

Estadística 1, Grupo H1

Proyecto final: Grupo 3

Cristian Rey, Fabian Perez, Santiago Bolaños

16 de febrero de 2023

Índice

1. Introducción	2
2. Metodología	2
3. Resultados	3
4. Discusión	9
5. Conclusión	9
6. Referencias	10
7. Anexos	10

1. Introducción

En la vida académica, los estudiantes a lo largo de su formación se ven puestos a prueba ante diferentes retos y al desarrollo de nuevos conocimientos, con los cuales se busca estar en una constante superación tanto personal como académica, es por esto que, para poder seguir en un constante aprendizaje, tienden a realizar ciertas acciones y hábitos con el fin de mejorar sus rendimientos en el área de estudio. Uno de esos hábitos es generar relaciones de amistad con sus demás compañeros de grupo, con el fin de apoyarse en determinados momentos en estos, con la petición de favores, estudio colaborativo, trabajos en grupo, etc. La finalidad del presente proyecto de investigación es determinar, a partir de un muestreo realizado en un salón de clase, una correlación entre las afinidades de un estudiante y como estas influyen en su rendimiento académico, ya sea de forma positiva o negativa. De igual manera, se realizaron una serie de preguntas individuales sobre la personalidad de los estudiantes en el ámbito académico.

2. Metodología

Los datos de estudio utilizados fueron tomados a través de una encuesta, que presentaba ciertas preguntas que ayudaban a responder la pregunta investigación. La encuesta fue respondida por 23 estudiantes, los cuales, a partir de una serie de opciones, registraron su afinidad con cada uno de sus compañeros de clase. Posteriormente, respondieron una serie de preguntas que buscaban saber su desempeño actual en la carrera que cursan, respondiendo preguntas como: ¿Que tan buena es su participación en clase?, ¿Cuál es su promedio ponderado de la carrera?, ¿Que tan motivado te sientes actualmente con la carrera que estudias?, ¿Qué tan seguido estudias con tus compañeros de clase fuera del horario escolar? y ¿Cuántas horas por semana dedicas a estudiar fuera de clase al día?

Posteriormente, los datos de las encuestas registradas fueron guardados en un archivo .csv, en el cual se trabajaron los datos. Uno de los requerimientos para nuestro proyecto era el de anonimizar los datos suministrados por cuestiones de privacidad de los estudiantes, proceso que se realizó asignándole un nombre por defecto a cada uno de los estudiantes (persona_1, persona_2, etc). Otro procedimiento que realizamos en la limpieza de datos, fue el de modificar el tipo de formato de los datos de promedio, de manera que no fueran datos tipo string, sino datos de tipo numérico.

En la etapa de construcción de modelos, se visualizaron los datos limpiados, se realizó su respectivo proceso de graficación, en donde se utilizaron diversas gráficas como diagramas de barras, diagramas de torta, diagramas de caja, etc, con el fin de ordenar la información y buscar tendencias y correlaciones entre variables. Finalmente, en la etapa de discusión

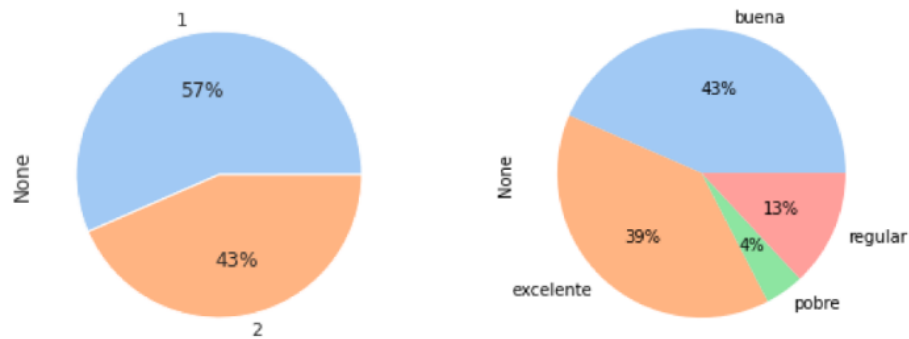
de resultados se responde a la pregunta central de la investigación y se plantean mejores soluciones para trabajos futuros.

3. Resultados

Al realizar un análisis exploratorio de datos (EDA), se obtuvo la siguiente tabla obtenida del archivo .csv, con la cual, se obtuvieron las gráficas expuestas a continuación:

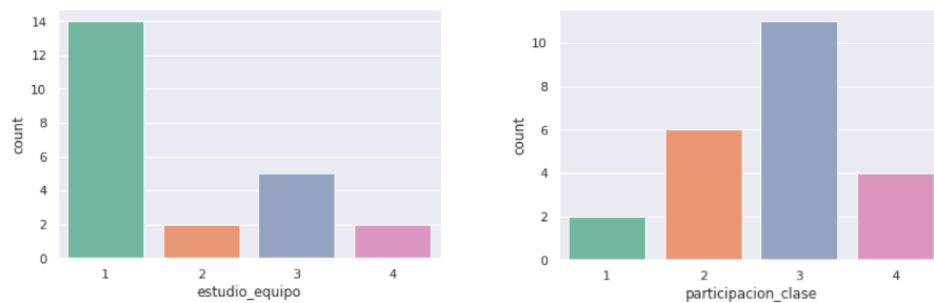
	Person_1	Hace cuanto lo/la conoce: (1)	Person_2	Hace cuanto lo/la conoce: (2)	Person_3	Hace cuanto lo/la conoce: (3)	Person_4	Hace cuanto lo/la conoce: (4)	Person_5	Hace cuanto lo/la conoce: (5)	...	Person_24	Hace cuanto lo/la conoce: (24)	Person_25	Hace cuanto lo/la conoce: (25)	promedio	estudio_indepe	teamwork	est
0	7	NaN	1	NaN	3	0.0	1	NaN	1	NaN	...	1	NaN	4	20.0	4.00	3	3	
1	1	NaN	7	NaN	1	NaN	3	10.0	1	NaN	...	1	NaN	1	NaN	3.90	2	3	
2	3	0.0	1	NaN	7	NaN	1	NaN	2	10.0	...	1	NaN	3	10.0	3.30	3	4	
3	1	NaN	2	10.0	1	NaN	7	NaN	3	40.0	...	1	NaN	1	NaN	4.03	0	4	
4	2	10.0	1	NaN	2	10.0	3	40.0	7	NaN	...	1	NaN	4	20.0	4.00	3	3	
5	1	NaN	1	NaN	1	NaN	2	10.0	2	20.0	...	1	NaN	1	NaN	4.10	2	2	
6	1	NaN	1	NaN	4	20.0	1	NaN	2	10.0	...	1	NaN	1	NaN	4.10	2	2	
7	1	NaN	1	NaN	1	NaN	6	40.0	2	10.0	...	1	NaN	1	NaN	4.10	1	3	
8	2	40.0	1	NaN	2	0.0	2	40.0	2	40.0	...	1	NaN	1	NaN	4.04	4	3	
9	1	NaN	1	NaN	1	NaN	1	NaN	1	NaN	...	3	0.0	1	NaN	3.20	0	2	
10	1	NaN	1	NaN	1	NaN	6	40.0	6	20.0	...	1	NaN	1	NaN	3.84	1	4	
11	3	20.0	1	NaN	4	40.0	3	10.0	5	20.0	...	1	NaN	1	NaN	4.30	3	4	
12	1	NaN	1	NaN	1	NaN	1	NaN	1	NaN	...	1	NaN	1	NaN	3.99	1	3	
13	2	10.0	2	0.0	1	NaN	2	40.0	3	40.0	...	2	0.0	2	10.0	4.45	2	3	
14	4	10.0	2	0.0	1	NaN	1	NaN	5	20.0	...	1	NaN	1	NaN	3.80	2	4	
15	3	40.0	1	NaN	3	0.0	2	0.0	2	0.0	...	1	NaN	1	NaN	4.09	2	3	
16	2	0.0	3	0.0	1	NaN	2	10.0	2	0.0	...	1	NaN	3	40.0	3.70	3	4	
17	5	40.0	1	NaN	1	NaN	1	NaN	5	20.0	...	1	NaN	1	NaN	4.10	3	3	
18	1	NaN	1	NaN	1	NaN	1	NaN	5	20.0	...	1	NaN	1	NaN	4.10	3	4	
19	1	NaN	1	NaN	1	NaN	1	NaN	1	NaN	...	2	10.0	1	NaN	4.20	2	4	
20	1	NaN	1	NaN	1	NaN	1	NaN	1	NaN	...	7	NaN	1	NaN	3.89	0	1	
21	1	NaN	1	NaN	3	10.0	1	NaN	3	10.0	...	1	NaN	7	NaN	3.89	2	4	

Basándonos en las preguntas "¿Qué tan motivado/a te sientes para seguir estudiando y alcanzar tus metas profesionales?" "¿Cómo describirías tus habilidades para trabajar en equipo?" se obtuvieron los siguientes diagramas de torta:

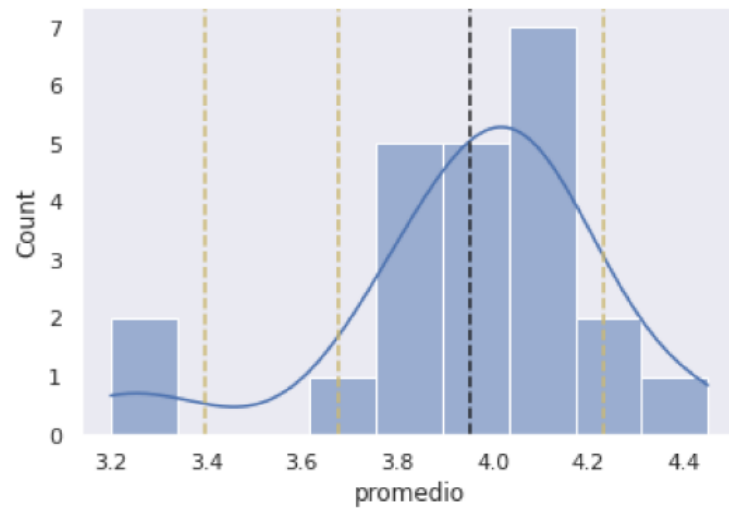


en donde se puede evidenciar solo dos respuestas registradas por parte de los estudiantes en la pregunta sobre motivación (regularmente motivado y muy motivado). Además, en la gráfica que representa la pregunta sobre el trabajo en equipo, podemos observar que la mayoría considera que tiene un buen o excelente desempeño trabajando en equipo, mientras que el 13% y el 4% de las respuestas consideran que sus habilidades en el trabajo grupal es regular y pobre respectivamente.

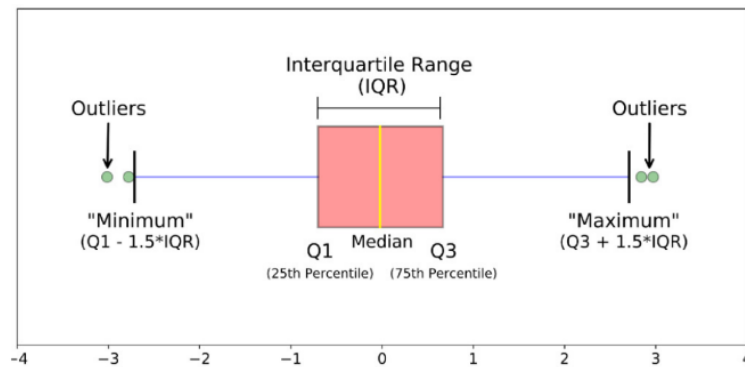
En la siguiente gráfica vemos dos diagramas de barras que representan la frecuencia de la participación en clase y la regularidad con la que los estudiantes estudian en equipo, siendo 1 el valor mas bajo (participación baja o estudio grupal nulo) y 4 el valor mas alto (participación alta o estudio grupal frecuente).



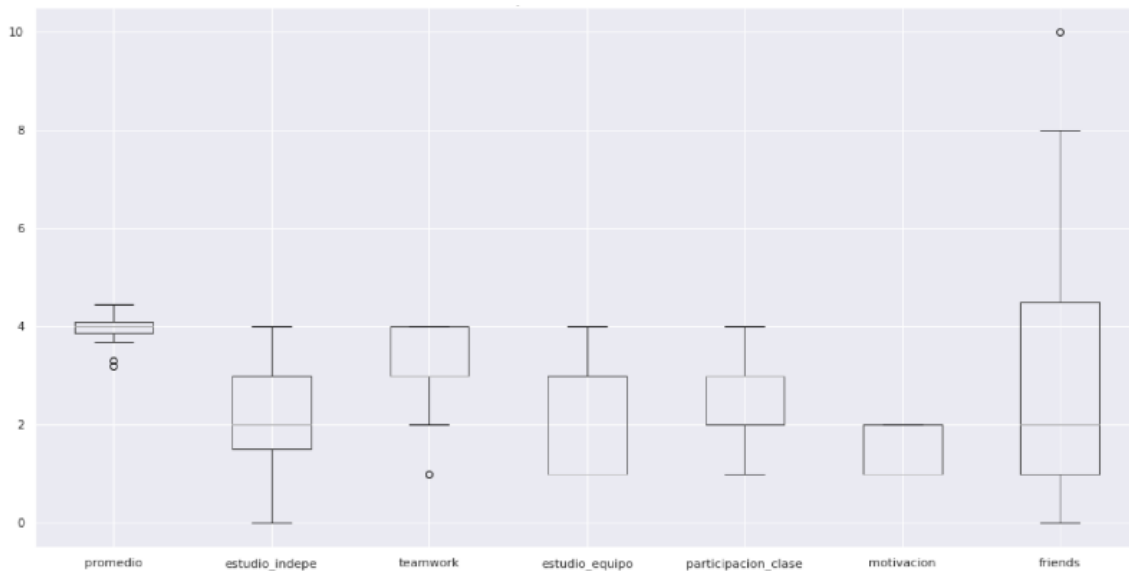
Continuando con las tablas de frecuencias, en la siguiente se registran los datos de la frecuencia del promedio, en donde la linea negra punteada representa la media y la gráfica, a pesar de que la muestra era muy pequeña, tiende a tener una forma normal.



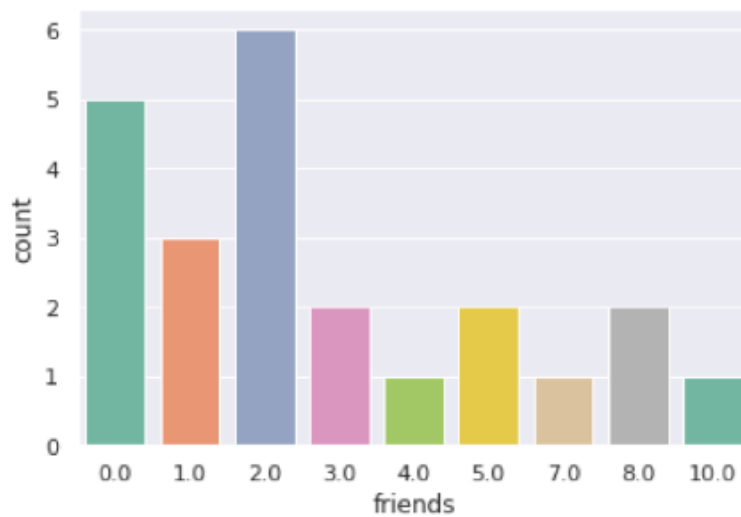
Un gráfico que nos permitió obtener información relevante sobre los outliers (datos atípicos) fue el diagrama de caja o boxplot. En la siguiente imagen se explica de manera general dicho gráfico. EL área de color rojo, representa el rango intercuartilico de los datos, lo que se encuentra a la izquierda de este, son los datos que se registran en el primer cuartil Q1 (percentil 25) y lo que se encuentra a la derecha son los datos ubicados en el tercer cuartil Q3 (percentil 75). De esta manera, todos los puntos que se encuentran afuera del primer o tercer cuartil, representan los outliers en la toma de datos.



Una vez entendido el diagrama de caja, podemos observar los diagramas resultantes de las variables planteadas en la encuesta.

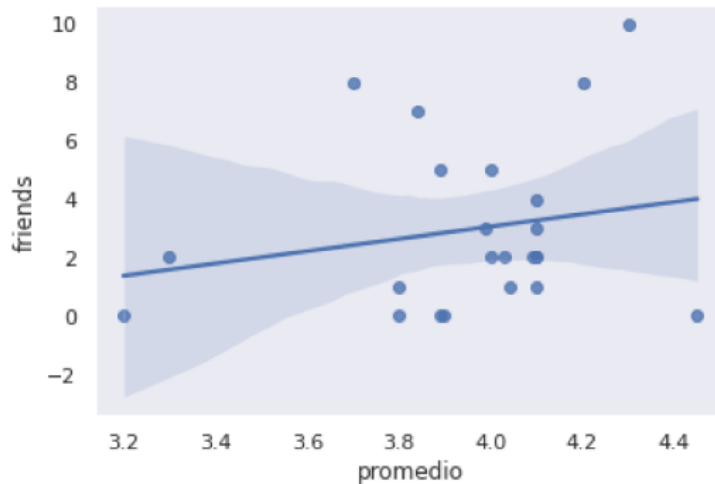


Teniendo en cuenta el concepto de "feature engineering", se creó una nueva variable llamada "friends" que representa el número de amigos que tiene el estudiante. A partir de esta, se creó un diagrama de barras que muestra la frecuencia de amigos en los estudiantes. Para esto se consideró como "amigo", las variables de las opciones: "Hablamos habitualmente", "Somos amigos", "Somos muy amigos".

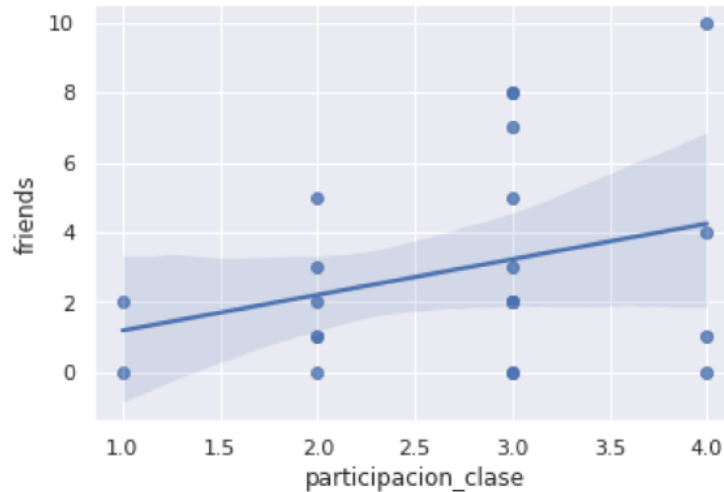


El siguiente diagrama de puntos muestra una correlación positiva débil entre la variable

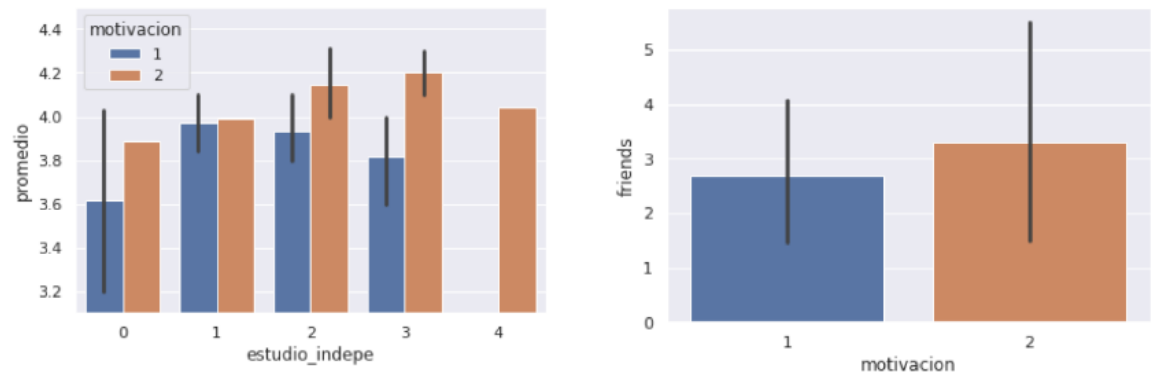
friends y el promedio académico de los estudiantes. A partir de una regresión lineal podemos observar mejor la tendencia de los datos, lo sombreado corresponde al intervalo de confianza.



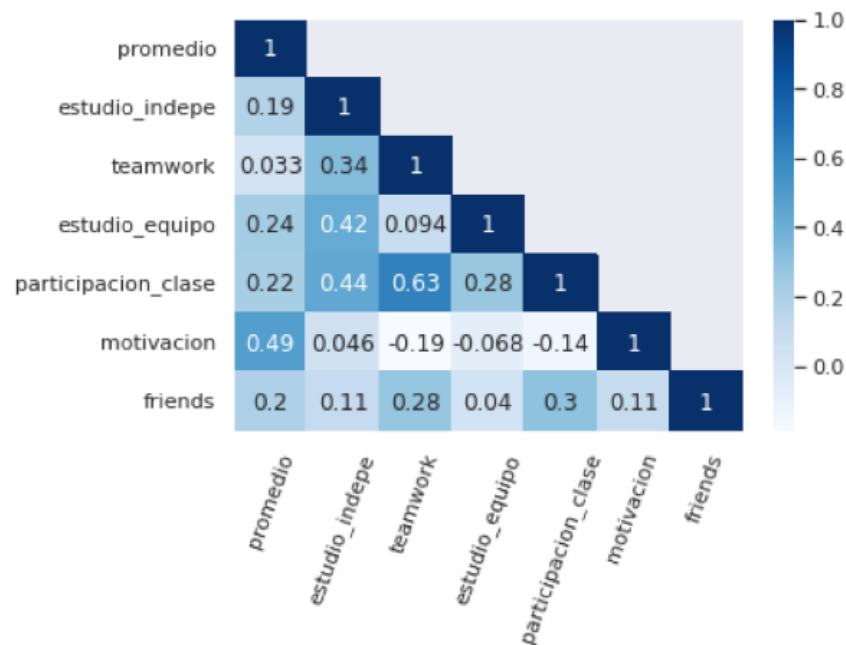
De la misma manera, se implementó el siguiente diagrama de puntos con la variable friends y la participación en clase.



Tomando en cuenta la variable de promedio, en donde solo se obtuvieron dos respuestas (muy motivado y regularmente motivado), se consideró un análisis con el promedio del estudiante, obteniendo el siguiente diagrama.



Finalmente, el diagrama de correlaciones presentado a continuación muestra que tan ligadas están las variables y a proporcionalidad de estas relaciones. Los valores del diagrama varían entre -1 y 1, porque se pueden tomar como una probabilidad. Por ejemplo, un valor de 0.2 representa una correlación positiva que representa un valor de 0.2 de relación. Entre mas cercano sea el valor a 1, implica que es mas probable que ocurra el evento. Este gráfico nos permite saber la tendencia general de la muestra tomada.



4. Discusión

Un factor importante a tener en cuenta es el hecho de que el tamaño de la muestra, al ser pequeño, representa un margen de error muy grande, lo cual puede implicar que las conclusiones obtenidas pueden ser erróneas, por esta razón se considera que sería mejor trabajar con una muestra de mayor tamaño de estudiantes para así llegar a obtener conclusiones mucho más precisas, se conservaron los outliers obtenidos debido a la falta de muestras.

Para el caso de estudio, la información obtenida a partir de los datos suministrados es coherente con el ámbito estudiado. A partir de esto se obtuvieron una serie de correlaciones que sustentan nuestra pregunta hipótesis, los datos sugieren que existe una relación positiva entre el número de amistades y el rendimiento académico, lo cual puede tener implicaciones para la forma en que las escuelas fomentan las relaciones sociales entre los estudiantes. También se obtuvieron datos interesantes que aportaron para ver el estado actual de los estudiantes sobre la carrera y que tanta motivación tienen por aprender. Como trabajo futuro, se plantea la idea de realizar un tratamiento diferente de datos, representándolos en grafos mediante una matriz de adyacencia.

5. Conclusión

En las respuestas conseguidas, se evidenció una concentración en la variable de motivación en valores altos, es decir, la todos los estudiantes se encuentran en un nivel de motivación alto o muy alto. Tras obtener la variable friends en el proceso de feature engineering, se visualiza una tendencia positiva débil en la relación entre el número de amigos y el promedio, por lo que se concluye que a mayor número de amigos tenga el estudiante, es probable que mayor sea su promedio.

De igual manera, se concluye también que a mayor número de amigos, también aumenta la participación en clase del estudiante. Una de las relaciones mas fuertes obtenidas a partir de las preguntas individuales realizadas, fue que entre mayor sea la participación en clase, mejores son las habilidades de trabajo en equipo, puesto que hay un valor de 0.63 registrado en el diagrama de correlaciones, siendo este el valor mas alto. El promedio académico también se ve afectado por el nivel de motivación, ya que a mayor nivel de motivación mayor será el promedio. En el diagrama de correlaciones se evidencian algunos valores negativos, lo que representa una relacion indirectamente proporcional, la correlacion negativa mas alta es la que existe entre el nivel de motivación y las habilidades para trabajar en equipo, esto indica que en un valor de 0.19, a mayor sea la motivación del estudiante, peores van a ser sus habilidades en el trabajo en equipo y viceversa.

6. Referencias

- Kramarae, C., & Treichler, P. A. (1990). Power relationships in the classroom.
- Rojas Ospina, David. (2019). Relaciones de Clase en el Sistema Universitario y su efecto sobre el Rendimiento Académico: El Caso de Bogotá.. Multidisciplinary Journal of Educational Research. 9. 1. 10.17583/remie.2019.3999
- Singh, A. K., Lakshmanan, K. (2021). PILHNB: Popularity, interests, locatio used hidden Naiver Bayesian-based model for link prediction in dynamic social networks. Neurocomputing, 461, 562-576. <https://doi.org/10.1016/j.neucom.2021.02.101>)

7. Anexos

- Repositorio de github con todos los recursos.