

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN

INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

Minería de Datos

Agrupamiento jerárquico y método de Ward

Práctica 4

Integrantes:

Jesús Tlamani Xochimitl

Francisco García Eduardo

Fecha de Entrega: 23 de mayo de 2024

Índice

1. DESCRIPCIÓN DEL ALGORITMO	2
2. METODOLOGÍA	2
3. PRUEBAS	3
3.0.1. Descripción del dendograma	3
3.1. Libros de excel creados apartir de los cortes del dendograma	3
3.1.1. Asignación de clusters de cada corte en los documentos de excel . .	4
3.2. Corte en altura 10	5
3.2.1. Para el grupo 1	6
3.2.2. Para el grupo 2	7
3.2.3. Para el grupo 3	8
3.2.4. Para el grupo 4	9
3.2.5. Para el grupo 5	10
3.2.6. Para el grupo 6	11
3.2.7. Para el grupo 7	12
3.2.8. Para el grupo 8	13
3.2.9. Para el grupo 9	14
3.3. Corte en altura 15	15
3.3.1. Para el grupo 1	15
3.3.2. Para el grupo 2	17
3.3.3. Para el grupo 3	19
3.3.4. Para el grupo 4	21
3.3.5. Para el grupo 5	23
3.4. Corte en altura 20	24
3.4.1. Para el grupo 1	25
3.4.2. Para el grupo 2	27
4. CONCLUSIONES	28
5. APÉNDICES	29
5.1. Código del Programa	29
5.2. Conceptos Usados	31
5.3. Manual de Usuario	32
5.4. Manual técnico	36
5.5. Referencias	37

1. DESCRIPCIÓN DEL ALGORITMO

El método de Ward es un algoritmo de agrupamiento jerárquico aglomerativo que se utiliza para identificar grupos naturales en un conjunto de datos. Se caracteriza por su enfoque en minimizar la varianza dentro de los clústeres al fusionarlos.

El algoritmo procede de la siguiente manera:

1. **Inicialización:** Cada punto de datos se considera un clúster individual.
2. **Cálculo de la distancia entre clústeres:** Se calcula una matriz de distancias entre todos los pares de clústeres, utilizando una medida de distancia adecuada, como la distancia euclidiana o la distancia de Manhattan.
3. **Fusión de clústeres:** Se fusionan los dos clústeres más cercanos en cada paso del algoritmo. La distancia entre dos clústeres se determina utilizando un criterio de fusión específico, que en el método de Ward se basa en la minimización de la suma de cuadrados de las diferencias entre los puntos y los centroides de los clústeres.
4. **Actualización de la matriz de enlace:** Se actualiza la matriz de enlace para reflejar la fusión de clústeres, lo que permite visualizar la jerarquía de agrupamiento mediante un dendrograma.
5. **Finalización:** El algoritmo continúa fusionando clústeres hasta que todos los puntos estén en un solo clúster.

El método de Ward es ampliamente utilizado debido a su capacidad para producir clústeres compactos y bien definidos, preservando la estructura de varianza de los datos originales. Esto lo hace útil en una variedad de aplicaciones, incluyendo la clasificación de datos biológicos, segmentación de clientes y análisis de datos médicos.

2. METODOLOGÍA

La metodología para realizar el análisis jerárquico de agrupamiento utilizando el método de Ward se describe a continuación:

1. **Carga de Datos:** Se inicia seleccionando un archivo CSV que contiene los datos de los estudiantes. Este archivo debe tener un formato tabular donde las filas representan a los estudiantes y las columnas representan las características de interés.
2. **Preprocesamiento de Datos:** Una vez cargados los datos, se realiza un preprocesamiento para seleccionar únicamente las columnas numéricas relevantes para el análisis de agrupamiento. Estas columnas pueden incluir la edad de los estudiantes, su género, su rendimiento académico, entre otras.
3. **Cálculo de Distancias:** Se calcula una matriz de distancias entre los puntos utilizando la distancia euclidiana al cuadrado entre las observaciones. Esto implica calcular la distancia entre cada par de puntos en el conjunto de datos.

4. **Algoritmo de Ward:** Se aplica el algoritmo de Ward para realizar el análisis jerárquico de agrupamiento. Este algoritmo fusiona iterativamente los clústeres más cercanos en función de la distancia entre ellos, manteniendo una jerarquía de agrupamiento.
5. **Generación de Grupos y Archivos Excel:** Después de aplicar el algoritmo de Ward, se generan grupos utilizando diferentes alturas de corte en el dendrograma. Para cada altura de corte, se asigna un grupo a cada punto en función de la altura de corte y se crean archivos Excel separados para cada grupo.
6. **Visualización:** Finalmente, se muestra el dendrograma con las líneas de corte y los grupos generados. Esto proporciona una visualización de la estructura jerárquica de los grupos identificados en los datos de los estudiantes.

3. PRUEBAS

3.0.1. Descripción del dendrograma

A continuación se presenta la imagen del dendrograma que nos arroja 3 cortes diferentes en distancias diversas, con los cuales obtendremos 3 nuevos documentos de Excel. Posteriormente, realizaremos los análisis correspondientes para cada línea que esté en los cortes.

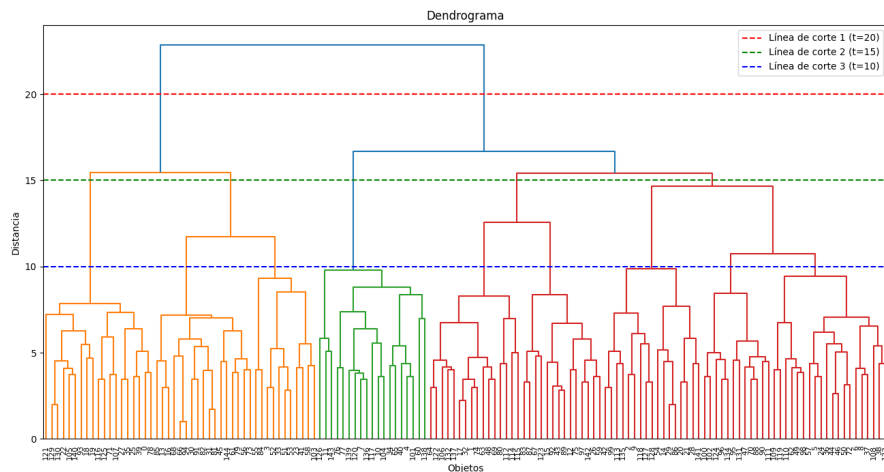


Figura 1: Imagen del dendrograma con los 3 cortes para realizar los analisis

3.1. Libros de excel creados apartir de los cortes del dendrograma

A continuación, se muestran los nuevos libros de Excel que han sido creados según las distancias en los cortes que hemos colocado en el código, las cuales son 10, 15 y 20. Estos documentos generados nos servirán para visualizar cómo se van asignando los datos a los clusters y cómo se agrupan los datos dependiendo de las líneas de corte que se tengan.

En cada línea de corte existen varios grupos, y cada grupo tiene su propia hoja de Excel, donde al final de la tabla se especifica a qué cluster pertenece.




 grupos_corte_10	13/05/2024 08:18 p. m.	Hoja de cálculo d...	60 KB
 grupos_corte_15	13/05/2024 08:18 p. m.	Hoja de cálculo d...	50 KB
 grupos_corte_20	13/05/2024 08:18 p. m.	Hoja de cálculo d...	41 KB

Figura 2: Imagen de los archivos de excel que se crearon apartir de los cortes en el dendograma

3.1.1. Asignación de clusters de cada corte en los documentos de excel

AG	AH	AI
GRADE	Grupo_20	
1	1	
1	1	
1	1	
1	1	
2	1	
3	1	
3	1	

Figura 3: Al final de cada tabla, se genera un nuevo atributo llamado Grupo, cada objeto comienza como un único cluster. En cada paso, se fusionan los dos clusters más similares hasta que todos los objetos están en un solo cluster, este proceso se repite en todos los documentos de excel,

3.2.1. Para el grupo 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	TUDENT ID	Student Age	Sex	hd high-scholarship	tyditional wdistic or spu	have a pialary	if avtion to theation type	ersae eduersae educr of sister/	rental statersae ocupae ocupatky study h(non-scieny (scientifici											
2	1	2	2	3	3	1	2	2	2	1	1	1	2	3	1	2	5	3	2	2
3	19	1	1	2	4	2	2	2	3	1	1	2	2	5	1	2	5	5	3	2
4	20	1	2	1	3	2	2	1	2	2	2	3	3	3	1	2	4	4	2	2
5	23	2	2	2	3	1	2	1	1	1	1	1	1	4	2	2	4	3	3	3
6	26	2	2	2	3	2	2	1	1	1	2	1	4	3	1	2	4	2	2	2
7	28	1	2	1	3	1	2	2	1	1	1	1	1	4	1	2	3	3	2	2
8	36	1	2	1	4	2	2	2	3	1	1	1	2	5	1	2	3	2	2	2
9	40	2	1	2	3	2	2	1	1	1	2	1	3	1	1	2	4	1	2	1
10	72	1	1	3	4	2	2	2	1	1	3	1	3	5	1	2	4	2	2	1
11	79	2	1	2	4	1	1	2	1	1	1	1	3	4	1	2	4	2	3	3
12	94	1	2	2	1	1	2	1	1	1	2	1	2	4	1	2	3	5	2	1
13	106	1	2	2	4	1	1	2	1	1	1	2	2	4	1	2	4	2	3	2
14	108	1	1	2	4	1	1	2	1	1	3	1	3	5	1	2	4	2	2	2
15	117	2	1	2	5	2	2	2	1	1	1	1	1	5	1	2	4	3	3	3
16	122	2	1	2	3	1	1	1	1	2	1	1	1	5	3	2	4	5	3	3
17	126	1	1	2	5	1	1	2	1	1	3	1	1	5	1	2	3	3	2	3
18	130	1	1	2	3	1	1	2	1	1	1	1	1	3	1	2	4	4	2	3
19	131	1	1	2	3	1	1	2	1	1	1	1	2	3	1	2	4	4	2	2
20	141	2	1	2	3	1	1	2	1	1	2	1	2	2	2	2	4	3	3	2
21																				
22																				
23																				

Figura 5: Resumen del grupo 1: Total de 19 personas, distribuidas en edades de 18 a 21 años (11 personas) y de 22 a 25 años (8 personas). La composición de género incluye 11 mujeres y 8 hombres. En cuanto a becas, 1 persona no tiene beca, 10 cuentan con una beca del 50 %, 6 con el 75 % y 2 con una beca completa. En relación a relaciones sentimentales, 6 alumnos tienen pareja y 13 no. Con respecto al estudio semanal, 1 persona no estudia, 6 estudian menos de 5 horas, 6 entre 6 y 10 horas, 3 entre 11 y 20 horas y 3 más de 20 horas. Sobre la asistencia a seminarios y clases, 14 asisten regularmente a seminarios, 14 siempre asisten a clases y 5 asisten a veces. En preparación para exámenes, 16 se preparan solos y 3 con amigos. Sobre la toma de apuntes, 1 nunca toma apuntes, 7 a veces y 11 siempre. Los promedios académicos del último semestre se distribuyen como sigue: 5 entre menor a 2.0, 9 de 2.0 a 2.49, 4 de 2.50 a 2.99, y 1 de 3.0 a 3.49. Las calificaciones esperadas al graduarse son: 5 menor a 2.0, 5 de 2.0 a 2.49, y 9 de 2.50 a 2.99. Finalmente, las calificaciones finales son: 1 Fail, 6 DD, 2 DC, 4 CC, 2 CB, 1 BB, 2 BA y 1 AA.

	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	conference	acts/activi	dance to ch	to midten	to midten	notes in en	ing in cla	interest aip	classroom	int average	de point a	COURSE ID	GRADE
2	1	1	1	1	1	3	2	1	2	1	1	1	1
3	2	1	2	1	1	3	1	3	3	3	3	1	2
4	1	2	1	1	1	3	2	2	3	2	3	1	3
5	1	1	1	1	2	3	1	2	3	3	3	1	3
6	1	1	1	1	1	2	1	3	2	1	2	1	3
7	1	1	2	1	1	3	1	2	1	2	1	1	1
8	1	1	2	1	1	2	1	3	1	1	1	1	1
9	1	1	1	1	1	2	2	2	3	2	1	1	1
10	1	1	1	1	1	2	3	3	3	2	2	3	6
11	1	1	2	1	1	3	3	2	2	2	2	4	4
12	1	1	1	1	1	3	3	3	2	2	2	6	4
13	1	1	2	1	2	2	1	3	1	3	3	7	7
14	1	1	1	2	1	2	3	2	1	4	2	7	6
15	2	1	1	1	1	3	3	3	1	2	3	8	1
16	1	3	1	2	1	1	1	1	1	1	1	8	0
17	2	1	1	1	2	3	2	3	1	1	3	9	2
18	2	1	1	2	2	3	2	3	2	2	3	9	3
19	2	1	1	1	2	3	1	3	2	2	3	9	1
20	1	1	1	1	1	2	1	2	1	3	3	9	5
21													
22													
23													
<	>	Grupo_1	Grupo_2	Grupo_3	Grupo_4	Grupo_5	Grupo_6	Grupo_7	Grupo_8	Grupo_9	+	:	◀

Figura 6: Continuacion de la figura 1

3.2.2. Para el grupo 2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	TUDENT ID	tudent Ag	Sex	ed high-scholarship	ditional wctistic or spu	have a salary if avtion to theation type	ersae educ	ersae educ	of sister/rental statersae ocupae ocupatly study h									
2	14	2	1	2	5	2	2	2	1	1	1	2	2	2	1	2	3	1
3	17	1	1	2	5	2	1	2	1	1	1	2	2	4	1	2	3	2
4	31	2	2	2	5	1	1	1	1	1	2	1	1	5	1	2	4	2
5	32	3	2	2	3	1	2	2	1	1	2	1	2	4	1	2	3	2
6	46	1	2	2	3	2	2	1	4	1	1	2	3	5	1	2	4	3
7	57	2	2	2	3	2	1	2	1	1	1	1	3	5	1	2	4	4
8	62	1	2	3	3	2	1	2	2	1	2	1	3	4	1	2	3	3
9	67	2	2	2	3	2	2	1	1	1	1	3	2	5	1	2	5	2
10	69	2	1	2	4	1	2	2	1	1	1	2	3	5	3	2	5	1
11	74	2	2	2	4	2	2	2	1	1	2	1	3	4	1	2	4	3
12	80	2	2	2	4	2	2	2	1	1	1	1	3	4	1	2	4	1
13	82	3	2	2	3	1	2	2	1	1	2	1	2	4	1	2	3	2
14	83	2	2	2	4	1	2	1	2	1	1	3	1	3	1	2	3	2
15	86	1	2	2	5	2	2	2	2	1	1	1	1	5	1	2	1	1
16	92	2	2	2	5	1	1	1	1	1	2	4	1	5	1	2	3	2
17	95	2	2	2	3	2	2	1	1	1	1	3	2	5	1	2	5	2
18	145	1	1	1	5	2	2	2	3	1	1	3	1	5	1	2	4	3
19																		
20																		
21																		
22																		
<	>	Grupo_1	Grupo_2	Grupo_3	Grupo_4	Grupo_5	Grupo_6	Grupo_7	Grupo_8	Grupo_9	+	:	◀					

Figura 7: Resumen del grupo 2: Total de 17 estudiantes, con una distribución de género de 4 femeninos y 13 masculinos. En cuanto a la edad, 5 están entre 18 a 21 años, 10 entre 22 a 25 años y 3 son mayores de 26 años. Sobre las becas, 7 tienen el 50 %, 4 el 75 % y 6 una beca completa. En relación sentimental, 6 estudiantes tienen pareja mientras que 11 no. Respecto a las horas de estudio semanal, 1 no estudia, 8 estudian menos de 5 horas, 4 entre 6 y 10 horas y 4 entre 11 y 20 horas. La asistencia a seminarios cuenta con 16 asistentes y 1 no asiste. Para la asistencia a clases, 13 siempre asisten y 4 a veces. En preparación para exámenes, 15 se preparan solos y 2 con amigos. En la toma de notas, 4 a veces toman notas y 13 siempre las toman. Los promedios del último semestre son: 9 entre 3.0 a 3.49 y 8 mayores que 3.49. Las calificaciones esperadas al graduarse son: 1 entre 2.0 a 2.49, 7 entre 2.50 a 2.99 y 9 entre 3.0 a 3.49. Las calificaciones finales se distribuyen como sigue: 2 DD, 5 CC, 1 CB, 5 BB, 3 BA y 1 AA.

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	(non-sciency (scientific) conferences/activities/dance to ch to midten to midten notes in ening in cl interest alp-classrooint average de point aCOURSE ID	GRADE													
2	2	1	1	1	2	1	1	3	2	3	3	4	2	1	1
3	2	2	1	1	2	1	1	3	2	3	3	4	3	1	1
4	2	2	1	1	1	2	1	2	3	3	3	5	4	1	5
5	2	3	1	1	1	1	1	3	3	2	3	4	3	1	3
6	2	2	1	1	1	1	1	2	2	2	1	4	3	1	3
7	2	3	1	1	1	1	2	3	2	3	3	5	4	1	5
8	1	2	1	3	2	1	1	3	2	2	2	4	4	1	5
9	2	2	1	1	1	1	1	3	2	2	3	5	4	2	5
10	2	2	1	1	2	1	1	2	2	3	2	4	3	3	5
11	1	1	2	1	1	1	2	3	2	3	1	5	3	3	6
12	2	2	1	3	1	1	1	3	3	3	2	4	4	4	3
13	2	2	1	1	1	1	1	3	3	2	3	5	4	5	3
14	2	3	1	1	1	1	1	3	3	2	3	4	3	5	7
15	1	1	1	1	1	1	1	3	2	3	3	4	3	5	4
16	2	2	1	1	1	1	1	2	3	3	2	5	4	6	6
17	2	3	1	1	1	1	1	3	2	2	3	5	4	6	6
18	1	1	1	1	1	2	1	3	2	3	1	5	4	9	3
19															
20															
21															
22															

Figura 8: Continuación del grupo 2

3.2.3. Para el grupo 3

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	TUDENT ID	tudent Ag	Sex	ed high-scholarship ty	ditional wctic or spu	have a	palary if avation to the	ation type	ersae eduersae educ	of sister/	rental statesae ocup					
2	4	1	1	1	3	1	2	1	2	1	2	1	2	5	1	2
3	24	3	2	2	2	1	1	2	5	1	1	1	4	3	1	2
4	34	2	1	2	3	1	2	1	1	1	1	1	2	5	1	2
5	42	3	2	2	3	1	2	2	2	1	2	1	4	2	1	2
6	52	2	1	3	3	1	1	2	1	1	1	1	2	5	1	2
7	53	2	1	2	3	1	2	2	2	1	1	1	3	4	1	2
8	54	2	1	2	3	2	1	2	2	1	1	1	1	5	1	2
9	56	3	2	2	3	1	2	1	4	1	2	1	1	5	3	2
10	59	3	2	2	3	2	2	1	3	1	1	1	3	5	2	2
11	85	3	2	3	3	1	2	1	3	1	2	4	2	5	3	2
12	104	1	2	1	4	2	2	1	5	1	2	1	3	4	1	2
13																
14																
15																
16																
17																
18																
19																
20																

Figura 9: Resumen del grupo 3: Consta de 11 estudiantes, 5 femeninos y 6 masculinos. Distribución por edades: 2 entre 18 a 21 años, 4 entre 22 a 25 años y 5 mayores de 26 años. En cuanto a becas, 9 tienen un 50 % de beca, 1 cuenta con el 75 % y 1 alumno no tiene beca. Relaciones sentimentales: 6 tienen pareja y 5 no. En estudio semanal, 4 no estudian ninguna hora, 2 estudian menos de 5 horas y 5 entre 6 y 10 horas. Todos los alumnos asisten a seminarios. Para la asistencia a clases, 7 siempre asisten y 4 asisten a veces. Preparación para exámenes: 9 se preparan solos y 2 nunca se preparan. En la toma de notas, 2 nunca toman notas, 4 a veces y 5 siempre. Los promedios del último semestre son: 1 entre menor a 2.0, 4 de 2.50 a 2.99, 3 de 3.0 a 3.49 y 3 mayores que 3.49. Las calificaciones esperadas al graduarse son: 3 de 2.0 a 2.49, 4 de 2.50 a 2.99 y 4 de 3.0 a 3.49. Las calificaciones finales son: 4 DD, 1 DC, 2 CC, 1 CB, 1 BB y 2 AA.

	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	ae ocupakly study h	non-sciety	(scientific	conferenc	cts/activid	ance to ch	to midten	to midten	notes in	ning in cla	interest ap	classroom	average	point a	COURSE ID	GRADE	
2	1	3	1	2	1	1	1	1	2	3	2	2	1	3	2	1	1
3	1	3	2	3	1	1	2	1	1	3	2	3	3	3	3	1	1
4	3	2	1	2	1	1	1	1	1	1	3	2	2	2	3	1	2
5	2	3	2	2	1	1	2	1	1	3	2	2	3	4	3	1	1
6	1	1	2	2	1	1	2	1	1	2	3	3	1	3	3	1	4
7	1	3	1	1	1	1	1	1	2	2	2	2	1	4	2	1	1
8	1	1	2	2	1	1	2	1	1	2	2	2	2	3	2	1	5
9	1	1	2	2	1	1	1	3	3	1	3	3	3	5	4	1	3
10	1	3	2	2	1	1	1	1	1	3	2	2	2	5	4	1	3
11	1	1	2	2	1	1	1	3	3	3	3	3	3	5	4	5	7
12	1	2	3	2	1	1	1	1	1	2	2	3	2	4	4	7	7
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	

Figura 10: Continuacion del grupo 3

3.2.4. Para el grupo 4

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	TUDENT ID	Student Age	Sex	ed high-scholarship	ditional wstic or spu	have a pialary if avtion to the	ation type	sae eduersae educ	of sister/	rental stater	sae ocupakly	study h						
2	5	2	2	1	3	2	2	1	3	1	4	3	3	2	1	2	4	2
3	8	1	1	2	3	1	1	1	2	2	3	4	3	1	1	4	3	1
4	11	1	1	1	3	2	2	2	3	2	3	3	4	2	1	3	2	1
5	12	1	1	1	4	1	1	2	4	2	3	5	5	1	1	3	2	3
6	35	1	2	1	3	2	2	1	2	1	1	3	3	3	1	2	4	2
7	41	1	2	1	3	2	2	2	2	1	1	2	3	1	1	2	5	1
8	61	2	1	2	3	2	2	2	5	2	1	6	1	3	1	2	3	1
9	66	1	2	2	3	2	2	1	3	1	1	2	4	3	1	2	4	1
10	77	2	2	1	2	2	1	2	2	1	1	5	4	3	1	4	3	2
11	78	1	2	1	2	2	2	1	2	2	2	6	5	2	1	4	2	2
12	102	1	2	2	4	2	2	1	2	1	3	2	3	1	1	2	4	2
13	105	1	1	2	3	2	2	2	1	2	2	4	4	1	1	3	2	2
14	118	3	1	1	3	2	1	2	1	1	2	4	4	1	1	1	3	2
15	121	2	1	1	3	1	1	1	2	2	3	3	3	2	1	3	4	1
16	127	1	1	1	4	1	1	1	3	2	3	4	6	2	1	4	3	4
17	137	1	1	2	3	1	1	1	1	2	3	3	3	3	1	3	4	2
18	139	1	1	2	4	1	1	1	5	2	3	3	1	1	2	4	5	2
19	140	1	1	2	4	1	2	1	2	2	3	3	3	1	1	5	4	1
20	144	2	1	2	4	1	1	1	5	2	3	4	4	1	1	3	3	2
21																		

Figura 11: Resumen del grupo 4: Total de 19 estudiantes, compuesto por 12 femeninos y 7 masculinos. Distribución etaria: 13 entre 18 a 21 años, 5 entre 22 a 25 años y 1 mayor de 26 años. En cuanto a becas, 2 tienen una del 25 %, 11 del 50 % y 6 del 75 %. Estado sentimental: 12 tienen pareja y 7 no. Horas de estudio semanal: 7 no estudian, 10 estudian menos de 5 horas y 1 entre 6 a 10 horas. Asistencia a seminarios: 14 asisten y 5 no. Asistencia a clases: 12 siempre asisten y 7 asisten a veces. Preparación para exámenes: 7 se preparan solos y 7 con amigos. Toma de notas: 1 nunca toma notas, 10 a veces y 8 siempre. Promedio acumulado de calificaciones del último semestre: 6 menores a 2.0, 10 entre 2.0 a 2.49, 2 entre 3.0 a 3.49 y 1 mayor a 3.49. Promedio esperado al graduarse: 6 menores a 2.0, 7 entre 2.0 a 2.49 y 6 entre 2.50 a 2.99. Calificaciones finales: 4 Fail, 4 DD, 5 DC, 2 CC, 2 CB y 2 AA.

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	non-sciency	scientific	conferences	acts/activities	ance to ch	to midten	to midten	notes in	ning in cla	interest ap	p-classro	int averag	de point a	COURSE ID	GRADE
2	1	1	1	1	1	2	1	2	2	2	1	2	2	1	1
3	2	2	1	1	1	3	1	3	2	2	1	1	1	1	2
4	1	1	1	1	2	1	1	2	2	2	2	1	1	1	2
5	3	3	1	3	1	3	2	3	1	3	3	4	3	1	0
6	1	1	1	1	2	1	1	2	2	3	1	4	1	1	2
7	2	2	1	1	1	1	1	3	1	2	1	2	3	1	1
8	2	2	1	1	1	1	1	1	3	3	1	2	1	1	2
9	1	2	1	3	2	2	1	2	1	2	1	2	2	1	2
10	3	2	1	1	1	3	2	2	1	2	3	2	2	4	4
11	2	2	2	1	2	3	1	2	2	2	2	1	1	4	7
12	2	2	1	1	1	1	1	3	3	2	1	2	3	7	7
13	2	2	1	1	1	1	1	2	3	2	2	1	1	7	3
14	2	3	2	1	2	1	1	2	3	2	1	2	3	8	1
15	1	1	1	1	2	2	1	3	3	3	2	2	2	8	1
16	2	3	1	1	1	2	1	3	3	3	1	2	2	9	3
17	2	2	2	1	1	3	1	3	2	2	1	2	2	9	0
18	2	3	2	3	1	2	1	3	2	3	1	1	3	9	0
19	1	2	2	1	2	2	1	2	3	2	1	1	2	9	0
20	2	1	1	1	1	2	1	2	1	2	1	5	3	9	4
21															

Figura 12: Continuación del grupo 4

3.2.5. Para el grupo 5

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	TUDENT ID	tudent Ag	Sex	ad high-scholarship	tydional w	cistic or spu	have a p	alary if av	ation to the	ation type	ersae edu	ersae educ	of sister/	rental st	ersae occu	
2	2	2	2	3	3	1	2	2	1	1	1	2	3	2	1	2
3	18	2	2	2	3	2	2	2	1	1	1	2	2	2	1	2
4	33	2	1	2	3	2	2	2	2	1	1	2	2	2	1	2
5	49	1	2	2	3	2	1	1	1	1	2	3	3	3	1	2
6	64	2	2	2	4	2	2	1	2	1	1	1	3	3	1	2
7	65	2	2	3	5	2	2	2	1	1	1	2	2	2	1	2
8	70	2	1	2	4	2	2	1	1	1	1	2	2	2	1	2
9	75	1	2	2	4	2	2	2	1	1	1	2	2	3	1	2
10	81	2	1	2	3	1	2	1	1	2	1	1	1	2	1	2
11	107	1	2	2	4	2	1	2	1	1	2	2	2	1	1	2
12	113	2	1	3	3	1	2	2	1	3	3	1	1	2	1	2
13	115	2	1	2	5	1	1	2	4	1	1	1	1	1	1	2
14	116	2	1	2	5	1	2	1	1	1	1	1	1	2	1	2
15	123	1	2	2	5	2	1	2	1	1	1	3	3	2	1	2
16	133	1	1	1	5	2	1	2	1	1	2	1	2	3	1	2
17	138	1	1	1	5	2	1	2	1	1	1	2	2	2	1	2
18																
19																
20																

Figura 13: Resumen del grupo 5: Compuesto por 16 estudiantes, con una distribución equitativa de género de 8 femeninos y 8 masculinos. Distribución por edades: 6 entre 18 a 21 años y 10 entre 22 a 25 años. Respecto a las becas, 6 tienen una beca del 50 %, 4 del 75 % y 6 cuentan con una beca completa. En relación sentimental, 5 tienen pareja y 11 no. Horas de estudio semanal: 6 no estudian ninguna hora, 8 estudian menos de 5 horas y 2 entre 6 a 10 horas. Asistencia a seminarios: 10 asisten y 6 no. Asistencia a clases: 14 siempre asisten y 2 asisten a veces. Preparación para los exámenes: 7 se preparan solos, 7 con amigos y 2 nunca se preparan. Toma de notas: 6 a veces y 10 siempre. Promedio acumulado del último semestre: 1 menor a 2.0, 6 de 2.0 a 2.49, 3 de 2.50 a 2.99, 5 de 3.0 a 3.49 y 1 mayor a 3.49. Promedio esperado al graduarse: 1 menor a 2.0, 6 entre 2.0 a 2.49, 6 de 2.50 a 2.99 y 3 de 3.0 a 3.49. Calificaciones finales: 4 DD, 5 DC, 3 CC, 3 CB, 5 BB y 7 AA.

	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	Occupational study	Non-scientific	Scientific	Conferences	Activities	to class	to midterms	to midterms	notes in class	interest	classroom	average	point	COURSE ID	GRADE		
2	1	2	2	2	1	1	1	1	1	3	2	3	2	2	3	1	1
3	1	2	2	2	2	1	2	1	1	2	2	2	2	2	2	1	2
4	1	2	2	2	1	1	2	1	1	2	2	2	1	2	3	1	1
5	2	1	1	2	1	1	1	1	1	3	2	3	2	2	2	1	1
6	1	1	1	2	1	1	1	3	1	2	2	2	1	3	3	1	5
7	1	1	2	2	1	3	1	2	1	3	2	3	1	4	3	1	3
8	3	1	1	2	1	1	1	2	1	3	2	3	1	3	2	3	5
9	1	2	2	2	2	1	1	2	1	3	2	3	2	3	3	3	7
10	3	1	1	1	1	3	1	1	1	2	3	1	2	4	2	5	4
11	1	3	2	2	1	1	1	3	2	2	2	2	1	4	4	7	7
12	1	1	1	1	2	3	1	2	1	3	3	3	1	4	2	8	2
13	1	2	2	2	2	2	1	1	1	2	2	1	2	1	1	8	2
14	2	3	3	2	2	3	1	2	2	3	2	3	2	2	2	8	1
15	1	2	1	2	1	2	1	2	1	3	2	2	1	4	3	8	2
16	1	2	2	3	2	1	1	2	1	3	2	3	1	5	4	9	3
17	1	2	2	2	1	1	1	1	1	3	1	3	1	2	4	9	2
18																	
19																	
20																	

Figura 14: Continuación del grupo 5

3.2.6. Para el grupo 6

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	TUDENT ID	Student Age	Sex	had high-scholarship	tydional wdistic or spu	have a pialary if avtion to theation type	ersae eduorsae educ of sister/	rental statesae ocu								
2	13	1	1	1	4	2	2	2	1	1	1	3	5	4	2	2
3	16	2	2	2	3	2	2	2	1	1	2	4	4	2	1	3
4	27	2	2	2	3	2	1	1	1	1	1	4	4	4	1	2
5	44	1	2	2	3	2	2	1	1	1	1	2	3	3	1	2
6	60	2	2	2	3	2	1	1	1	1	1	4	4	3	1	2
7	68	2	2	3	3	1	1	2	3	1	2	3	4	4	2	2
8	76	1	2	2	4	2	1	2	1	1	3	1	5	4	1	2
9	84	2	2	3	3	2	2	2	3	1	1	4	4	2	1	1
10	88	2	2	2	3	2	1	2	2	1	1	3	4	2	1	1
11	90	2	2	2	3	2	2	2	1	1	1	3	5	3	1	2
12	93	1	2	2	3	2	2	2	1	1	1	3	4	4	1	3
13	98	1	2	2	4	1	2	2	1	1	3	3	3	4	1	2
14	124	2	1	3	3	2	2	2	3	1	1	3	2	4	1	1
15	143	1	1	1	4	2	2	2	1	1	1	3	4	4	1	2
16																
17																
18																
19																

Figura 15: Resumen del grupo 6: Compuesto por 14 alumnos, con una mayoría de 11 masculinos y 3 femeninos. Respecto a la edad, 6 tienen entre 18 a 21 años y 8 entre 22 a 25 años. En cuanto a becas, 10 tienen una del 50 % y 4 del 75 %. En relación sentimental, 3 tienen pareja y 11 no. Horas de estudio semanal: 1 no estudia ninguna hora, 9 estudian menos de 5 horas y 4 entre 6 a 10 horas. Asistencia a seminarios: 12 asisten y solo 2 no. Asistencia a clases: 10 siempre asisten y 4 a veces. Preparación para exámenes parciales: 12 se preparan solos, 1 con amigos y 1 nunca se prepara. Toma de notas en clase: 4 a veces y 10 siempre. Promedio acumulado del último semestre: 3 menores a 2.0, 4 de 2.0 a 2.49, 3 de 2.50 a 2.99 y 4 de 3.0 a 3.49. Promedio esperado al graduarse: 4 menores a 2.0, 6 de 2.0 a 2.49 y 4 de 2.50 a 2.99. Calificaciones finales: 1 Fail, 4 DD, 1 DC, 1 CB, 1 BB, 3 BA y 3 AA.

	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	sae ocupatky study h(non-sciety (scientific)conferencets/actividance to ch to midten to midtex notes in ening in cla interest aip-classroom averagede point aCOURSE ID																GRADE
2	2	3	2	2	1	1	1	1	1	2	2	2	3	4	2	1	0
3	1	2	2	2	2	3	2	1	1	3	2	2	3	2	2	1	2
4	4	2	3	2	1	1	1	1	1	3	3	3	2	2	1	1	1
5	1	2	2	2	1	1	2	1	1	3	1	3	3	3	2	1	4
6	3	3	2	2	1	1	1	1	1	3	3	2	2	4	3	1	5
7	2	2	2	2	1	1	2	1	1	2	2	1	2	2	3	2	1
8	2	3	3	1	1	1	1	1	1	2	2	2	2	4	1	3	7
9	2	1	2	2	1	1	1	3	1	3	2	3	3	1	2	5	7
10	1	2	2	2	1	1	1	2	1	2	2	2	2	1	1	6	6
11	2	2	2	2	1	1	2	1	1	3	1	3	3	3	2	6	6
12	2	3	2	2	1	1	1	1	1	3	2	3	3	2	2	6	7
13	3	2	2	2	1	1	1	1	1	3	2	2	1	3	3	7	6
14	1	2	2	3	2	1	1	1	1	3	3	2	1	1	1	8	1
15	4	2	2	2	1	1	1	1	1	3	3	2	1	4	3	9	1
16																	
17																	
18																	
19																	
20																	

Figura 16: Continuación del grupo 6

3.2.7. Para el grupo 7

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	TUDENT ID	tudent Age	Sex	ed high-scholarship	ditional wctistic or spu have a	alary If avtion to theation type	ersae eduorsae educr of sister/	rental statesae ocupae ocupatky study f										
2	3	2	2	2	3	2	2	2	2	4	2	2	2	2	1	2	1	2
3	10	2	1	2	3	2	2	1	3	4	2	1	2	3	1	2	3	2
4	15	3	2	2	4	1	1	2	3	4	2	3	1	2	1	4	1	2
5	21	1	2	2	5	1	2	1	1	4	2	3	3	3	2	2	3	4
6	22	1	2	2	5	2	2	1	1	4	2	2	2	4	1	2	3	3
7	29	3	2	2	3	2	2	1	1	4	2	2	2	4	1	2	1	2
8	30	2	2	3	4	2	2	2	1	4	2	3	3	2	1	4	1	2
9	43	2	2	2	3	2	1	2	1	4	2	4	4	2	1	3	1	2
10	55	2	2	2	3	2	2	2	3	4	2	1	4	2	1	4	2	2
11	87	2	2	2	4	2	2	2	1	4	2	3	3	2	1	4	1	2
12	100	2	2	2	3	2	2	2	1	4	2	3	4	2	1	2	1	3
13	114	2	1	3	3	2	2	2	1	4	2	4	4	1	1	3	2	2
14	119	1	2	2	5	2	1	1	1	4	2	1	2	3	1	4	4	1
15	128	1	1	2	4	2	2	2	1	4	3	3	2	2	1	3	4	2
16	129	1	1	2	4	2	1	1	1	4	3	3	4	2	1	4	2	2
17	136	2	1	2	3	1	1	2	1	4	2	3	3	1	1	2	1	2
18	142	1	1	2	4	2	2	2	1	4	2	1	1	5	1	2	1	3
19																		
20																		
21																		
22																		

Figura 17: Resumen del grupo 7: Conformado por 17 estudiantes, de los cuales 6 son femeninos y 11 masculinos. Respecto a la edad, 6 tienen entre 18 a 21 años, 9 entre 22 a 25 años y 2 son mayores de 26 años. En relación a las becas, 8 tienen una del 50 %, 6 del 75 % y 3 tienen una beca completa. Estado sentimental: 6 tienen pareja y 11 no. Horas de estudio semanal: 1 no estudia ninguna hora, 12 estudian menos de 5 horas, 3 entre 6 a 10 horas y 1 de 11 a 20 horas. Asistencia a seminarios: 13 asisten y 5 no. Asistencia a clases: 14 siempre asisten y 3 a veces. Preparación para exámenes parciales: 15 se preparan solos, 1 con amigos y 1 nunca se prepara. Toma de notas en clase: 5 a veces y 12 siempre. Promedio acumulado del último semestre: 2 menores de 2.0 a 2.49, 5 de 2.0 a 2.49, 3 de 2.50 a 2.99, 3 de 2.0 a 3.49 y 4 mayores a 3.49. Promedio esperado al graduarse: 7 de 2.0 a 2.49, 6 de 2.50 a 2.99 y 4 de 3.0 a 3.49. Calificaciones finales: 2 Fail, 1 DD, 2 DC, 3 CC, 5 BB y 7 AA.

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	non-sciency	(scientific)	conferences	acts/activi	dance to ch	to midten	to midten	notes in c	ning in cla	interest alp	alp-classroom	int averaged	ede point a	COURSE ID	GRADE
2	1	2	1	1	1	1	1	2	2	1	1	2	2	1	1
3	2	2	1	1	2	1	1	2	2	2	2	1	2	1	0
4	2	2	1	1	1	1	1	2	3	2	1	4	4	1	2
5	2	2	2	1	1	1	1	2	3	1	2	3	4	1	1
6	2	2	1	1	1	1	1	3	1	3	3	3	3	1	1
7	2	2	1	1	1	1	1	3	2	3	3	5	4	1	3
8	2	2	1	1	1	1	1	3	2	2	2	4	3	1	5
9	2	1	1	1	1	1	1	2	1	3	1	2	3	1	1
10	1	1	1	1	1	1	2	3	1	3	1	5	3	1	3
11	2	2	1	1	1	1	1	3	2	3	2	5	4	5	5
12	2	2	1	1	1	1	1	3	2	3	2	2	2	7	7
13	2	3	2	3	2	2	1	2	2	1	1	1	2	8	1
14	2	2	1	1	1	3	1	3	3	2	2	3	3	8	1
15	1	1	2	1	1	1	1	3	2	2	1	2	2	9	1
16	1	1	2	1	1	1	1	3	2	2	1	2	2	9	0
17	3	2	1	3	1	1	1	3	3	2	1	3	2	9	2
18	2	2	2	1	2	1	1	3	2	2	1	5	3	9	5
19															
20															
21															
22															

Figura 18: Continuación del grupo 7

3.2.8. Para el grupo 8

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	TUDENT ID	Student Age	Sex	ed high-scholarship	tyditiolnal wctistic or spu	have a pialary	If avation to the	ation type	ersae educ	ersae educ	of sister/	rental statesae	ocupae ocupatly	study h				
2	48	2	2	2	3	2	2	1	1	1	2	4	3	1	1	5	4	2
3	71	1	2	2	4	2	1	1	1	1	1	4	3	1	2	4	2	2
4	89	1	2	2	4	2	1	1	1	1	1	4	4	1	1	3	2	2
5	91	2	1	2	3	2	1	1	1	2	3	3	3	1	1	4	2	2
6	96	1	2	3	5	1	1	1	1	2	3	3	4	1	1	2	1	2
7	97	1	2	2	4	2	1	1	1	2	3	2	2	3	1	4	2	2
8	101	1	2	2	4	2	2	2	1	2	3	4	1	2	1	3	3	2
9	103	1	2	2	3	2	2	1	1	1	2	3	1	1	2	4	4	2
10	112	1	1	1	5	2	1	2	1	1	2	3	3	1	1	3	2	1
11	125	1	1	2	4	1	1	1	1	1	3	2	3	2	1	4	3	4
12	132	1	1	1	5	2	1	2	1	2	2	4	3	1	1	3	2	4
13	135	1	1	2	4	2	1	2	1	2	3	1	1	2	1	4	3	2
14																		
15																		
16																		
17																		
18																		
19																		
20																		
21																		
22																		

Figura 19: Resumen del grupo 8: Compuesto por 12 alumnos, de los cuales 10 tienen entre 18 a 21 años y 2 entre 22 a 25 años. En cuanto al género, 5 son femeninos y 7 masculinos. Respecto a las becas, 3 tienen el 50 %, 6 el 75 % y 3 tienen una beca completa. Estado sentimental: 8 tienen pareja y 4 no. Horas de estudio semanal: 1 no hace ninguna hora, 9 hacen menos de 5 horas y 2 entre 11 a 20 horas. Asistencia a seminarios: 7 asisten y 5 no. Asistencia a clases: 11 siempre asisten y 1 solo a veces. Preparación para exámenes parciales: 11 se preparan solos y solo 1 con amigos. Toma de notas en clase: 5 a veces y 7 siempre. Promedio acumulado del último semestre: 2 de 2.0 a 2.49, 3 de 2.50 a 2.99, 3 de 3.0 a 3.49 y 5 mayores a 3.49. Promedio esperado al graduarse: 1 de 2.0 a 2.49, 9 de 2.50 a 2.99 y 2 de 3.0 a 3.49. Calificaciones finales: 1 DD, 1 DC, 2 CC, 1 CB, 1 BB, 3 BA y 7 AA.

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	(non-scientific)	(scientific)	conferences/activities	attendance to class	attendance to midterms	attendance to midterms	attendance to midterms	attendance to midterms	attendance to midterms	attendance to midterms	attendance to midterms	attendance to midterms	attendance to midterms	attendance to midterms	attendance to midterms
2	2	2	1	1	2	1	1	2	1	3	2	5	3	1	3
3	2	2	1	1	1	1	1	2	2	3	1	5	4	3	7
4	2	2	1	1	1	1	2	3	1	2	3	5	3	6	6
5	2	2	1	1	1	1	1	2	3	2	3	4	2	6	6
6	2	3	1	1	1	1	1	3	2	2	1	4	3	7	5
7	3	2	1	1	1	1	1	3	2	3	2	2	3	7	7
8	2	2	1	1	1	1	1	2	2	2	1	3	3	7	6
9	2	2	2	1	1	1	1	3	1	2	1	3	4	7	7
10	2	2	2	2	1	1	1	3	1	3	3	4	3	8	2
11	2	2	2	1	1	2	2	3	3	2	1	3	3	9	3
12	1	3	2	1	1	1	2	3	2	3	1	5	3	9	4
13	2	2	2	1	1	1	1	2	1	2	1	2	3	9	1
14															
15															
16															
17															
18															
19															
20															
21															
22															

Figura 20: Continuación del grupo 8

3.2.9. Para el grupo 9

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	TUDENT ID	Student Age	Sex	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship
2	6	2	2	2	3	2	2	2	2	1	1	3	3	2	1	2	3	1	1
3	7	1	2	2	4	2	2	2	1	1	3	1	3	1	1	2	4	2	2
4	9	2	1	3	3	2	1	1	1	1	3	2	4	2	1	2	4	1	2
5	25	2	2	2	3	2	2	2	2	1	1	3	1	3	1	2	4	1	2
6	37	2	2	3	4	1	2	1	3	1	1	1	1	2	2	4	3	1	1
7	38	2	2	2	3	1	1	1	2	2	3	3	3	1	1	2	3	2	2
8	39	2	2	2	5	2	2	2	1	1	1	3	3	1	1	2	4	3	2
9	45	2	2	3	3	2	2	1	1	1	1	1	3	2	1	2	5	2	2
10	47	2	2	2	3	2	2	1	1	1	1	1	2	3	1	2	5	2	2
11	50	1	2	1	4	2	2	2	2	4	1	2	3	2	1	2	4	2	2
12	51	2	2	2	3	2	2	1	2	2	1	1	1	2	1	2	3	2	2
13	58	2	2	2	3	1	1	2	1	1	1	4	2	3	1	3	3	2	3
14	63	2	2	2	3	2	2	2	1	4	2	2	3	2	1	2	4	3	2
15	73	1	2	2	3	2	1	1	1	1	2	1	1	2	1	2	3	2	3
16	99	1	2	2	4	2	1	2	1	4	2	2	4	2	1	2	5	2	3
17	109	2	1	1	5	2	1	2	2	2	1	1	2	1	3	1	5	3	3
18	110	1	2	1	3	2	1	1	1	2	2	4	4	2	3	2	5	3	2
19	111	2	2	2	3	2	2	2	2	4	2	4	4	2	2	1	5	2	2
20	120	2	1	2	4	2	1	2	1	1	2	3	3	2	3	2	5	5	2
21	134	1	1	2	5	2	2	1	1	1	1	1	1	1	1	2	3	3	1
22																			
23																			

Figura 21: Resumen del grupo 9: Compuesto por 20 estudiantes, de los cuales 4 son femeninos y 16 masculinos. Respecto a la edad, 6 tienen entre 18 a 21 años y 14 entre 22 a 25 años. En cuanto a las becas, 12 tienen un 50 %, 5 un 75 % y 3 tienen una beca completa. Estado sentimental: 9 tienen pareja y 11 no. Horas de estudio semanal: 4 no estudian ninguna hora, 10 estudian menos de 5 horas, 5 de 6 a 10 horas y 1 de 11 a 20 horas. Asistencia a seminarios: 18 asisten y 2 no. Asistencia a clases: 15 siempre asisten y 5 a veces. Preparación para exámenes parciales: 15 lo hicieron solos y 2 con amigos. Toma de notas en clase: 1 nunca toma notas, 11 a veces y 8 siempre. Promedio acumulado del último semestre: 1 es de 2.0 a 2.49, 5 son de 2.50 a 2.99, 10 de 3.0 a 3.49 y 4 mayores a 3.49. Promedio esperado al graduarse: 2 son de 2.0 a 2.49, 10 de 2.50 a 2.99 y 8 de 3.0 a 3.49. Calificaciones finales: 3 DD, 7 DC, 2 CC, 1 CB, 3 BB, 2 BA y 2 AA.

	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	y (scientific)	conferenc	ects/activi	dance to ch	to midten	to midten	notes in	ning in cla	interest aip	ip-classroo	int average	de point a	COURSE ID	GRADE
2	2	1	1	1	1	1	1	2	1	2	4	4	1	2
3	2	2	1	2	1	1	3	3	3	3	4	4	1	5
4	2	1	1	1	1	1	3	2	2	2	4	3	1	5
5	2	1	1	1	1	1	2	1	3	2	4	4	1	2
6	3	1	1	2	2	1	2	1	2	1	4	3	1	2
7	2	1	1	1	1	1	3	2	3	3	5	4	1	1
8	2	1	1	1	2	1	2	2	2	2	4	3	1	2
9	2	1	1	1	2	1	3	2	2	1	4	3	1	1
10	2	1	1	1	1	1	2	2	3	1	4	2	1	5
11	2	1	1	1	2	1	2	2	2	1	3	2	1	2
12	2	1	1	2	1	1	2	2	2	3	3	3	1	1
13	2	1	1	1	1	1	3	2	3	1	5	4	1	4
14	2	1	1	1	1	1	2	3	3	2	5	4	1	3
15	3	1	1	1	1	1	2	2	3	2	3	3	3	6
16	2	1	1	1	1	1	3	1	2	1	4	4	7	7
17	2	1	1	2	1	1	2	3	1	2	3	3	7	6
18	2	1	1	1	2	2	3	2	3	1	4	4	7	7
19	2	1	1	2	1	1	2	2	2	1	2	3	8	2
20	2	1	1	1	1	2	2	2	3	1	3	3	8	2
21	2	2	1	1	1	1	3	3	3	2	5	3	9	3
22														
23														
24														

Figura 22: Continuación del grupo 9

3.3. Corte en altura 15

Para el corte 15 tenemos un total de 5 grupos, cada grupo esta compuesto por una cantidad distinta de objetos que se muestran en una hoja diferente de excel

3.3.1. Para el grupo 1

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
TUDENT	udent Ag	Sex	id high-scholarship	ditional wetic or spu	have a plary if avtion to theation typeersae educsae educ of sister/rental statesae ocupae ocupakly study if non-sciety (scientific)	conferenc	ects/activi	dance to ch	to midten	to midten	notes in	ning in cla	interest aip	ip-classroo	int average	de point a	COURSE ID	GRADE					
1	2	2	3	3	1	2	2	1	1	1	1	2	3	1	2	5	3	2	2	1	1	1	1
19	1	1	2	4	2	2	2	3	1	1	2	2	5	1	2	5	5	3	2	2	1	2	1
20	1	2	1	3	2	2	1	2	2	2	3	3	3	1	2	4	4	2	2	1	2	1	1
23	2	2	2	3	1	2	1	1	1	1	1	4	2	2	4	3	3	3	1	1	1	1	1
26	2	2	2	3	2	2	1	1	1	2	1	4	3	1	2	4	2	2	2	1	1	1	1
28	1	2	1	3	1	2	2	1	1	1	1	1	4	1	2	3	3	2	2	1	1	2	1
36	1	2	1	4	2	2	2	3	1	1	1	2	5	1	2	3	2	2	2	1	1	2	1
40	2	1	2	3	2	2	1	1	1	2	1	3	1	1	2	4	1	2	1	1	1	1	1
72	1	1	3	4	2	2	2	1	1	3	1	3	5	1	2	4	2	2	1	1	1	1	1
79	2	1	2	4	1	1	2	1	1	1	1	3	4	1	2	4	2	3	3	1	1	2	1
94	1	2	2	1	1	2	1	1	1	2	1	2	4	1	2	3	5	2	1	1	1	1	1
106	1	2	2	4	1	1	2	1	1	1	2	2	4	1	2	4	2	3	2	1	1	2	1
108	1	1	2	4	1	1	2	1	1	3	1	3	5	1	2	4	2	2	2	1	1	1	2
117	2	1	2	5	2	2	2	1	1	1	1	1	5	1	2	4	3	3	3	2	1	1	1
122	2	1	2	3	1	1	1	1	2	1	1	1	5	3	2	4	5	3	3	1	3	1	2
126	1	1	2	5	1	1	2	1	1	3	1	1	5	1	2	3	3	2	3	2	1	1	1
130	1	1	2	3	1	1	2	1	1	1	1	1	3	1	2	4	4	2	3	2	1	1	2
131	1	1	2	3	1	1	2	1	1	1	1	2	3	1	2	4	4	2	2	2	1	1	1
141	2	1	2	3	1	1	2	1	1	2	1	2	2	2	2	4	3	3	2	1	1	1	1

Figura 23: Resumen del grupo 1: Los estudiantes tienen una edad promedio de entre 1 y 2 años, con un grupo equitativo de 10 estudiantes masculinos y 10 estudiantes femeninos. No se proporcionaron datos sobre el tipo de beca que reciben los estudiantes. En cuanto al estado civil, 8 estudiantes tienen pareja, mientras que 12 estudiantes no tienen pareja. En promedio, los estudiantes dedican entre 1 y 2 horas por semana al estudio. Todos los estudiantes asisten regularmente a seminarios y conferencias relacionadas con el departamento, así como a clases. La mayoría de los estudiantes se preparan para los exámenes parciales. Todos los estudiantes toman apuntes durante las clases. Los promedios acumulativos en el último semestre varían entre 1 y 9, mientras que los promedios esperados en la graduación van desde 1 hasta 5 Las calificaciones finales oscilan entre 0 y 7.

	Z	AA	AB	AC	AD	AE	AF	AG	
1	3	2	1	2	1	1	1	1	1
1	3	1	3	3	3	3	1	2	
1	3	2	2	3	2	3	1	3	
2	3	1	2	3	3	3	1	3	
1	2	1	3	2	1	2	1	3	
1	3	1	2	1	2	1	1	1	
1	2	1	3	1	1	1	1	1	
1	2	2	2	3	2	1	1	1	
1	2	3	3	3	2	2	3	6	
1	3	3	2	2	2	2	4	4	
1	3	3	3	2	2	2	6	4	
2	2	1	3	1	3	3	7	7	
1	2	3	2	1	4	2	7	6	
1	3	3	3	1	2	3	8	1	
1	1	1	1	1	1	1	8	0	
2	3	2	3	1	1	3	9	2	
2	3	2	3	2	2	3	9	3	
2	3	1	3	2	2	3	9	1	
1	2	1	2	1	3	3	9	5	

Figura 24: Continuacion de la figura 1

3.3.2. Para el grupo 2

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
STUDENT ID	Student Age	Sex	scholarship	high-scholarship	traditional	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship	scholarship
4	1	1	1	3	1	2	1	2	1	2	1	2	5	1	2	1	3	1	2	1	1	1	1	2	3
14	2	1	2	5	2	2	2	1	1	1	2	2	2	1	2	3	1	2	1	1	1	2	1	1	3
17	1	1	2	5	2	1	2	1	1	1	2	2	4	1	2	3	2	2	2	1	1	2	1	1	3
24	3	2	2	2	1	1	2	5	1	1	1	4	3	1	2	1	3	2	3	1	1	2	1	1	3
31	2	2	2	5	1	1	1	1	1	2	1	1	5	1	2	4	2	2	2	1	1	1	2	1	2
32	3	2	2	3	1	2	2	1	1	2	1	2	4	1	2	3	2	2	3	1	1	1	1	1	3
34	2	1	2	3	1	2	1	1	1	1	1	2	5	1	2	3	2	1	2	1	1	1	1	1	1
42	3	2	2	3	1	2	2	2	1	2	1	4	2	1	2	2	3	2	2	1	1	2	1	1	3
46	1	2	2	3	2	2	1	4	1	1	2	3	5	1	2	4	3	2	2	1	1	1	1	1	2
52	2	1	3	3	1	1	2	1	1	1	1	2	5	1	2	1	1	2	2	1	1	2	1	1	2
53	2	1	2	3	1	2	2	2	1	1	1	3	4	1	2	1	3	1	1	1	1	1	1	2	2
54	2	1	2	3	2	1	2	2	1	1	1	1	5	1	2	1	1	2	2	1	1	2	1	1	2
56	3	2	2	3	1	2	1	4	1	2	1	1	5	3	2	1	1	2	2	1	1	1	3	3	1
57	2	2	2	3	2	1	2	1	1	1	1	3	5	1	2	4	4	2	3	1	1	1	1	2	3
59	3	2	2	3	2	2	1	3	1	1	1	3	5	2	2	1	3	2	2	1	1	1	1	1	3
62	1	2	3	3	2	1	2	2	1	2	1	3	4	1	2	3	3	1	2	1	3	2	1	1	3
67	2	2	2	3	2	2	1	1	1	1	3	2	5	1	2	5	2	2	2	1	1	1	1	1	3
69	2	1	2	4	1	2	2	1	1	1	2	3	5	3	2	5	1	2	2	1	1	2	1	1	2
74	2	2	2	4	2	2	2	1	1	2	1	3	4	1	2	4	3	1	1	2	1	1	1	1	2
80	2	2	2	4	2	2	2	1	1	1	1	3	4	1	2	4	1	2	2	1	3	1	1	1	3
82	3	2	2	3	1	2	2	1	1	2	1	2	4	1	2	3	2	2	2	1	1	1	1	1	3
83	2	2	2	4	1	2	1	2	1	1	3	1	3	1	2	3	2	2	3	1	1	1	1	1	3
85	3	2	3	3	1	2	1	3	1	2	4	2	5	3	2	1	1	2	2	1	1	1	3	3	3
86	1	2	2	5	2	2	2	2	1	1	1	1	5	1	2	1	1	1	1	1	1	1	1	1	3
92	2	2	2	5	1	1	1	1	1	2	4	1	5	1	2	3	2	2	2	1	1	1	1	1	2
95	2	2	2	3	2	2	1	1	1	1	3	2	5	1	2	5	2	2	3	1	1	1	1	1	3
104	1	2	1	4	2	2	1	5	1	2	1	3	4	1	2	1	2	3	2	1	1	1	1	1	2
145	1	1	1	5	2	2	2	3	1	1	3	1	5	1	2	4	3	1	1	1	1	1	2	1	3

Figura 25: Los estudiantes tienen edades que varían entre 1 y 3 años. Hay una distribución equitativa entre estudiantes masculinos y femeninos. No se proporcionaron datos sobre el tipo de beca que reciben los estudiantes. En cuanto al estado civil, la mitad de los estudiantes tienen pareja y la otra mitad no. El promedio de horas de estudio semanal oscila entre 1 y 5 horas. Todos los estudiantes asisten a los seminarios y conferencias relacionados con el departamento. Todos los estudiantes asisten regularmente a clases. La mayoría de los estudiantes se preparan para los exámenes parciales. Todos los estudiantes toman apuntes durante las clases. Los promedios acumulativos en el último semestre varían entre 1 y 5, mientras que los promedios esperados en la graduación van desde 1 hasta 7. Las calificaciones finales oscilan entre 1 y 9.

AA	AB	AC	AD	AE	AF	AG	
Learning in class	Interest in class	Pre-classroom	Post-classroom	Average	Grade point average	COURSE ID	GRADE
2	2	1	3	2	1	1	
2	3	3	4	2	1	1	
2	3	3	4	3	1	1	
2	3	3	3	3	1	1	
3	3	3	5	4	1	5	
3	2	3	4	3	1	3	
3	2	2	2	3	1	2	
2	2	3	4	3	1	1	
2	2	1	4	3	1	3	
3	3	1	3	3	1	4	
2	2	1	4	2	1	1	
2	2	2	3	2	1	5	
3	3	3	5	4	1	3	
2	3	3	5	4	1	5	
2	2	2	5	4	1	3	
2	2	2	4	4	1	5	
2	2	3	5	4	2	5	
2	3	2	4	3	3	5	
2	3	1	5	3	3	6	
3	3	2	4	4	4	3	
3	2	3	5	4	5	3	
3	2	3	4	3	5	7	
3	3	3	5	4	5	7	
2	3	3	4	3	5	4	
3	3	2	5	4	6	6	
2	2	3	5	4	6	6	
2	3	2	4	4	7	7	
2	3	1	5	4	9	3	

Figura 26: Continuación del grupo 2

3.3.3. Para el grupo 3

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
TUDENT ID	Student Age	Sex	aid	high-scholarship	ditional	uistic or spu	have a	alary if avtion to th	tion type	ersae educ	of sister/	rental st	ersae ocugae	occupatly study	non-scientific	conference				
5	2	2	1	3	2	2	1	3	1	4	3	3	2	1	2	4	2	1	1	1
8	1	1	2	3	1	1	1	2	2	3	4	3	1	1	4	3	1	2	2	1
11	1	1	1	3	2	2	2	3	2	3	3	4	2	1	3	2	1	1	1	1
12	1	1	1	4	1	1	2	4	2	3	5	5	1	1	3	2	3	3	3	1
35	1	2	1	3	2	2	1	2	1	1	3	3	3	1	2	4	2	1	1	1
41	1	2	1	3	2	2	2	2	1	1	2	3	1	1	2	5	1	2	2	1
61	2	1	2	3	2	2	2	5	2	1	6	1	3	1	2	3	1	2	2	1
66	1	2	2	3	2	2	1	3	1	1	2	4	3	1	2	4	1	1	2	1
77	2	2	1	2	2	1	2	2	1	1	5	4	3	1	4	3	2	3	2	1
78	1	2	1	2	2	2	1	2	2	2	6	5	2	1	4	2	2	2	2	2
102	1	2	2	4	2	2	1	2	1	3	2	3	1	1	2	4	2	2	2	1
105	1	1	2	3	2	2	2	1	2	2	4	4	1	1	3	2	2	2	2	1
118	3	1	1	3	2	1	2	1	1	2	4	4	1	1	1	3	2	2	3	2
121	2	1	1	3	1	1	1	2	2	3	3	3	2	1	3	4	1	1	1	1
127	1	1	1	4	1	1	1	3	2	3	4	6	2	1	4	3	4	2	3	1
137	1	1	2	3	1	1	1	1	2	3	3	3	3	1	3	4	2	2	2	2
139	1	1	2	4	1	1	1	5	2	3	3	1	1	2	4	5	2	2	3	2
140	1	1	2	4	1	2	1	2	2	3	3	3	1	1	5	4	1	1	2	2
144	2	1	2	4	1	1	1	5	2	3	4	4	1	1	3	3	2	2	1	1

Figura 27: Resumen del grupo 3: Edad de los estudiantes: La mayoría de los estudiantes en este conjunto de datos se encuentran en el rango de edad de 18 a 21 años, lo que sugiere que son estudiantes jóvenes recién ingresados a la universidad. Género: Hay una distribución equilibrada entre estudiantes mujeres y hombres en este conjunto de datos. Becas: Varios estudiantes cuentan con becas parciales o completas, lo que podría influir en su desempeño académico al tener menos preocupaciones financieras. Parejas: Pocos estudiantes reportan tener una pareja, lo que podría permitirles dedicar más tiempo a sus estudios. Horas de estudio semanales: La mayoría de los estudiantes estudian menos de 10 horas a la semana, lo que podría ser insuficiente para un buen desempeño académico. Asistencia a seminarios/conferencias: Pocos estudiantes asisten a seminarios o conferencias relacionadas con su carrera, lo que podría limitar su exposición a conocimientos y oportunidades adicionales. Asistencia a clases: Mientras algunos estudiantes asisten siempre a clases, otros lo hacen solo a veces o nunca, lo que podría afectar su rendimiento. Preparación para exámenes parciales: Algunos estudiantes se preparan solos o con amigos, mientras que otros nunca se preparan, lo que podría influir en sus calificaciones. Toma de notas: Hay una distribución variada en cuanto a la frecuencia de toma de notas en clase, lo que podría afectar la retención de conocimientos. Calificaciones promedio del último semestre y esperadas al graduarse: La mayoría de los estudiantes tienen calificaciones promedio en el rango de 2.50 a 3.49, lo que sugiere un desempeño académico moderado a bueno. Calificaciones finales del curso: Las calificaciones finales del curso varían desde "Fail" hasta ".^A", lo que indica una amplia gama de desempeño académico en este conjunto de datos.

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	COURSE ID	GRADE
1	1	2	1	2	2	2	1	2	2	1	1	1	1
1	1	3	1	3	2	2	1	1	1	1	1	2	2
1	2	1	1	2	2	2	2	1	1	1	1	2	2
3	1	3	2	3	1	3	3	4	3	1	1	0	0
1	2	1	1	2	2	3	1	4	1	1	1	2	2
1	1	1	1	3	1	2	1	2	3	1	1	1	1
1	1	1	1	1	3	3	1	2	1	1	1	2	2
3	2	2	1	2	1	2	1	2	2	1	1	2	2
1	1	3	2	2	1	2	3	2	2	2	4	4	4
1	2	3	1	2	2	2	2	1	1	4	7	7	7
1	1	1	1	3	3	2	1	2	3	7	7	7	7
1	1	1	1	2	3	2	2	1	1	7	3	3	3
1	2	1	1	2	3	2	1	2	3	8	1	1	1
1	2	2	1	3	3	3	2	2	2	8	1	1	1
1	1	2	1	3	3	3	1	2	2	9	3	3	3
1	1	3	1	3	2	2	1	2	2	9	0	0	0
3	1	2	1	3	2	3	1	1	3	9	0	0	0
1	2	2	1	2	3	2	1	1	2	9	0	0	0
1	1	2	1	2	1	2	1	5	3	9	4	4	4

Figura 28: Continuacion del grupo 3

3.3.4. Para el grupo 4

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
STUDENT ID	Student Age	Sex	had high-scholarship	typical	scholarship	typical	scholarship	typical	scholarship	typical	scholarship	typical	scholarship	typical	scholarship	typical	scholarship	typical	scholarship	typical	scholarship	typical	scholarship	typical	scholarship
2	2	2	3	3	1	2	2	1	1	1	2	3	2	1	2	1	2	2	2	1	1	1	1	1	3
13	1	1	1	4	2	2	2	1	1	1	3	5	4	2	2	2	3	2	2	1	1	1	1	1	2
16	2	2	2	3	2	2	2	1	1	2	4	4	2	1	3	1	2	2	2	2	3	2	1	1	3
18	2	2	2	3	2	2	2	1	1	1	2	2	2	1	2	1	2	2	2	2	1	2	1	1	2
27	2	2	2	3	2	1	1	1	1	1	4	4	4	1	2	4	2	3	2	1	1	1	1	1	3
33	2	1	2	3	2	2	2	2	1	1	2	2	2	1	2	1	2	2	2	1	1	2	1	1	2
44	1	2	2	3	2	2	1	1	1	1	2	3	3	1	2	1	2	2	2	1	1	2	1	1	3
49	1	2	2	3	2	1	1	1	1	2	3	3	3	1	2	2	1	1	2	1	1	1	1	1	3
60	2	2	2	3	2	1	1	1	1	1	4	4	3	1	2	3	3	2	2	1	1	1	1	1	3
64	2	2	2	4	2	2	1	2	1	1	1	3	3	1	2	1	1	1	2	1	1	1	3	1	2
65	2	2	3	5	2	2	2	1	1	1	2	2	2	1	2	1	1	2	2	1	3	1	2	1	3
68	2	2	3	3	1	1	2	3	1	2	3	4	4	2	2	2	2	2	2	1	1	2	1	1	2
70	2	1	2	4	2	2	1	1	1	1	2	2	2	1	2	3	1	1	2	1	1	1	2	1	3
75	1	2	2	4	2	2	2	1	1	1	2	2	3	1	2	1	2	2	2	2	1	1	2	1	3
76	1	2	2	4	2	1	2	1	1	3	1	5	4	1	2	2	3	3	1	1	1	1	1	1	2
81	2	1	2	3	1	2	1	1	2	1	1	1	2	1	2	3	1	1	1	1	3	1	1	1	2
84	2	2	3	3	2	2	2	3	1	1	4	4	2	1	1	2	1	2	2	1	1	1	3	1	3
88	2	2	2	3	2	1	2	2	1	1	3	4	2	1	1	1	2	2	2	1	1	1	2	1	2
90	2	2	2	3	2	2	2	1	1	1	3	5	3	1	2	2	2	2	2	1	1	2	1	1	3
93	1	2	2	3	2	2	2	1	1	1	3	4	4	1	3	2	3	2	2	1	1	1	1	1	3
98	1	2	2	4	1	2	2	1	1	3	3	3	4	1	2	3	2	2	2	1	1	1	1	1	3
107	1	2	2	4	2	1	2	1	1	1	2	2	2	1	1	2	1	3	2	2	1	1	1	3	2
113	2	1	3	3	1	2	2	1	3	3	1	1	2	1	2	1	1	1	1	2	3	1	2	1	3
115	2	1	2	5	1	1	2	4	1	1	1	1	1	1	2	1	2	2	2	2	2	1	1	1	2
116	2	1	2	5	1	2	1	1	1	1	1	1	2	1	2	2	3	3	2	2	3	1	2	2	3
123	1	2	2	5	2	1	2	3	1	1	3	3	2	1	2	1	2	1	2	1	2	1	2	1	3
124	2	1	3	3	2	2	2	3	1	1	3	2	4	1	1	1	2	2	3	2	1	1	1	1	3
133	1	1	1	5	2	1	2	1	1	2	1	2	3	1	2	1	2	2	3	2	1	1	2	1	3

Figura 29: Resumen del grupo 4: Edad de los estudiantes: La mayoría de los estudiantes se encuentran en el rango de edad de 22 a 25 años, lo que sugiere que son estudiantes de cursos superiores. Género: Hay una distribución equilibrada entre estudiantes mujeres y hombres. Becas: Varios estudiantes cuentan con becas parciales o completas, incluyendo becas completas del 100. Parejas: Pocos estudiantes reportan tener una pareja. Horas de estudio semanales: La mayoría de los estudiantes estudian entre 6 y 10 horas a la semana, aunque hay algunos que estudian más de 20 horas. Asistencia a seminarios/conferencias: Pocos estudiantes asisten a seminarios o conferencias relacionadas con su carrera. Asistencia a clases: Mientras algunos estudiantes asisten siempre a clases, otros lo hacen solo a veces o nunca. Preparación para exámenes parciales: La mayoría de los estudiantes se preparan solos o con amigos para los exámenes parciales. Toma de notas: Hay una distribución variada en cuanto a la frecuencia de toma de notas en clase. Calificaciones promedio del último semestre: Las calificaciones promedio del último semestre se distribuyen en todo el rango, desde menos de 2.00 hasta más de 3.49. Calificaciones promedio esperadas al graduarse: Las calificaciones promedio esperadas al graduarse también se distribuyen en todo el rango. Calificaciones finales del curso: Las calificaciones finales del curso varían desde "Fail" hasta "A", lo que indica una amplia gama de desempeño académico.

AA	AB	AC	AD	AE	AF	AG
ening in cla	interest a	p-classroo	int average	de point a	COURSE ID	GRADE
2	3	2	2	3	1	1
2	2	3	4	2	1	0
2	2	3	2	2	1	2
2	2	2	2	2	1	2
3	3	2	2	1	1	1
2	2	1	2	3	1	1
1	3	3	3	2	1	4
2	3	2	2	2	1	1
3	2	2	4	3	1	5
2	2	1	3	3	1	5
2	3	1	4	3	1	3
2	1	2	2	3	2	1
2	3	1	3	2	3	5
2	3	2	3	3	3	7
2	2	2	4	1	3	7
3	1	2	4	2	5	4
2	3	3	1	2	5	7
2	2	2	1	1	6	6
1	3	3	3	2	6	6
2	3	3	2	2	6	7
2	2	1	3	3	7	6
2	2	1	4	4	7	7
3	3	1	4	2	8	2
2	1	2	1	1	8	2
2	3	2	2	2	8	1
2	2	1	4	3	8	2
3	2	1	1	1	8	1
2	3	1	5	4	9	3

Figura 30: Continuación del grupo 4

3.3.5. Para el grupo 5

STUDENT ID	Student Age	Sex	hd	high-scholarship	ditional	uistic or spki	have a salary	if exition to th	tion type	ersee edu	ersee edu	educ of sister	rental stat	ersee cou	ersee cou	study H	(non-sciety	(scientific	conferen	jects/acti	ance to q	to midte	to midte	notes in d
3	2	2	2	2	3	2	2	2	2	4	2	2	2	2	1	2	1	2	1	2	1	1	1	2
6	2	2	2	2	3	2	2	2	2	1	1	3	3	2	1	2	3	1	1	2	1	1	1	1
7	1	2	2	4	2	2	2	1	1	3	1	3	1	1	2	4	2	2	2	2	1	2	1	3
9	2	2	3	3	2	1	1	1	1	3	1	2	4	2	1	4	2	2	2	1	1	1	1	3
10	2	1	2	3	2	2	1	3	4	2	1	2	3	1	2	3	2	2	2	1	1	2	1	2
15	3	2	2	4	1	1	2	3	4	2	3	1	2	1	4	1	2	2	2	1	1	1	1	2
21	1	2	2	5	1	2	1	1	4	2	3	3	3	2	2	3	4	2	2	2	1	1	1	3
22	1	2	2	5	2	2	1	1	4	2	2	2	4	1	2	3	3	2	2	1	1	1	1	3
25	2	2	2	3	2	2	2	2	1	1	3	1	3	1	2	4	1	2	2	1	1	1	1	2
29	3	2	2	3	2	2	1	1	4	2	2	2	4	1	2	1	2	2	2	1	1	1	1	3
30	2	2	3	4	2	2	2	1	4	2	3	3	2	1	4	1	2	2	2	1	1	1	1	3
37	2	2	3	4	1	2	1	3	1	1	1	1	2	2	4	3	1	1	3	1	1	2	2	1
38	2	2	2	3	1	1	1	2	2	3	3	3	1	1	2	3	2	2	2	1	1	1	1	3
39	2	2	2	5	2	2	2	1	1	1	3	3	1	1	2	4	3	2	2	1	1	1	2	1
43	2	2	2	3	2	1	2	1	4	2	4	4	2	1	3	1	2	2	1	1	1	1	1	2
45	2	2	3	3	2	2	1	1	1	1	1	3	2	1	2	5	2	2	2	1	1	1	2	1
47	2	2	2	3	2	2	1	1	1	1	2	3	2	1	2	5	2	2	2	1	1	1	1	2
48	2	2	2	3	2	2	1	1	1	2	4	3	1	1	5	4	2	2	2	1	1	2	1	2
50	1	2	1	4	2	2	2	2	4	1	2	3	2	1	2	4	2	2	2	1	1	1	2	1
51	2	2	2	3	2	2	1	2	2	1	1	1	2	1	2	3	2	2	2	1	1	2	1	2
55	2	2	2	3	2	2	2	3	4	2	1	4	2	1	4	2	2	1	1	1	1	1	1	3
58	2	2	2	3	1	1	2	1	1	1	1	2	3	1	3	3	2	3	2	1	1	1	1	3
63	2	2	2	3	2	2	2	1	4	2	2	3	2	1	2	4	3	2	2	1	1	1	1	2
71	1	2	2	4	2	1	1	1	1	1	4	3	1	2	4	2	2	2	2	1	1	1	1	2
75	1	2	2	3	2	1	1	1	1	2	1	1	2	1	2	3	2	3	3	1	1	1	1	2
87	2	2	2	4	2	2	2	1	4	2	3	3	2	1	4	1	2	2	2	1	1	1	1	3
89	1	2	2	4	2	2	1	1	1	1	4	4	1	1	3	2	2	2	2	1	1	1	1	2
91	2	1	2	3	2	1	1	1	1	2	3	3	3	1	1	4	2	2	2	2	1	1	1	1

Figura 31: Resumen del grupo 5: Compuesto por Mujeres: 27 estudiantes Hombres: 30 estudiantes Tener pareja: Sí tienen pareja: 9 estudiantes No tienen pareja: 48 estudiantes Horas de estudio semanales: Menos de 5 horas: 12 estudiantes 6-10 horas: 28 estudiantes 11-20 horas: 11 estudiantes Más de 20 horas: 6 estudiantes Asistencia a seminarios/conferencias: Asisten: 13 estudiantes No asisten: 44 estudiantes Asistencia a clases: Siempre asisten: 23 estudiantes A veces asisten: 25 estudiantes Nunca asisten: 9 estudiantes Preparación para exámenes parciales: Solos: 31 estudiantes Con amigos: 13 estudiantes No se aplica: 13 estudiantes Toma de notas en clase: Nunca: 4 estudiantes A veces: 27 estudiantes Siempre: 26 estudiantes

AA	AB	AC	AD	AE	AF	AG	
ening in cla	interest a	p-classroo	int average	de point a	COURSE ID	GRADE	C
2	1	1	2	2	1	1	
2	1	2	4	4	1	2	
3	3	3	4	4	1	5	
2	2	2	4	3	1	5	
2	2	2	1	2	1	0	
3	2	1	4	4	1	2	
1	2	3	4	4	1	1	
1	3	3	3	3	1	1	
1	3	2	4	4	1	2	
2	3	3	5	4	1	3	
2	2	2	4	3	1	5	
1	2	1	4	3	1	2	
2	3	3	5	4	1	1	
2	2	2	4	3	1	2	
1	3	1	2	3	1	1	
2	2	1	4	3	1	1	
2	3	1	4	2	1	5	
1	3	2	5	3	1	3	
2	2	1	3	2	1	2	
2	2	3	3	3	1	1	
1	3	1	5	3	1	3	
2	3	1	5	4	1	4	
3	3	2	5	4	1	3	
2	3	1	5	4	3	7	
2	3	2	3	3	3	6	
2	3	2	5	4	5	5	
1	2	3	5	3	6	6	
3	2	3	4	2	6	6	

Figura 32: Continuación del grupo 5

3.4. Corte en altura 20

Para el corte 20 tenemos un total de 2 grupos, cada grupo esta compuesto por una cantidad distinta de objetos que se muestran en una hoja diferente de excel

3.4.1. Para el grupo 1

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
STUDENT ID	Student Age	Sex	had high school scholarship	additional academic or sports	have a partner	if available to the student type	of sister/	rental	staff	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/	of sister/
1	2	2	3	3	1	2	2	1	1	1	1	2	3	1	2	5	3	2	2	1	1	1	1	1	3
4	1	1	1	3	1	2	1	2	1	2	1	2	5	1	2	1	3	1	2	1	1	1	1	1	3
14	2	1	2	5	2	2	2	1	1	1	2	2	2	1	2	3	1	2	1	1	1	2	1	1	3
17	1	1	2	5	2	1	2	1	1	1	2	2	4	1	2	3	2	2	2	1	1	2	1	1	3
19	1	1	2	4	2	2	2	3	1	1	2	2	5	1	2	5	5	3	2	2	1	2	1	1	3
20	1	2	1	3	2	2	1	2	2	2	3	3	3	1	2	4	4	2	2	1	2	1	1	1	3
23	2	2	2	3	1	2	1	1	1	1	1	1	4	2	2	4	3	3	3	1	1	1	1	1	3
24	3	2	2	2	1	1	2	5	1	1	1	4	3	1	2	1	3	2	3	1	1	2	1	1	3
26	2	2	2	3	2	2	1	1	1	2	1	4	3	1	2	4	2	2	2	1	1	1	1	1	2
28	1	2	1	3	1	2	2	1	1	1	1	1	4	1	2	3	3	2	2	1	1	2	1	1	3
31	2	2	2	5	1	1	1	1	1	2	1	1	5	1	2	4	2	2	2	1	1	1	2	1	2
32	3	2	2	3	1	2	2	1	1	2	1	2	4	1	2	3	2	2	3	1	1	1	1	1	3
34	2	1	2	3	1	2	1	1	1	1	1	2	5	1	2	3	2	1	2	1	1	1	1	1	1
36	1	2	1	4	2	2	2	3	1	1	1	2	5	1	2	3	2	2	2	1	1	2	1	1	2
40	2	1	2	3	2	2	1	1	1	2	1	3	1	1	2	4	1	2	1	1	1	1	1	1	2
42	3	2	2	3	1	2	2	2	1	2	1	4	2	1	2	2	3	2	2	1	1	2	1	1	3
46	1	2	2	3	2	2	1	4	1	1	2	3	5	1	2	4	3	2	2	1	1	1	1	1	2
52	2	1	3	3	1	1	2	1	1	1	1	2	5	1	2	1	1	2	2	1	1	2	1	1	2
53	2	1	2	3	1	2	2	2	1	1	1	3	4	1	2	1	3	1	1	1	1	1	1	2	2
54	2	1	2	3	2	1	2	2	1	1	1	1	5	1	2	1	1	2	2	1	1	2	1	1	2
56	3	2	2	3	1	2	1	4	1	2	1	1	5	3	2	1	1	2	2	1	1	1	3	3	1
57	2	2	2	3	2	1	2	1	1	1	1	3	5	1	2	4	4	2	3	1	1	1	1	2	3
59	3	2	2	3	2	2	1	3	1	1	1	3	5	2	2	1	3	2	2	1	1	1	1	1	3
62	1	2	3	3	2	1	2	1	2	1	3	3	4	1	2	3	3	1	2	1	3	2	1	1	3
67	2	2	2	3	2	2	1	1	1	1	3	2	5	1	2	5	2	2	2	1	1	1	1	1	3
69	2	1	2	4	1	2	2	1	1	1	2	3	5	3	2	5	1	2	2	1	1	2	1	1	2
72	1	1	3	4	2	2	2	1	1	3	1	3	5	1	2	4	2	2	1	1	1	1	1	1	2
74	2	2	2	4	2	2	2	1	1	2	1	3	4	1	2	4	3	1	1	2	1	1	1	2	3

Figura 33: Resumen del grupo 1: En cuanto al género de los estudiantes, hay 27 mujeres y 30 hombres en este conjunto de datos. Respecto a tener una pareja, 9 estudiantes reportan tener una pareja, mientras que 48 estudiantes no tienen una pareja actualmente. En relación con las horas de estudio semanales, 12 estudiantes estudian menos de 5 horas a la semana, 28 estudiantes estudian entre 6 y 10 horas semanales, 11 estudiantes dedican entre 11 y 20 horas semanales al estudio, y 6 estudiantes estudian más de 20 horas por semana. En cuanto a la asistencia a seminarios o conferencias relacionadas con su carrera, 13 estudiantes asisten a estos eventos, mientras que 44 estudiantes no asisten. Respecto a la asistencia a clases, 23 estudiantes asisten siempre, 25 estudiantes asisten solo a veces, y 9 estudiantes nunca asisten a clases. Para la preparación de exámenes parciales, 31 estudiantes se preparan solos, 13 estudiantes se preparan con amigos, y para 13 estudiantes no se aplica esta categoría. En cuanto a la toma de notas en clase, 4 estudiantes nunca toman notas, 27 estudiantes toman notas a veces, y 26 estudiantes toman notas siempre durante las clases.

AA	AB	AC	AD	AE	AF	AG	
Learning in class	Interest in class	Pre-classroom	Post-classroom	Int. average	Grade point average	COURSE ID	GRADE
2	1	2	1	1	1	1	
2	2	1	3	2	1	1	
2	3	3	4	2	1	1	
2	3	3	4	3	1	1	
1	3	3	3	3	1	2	
2	2	3	2	3	1	3	
1	2	3	3	3	1	3	
2	3	3	3	3	1	1	
1	3	2	1	2	1	3	
1	2	1	2	1	1	1	
3	3	3	5	4	1	5	
3	2	3	4	3	1	3	
3	2	2	2	3	1	2	
1	3	1	1	1	1	1	
2	2	3	2	1	1	1	
2	2	3	4	3	1	1	
2	2	1	4	3	1	3	
3	3	1	3	3	1	4	
2	2	1	4	2	1	1	
2	2	2	3	2	1	5	
3	3	3	5	4	1	3	
2	3	3	5	4	1	5	
2	2	2	5	4	1	3	
2	2	2	4	4	1	5	
2	2	3	5	4	2	5	
2	3	2	4	3	3	5	
3	3	3	2	2	3	6	
2	3	1	5	3	3	6	

Figura 34: Continuacion de la figura 1

3.4.2. Para el grupo 2

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
STUDENT ID	Student Age	Sex	hd high scholarship	ditional	edistic or spu	have a palar	if avtion to the	tion type	ersee	edubae	educ	of sister/	rental	stafersae	ocugae	ocupatly	study	hnon-sdienly	scientifi	conferenc	ect/activ	dance to ch	to midte	notes in e	
2	2	2	3	3	1	2	2	1	1	1	2	3	2	1	2	1	2	2	2	1	1	1	1	1	3
3	2	2	2	3	2	2	2	2	4	2	2	2	2	1	2	1	2	1	2	1	1	1	1	1	2
5	2	2	1	3	2	2	1	3	1	4	3	3	2	1	2	4	2	1	1	1	1	1	2	1	2
6	2	2	2	3	2	2	2	2	1	1	3	3	2	1	2	3	1	1	2	1	1	1	1	1	1
7	1	2	2	4	2	2	2	1	1	3	1	3	1	1	2	4	2	2	2	2	1	2	1	1	3
8	1	1	2	3	1	1	1	2	2	3	4	3	1	1	4	3	1	2	2	1	1	1	3	1	3
9	2	1	3	3	2	1	1	1	1	3	2	4	2	1	2	4	1	2	2	1	1	1	1	1	3
10	2	1	2	3	2	2	1	3	4	2	1	2	3	1	2	3	2	2	2	1	1	2	1	1	2
11	1	1	1	3	2	2	2	3	2	3	3	4	2	1	3	2	1	1	1	1	1	2	1	1	2
12	1	1	1	4	1	1	2	4	2	3	5	5	1	1	3	2	3	3	3	1	3	1	3	2	3
13	1	1	1	4	2	2	2	1	1	1	3	5	4	2	2	2	3	3	2	2	1	1	1	1	2
15	3	2	2	4	1	1	2	3	4	2	3	1	2	1	4	1	2	2	2	1	1	1	1	1	2
16	2	2	2	3	2	2	2	1	1	2	4	4	2	1	3	1	2	2	2	2	3	2	1	1	3
18	2	2	2	3	2	2	2	1	1	1	2	2	2	1	2	1	2	2	2	2	1	2	1	1	2
21	1	2	2	5	1	2	1	1	4	2	3	3	3	2	2	3	4	2	2	2	1	1	1	1	3
22	1	2	2	5	2	2	1	1	4	2	2	2	4	1	2	3	3	2	2	1	1	1	1	1	3
25	2	2	2	3	2	2	2	2	1	1	3	1	3	1	2	4	1	2	2	1	1	1	1	1	2
27	2	2	2	3	2	1	1	1	1	1	4	4	4	1	2	4	2	3	2	1	1	1	1	1	3
29	3	2	2	3	2	2	1	1	4	2	2	2	4	1	2	1	2	2	2	2	1	1	1	1	3
30	2	2	3	4	2	2	2	1	4	2	3	3	2	1	4	1	2	2	2	2	1	1	1	1	3
33	2	1	2	3	2	2	2	2	1	1	2	2	2	1	2	1	2	2	2	1	1	2	1	1	2
35	1	2	1	3	2	2	1	2	1	1	3	3	3	1	2	4	2	1	1	1	1	2	1	1	2
37	2	2	3	4	1	2	1	3	1	1	1	1	2	2	4	3	1	1	3	1	1	2	2	1	2
38	2	2	2	3	1	1	1	2	2	3	3	3	1	1	2	3	2	2	2	1	1	1	1	1	3
39	2	2	2	5	2	2	2	1	1	1	3	3	1	1	2	4	3	2	2	1	1	1	2	1	2
41	1	2	1	3	2	2	2	2	1	1	2	3	1	1	2	5	1	2	2	1	1	1	1	1	3
43	2	2	2	3	2	1	2	1	4	2	4	4	2	1	3	1	2	2	1	1	1	1	1	1	2
44	1	2	2	3	2	2	1	1	1	1	2	3	3	1	2	1	2	2	2	1	1	2	1	1	3

Figura 35: En cuanto al género de los estudiantes, se observa que hay una cantidad ligeramente mayor de hombres que de mujeres en este conjunto de datos. La distribución de edades muestra que la mayoría de los estudiantes se encuentran en el rango de 22 a 25 años, lo que sugiere que son estudiantes de cursos superiores o de posgrado. Respecto a las becas, una cantidad significativa de estudiantes cuenta con becas parciales o completas, incluyendo un grupo importante que tiene becas que cubren el 100%. En relación con tener una pareja, la gran mayoría de los estudiantes no reporta tener una pareja actualmente, lo que les permitiría dedicar más tiempo a sus estudios sin tener tantas responsabilidades adicionales. Sobre las horas de estudio semanales, se observa una distribución variada, con una proporción considerable de estudiantes que estudian entre 6 y 10 horas a la semana, mientras que otros dedican menos de 5 horas o más de 20 horas semanales al estudio. Esta variación podría tener un impacto en su rendimiento académico. En cuanto a la asistencia a seminarios o conferencias relacionadas con su carrera, la mayoría de los estudiantes no asiste a estos eventos, lo que podría limitar su exposición a conocimientos y oportunidades adicionales en su área de estudio. Respecto a la asistencia a clases, hay una distribución más equilibrada entre los estudiantes que asisten siempre, a veces o nunca, lo que podría influir en su desempeño académico. Para la preparación de exámenes parciales, la mayoría de los estudiantes se prepara sola o con amigos, mientras que para algunos no se aplica esta categoría, posiblemente porque no tienen exámenes parciales o porque utilizan otras estrategias de preparación. En cuanto a la toma de notas en clase, hay una distribución variada, con una proporción significativa de estudiantes que toma notas a veces o siempre, y una menor cantidad que nunca toma notas. Sobre las calificaciones promedio del último semestre y las esperadas al graduarse, se observa una amplia distribución en todo el rango, desde calificaciones más bajas hasta calificaciones más altas, lo que indica una diversidad en el desempeño académico de los estudiantes. Finalmente, las calificaciones finales del curso también muestran una gran variedad, desde "Fail" hasta "A", lo que refleja una amplia gama de resultados académicos en este conjunto de datos.

AA	AB	AC	AD	AE	AF	AG	
ening in cla	interest a	p-classroom	int averaged	de point a	COURSE ID	GRADE	G
2	3	2	2	3	1	1	
2	1	1	2	2	1	1	
2	2	1	2	2	1	1	
2	1	2	4	4	1	2	
3	3	3	4	4	1	5	
2	2	1	1	1	1	2	
2	2	2	4	3	1	5	
2	2	2	1	2	1	0	
2	2	2	1	1	1	2	
1	3	3	4	3	1	0	
2	2	3	4	2	1	0	
3	2	1	4	4	1	2	
2	2	3	2	2	1	2	
2	2	2	2	2	1	2	
1	2	3	4	4	1	1	
1	3	3	3	3	1	1	
1	3	2	4	4	1	2	
3	3	2	2	1	1	1	
2	3	3	5	4	1	3	
2	2	2	4	3	1	5	
2	2	1	2	3	1	1	
2	3	1	4	1	1	2	
1	2	1	4	3	1	2	
2	3	3	5	4	1	1	
2	2	2	4	3	1	2	
1	2	1	2	3	1	1	
1	3	1	2	3	1	1	
1	3	3	3	2	1	4	

Figura 36: Continuación del grupo 2

4. CONCLUSIONES

Basadas en el análisis jerárquico de agrupamiento utilizando el método de Ward, se pueden extraer varias conclusiones:

1. **Identificación de Grupos Naturales:** El análisis permite identificar grupos naturales dentro de los datos de los estudiantes, representando diferentes perfiles basados

en características como la edad, género y rendimiento académico.

2. **Estructura Jerárquica:** El dendrograma muestra la estructura jerárquica de los grupos, proporcionando información sobre la similitud entre ellos y la relación de inclusión.
3. **Interpretación de Resultados:** La interpretación del dendrograma y grupos permite comprender las características que definen a cada grupo, útil para identificar patrones de comportamiento o necesidades específicas.
4. **Segmentación de Estudiantes:** Los grupos identificados pueden usarse para segmentar a los estudiantes en función de características similares, con aplicaciones en la personalización de la enseñanza o el diseño de programas educativos.
5. **Validación y Evaluación:** Es importante validar los resultados mediante técnicas de validación interna o externa para asegurar su robustez y relevancia.

En resumen, el análisis jerárquico de agrupamiento utilizando el método de Ward proporciona una herramienta poderosa para explorar y comprender la estructura subyacente en los datos de los estudiantes, lo que puede llevar a conclusiones significativas y acciones prácticas en el ámbito educativo.

5. APÉNDICES

5.1. Código del Programa

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import openpyxl
5 import tkinter as tk
6 from tkinter import filedialog
7
8 def cargar_csv():
9     # Abrir cuadro de diálogo para seleccionar archivo CSV
10    archivo_csv = filedialog.askopenfilename(
11        title="Seleccionar archivo CSV", filetypes=[("Archivos CSV", "*.csv")]
12    )
13    if archivo_csv:
14        cargar_datos(archivo_csv)
15    else:
16        mensaje.set("No se seleccionó ningún archivo CSV.")
17
18 def cargar_datos(archivo_csv):
19     # Cargar los datos del archivo CSV
20     datos = pd.read_csv(archivo_csv)
21
22     # Definir las columnas numéricas relevantes para el análisis
23     columnas_numericas = [
24         "Student Age",
25         "Sex",
26         "Graduated high-school type",
27         "scholarship type",
28         "additional work",
29         "regular artistic or sport activity",
30         "do you have a partner",
31         "total salary if available",
32         "transportation to the university",
33         "accommodation type in cyprus",
34         "mothersae education",
35         "fathersae education",
36         "number of sister/brother",
37         "parental status",
38         "mothersae occupation",
39         "farsae occupation",
40         "weekly study hours",
41         "reading frequency (non-scientify book/journals)",
42         "reading frequency (scientific books/journals)",
43         "Attendance to the seminars/conferences related to the department",
44         "Impact of your projects/activities on your success",
45         "Attendance to classes",
46         "Preparation to midterm exams 1",
47         "Preparation to midterm exams 2",
48         "Taking notes in classes",
49         "Listening in classes",
50         "Discussion improves my interest and success in the course",
51         "Flip-classroom",
```

Figura 37:

```

52     "Cumulative grade point average in the last semester",
53     "Expected Cumulative grade point average in the graduation",
54 ]
55
56 # Filtrar las columnas que existen en el archivo CSV
57 columnas_existentes = datos.columns.tolist()
58 columnas_numericas = [
59     col for col in columnas_numericas if col in columnas_existentes
60 ]
61
62 # Obtener los datos numéricos y eliminar filas con valores faltantes
63 datos_numericos = datos[columnas_numericas].dropna()
64
65 # Convertir los datos a una matriz de distancias
66 matriz_distancias = datos_numericos.values
67
68 # Aplicar el método de Ward
69 Z = []
70 clusters = {i: [i] for i in range(len(matriz_distancias))} # Inicializar cada punto como un cluster individual
71 distancias = np.array([[np.linalg.norm(matriz_distancias[i] - matriz_distancias[j]) for j in range(len(matriz_distancias))] for i in range(len(matriz_distancias))])
72 np.fill_diagonal(distancias, np.inf) # Llenar la diagonal con infinito para evitar que un punto se fusione consigo mismo
73
74 while len(clusters) > 1:
75     # Encontrar los dos clusters más cercanos
76     min_dist = np.inf
77     to_merge = None
78     for i in clusters:
79         for j in clusters:
80             if i < j:
81                 # Calcular la distancia entre clusters usando el criterio de Ward
82                 dist = sum(distancias[p][q] ** 2 for p in clusters[i] for q in clusters[j])
83                 dist /= len(clusters[i]) * len(clusters[j])
84                 if dist < min_dist:
85                     min_dist = dist
86                     to_merge = (i, j)
87
88     if to_merge:
89         i, j = to_merge
90         new_cluster = clusters[i] + clusters[j] # Fusionar los clusters
91         for k in clusters:
92             if k != i and k != j:
93                 # Actualizar las distancias entre el nuevo cluster y los demás clusters
94                 dist = sum(distancias[p][q] ** 2 for p in new_cluster for q in clusters[k])
95                 dist /= len(new_cluster) * len(clusters[k])
96                 distancias[new_cluster[0]][clusters[k][0]] = dist
97                 distancias[clusters[k][0]][new_cluster[0]] = dist
98
99         # Añadir la nueva combinación de clusters a la lista Z
100         Z.append([i, j, np.sqrt(min_dist), len(new_cluster)])
101         del clusters[i]
102         del clusters[j]

```

Figura 38:

```

103     clusters[min(clusters.keys())] = new_cluster # Añadir el nuevo cluster al diccionario
104     distancias[new_cluster[0], new_cluster[0]] = np.inf # Evitar que el nuevo cluster se fusione consigo mismo
105
106 Z = np.array(Z) # Convertir Z a un array de numpy para su posterior uso
107
108 # Graficar el dendrograma
109 plt.figure(figsize=(25, 10))
110 plt.title("Dendrograma")
111 plt.xlabel("Objetos")
112 plt.ylabel("Distancia")
113 plot_dendrogram(Z) # Función que utiliza dendrograma de scipy para graficar
114
115 # Obtener los límites del eje y para determinar la altura máxima
116 ylim = plt.ylim()
117 altura_maxima = ylim[1]
118
119 # Dibujar las líneas de corte en el dendrograma
120 lineas_corte_altura = [20, 15, 10]
121 colores = ["r", "g", "b"]
122 for i, altura in enumerate(lineas_corte_altura):
123     plt.axhline(
124         y=altura,
125         c=colores[i],
126         linestyle="--",
127         label=f"Línea de corte {i+1} (t={altura})",
128     )
129
130 # Ajustar los límites del eje y para incluir las líneas de corte
131 plt.ylim(0, altura_maxima)
132
133 # Agregar una leyenda
134 plt.legend()
135
136 # Generar las líneas de corte y los archivos Excel
137 for altura in [20, 15, 10]:
138     grupo = fcluster_manual(Z, t=altura) # Asignar clusters manualmente con el método definido
139     datos[f"Grupo_{altura}"] = grupo # Añadir la columna de grupo al dataframe
140     grupos = datos.groupby(f"Grupo_{altura}") # Agrupar los datos por cluster
141
142     # Crear un libro de Excel
143     escritor = pd.ExcelWriter(f"grupos_corte_{altura}.xlsx", engine="openpyxl")
144
145     # Agregar una hoja para cada grupo
146     for nombre_grupo, grupo_datos in grupos:
147         grupo_datos.to_excel(
148             escritor, sheet_name=f"Grupo_{nombre_grupo}", index=False
149         )
150
151     # Guardar el libro de Excel
152     escritor.save()
153     escritor.close()

```

Figura 39:

```

154     mensaje.set(f"Libros de Excel creados correctamente :D.")
155
156     plt.show()
157
158     def plot_dendrogram(Z):
159         from scipy.cluster.hierarchy import dendrogram
160         dendrogram(Z) # Utiliza la función dendrogram de scipy para graficar
161
162     def fcluster_manual(Z, t):
163         # Asignar clusters manualmente basándose en una altura de corte t
164         clusters = {i: [i] for i in range(len(Z) + 1)}
165         for row in Z:
166             if row[2] < t:
167                 i, j = int(row[0]), int(row[1])
168                 new_cluster = clusters.pop(i) + clusters.pop(j)
169                 clusters[min(clusters.keys())] = new_cluster
170
171         cluster_ids = np.zeros(len(Z) + 1)
172         for cluster_id, indices in clusters.items():
173             for index in indices:
174                 cluster_ids[index] = cluster_id
175
176         return cluster_ids
177
178     def salir():
179         ventana.destroy()
180
181     # Crear ventana principal
182     ventana = tk.Tk()
183     ventana.title("AGRUPAMIENTO JERARQUICO")
184     ventana.geometry("800x600") # Tamaño de la ventana
185
186     # Colores para la GUI
187     color_fondo = "#FFD700" # Amarillo
188     color_botones = "#FF6347" # Rojo
189
190     # Configurar color de fondo
191     ventana.configure(bg=color_fondo)
192
193     # Crear título
194     etiqueta_titulo = tk.Label(ventana, text="AGRUPAMIENTO JERARQUICO", font=("Arial", 20, "bold"), bg=color_fondo)
195     etiqueta_titulo.pack(pady=20)
196
197     # Crear botón para cargar CSV
198     boton_cargar_csv = tk.Button(ventana, text="Cargar CSV", command=cargar_csv, bg=color_botones, fg="white", font=("Arial", 14))
199     boton_cargar_csv.pack(pady=20)
200
201     # Crear botón para salir del programa
202     boton_salir = tk.Button(ventana, text="SALIR", command=salir, bg=color_botones, fg="white", font=("Arial", 14))
203     boton_salir.pack(pady=10)
204

```

Figura 40:

```

204     boton_salir.pack(pady=10)
205
206     # Mensaje para mostrar el resultado
207     mensaje = tk.StringVar()
208     etiqueta_mensaje = tk.Label(ventana, textvariable=mensaje, font=("Arial", 12), bg=color_fondo)
209     etiqueta_mensaje.pack(pady=10)
210
211     # Ejecutar el bucle principal de la ventana
212     ventana.mainloop()

```

Figura 41:

5.2. Conceptos Usados

Distancia Euclidiana: La distancia euclidiana es una medida de disimilitud comúnmente utilizada en el análisis de clústeres. Calcula la distancia geométrica entre dos puntos en un espacio multidimensional. Para dos puntos $X = (x_1, x_2, \dots, x_n)$ e $Y = (y_1, y_2, \dots, y_n)$ en un espacio n -dimensional, la distancia euclidiana se define como:

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

Fórmula de Ward: El método de Ward se basa en la idea de minimizar la suma de las desviaciones cuadradas dentro de los clústeres. En cada paso, se combinan los dos clústeres que resulten en el mínimo aumento de la suma total de las desviaciones cuadradas dentro

de los clústeres. La fórmula utilizada por el método de Ward para calcular la distancia entre dos clústeres A y B es:

$$d(A, B) = \sqrt{\frac{|A| + |B|}{|A| + |B|}} \cdot d(\text{centroide}(A), \text{centroide}(B))$$

Donde:

$|A|$ y $|B|$ son el número de observaciones en los clústeres A y B , respectivamente.

$d(\text{centroide}(A), \text{centroide}(B))$ es la distancia euclidiana entre los centroides de los clústeres A y B . Esta fórmula tiene en cuenta tanto la distancia entre los centroides como el tamaño de los clústeres. Tiende a producir clústeres compactos y de tamaños similares.

Dendrograma: El resultado del agrupamiento jerárquico se suele representar gráficamente mediante un dendrograma, que es una representación en forma de árbol que muestra cómo se agrupan las observaciones en clústeres. Los clústeres se fusionan de forma ascendente, comenzando con cada observación como un clúster individual y combinándolos sucesivamente hasta que todas las observaciones forman un único clúster.

5.3. Manual de Usuario

A continuación se muestra paso a paso como utilizar el programa

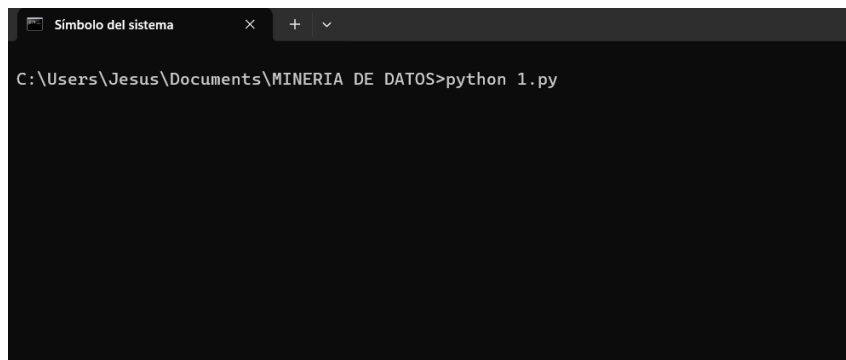


Figura 42: Se ejecuta la terminal y se navega a la dirección en la que se encuentra el programa. Estando ya dentro, se ejecuta la siguiente instrucción: "python 1.py". Se tiene que poner "python" seguido del nombre de tu programa, para que la terminal sepa que va a ejecutar un archivo de lenguaje Python y lo pueda leer sin problema.



Figura 43: A continuación se abre la interfaz, y lo que haremos será darle click en **CARGAR CSV**.

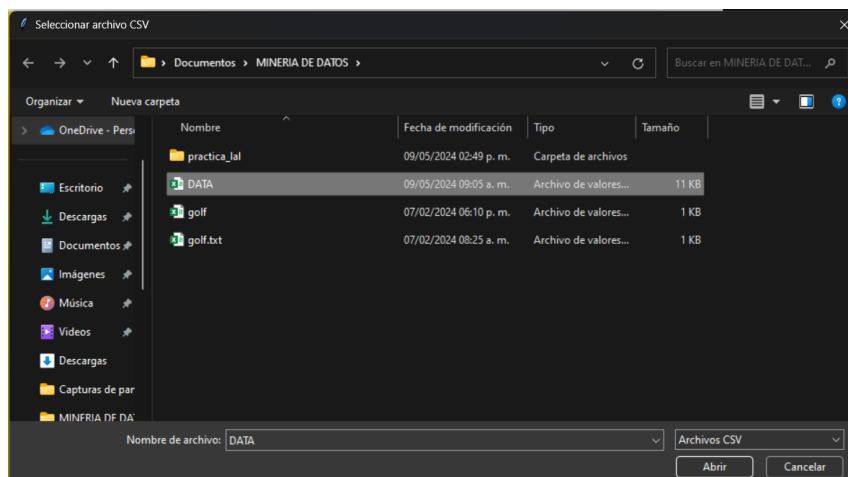


Figura 44: Seleccionamos DATA y lo abrimos



Figura 45: El programa nos mandará un mensaje confirmando que los Excel han sido creados

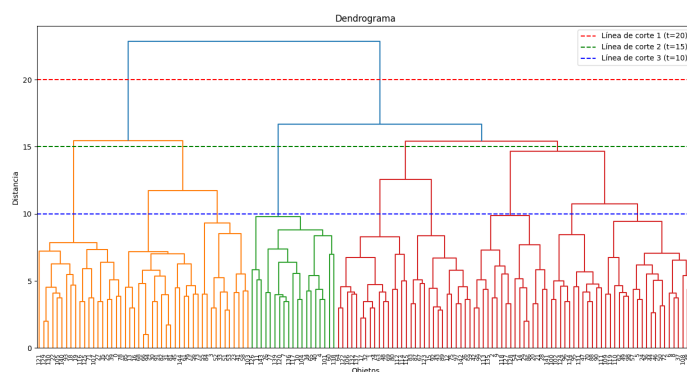


Figura 46: Aparte del mensaje, se nos abrirá una nueva ventana en la que se muestra la imagen del dendograma y sus líneas de corte




	grupos_corte_10	11/05/2024 06:37 p. m.	Hoja de cálculo d...	63 KB
	grupos_corte_15	11/05/2024 06:37 p. m.	Hoja de cálculo d...	47 KB
	grupos_corte_20	11/05/2024 06:37 p. m.	Hoja de cálculo d...	41 KB

Figura 47: Verificamos en la carpeta donde está guardado nuestro programa que ahora también ya se encuentran 3 nuevos documentos de Excel, que son los que se acaban de crear cuando cargamos el data a nuestro programa

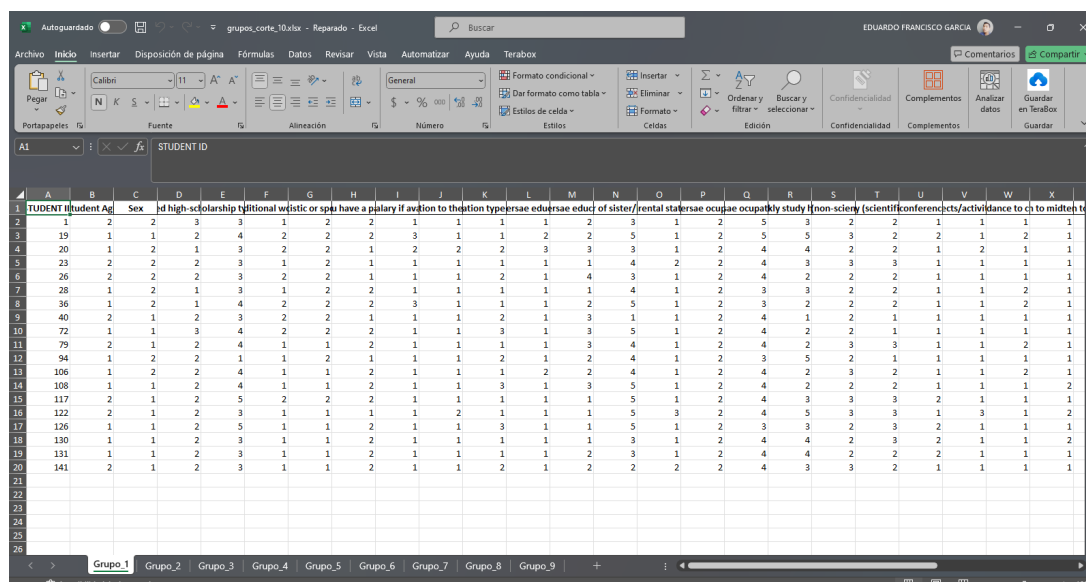


Figura 48: CORTE 10 (9 GRUPOS).

Figura 49: CORTE 15 (5 GRUPOS).

Figura 50: CORTE 20 (2 GRUPOS)

Abrimos un Excel y observaremos que se crean varias hojas, cada hoja representa un grupo. Cada Excel tiene diferentes hojas ya que cada Excel representa un corte específico y cada corte tiene diferentes grupos.

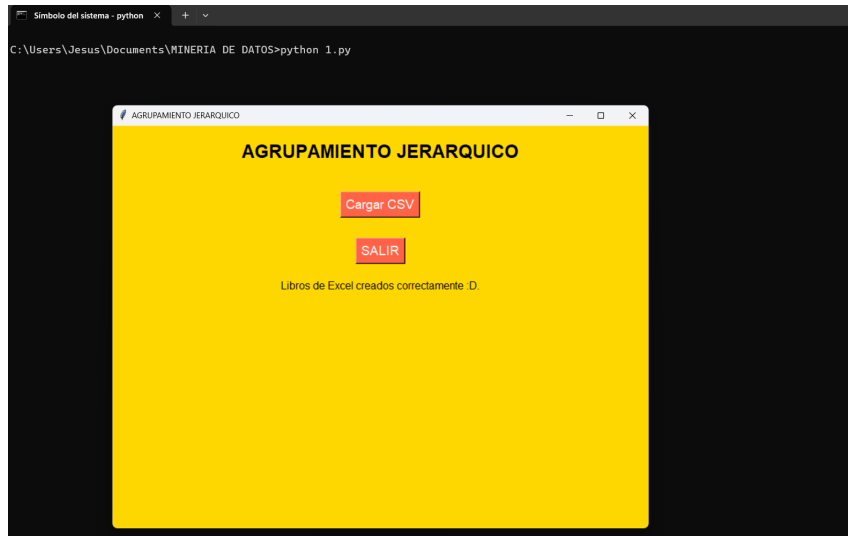


Figura 51: Regresamos a nuestra interfaz y ahora pulsaremos el botón SALIR.

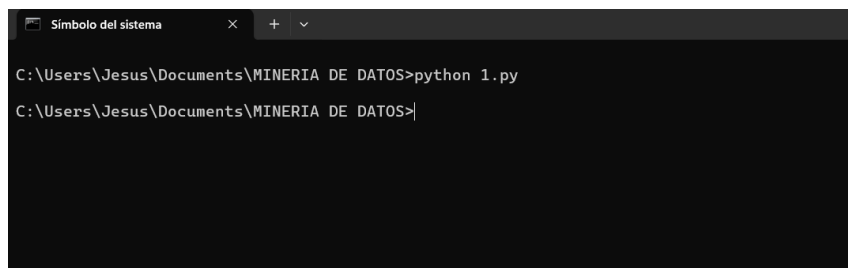


Figura 52: Verificamos en la terminal que, en efecto, el programa ya finalizó.

5.4. Manual técnico

Instalar Python

Para instalar Python en tu sistema Windows, sigue estos pasos:

1. **Descargar el instalador de Python:** Ve al sitio web oficial de Python (<https://www.python.org/downloads/>) y descarga el instalador adecuado para tu sistema Windows. Se recomienda descargar la versión más reciente de Python, que esté marcada como estable.
2. **Ejecutar el instalador:** Una vez que se haya completado la descarga, debes hacer doble clic en el archivo descargado para ejecutar el instalador de Python. Asegúrate de marcar la opción *Agregar Python X.X a PATH* durante la instalación para que Python se agregue automáticamente a la variable de entorno PATH del sistema.
3. **Completar la instalación:** Sigue las instrucciones en pantalla para completar la instalación de Python. Asegúrate de marcar la opción para instalar *pip*, que es un administrador de paquetes de Python utilizado para instalar y administrar bibliotecas y paquetes de terceros.

4. **Verificar la instalación:** Una vez que la instalación haya finalizado, abre una ventana de comandos (CMD) y escribe el siguiente comando para verificar que Python se haya instalado correctamente:

```
python --version
```

Deberías ver la versión de Python instalada en tu sistema.

5. **Instalar un entorno de desarrollo (opcional):** Si planeas desarrollar en Python, considera instalar un entorno de desarrollo integrado (IDE) como Spyder, PyCharm o Visual Studio Code. Estos IDEs proporcionan características adicionales, como resaltado de sintaxis, depuración de código y administración de proyectos.

Adicionalmente, para una correcta ejecución es necesario importar las siguientes librerías, como **Pandas** para el manejo de datos, **NumPy** para operaciones matemáticas, **customtkinter** para la creación de la interfaz gráfica de usuario y **scipy** para la implementación del Dendrograma. Es importante mencionar que los comandos proporcionados a continuación tienen que ser ejecutados desde la terminal (Cmd):

```
# Instalar pandas
pip install pandas
```

```
# Instalar numpy
pip install numpy
```

```
# Instalar customtkinter
pip install customtkinter
```

```
# Instalar scipy
pip install scipy
```

5.5. Referencias

Página de descarga de la base de datos:

<https://archive.ics.uci.edu/dataset/856/higher+education+students+performance+evaluation>