The main goal of my project was to find the overlap in 2 exams (or more) and categorizing the shared students, then tracking each group to see how their performance changed by visualizing each DataFrame.

First I decided to divide this project into four different segments:

1. Comparing exams that were held before Eid (with the approximate distance of three months)
2. Comparing exams that were held after Eid (with the approximate distance of two months)
3. Comparing two exams with distance of four to five months (Both held either before Eid or After Eid)
4. Comparing two exams, one chosen before Eid and the other one after Eid

For starting I had to find a number that gave me a sense about how many students each of these exams had in common. Therefore I thought the best way to calculate such number is to divide the number of shared students by the number of all of the students that participated in both exams. In other words, if we consider a Venn diagram of both exams, what I believed to be correct was to divide their intersection by their union and converting it to percentage:

But the numbers calculated were not an accurate description of the overlaps.

So by rethinking and consulting with my co-worker (Mr.Tehranchi) we came up with a better approach which made more sense regarding our needs.

We decided that we should change the denominator to the mean of the number of participant of both exams, because that way the number calculated (which I called Overlapped Students Percentage or OSP) multiplied by the number of participants in either exams will result in approximate number of overlapped students. The formula is:

After choosing the best number to represent the percentage of overlapped students I carried on with computing it in the first segment of the project by choosing two exams with the distance of three months (or near). The results were acceptable as it is shown in the table below:

|  |  |
| --- | --- |
| **First Segment** | |
| **Exams** | **OSP** |
| *Mehr 23rd and Dey 24th* | 64.79317433052685 |
| *Aban 14th and Bahman 15th* | 69.42827798699076 |
| *Azar 5th and Esfand 6th* | 69.42827798699076 |
| *Azar 26th and Esfand 20th* | 68.03079416531604 |

On to the next segments:

|  |  |
| --- | --- |
| **Second Segment** | |
| **Exams** | **OSP** |
| *Farvardin 19th and khordad 20th* | 54.86653396621243 |
| *Ordibehesht 2nd and tir 3rd* | 52.67382702091577 |