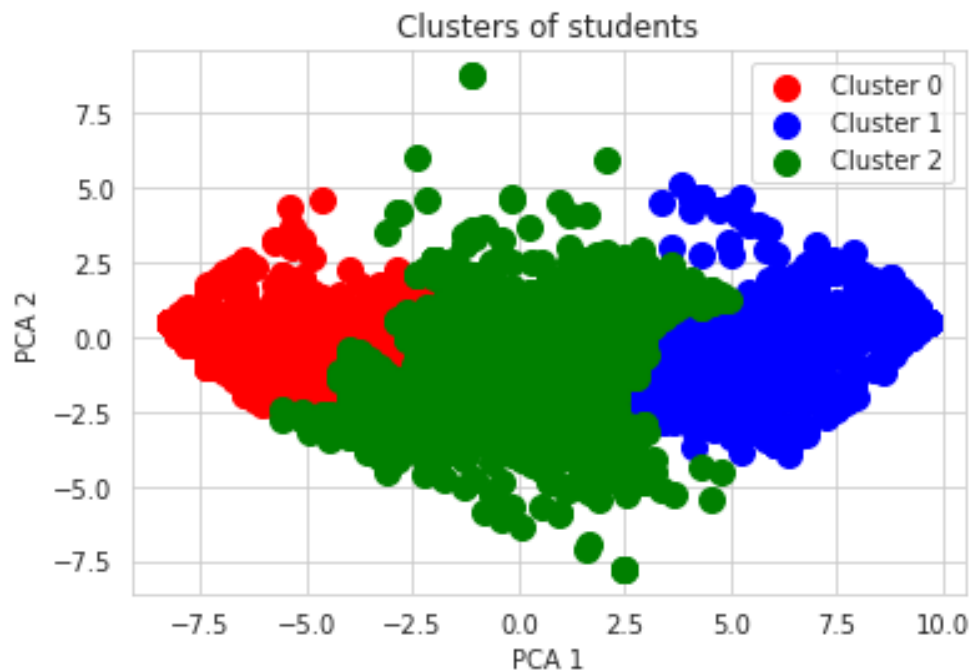▶ silhouette score: 0.571

▶ Counter ({0: 2389, 2: 2158, 1: 1273})

Dendogram

- We plotted dendrogram to see how many clusters should we use.
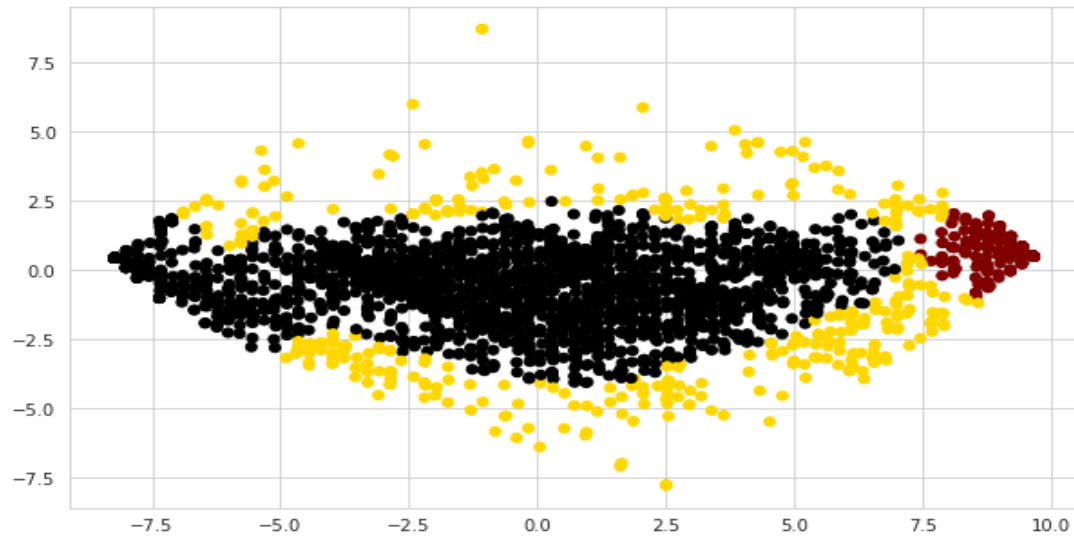- We decide to take number of clusters as 3.

Agglomerative Clustering

▶ Counter ({2: 2397, 0: 2011, 1: 1412})

▶ silhouette score: 0.571

Clusters of students

DB SCAN

- number of clusters: 3
- Counter({0: 4735, 1: 691, -1: 394})
- unclustered percentage: 0.06769759450171821
- silhouette score: 0.327

COMPARISION OF CLUSTERING METHODS

Clustering methods	Silhouette score
K-Means Clustering	0.571
Agglomerative Clustering	0.571
DB SCAN	0.327

REFERENCES

- https://github.com/naveen-chauhan/Turkiye-Student-Evaluation/blob/master/turkiye%2Bstudent%2Bevaluation.ipynb
- https://www.kaggle.com/harishvutukuri/turkiye-student-evaluation