

Analysis through machine learning on influence of COVID-19 on higher education before, during and after pandemic

Carlos Andres Granados Ortiz*
Sweney Oriana Giraldo Aristizabal**
Ivan Andres Padilla Escorsia***

Received: 07-03-2025

Accepted: 02-05-2025

Quote as: Granados, C., Giraldo, S. y Padilla, I. (2025). Analysis through machine learning on influence of COVID-19 on higher education before, during and after pandemic. *Revista Interamericana de Investigación, Educación y Pedagogía*, 18(2), 235-263. <https://doi.org/10.15332/25005421.XXXX>

Abstract

Data mining, a process of uncovering silent characteristics of big data is one of such techniques which have nowadays become more popular for treating massive volume of data set information. In the current study, we apply a machine learning analysis, one of the data mining techniques to classify grades obtained by students before, during and after Covid-19 pandemic. The results obtained permit us to have a sense of knowing if pandemic affected the grades of the students or helped the students to improve their average. The main objective of this study is to optimize monitoring techniques in a public university located in Colombia which will be very valuable to

* Magister en Ciencias Matemáticas, Universidad de Antioquia, Medellín, Colombia
Email: carlosgranadosortiz@outlook.es (Corresponding author)

** Magister en Neuropedagogía, Universidad del Atlántico, Barranquilla, Colombia
Email: sweneyoriana@gmail.com

*** Magister en Educación, Universidad del Atlántico, Barranquilla, Colombia.
Email: iapadilla@uninorte.edu.co

the government, educators, students, researchers and others involved in understanding seriousness of the problem that Covid-19 present in the education.

Keywords: _____

Análisis mediante aprendizaje automático de la influencia del COVID-19 en la enseñanza superior antes, durante y después de la pandemia

Resumen

La minería de datos, un proceso para descubrir características silenciosas de los big data, es una de esas técnicas que hoy en día se han hecho más populares para tratar volúmenes masivos de información de conjuntos de datos. En el presente estudio, aplicamos un análisis de aprendizaje automático, una de las técnicas de minería de datos, para clasificar las notas obtenidas por los estudiantes antes, durante y después de la pandemia de Covid-19. Los resultados obtenidos nos permiten tener una idea de si la pandemia afectó a las notas de los alumnos o les ayudó a mejorar su media. El objetivo principal de este estudio es optimizar las técnicas de monitoreo en una universidad pública ubicada en Colombia lo cual será muy valioso para el gobierno, educadores, estudiantes, investigadores y demás involucrados en entender la gravedad del problema que Covid-19 presenta en la educación.

Palabras clave: _____

Análise de aprendizado de máquina da influência da COVID-19 no ensino superior antes, durante e depois da pandemia

Resumo

A mineração de dados, um processo para descobrir características silenciosas de big data, é uma das técnicas que se tornaram mais populares hoje em dia para lidar com volumes maciços de informações de conjuntos de dados. No presente estudo, aplicamos a análise de aprendizado de máquina, uma das técnicas de mineração de dados, para classificar as notas obtidas pelos alunos antes, durante e depois da pandemia de Covid-19. Os resultados obtidos nos permitem ter uma ideia se a pandemia afetou as notas dos alunos ou os ajudou a melhorar sua média. O principal objetivo deste estudo é otimizar as técnicas de monitoramento em uma universidade pública localizada na Colômbia, o que será muito valioso para o governo, educadores, estudantes, pesquisadores e outros envolvidos na compreensão da gravidade do problema que a Covid-19 apresenta na educação.

Palavras-chave: _____

Introduction

The Covid-19 virus has caused the evaluation processes in basic primary, secondary and higher education in Colombia to vary. The flexibility in the contents, the teaching modality, in some cases through asynchronous encounters has brought implications in the

academic performance of the students, this due to the fact that the majority of educational institutions of the public sector suspended all kinds of activities due to the scarce lack of technological resources for teaching during virtual teaching. This had a negative influence on the teaching-learning process, due to the fact that for many years it was not invested in the adaptation of virtual campuses, institutional websites, training for teachers and students in the management of ICT (Ríos, 2020). This made the transition from face-to-face teaching to distance work (Gewin, 2020) to be complex.

On the other hand, the scarce experience in digital education, little training in its didactic use in learning networks, online collaborative work, and change in educational evaluation by teachers (Conde and Padilla, 2021) was not easy, since given the distance modality, the teaching and learning processes are carried out completely remotely through the support of Information and Communication Technologies (ICT).

Now, regarding the instruments used to evaluate at this time, they were: a registry called "E-learning", which allowed to register attendance and participation, as well as activities in a quantitative way; the rubrics by which each teacher qualitatively evaluated the activities carried out by the students (freely, according to the criteria of each teacher), and the exams carried out within the school's platform and free educational platforms (Núñez, 2020).

Sarker (2022) mentioned that: we live in the digital age, which, like anything else, has its upsides and downsides. The main drawback is the security risk (Sarker, 2022). As more of our sensitive information transfers to the digital arena, security breaches are becoming more common and catastrophic. Cyber-criminals are growing more adept in their attempts to avoid detection, and many newer malware kits are already incorporating new ways to get out of antivirus and other threat detection systems. Cybersecurity, on the other hand, is at a crossroads, and future research efforts should be focused on cyber-

attack prediction systems that can foresee important scenarios and consequences, rather than depending on defensive solutions and focusing on mitigation. Systems that are based on a complete, predictive study of cyber risks are required all around the world. The key functionalities in cybersecurity such as prediction, prevention, identification or detection as well as corresponding incident response should be done intelligently and automatically. Artificial intelligence (AI), which is based primarily on Machine Learning (ML) (Sarker, 2021).

Taking into account mentioned above, in this study the objective of this paper is to study through machine learning how the process on the evaluation in a public university located in Barranquilla-Colombia in times of pandemic. For that we do a statistic analyze about the grades that 277 students had before of Covid-19, during Covid-19 and after the Covid-19. Those moments are divided into two parts; it means that each year is divided into two semesters, thus grades of students before, during and after Covid-19 were chosen from 2018, 2019 and 2020, and 2021 respectively; and each year is divided by first semester and second semester.

Methodology

First at all, we collect the data from the Campus-it which is the system that has all grades report from whole the students who belongs to the university to any program. In this case, we collect the information from 276 students who took courses related to Basic Mathematics, Calculus I (Differential calculus), Calculus II (Integral calculus), Calculus III (Multivariate calculus), Calculus IV (Ordinary Differential Equations), Calculus V (Partial Differential Equations), Numerical Methods and Analysis from an engineering's programs. We shall recall that all 277 students are the same who did not fail any course and continued their studies. To see the data, we refer the reader to appendix 1.

As we mentioned before, we take data from 2018, 2019, 2020 and 2021; each course is related in each year as follows:

Table 1. Relation year, semester and course.

Year	Semester	Course
2018	First semester	Basic Mathematics
	Second semester	Differential calculus
2019	First semester	Integral calculus
	Second semester	Multivariate calculus
2020	First semester	Ordinary Differential Equations
	Second semester	Partial Differential Equations
2021	First semester	Numerical Methods
	Second semester	Analysis

It is well-known that all notions from mathematics are articulated, for instance a mathematician's student has to know about derivate to find the integral o multivariate derivate, and the same happened with the course describe in table 1, to have a good knowledge and to understand the next one, the student has to have a good knowing about the previous course.

Secondly, to make data analysis, we use Python language program to show the results that we obtain, further with python better figure can be gotten to interpretive the data and thus better conclusion can be given.

Data Analysis

The source python code that we use for making the study can be found in appendix 2, first me make some change of the name of the columns for better analysis, 2018 first semester, 2018 second semester, 2019 first semester, 2019 second semester, 2019 first semester, 2019 second semester, 2021 first semester and 2021 second semester are changed by ps18, ss18, ps19, ss19, ps20, ss20, ps21 and ss21, respectively. Then, descriptive statistical analysis is showed:

Figure 1. Descriptive statistical analysis

	ps18	ss18	ps19	ss19	ps20	ss20	ps21	ss21
count	277.00	277.00	277.00	277.00	277.00	277.00	277.00	277.00
mean	2.96	3.51	4.43	4.25	3.47	3.46	2.90	3.01
std	0.59	0.83	0.35	0.26	0.29	0.41	0.52	0.59
min	2.00	2.01	3.80	3.80	3.00	2.80	2.00	2.00
25%	2.47	2.88	4.12	4.04	3.21	3.11	2.48	2.50
50%	2.93	3.49	4.46	4.26	3.45	3.44	2.87	3.01
75%	3.48	4.16	4.72	4.47	3.73	3.83	3.36	3.53
max	3.99	5.00	5.00	4.70	4.00	4.20	3.80	4.00

At university, the minimum value that a student can get is 0 and the maximum value that a student can get is 5, in this case the students can obtain any grade from [0,5]. The grade that a student needs to pass the course is 3. Thus, as figure 1 we can see that before pandemic the average was in 2.96 and 3.5 which mean that to many students obtained bad grades. Otherwise, when pandemic began the average was in 4.4 and 4.2 which means that virtual classes helped the student to obtained better grades. Unfortunately, the average decreased up to 3. Nevertheless, during pandemic standard deviation was lesser than before and after pandemic. So far, we can see that virtual classes were good for the students to improve their average, we shall remain that average could help students to obtain scholarship.

Now, we show a bar chart of each semester and their density.

Figure 2. Bar chart 2018 first semester

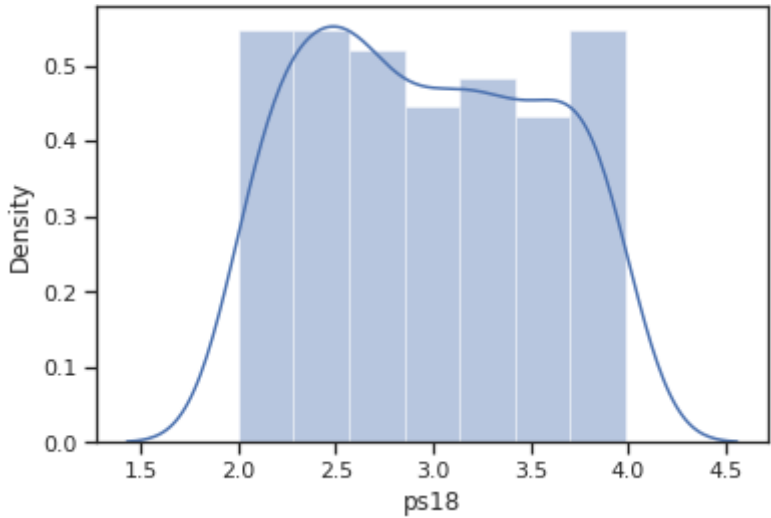


Figure 3. Bar chart 2018 second semester

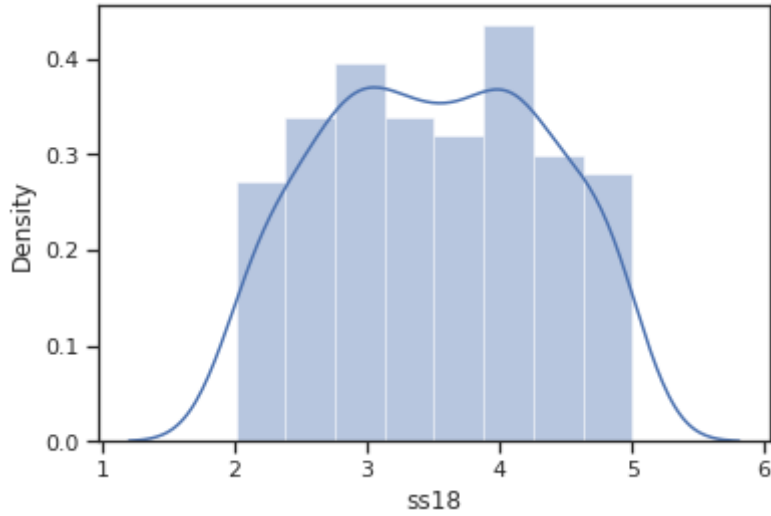


Figure 4. Bar chart 2019 first semester

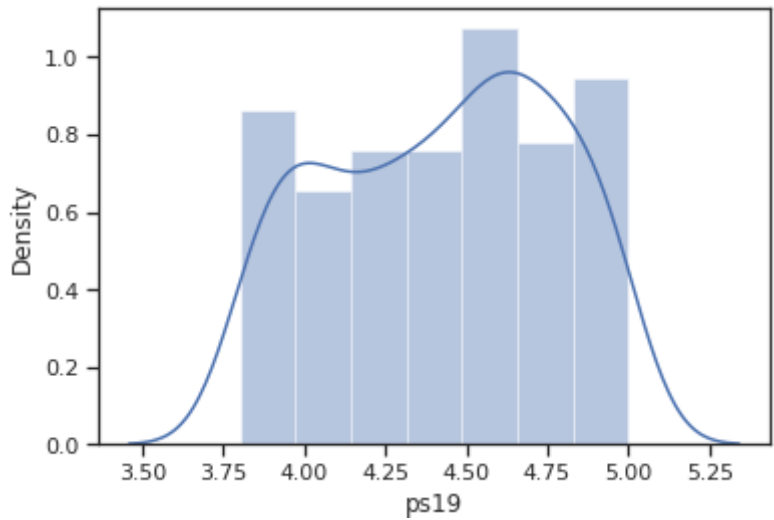


Figure 5. Bar chart 2019 second semester

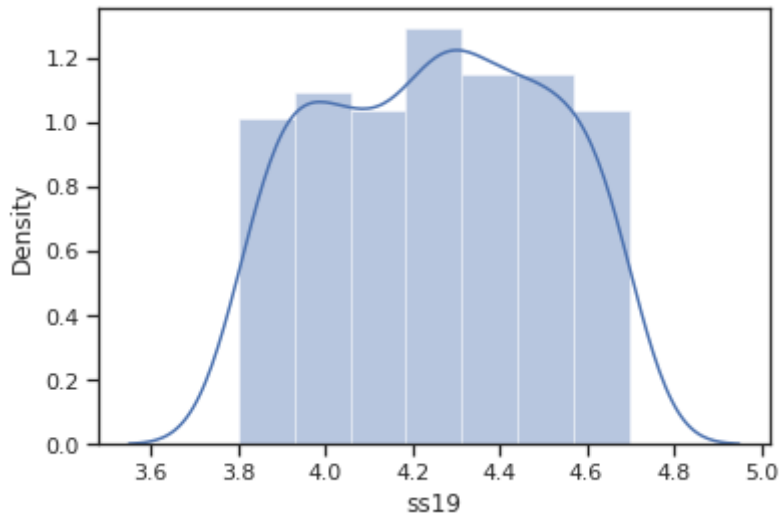


Figure 6. Bar chart 2020 first semester

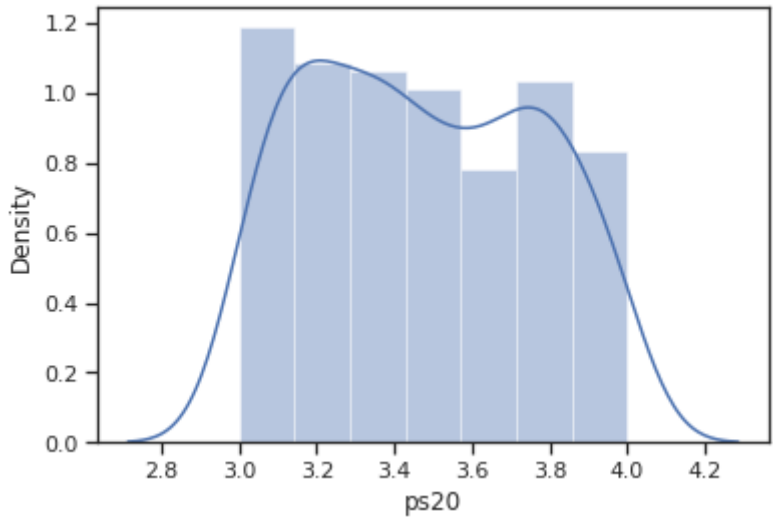


Figure 7. Bar chart 2020 second semester

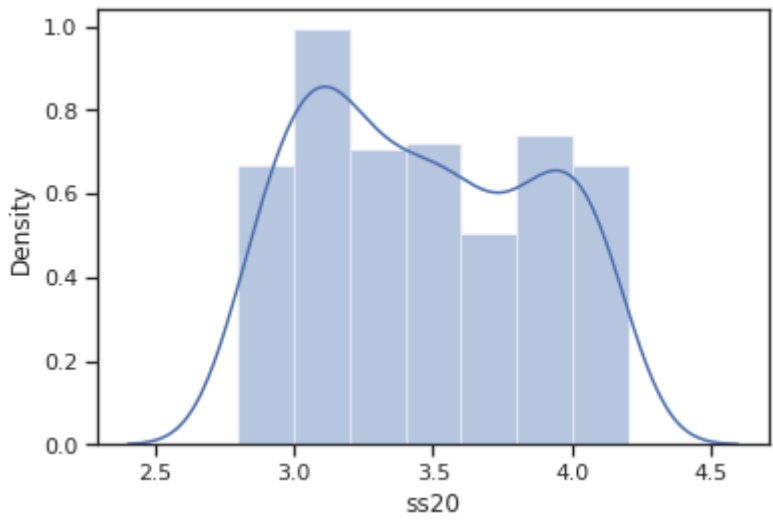


Figure 8. Bar chart 2021 first semester

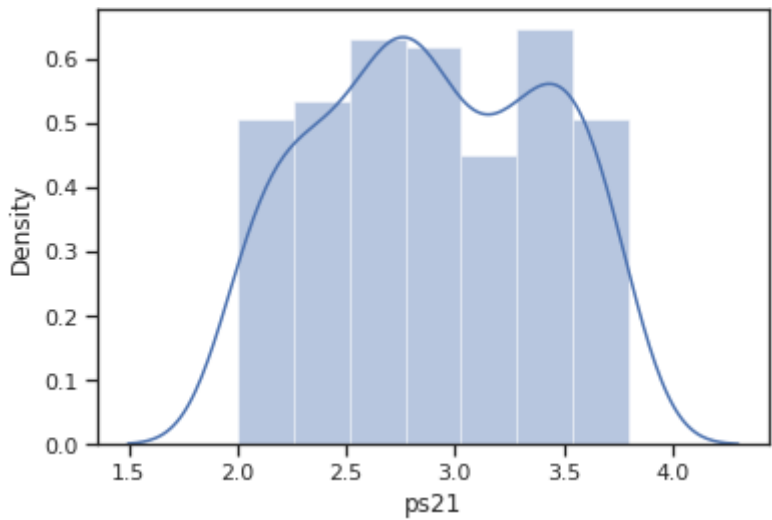
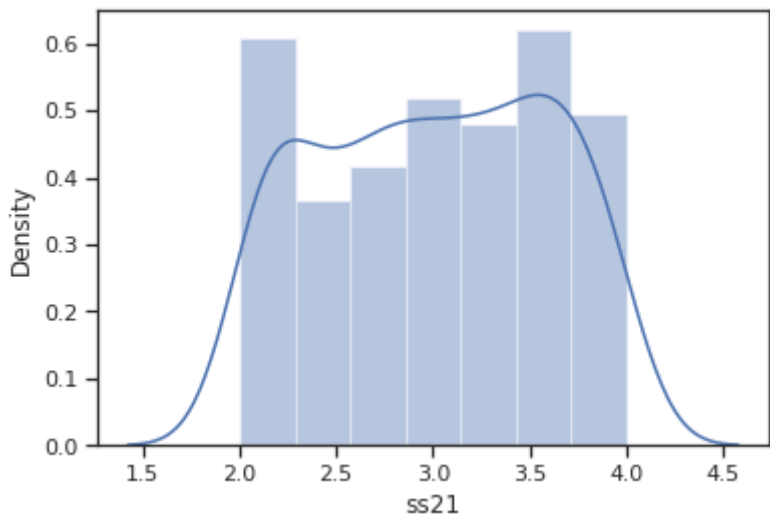
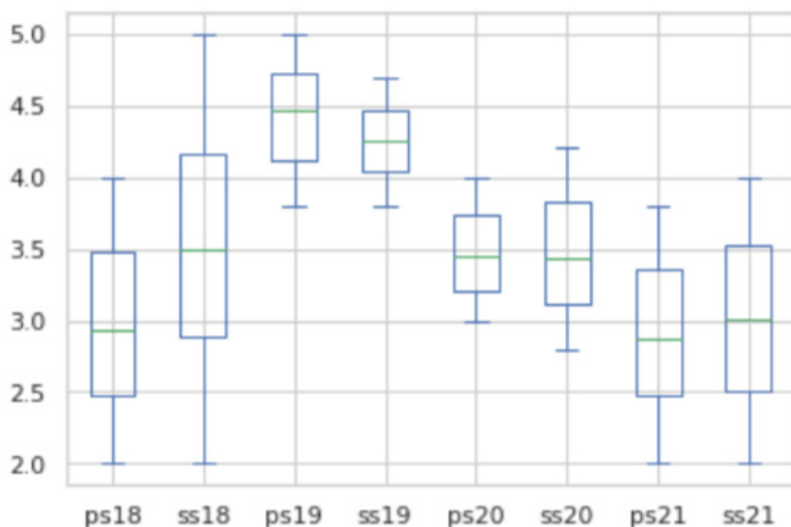


Figure 9. Bar chart 2021 second semester



Next, we show a box plot we show the behavior of the data set:

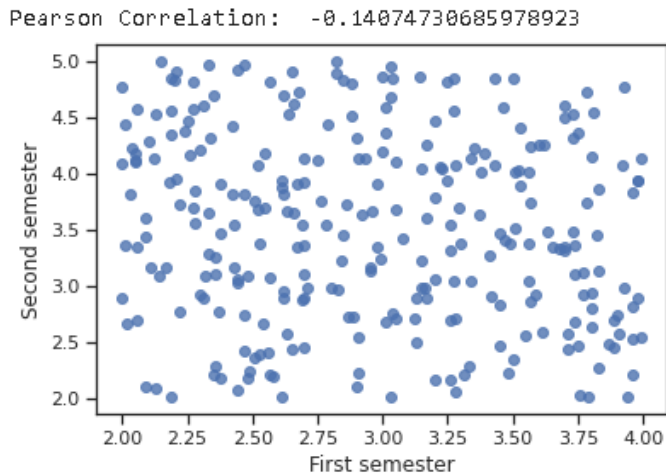
Figure 10. Box plot of data set



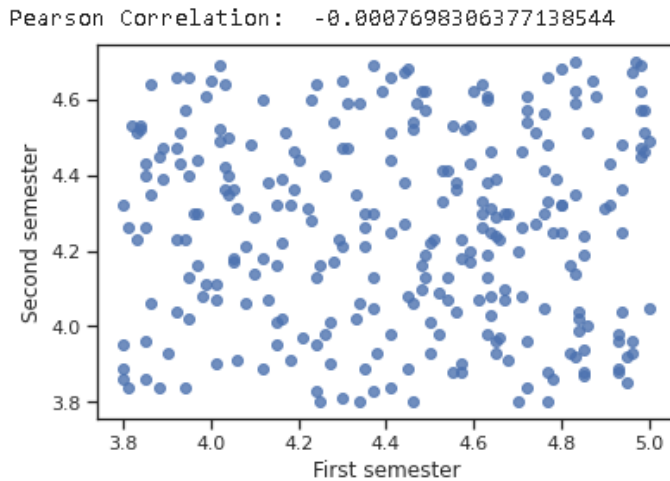
In figure 10, we can see that first semester and second semester of 2019, and first of 2020, data set are homogenous, but the rest of the semester before and after Covid-19 are similar, for this we can imply that when class returned face to face, the grades returned as they used to be. During pandemic grades used to be higher and we expected that grades continued in the same way, but we can see that it was not.

Below an inferential analysis of the data using Pearson's correlation method is showed by comparing year by year. For each correlation we define the following hypothesis:

H_0 : There is no linear relationship between the grades.

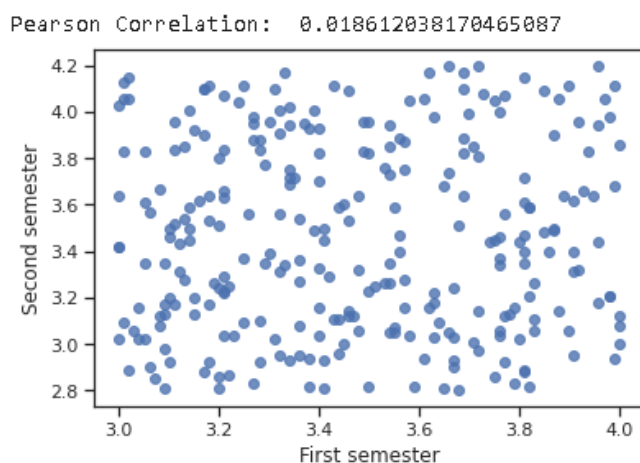
Figure 11. Correlation between 2018

From figure 11, when we calculate p which gives signification level, this is below to 0.05 which means that H_0 should be eliminated and there is a correlation between first semester and second semester.

Figure 12. Correlation between 2019

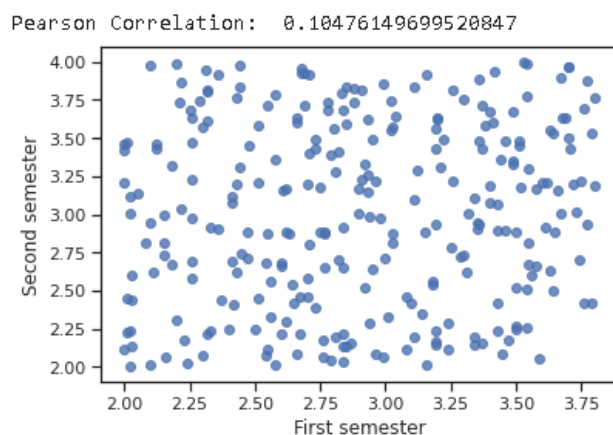
From figure 12, when we calculate p which gives signification level, this is above to 0.05 which means that H_0 should be accepted and there is no correlation between first semester and second semester.

Figure 13. Correlation between 2020



From figure 13, when we calculate p which gives signification level, this is above to 0.05 which means that H_0 should be accepted and there is no correlation between first semester and second semester.

Figure 14. Correlation between 2021

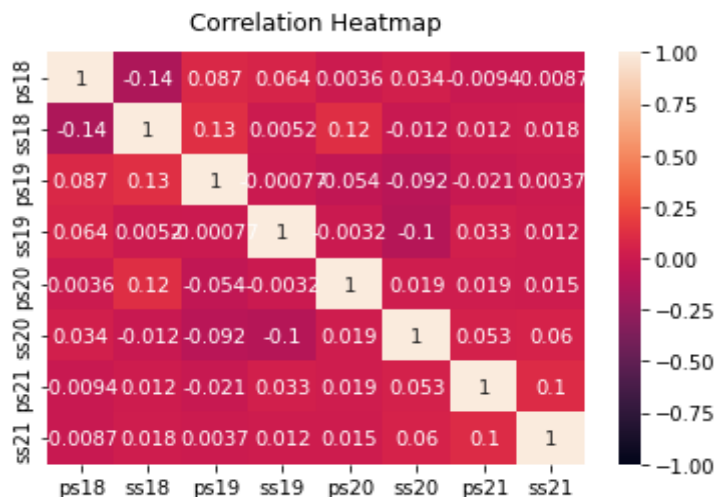


From figure 14, when we calculate p which gives signification level, this is above to 0.05 which means that H_0 should be accepted and there is no correlation between first semester and second semester.

Next, we find the correlation of grades before, during and after Covid-19. Thus, we define the following hypothesis:

H_0 : There is no linear relationship between the grades before, during and after Covid-19.

Figure 15. Correlation between grades before, during and after Covid-19



In figure 15, we can see that the relation between each semester respected each year is very low. Nonetheless, there is a small relationship between the pandemic periods (2019-2020) when the classes were developed in virtual mode, and those moments in which the classes were developed in face-to-face mode 2018 and 2021.

Conclusion

In this paper, we have carried out a statistical study using machine learning to observe the behavior of the grades before, during and after Covid-19, as well as to study if there is an existing relationship between them. From this, we were able to observe that during the pandemic when the classes were held in virtual mode, the students obtained good grades, compared to before and after the grades are far from these. Based on this, it is recommended to carry out a qualitative study where the causes of these are determined, in turn it can be argued that there is no relationship between before and after with the during Covid-19, as can be seen in the box plot and correlation diagrams.

On the other hand, it is well-known that these notes could have presented a degree of uncertainty, which is why the reader is invited to carry out this study applying neutrosophic statistics and thus compare these results with obtained in this paper.

Conflict of interest

Authors have no conflict of interest.

Data availability

All data generated or analyzed during this study are included in this published article.

References

- F, Núñez, Evaluar en tiempos de pandemia. Experiencias desde el contexto virtual, *Universciencia*, 57(2021), 23-33.
- R. Conde and I. Padilla, Aprender matemáticas en tiempos del COVID-19: Un estudio de caso con estudiantes universitarias, *Educación y Humanismo*, 23(2021), 1-17.
- V. Gewin, Five tips for moving teaching online as COVID-19 takes hold, *Nature*, 580(2020), 295-296.
- C. Ríos, Educación Superior Universitaria Pública del Perú, *Revista Clake Education*, 1(2020), 1-12.
- I. H. Sarker, Smart city data science: towards data-driven smart cities with open research issues. *Internet Things* 19 (2022), 100528.
- I. H. Sarker, Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci* 2(3) (2021), 1-21.
- I. H. Sarker, Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. *Ann. Data. Sci.* (2022), 1-26.

**Pasar las
referencias a
norma APA**

Appendices

Appendix 1. Grades of students before, during and after covid-19.

Table 2. Data set

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
3,92	4,8	4,06	4,51	3,67	3,63	3,52	3,74
2,56	2,94	4,97	4,13	3,22	3,95	3,54	2,17
3,71	3,49	4,96	4,21	3,93	2,87	2,89	2,95
2,92	4,52	4,47	3,9	3,57	3,46	3,35	2,91
3,06	4,62	4,63	4,56	3,05	3	2,59	3,89
2,85	3,41	3,89	4,21	3,02	3,2	3,28	3,88
2,81	2,9	4,38	4,62	3,82	3,3	3,17	3,21
3,45	3,52	4,33	4,69	3,58	3,4	2,02	2,6
3,62	3,02	4,72	4,46	3,52	2,91	3,12	2,43
2,5	2,92	3,89	4,22	3,1	4,06	3,44	3
2,95	4,48	4,53	4,67	3,5	3,26	2,73	3,36
3,49	2,05	4,55	4,44	3,92	3,03	2,69	3,76
2,69	2	4,45	4,7	3,53	3,46	2,97	3,58
3,65	4,75	4,13	4,05	3,08	4,04	3,09	2,13
2,4	2,04	4,34	4,44	3,34	4,01	2,02	2,09
2,07	2,59	4,3	4,58	3,14	4,2	2,36	3,96
2,79	3,01	4,9	4,35	3,25	3,79	3,79	3,11
3,48	5	3,87	3,81	3,15	3,54	2,93	2,88
3,64	4,55	4,41	3,82	3,98	3,02	3,61	3,61
2,31	4,73	4,01	3,93	3,06	4,01	3,31	2,08
3,07	2,43	4,94	4,34	3,78	4,02	3,7	3,86
2,56	3,77	4,55	3,95	3,42	3,04	2,7	3,59
3,37	4,47	4,22	4,42	3,38	4,04	2,89	2,17
3,07	2,2	4,66	4,53	3,54	3,46	2,08	3,52
3,07	4,91	4,47	3,85	3,07	4,16	2,22	3,65
2,31	3,79	4,71	4,37	3,82	3,92	2,34	3,19

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
2,4	3,3	4,33	4,17	3,09	3,5	2,12	3,9
3,23	4,18	4,21	4,59	3,39	3,77	3,17	2,98
2,57	2,04	4,23	4,53	3,09	3,82	3,35	3,9
2,66	3,84	4,03	4,63	3,03	4,03	2,2	3,46
3,46	3,1	4,75	4,41	3,47	4,09	2,38	3,91
2,5	2,16	4,45	3,87	3,99	2,98	3,07	3,44
3,84	2,61	4,23	4,13	3,39	3,44	2,11	2,27
2,13	4,86	4,91	4,01	3,35	4,04	2,13	2,26
2,76	4,13	4,09	3,93	3,85	3,58	3,42	3,33
2,03	3,85	4,6	4,69	3,49	3,3	2,08	2,44
3,35	2,17	4,32	4,56	3,79	4,03	2,64	2,31
3,4	4,01	3,89	3,98	3,74	3,93	2,82	2,91
2,91	2,32	4,44	4,3	3,65	3,24	3,73	2,34
3,02	3,71	4,33	4,41	3,59	4,2	2,08	3,17
3,94	4,65	4,28	4,01	3,4	3,14	2,14	3,57
3,93	2,41	4,83	3,83	3,11	3,37	2,03	3,49
2,9	2,82	4,27	4,23	3,04	3,72	2,13	3,66
2,72	2,39	4,29	4,4	3,7	3,58	2,2	2,7
3,1	3,76	3,84	3,83	3,45	3,45	3,25	3,47
2,08	4,56	3,94	4,43	3,75	4,13	3,52	3,53
3,78	3,37	4,67	4,45	3,07	2,97	3,17	3,71
3,32	2,95	4,74	4,55	3,23	3,38	2,33	2,5
3,94	2,33	4,83	4,05	3,63	2,89	2,55	2,06
3,29	2,92	4,65	4,08	3,06	4,19	2	2,44
2,91	2,08	4,37	3,85	3,11	3,06	3,55	2,97
2,29	4,88	3,86	4,24	3,6	3,89	3,39	2,39
2,88	2,3	4,82	4,11	3,96	4,17	2,69	3,99
2,8	2,52	4,55	4,39	3,29	2,97	2,68	2,4
3,45	2,64	4,5	3,91	3,2	3,22	3,67	3,51
2,53	4,79	3,89	4,42	3	3,73	2,41	3,98

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
3,26	4,66	4,57	4,05	3,92	3,25	3,02	3,73
3,91	3,38	4,38	4,04	3,26	3,68	3,2	3,98
2,64	2,97	4,5	4,21	3,86	3,33	2,5	2,96
3,22	4,65	4,86	4,69	3,96	3,66	3,77	2,49
2,38	2,87	4,96	3,81	3,22	4,2	2,45	3,18
3,84	2,25	4,47	4,54	3,62	3,74	2,42	3,85
2,67	4	4,23	4,02	3,12	3,68	2,89	3,18
3,72	2,08	4,67	4,06	3,55	3,14	2,72	3,48
3,23	2,41	4,85	4,03	3,28	2,96	2,9	3,64
3,55	2,13	4,56	4,03	3,59	2,91	3,7	2,08
3,66	3,02	3,82	4,29	3,08	2,86	3,69	2,17
3,2	2,41	4,25	4,09	3,38	4,02	3,74	2,19
3,03	3,32	3,91	4,02	3,94	3,73	3,7	2,94
2,52	3,26	4,88	4,42	3,16	3,88	2,4	3,98
2,91	4,38	4,25	4,43	3,26	2,92	2,86	2,54
3,41	3,56	4,27	3,94	3,89	4,17	3,46	3,85
3,33	3,56	4,44	4,57	3,57	3,99	2,77	3,3
3,33	3,48	4,51	4	3,96	3,34	2,61	2,48
3,29	4,59	4,76	4,17	3,23	3,58	3,55	3,85
2,3	4,03	4,57	3,88	3,46	2,9	2,95	2,7
3,45	2,71	4,18	4,41	3,68	3,92	3,21	3,68
3,48	3,97	3,81	4,35	3	3,29	3,05	3,24
2,3	3,45	4,45	3,92	3,47	3,53	3,8	2,35
2,47	2,68	4,96	4,07	3,41	4,05	3,07	3,1
2,71	3,59	4,77	4,39	3	2,84	2,99	2,83
2,94	2,45	4,81	4,04	3,68	3	2,69	3,95
3,41	4,8	4,95	4,69	3,35	3,54	2,22	3,64
2,08	4,96	3,83	4,19	3,22	3,11	2,01	3,9
2,52	2,66	4,35	4,03	3,21	3,1	2,68	2,98
2,36	3,15	4,18	3,89	3,65	3,71	2,69	2,27

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
3,95	4,24	4,59	4,59	3,4	4,04	3,13	3,09
2,25	2,51	4,88	4,14	3,29	3,56	2,67	3,41
3,26	4,72	4,91	4,18	3,91	3,37	2,78	2,45
3,56	2,51	4,05	4,27	3,87	3,16	2,47	3,1
2,84	4,86	4,46	4,08	3,14	3,02	2,64	2,85
3,78	3,46	4,84	3,81	3,21	3,95	2,99	3,2
3,73	3,03	4,47	4,67	3,72	3,75	3,42	2,47
3,83	3,12	4,46	4,16	3,75	4,05	3,77	2,87
2,87	2,77	4,07	4,03	3,35	3,3	2,95	3,23
2,24	2,31	4,25	4,04	3,38	3,4	2,65	3,17
3,11	2,94	4,21	4,16	3,14	3,6	3,53	3,71
2,19	2,94	4,54	4,53	3,91	3,04	3,55	3,79
3,16	2,89	4,75	4,07	3,89	3,03	2,88	3,53
3,16	3,99	4,77	4,52	3,4	3,62	3,36	2,21
2,78	3,29	3,83	4,69	3,49	3,52	2,27	2,63
2,95	3,96	4,32	3,95	3,49	3,78	2,73	2,69
2,57	4,27	4,77	4,41	3,37	3,07	2,57	2,34
2,19	2,31	4,5	4,14	3,3	3,29	3,47	3,44
2,22	4,32	4,86	3,88	3,24	3,52	2,8	3,48
2,18	4,39	4,18	4,53	3,46	4,12	2,9	2,71
3,61	3,41	4,92	4,1	3,51	3,67	3,11	2,07
2,25	3,18	4,17	4,09	3,95	3,29	3,56	3,69
3,47	3,33	4,27	4,49	3,22	4,04	3,37	2
3,97	3,32	4,48	4,58	3,55	2,97	3,27	2,44
2,12	2,38	3,86	4,57	3,11	4,19	2,12	3
3,98	3,01	4,99	4,14	3,54	2,9	3,11	2,02
3,69	3,14	4,89	4	3,16	4,18	3,52	2,53
2,62	3,67	4,43	4,39	3,54	3,51	2,32	3,57
2	3,05	4,46	3,9	3,04	3,23	3,39	2,6
2,36	2,62	4,01	3,87	3,28	2,91	2,06	3,42

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
2,61	4,63	4,78	4,59	3,83	2,93	2,6	3,59
2,28	4,84	4,74	3,92	3,03	3,42	2,9	3
3,99	2,6	4,94	4,4	3,57	3,86	2	2,04
3,12	2,88	4,4	3,97	3,44	3,73	2,26	3,53
3,28	2,62	4,63	4,52	3,1	3,46	2,08	2,65
2,21	3,84	4,83	4,01	3,53	2,98	2,3	2,19
2,42	2,41	4,54	4,34	3,24	3,29	3,69	3,79
3,63	3,27	3,99	4,01	3,19	4,12	2,74	2,42
4	4,14	4,65	4,28	3,61	3,75	2,01	3,81
2,75	2,7	4,23	4,08	3,66	3,02	2,78	2,47
2,5	3,71	4,81	4,63	3	2,98	3,1	2,04
2,27	2,32	4,84	4,26	3,58	3,08	3,47	3,64
2,27	3,54	4,72	4,14	3,85	3,07	2,64	2,87
3,74	3,18	4,96	4,29	3,33	3,97	2,27	2,05
3,35	4,76	4,18	4,56	3,14	2,99	3,19	3,41
2,65	3	4,09	4,38	3,83	4,08	3,75	3,22
3,29	4,11	3,97	4,12	3,1	4,01	2,95	3,81
3,3	4,38	4,76	4,59	3,1	3,85	2,25	3,48
2,18	4,95	4,19	4,17	3,52	3,36	2,88	2,62
3,54	4,29	4,59	3,85	4	4,07	3,44	2,91
2,75	3,77	4,54	3,95	3,66	3,54	2,27	2,02
3,29	2,47	4,29	4,6	3,88	3,13	2,65	3,59
2,18	2,25	4,46	3,96	3,89	4,11	2,12	3,37
2,3	4,9	4,74	3,94	3,83	3,39	2,76	2,41
2,79	4,86	4,99	4,27	3,89	4,01	3,31	3,66
3,63	3,38	3,81	4,11	3,42	3,08	3,66	3,67
2,68	2,32	4,11	4,25	3,11	3,14	3,76	3,46
2,91	2,86	4,63	4,17	3	3,85	3,7	3,48
3,39	4,66	4,06	4,12	3,58	3,62	2,28	3,13
3,51	4,92	4,44	4,59	3,77	3,17	2,63	3,74

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
2,87	4,93	4,61	3,87	3,95	3,65	2,23	2,31
2,4	2,16	4,92	4,13	3,9	4,06	3,12	2,53
2,92	2,09	4,71	4,36	3,29	2,84	3,1	2,02
3,3	4,41	4,07	3,97	3,81	3,27	2,66	2,8
2,67	4,19	3,87	3,8	3,1	3,5	2,72	3,5
2,88	2,82	4,77	3,95	3,81	3,01	3,58	2,64
2,25	2,8	3,98	4,52	3,61	4,12	2,28	3,82
2,17	4,41	4,97	4,12	3,75	3,45	2,64	2,08
3,46	3,29	4,72	4,27	3,65	3,4	2,91	3,69
2,17	2,69	4,87	4,24	3,7	3,1	2,35	2,14
2,42	2,51	4,74	4,45	3,72	3,87	2,85	2,86
3,89	2,54	4,91	4,22	3,14	3,34	2,9	2,42
2,57	3,98	4,32	3,82	3,36	2,99	3,05	3,39
3,27	2,85	4,16	4,36	3,01	3,34	3,23	2,15
3,34	3,54	4,29	3,97	3,36	3,99	2,43	3,25
2,2	4,06	4,89	4,61	3,29	2,86	2,71	3,12
3,22	2,6	4,5	4,53	3,72	3,9	2,56	3,42
2,29	3,85	4,83	4,67	3,76	3,56	2,87	2,36
3,18	2,46	4,01	4,23	3	3,91	3,58	2,91
2,33	2,11	4,42	4,16	3,08	3,52	2,37	2,39
3,09	4,77	3,85	3,83	3,77	3,38	3,5	2,55
3,43	4,96	4,36	4,15	3,83	3,69	3,55	3,88
2,33	3,26	3,84	3,92	3,92	3,14	2,02	2,77
3,39	2,21	4,32	4,02	3,49	3,87	3,03	3,86
3,46	3,13	4,92	3,86	3,06	3,3	3,05	3,89
2,47	2,79	4,76	4,66	3,73	4,18	3,64	3,74
2,4	2,1	4,51	4,08	3,09	3,04	2,84	2,4
3,43	2,53	4,69	3,94	3,76	2,87	2,09	3,37
2,86	2,08	4,66	4,2	3,71	4	3,13	2,9
2,91	2,31	4,92	4,32	3,48	4,2	2,48	3,7

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
3,61	2,44	4,1	4,46	3,02	3,44	2,68	3,81
3,06	4,36	4,39	3,85	4	3,3	2,4	3,39
2,39	2,75	4,01	4,69	3,19	2,9	2,73	3,51
3,3	2,63	4,33	4,68	3,4	4,14	2,57	2,53
3,45	4,45	4,78	4,66	3,87	3,02	3,79	3,71
3,45	2,08	4,79	4,22	3,18	3,78	3,34	2,45
2,84	3,12	4,22	4,18	3,5	3,43	3,23	3,74
2,69	3,54	4,87	4,19	3,16	3,91	2,54	3,45
2,66	3,85	4,65	4,46	3,03	3,48	3,5	3,17
2,95	2,55	4,01	4,59	3,83	2,97	3,51	2,75
3,2	2,95	3,86	4,15	3,91	3,95	3,53	3,61
3,31	5	3,81	3,83	3,03	3,41	2,11	3,58
2,76	3,86	4,71	4,68	3,02	3,84	3,61	3,37
3,81	2,62	3,8	4,54	3,93	2,87	3,46	2,09
3,98	2,14	4,25	4,31	3,32	3,29	2,31	2,61
3,1	2,16	4,29	4,28	3,63	4,13	3,08	3,61
3,19	2,43	4,97	4,14	3,03	3,04	2,67	2,32
2,43	3,45	4,49	4,54	3,17	3,88	3,7	2,37
2,89	2,96	3,82	4,58	3,09	3,96	2,36	2,79
3,07	5	4,63	4,07	3,13	3,09	3,37	3,47
3,24	2,27	4,13	4,68	3,18	3,56	2,68	3,42
3,89	4,61	4,93	4,16	3,86	2,92	2,55	2,55
2,68	3,96	4,68	4,57	3,76	3,98	2,3	3,93
3,42	4,73	4,1	3,85	3,01	3,34	3,64	3,66
3,88	3,81	4,45	4,02	3,05	3,78	2,82	3,74
2,36	2,97	4,13	4,59	3,48	3,87	2,7	2,82
3,82	3,49	4,83	4,16	3,91	3,07	2,48	3,6
2,11	4,61	4,77	3,88	3,59	4,1	3,77	2,48
2,6	2,02	4,87	4,35	3,81	3,27	3,63	2,17
2,76	3,82	3,84	3,92	3,82	3,54	3,03	2,89

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
3,57	4,59	4,39	3,81	3,08	3,4	3,27	2,91
2,47	3,88	3,85	4,39	3,83	3,73	2	3,67
3,13	4,37	4,7	3,8	3,55	3,02	3,06	3,92
2,78	2,1	4,14	4,26	3,52	3,84	3,17	2,87
2,56	3,5	4,05	4,5	3,58	4,02	3,02	2,6
3,62	4,85	4,04	4,28	3,91	3,71	2,65	3,94
2	2,99	4,33	4,43	3,4	3,48	2,71	3,54
2,15	4,87	4,75	4,05	3,56	3,03	2,64	3,87
3,28	4,56	4,57	3,83	3,1	3,17	3,78	3,82
2,06	2,27	4,69	4,51	3,93	3,08	3,19	3,28
3,98	4,38	4,45	4,46	3,22	3,15	2,97	3,24
3,7	2,54	4,91	3,8	3,35	3,72	2,28	3,73
3,19	3,91	4,17	4,02	3,62	2,96	3,3	3,73
3,74	3,44	4,18	4,27	3,9	2,89	3,62	2,27
2,45	2,66	4,01	4,62	3,3	4,01	3,23	4
3,16	4,25	4,56	4,24	3,36	3,73	3,54	2,69
3,42	4,52	4,67	4,3	3,91	3,37	3,05	2,91
3,64	2,22	4,46	4,6	3,02	2,82	2,44	2,39
2,07	3,69	4,99	4,01	3,12	3,06	2,34	3,18
3,16	2,98	3,81	4,54	3,07	3,94	3,39	2,4
2,23	3,06	3,8	4,39	3,23	3,35	2,95	3,02
3,22	4,81	4,88	4,29	3,98	3,66	2,64	3,74
3,31	4,09	4,73	4,7	3,87	3,09	3,68	3,32
2,11	4,47	4,93	4,23	3,2	3,29	2,86	3,99
3,8	2,87	4,14	4,25	3,97	3,3	3,28	2,1
2,88	2,49	4,98	4,24	3,05	3,8	2,57	2,81
2,35	4,86	4,8	4,35	3,46	3,52	3,06	3,97
2,87	2,54	4,31	4,55	3,02	4,11	3,5	2
3,8	2,42	4,88	3,92	3,1	4,15	2,51	3,49
3,47	4,43	4,38	4,31	3,17	3,02	2,2	3,85

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
3,45	4,95	4,19	4,13	3,83	2,8	2,23	3,17
3,33	4,69	3,85	4,3	3,24	4,16	2,66	2
2,87	3,97	4,21	4,67	3,49	2,98	3,65	2,99
3,88	2,45	4,44	3,86	3,36	2,82	2,45	2,28
2,01	4,68	4,62	4,53	3,58	3,33	3,15	2,41
2,5	3,28	4,64	4,03	3,76	4,02	2,85	3,18
2,95	2,92	4,14	4,62	4	3,91	2,8	3,43
3,64	4,07	4,89	3,92	3,23	2,89	3,5	2,44
2,29	4,28	4,25	4,34	3,39	3,1	3,49	2,6
2,81	3,76	4,56	4,31	3,1	3,5	2	3,3
3,42	3,76	4,14	4,1	3,93	3,85	3,39	3,75
2,41	2,65	4,08	4,34	3,7	3,34	3,71	2,94
2,46	3,79	4,4	3,85	3,21	3,32	2,95	3,91
2,08	2	4,29	4,38	3,92	3,72	2,28	3,12
2,72	2,8	3,8	4,6	3,78	3,58	3,47	3,8
3,76	3,44	4,48	4,04	3,57	3,77	3,38	3,74
2,38	4,39	4,25	3,95	3,01	3,48	3,26	3,66
3,83	4,63	4,62	3,91	3,27	3,5	2,76	2,03
3,71	2,05	3,85	3,96	3,91	3,78	2,66	2,07
2,89	3,56	3,91	4,47	3,35	3,16	3,77	2,79
3,06	4,46	4,33	4,22	3,6	2,93	3,63	3,13
2,5	3,9	4,39	3,97	3,86	4,2	2,93	3,59
2,07	2,87	4,45	3,93	3,66	3,93	3,16	2,93
3,68	2,05	4,06	4,62	3,91	3,75	3,25	3
3,58	4,15	4,11	4,13	3,14	3,85	3,26	3,12
3,4	3,55	4,81	4,11	3,4	3,54	3,51	2,05
2,81	3,9	4,17	4,1	3,23	2,98	3,3	3,42
3,73	4,94	3,82	4,04	3,86	2,8	3,53	3,21
3,99	3,84	4,26	3,81	3,38	2,85	2,38	2,79
2,46	4,41	4,53	4,04	3,66	3,37	2,08	3,43

2018 (1st semester)	2018 (2nd semester)	2019 (1st semester)	2019 (2nd semester)	2020 (1st semester)	2020 (2nd semester)	2021 (1st semester)	2021 (2nd semester)
2,97	4,59	4,99	4,19	3,07	3,52	2,7	3,32
3,75	3,71	4,42	4,64	3,89	3,64	3,15	2,17
3,66	4,66	3,82	4,22	3,3	3,12	3,43	2,52
3,63	2,47	3,89	3,88	3,56	3,65	3,37	2,89
3,53	3,51	3,97	3,98	3,54	3,44	2,75	2,56
2,54	3,06	3,8	4,02	3,3	3,83	2,14	3,99
2,34	2,48	4,46	4,56	3,39	3,31	3,69	3,79
3,15	3,62	4,21	4,68	3,88	3,52	2,36	3,85
3,04	2,41	4,6	3,99	3,22	3,76	2,29	3,69
3,96	4,49	4,42	3,81	3,98	3,96	3,78	3,42
2,43	3,96	4,24	3,99	3,3	3,29	3,05	3,45

Appendix 2. Source code

```

import pandas as pd
import numpy as np
import statistics as st
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import warnings
warnings.filterwarnings('ignore')
from sklearn.datasets import load_diabetes
from matplotlib import style
from scipy import stats
from scipy.stats import pearsonr
# Import data frame
df=pd.read_excel('Appendix 1')
print(df)
# Change the names of columns for better analysis and use of python code
df.columns=["ps18","ss18","ps19","ss19","ps20","ss20","ps21","ss21"]
# Descriptive statistic analysis
round(df.describe(),2)
# Box plot figure

```

```
df.plot.box()

# Inferencial analysis :

# 2018

# Dispersion diagram:

print('Pearson Correlation: ', df['ps18'].corr(df['ss18'],
method='pearson'))

fig, ax = plt.subplots(1, 1, figsize=(6,4))
ax.scatter(x=df.ps18, y=df.ss18, alpha= 0.8)
ax.set_xlabel('First semester')
ax.set_ylabel('Second semester');

#Pearson

pearson_r,pearson_p = stats.pearsonr(df["ps18"], df["ss18"])
print(f"r:{pearson_r} - p:{pearson_p}")

#Spearman

spearman_r,spearman_p = stats.spearmanr(df["ps18"], df["ss18"])
print(f"rho:{spearman_r} - p:{spearman_p}")

# 2019

# Dispersion diagram:

print('Pearson Correlation: ', df['ps19'].corr(df['ss19'],
method='pearson'))

fig, ax = plt.subplots(1, 1, figsize=(6,4))
ax.scatter(x=df.ps19, y=df.ss19, alpha= 0.8)
ax.set_xlabel('First semester')
ax.set_ylabel('Second semester');

#Pearson

pearson_r,pearson_p = stats.pearsonr(df["ps19"], df["ss19"])
print(f"r:{pearson_r} - p:{pearson_p}")

#Spearman

spearman_r,spearman_p = stats.spearmanr(df["ps19"], df["ss19"])
print(f"rho:{spearman_r} - p:{spearman_p}")

# 2020

# Dispersion diagram:

print('Pearson Correlation: ', df['ps20'].corr(df['ss20'],
method='pearson'))

fig, ax = plt.subplots(1, 1, figsize=(6,4))
ax.scatter(x=df.ps18, y=df.ss18, alpha= 0.8)
ax.set_xlabel('First semester')
ax.set_ylabel('Second semester');

#Pearson

pearson_r,pearson_p = stats.pearsonr(df["ps20"], df["ss20"])
```

```

print(f"r:{pearson_r} - p:{pearson_p}")
#Spearman
spearman_r,spearman_p = stats.spearmanr(df["ps20"], df["ss20"])
print(f"rho:{spearman_r} - p:{spearman_p}")
# 2021
# Dispersion diagram:
print('Pearson Correlation: ', df['ps21'].corr(df['ss21'],
method='pearson'))
fig, ax = plt.subplots(1, 1, figsize=(6,4))
ax.scatter(x=df.ps21, y=df.ss21, alpha= 0.8)
ax.set_xlabel('First semester')
ax.set_ylabel('Second semester');
#Pearson
pearson_r,pearson_p = stats.pearsonr(df["ps21"], df["ss21"])
print(f"r:{pearson_r} - p:{pearson_p}")
#Spearman
spearman_r,spearman_p = stats.spearmanr(df["ps21"], df["ss21"])
print(f"rho:{spearman_r} - p:{spearman_p}")
# General correlation analysis
heatmap=sns.heatmap(df.corr(), vmin=-1, vmax=1, annot=True)
heatmap.set_title('Correlation Heatmap', fontdict={'fontsize':12},
pad=12)
# Correlation matrix
df.corr()

```

