



Instituto Politécnico Nacional

Escuela Superior de Cómputo

Minería de datos

Exposición: Clustering

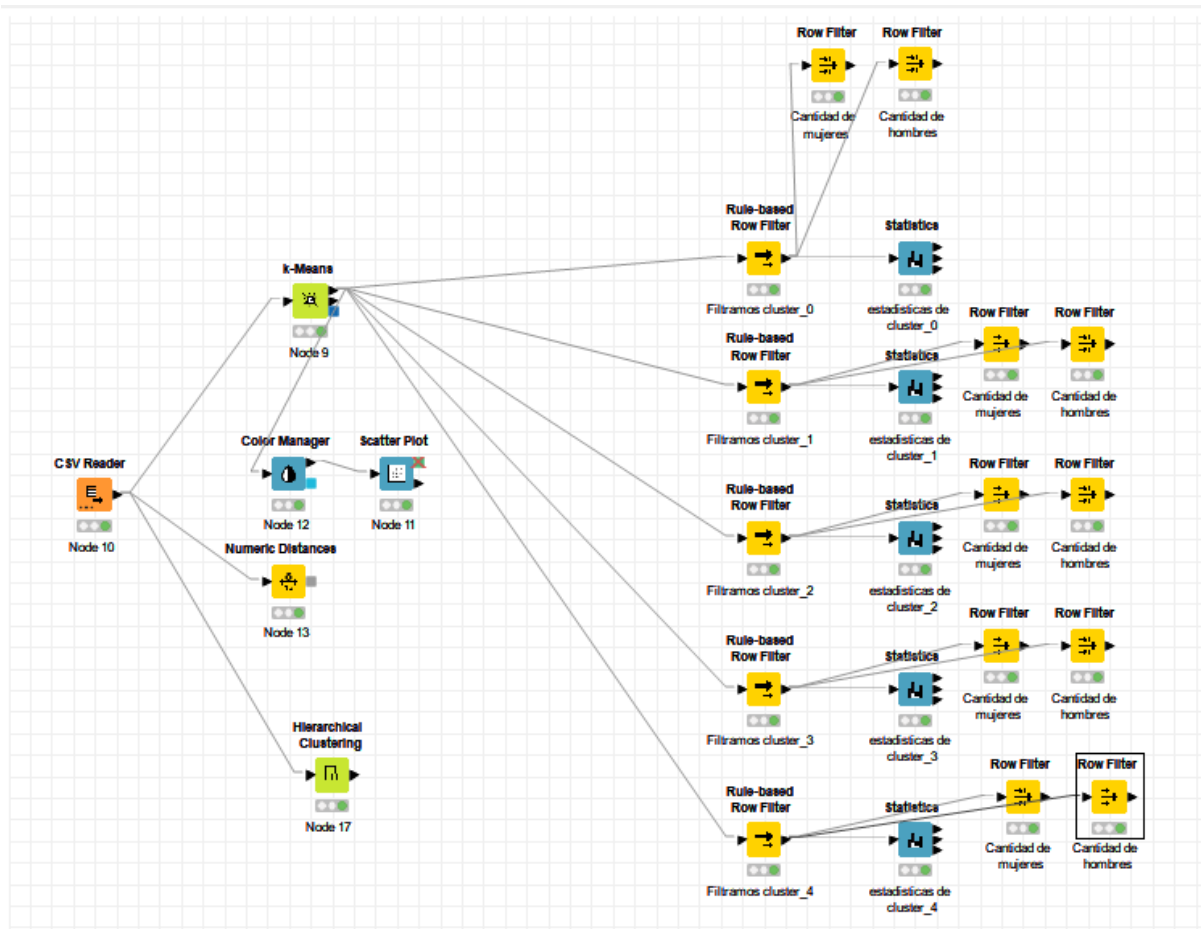
Equipo: 4

Alumnos: Becerril Hernández Aldo  
Lopez Garcia Felipe de Jesus

Grupo: 3CV19

Fecha: 25-11-2021

# Clustering



DataSet obtenido de Kaggle:

<https://www.kaggle.com/roshansharma/mall-customers-clustering-analysis/data>

+

 Create

Home

Competitions

Datasets

<> Code

Discussions

Courses

More

Search

Sign InRegister

ROSHAN SHARMA · 1Y AGO · 107,198 VIEWS

574Copy & Edit1355

# Mall Customers Clustering Analysis

Python · [Mall Customer Segmentation Data](#)

NotebookDataLogsComments (126)

## Data

Mall\_Customers.csv (3.98 kB)

DetailCompactColumn

5 of 5 columns

About this file

This file contains the basic information (ID, age, gender, income, spending score) about the customers

CustomerID

Gender

Age

Annual Income (k\$)

Spending Score (1-10)

Input (3.98 kB)

Data Sources

Mall Customer Segment...

- Mall\_Customers.csv

View Active Events

+

 Create

Home

Competitions

Datasets

<> Code

Discussions

Courses

More

Search

Sign InRegister

Mall Customers Clustering Analysis

574Copy & Edit1355

NotebookDataLogsComments (126)

DetailCompactColumn

5 of 5 columns

About this file

This file contains the basic information (ID, age, gender, income, spending score) about the customers

CustomerID

Gender

Age

Annual Income (k\$)

Spending Score (1-10)

Female 56%

Male 44%

1870

15137

139

177

176

CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-10)
1	Male	19	15	39
2	Male	21	15	81
3	Female	28	16	6
4	Female	23	16	77
5	Female	31	17	48
6	Female	22	17	76

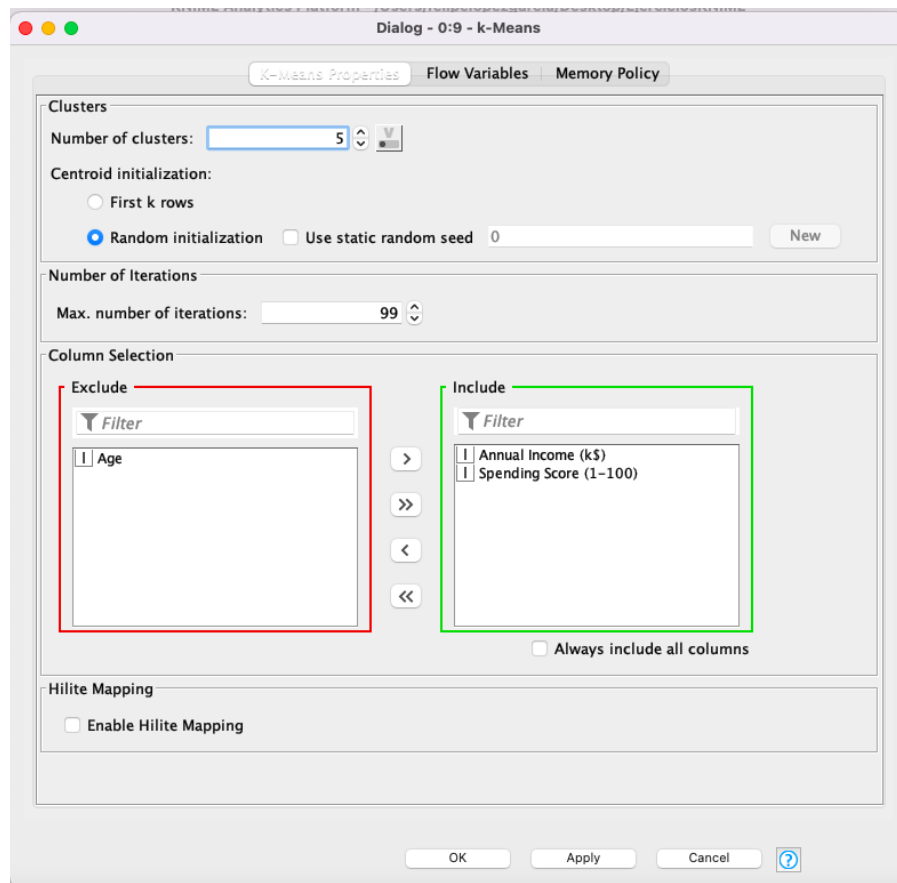
View Active Events

Mall Customer Segment...

Mall\_Customers.csv

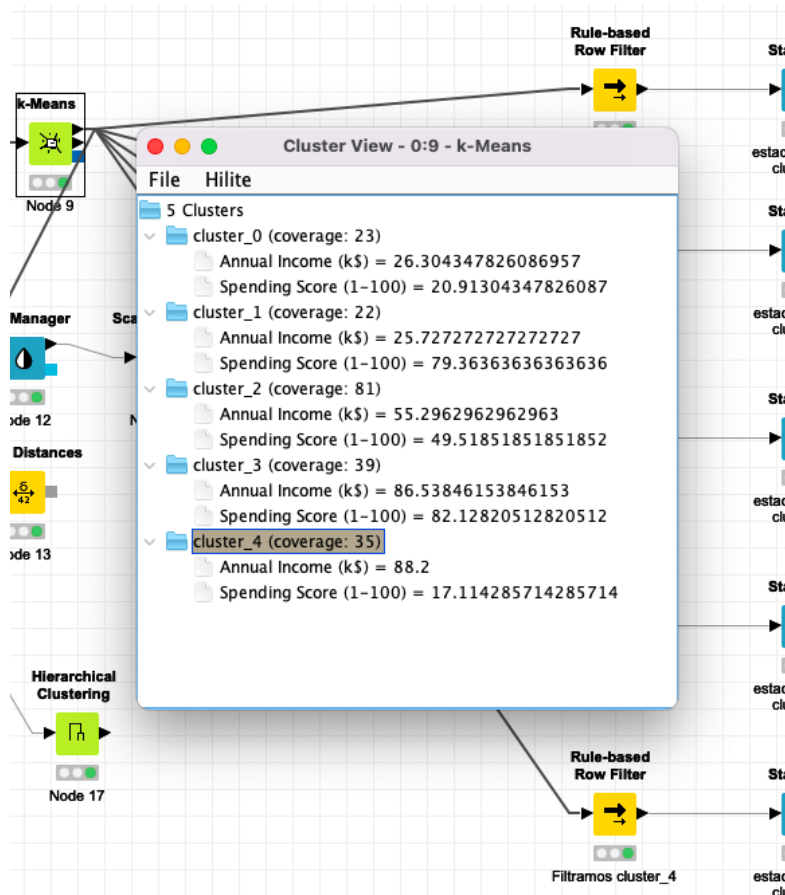
# K-Mean (K-Medias)

Cargamos los datos con el Módulo “CSV Reader”, y configuramos el nodo “K-Means”:



Definimos el número de clusters a 5. K-Means se basa en la idea de colocar unos centroides, a partir de los cuales mide la distancia de todos los demás nodos, hace una agrupación por cercanía. La forma más común de inicializar estos centroides es de manera aleatoria. Desactivamos la opción “Use static random seed” para que genere su propia semilla. Y para número de iteraciones será 99.

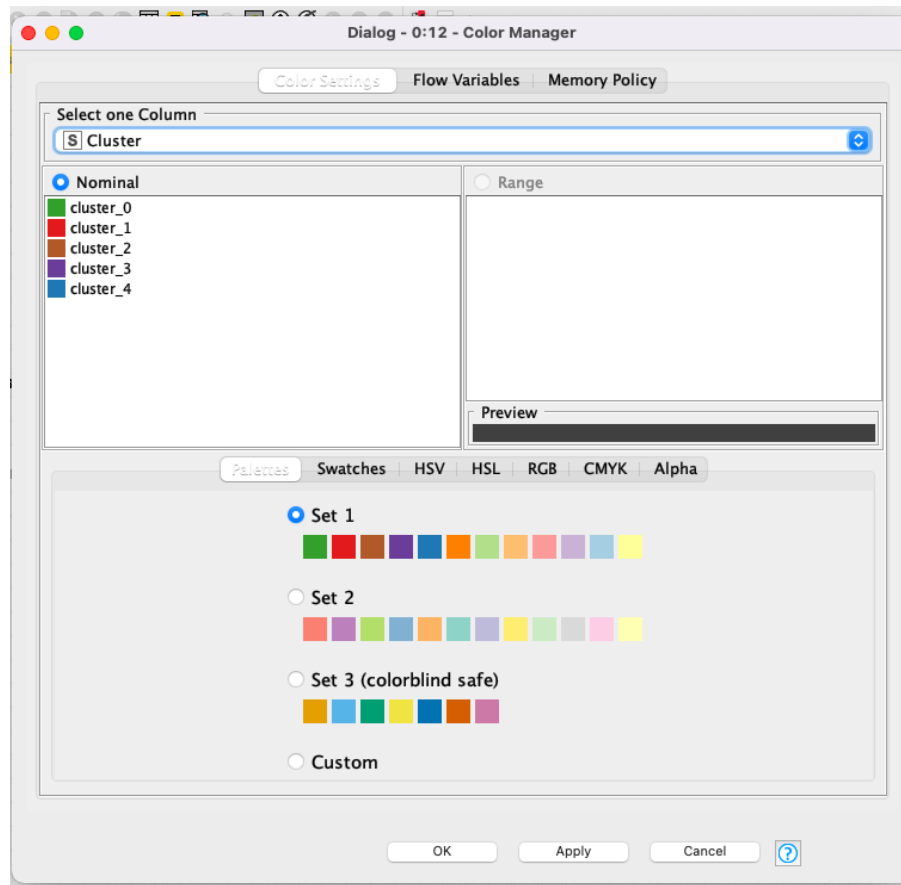
Además decidimos qué valores serán los cuales van a servir para agrupar y crear los clusters. En este ejemplo tomamos los valores “Ingresos Anuales” y “Gasto Total”, los cuales vienen en K’s (Miles de datos por unidad).



Se ejecuta el nodo y podemos visualizar el resultado. Produce los 5 clusters, de principio no sabe que nombre darle así que les pone uno por default. El parametro coverage nos dice que:

- Para el cluster\_0 agrupó 23 elementos
- Para el cluster\_1 agrupó 22 elementos
- Para el cluster\_2 agrupó 81 elementos
- Para el cluster\_3 agrupó 39 elementos
- Para el cluster\_4 agrupó 35 elementos

Si abrimos las ramas como en la ilustración, observamos que nos devuelve las coordenadas de cada uno de los centroides.



El nodo Color Manager nos permite definir colores para representar los clusters.

Dialog - 0:11 - Scatter Plot

Options | Axis Configuration | General Plot Options | View Controls | Flow Variables | Memory Policy

☐ Create image at output

Maximum number of rows: 2,500

Selection column name: Selected (Scatter Plot)

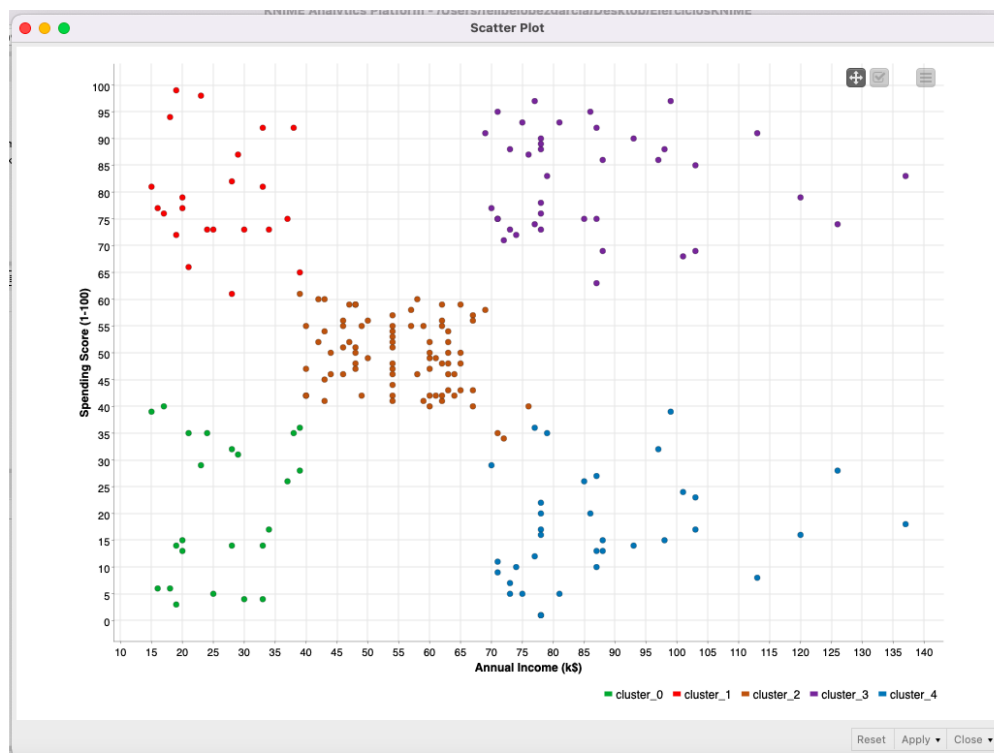
Choose column for x axis  
Annual Income (k\$)

Choose column for y axis  
Spending Score (1-100)

☒ Report on missing values

OK Apply Cancel ?

Para plotear el resultado del K-Means empatamos los valores “Ingresos Anuales” y “Gasto Total”



Ploteamos los resultados y observamos los clusters resultantes.

# Estadísticas de los clusters

La siguiente imagen muestra el conjunto de datos generado por el cluster\_0

▲ Filtered - 0:19 - Rule-based Row Filter (Filtramos cluster\_0)

File Edit Hilite Navigation View

Table "default" - Rows: 22 Spec - Columns: 5 Properties Flow Variables

Row ID	S Gender	I Age	I Annual ...	I Spendi...	S Cluster
2	Male	21	15	81	cluster_0
4	Female	23	16	77	cluster_0
6	Female	22	17	76	cluster_0
8	Female	23	18	94	cluster_0
10	Female	30	19	72	cluster_0
12	Female	35	19	99	cluster_0
14	Female	24	20	77	cluster_0
16	Male	22	20	79	cluster_0
18	Male	20	21	66	cluster_0
20	Female	35	23	98	cluster_0
22	Male	25	24	73	cluster_0
24	Male	31	25	73	cluster_0
26	Male	29	28	82	cluster_0
28	Male	35	28	61	cluster_0
30	Female	23	29	87	cluster_0
32	Female	21	30	73	cluster_0
34	Male	18	33	92	cluster_0
36	Female	21	33	81	cluster_0
38	Female	30	34	73	cluster_0
40	Female	20	37	75	cluster_0
42	Male	24	38	92	cluster_0
46	Female	24	39	65	cluster_0

En la imagen anterior podemos observar que contiene 22 elementos, y cada una de las filas contiene los datos de género, edad, ingresos anuales y puntuación de gastos de cada cliente

A continuación se muestran las estadísticas del cluster\_0

▲ Statistics Table - 0:20 - Statistics (estadísticas de)

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 16 Properties Flow Variables

Row ID	S Column	D Min	D Max	D Mean	D Std. de...	D Variance
Age	Age	18	35	25.273	5.257	27.636
Annual Incom...	Annual Inco...	15	39	25.727	7.567	57.255
Spending Sco...	Spending Sc...	61	99	79.364	10.504	110.338



En las estadísticas podemos observar que podemos observar que:

**La edad mínima** de las personas dentro de este cluster es **18** años.

**La edad máxima** de las personas dentro de este cluster es **35** años.

**El ingreso anual mínimo** de las personas de este cluster es **15**.

**El ingreso anual máximo** de las personas de este cluster es **39**.

**La puntuación mínima de gastos** de las personas de este cluster es **61**.

**La puntuación máxima de gastos** de las personas de este cluster es **99**.

**La media de edad** en este cluster es **25.27** años.

**La media de ingresos anuales** en este cluster es **25.72**.

**La media de puntuación de gastos** en este cluster es de 79.36 puntos.

**La desviación estándar de la edad** de las personas en este cluster es de **5.25**.

**La desviación estándar de los ingresos anuales** de las personas en este cluster es de **7.56**.

**La desviación estándar de la puntuación de gastos** de las personas del cluster es de **5.25**.

**La varianza de la edad** de las personas en este cluster es de **27.63**.

**La varianza de los ingresos anuales** de las personas en este cluster es de **57.25**.

**La varianza de la puntuación de gastos** de las personas en este cluster es de **110.33**.

Por último para el cluster\_0 se muestra la cantidad de hombres y mujeres que hay

Filtered - 0:29 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 13 Spec - Columns: 5 Properties Flow Variables

Row ID	S Gender	I Age	I Annual ...	I Spend...	S Cluster
4	Female	23	16	77	cluster_0
6	Female	22	17	76	cluster_0
8	Female	23	18	94	cluster_0
10	Female	30	19	72	cluster_0
12	Female	35	19	99	cluster_0
14	Female	24	20	77	cluster_0
20	Female	35	23	98	cluster_0
30	Female	23	29	87	cluster_0
32	Female	21	30	73	cluster_0
36	Female	21	33	81	cluster_0
38	Female	30	34	73	cluster_0
40	Female	20	37	75	cluster_0
46	Female	24	39	65	cluster_0

Como se observa en la imagen en este cluster se tienen **13** filas en las que el atributo género tiene la clase mujer.

Filtered - 0:31 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 9 Spec - Columns: 5 Properties Flow Variables

Row ID	S Gender	I Age	I Annual ...	I Spendi...	S Cluster
2	Male	21	15	81	cluster_0
16	Male	22	20	79	cluster_0
18	Male	20	21	66	cluster_0
22	Male	25	24	73	cluster_0
24	Male	31	25	73	cluster_0
26	Male	29	28	82	cluster_0
28	Male	35	28	61	cluster_0
34	Male	18	33	92	cluster_0
42	Male	24	38	92	cluster_0

Como se observa en la imagen en este cluster se tienen 9 filas en las que el atributo género tiene la clase hombre.

La siguiente imagen muestra el conjunto de datos generado por del cluster\_1

Filtered - 0:21 - Rule-based Row Filter (Filtramos cluster\_1)

File Edit Hilite Navigation View

Table "default" - Rows: 23 Spec - Columns: 5 Properties Flow Variables

Row ID	S Gender	I Age	I Annual ...	I Spendi...	S Cluster
1	Male	19	15	39	cluster_1
3	Female	20	16	6	cluster_1
5	Female	31	17	40	cluster_1
7	Female	35	18	6	cluster_1
9	Male	64	19	3	cluster_1
11	Male	67	19	14	cluster_1
13	Female	58	20	15	cluster_1
15	Male	37	20	13	cluster_1
17	Female	35	21	35	cluster_1
19	Male	52	23	29	cluster_1
21	Male	35	24	35	cluster_1
23	Female	46	25	5	cluster_1
25	Female	54	28	14	cluster_1
27	Female	45	28	32	cluster_1
29	Female	40	29	31	cluster_1
31	Male	60	30	4	cluster_1
33	Male	53	33	4	cluster_1
35	Female	49	33	14	cluster_1
37	Female	42	34	17	cluster_1
39	Female	36	37	26	cluster_1
41	Female	65	38	35	cluster_1
43	Male	48	39	36	cluster_1
45	Female	49	39	28	cluster_1

En la imagen anterior podemos observar que contiene 23 elementos, y cada una de las filas contiene los datos de género, edad, ingresos anuales y puntuación de gastos de cada cliente

A continuación se muestran las estadísticas del cluster\_1

Statistics Table - 0:22 - Statistics (estadísticas de)

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 16 Properties Flow Variables

Row ID	S Column	D Min	D Max	D Mean	D Std. de...	D Variance
Age	Age	19	67	45.217	13.229	174.996
Annual Incom...	Annual Inco...	15	39	26.304	7.894	62.312
Spending Sco...	Spending Sc...	3	40	20.913	13.017	169.447

En las estadísticas podemos observar que podemos observar que:

**La edad mínima** de las personas dentro de este cluster es **19** años.

**La edad máxima** de las personas dentro de este cluster es **67** años.

**El ingreso anual mínimo** de las personas de este cluster es **15**.

**El ingreso anual máximo** de las personas de este cluster es **39**.

**La puntuación mínima de gastos** de las personas de este cluster es **3**.

**La puntuación máxima de gastos** de las personas de este cluster es **40**.

**La media de edad** en este cluster es **45.21** años.

**La media de ingresos anuales** en este cluster es **26.30**.

**La media de puntuación de gastos** en este cluster es de **20.91** puntos.

**La desviación estándar de la edad** de las personas en este cluster es de **13.22**.

**La desviación estándar de los ingresos anuales** de las personas en este cluster es de **7.89**.

**La desviación estándar de la puntuación de gastos** de las personas del cluster es de **13.01**.

**La varianza de la edad** de las personas en este cluster es de **174.99**.

**La varianza de los ingresos anuales** de las personas en este cluster es de **62.31**.

**La varianza de la puntuación de gastos** de las personas en este cluster es de **169.44**.

Por último para el cluster\_1 se muestra la cantidad de hombres y mujeres que hay

▲ Filtered - 0:33 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 14 Spec - Columns: 5 Properties Flow Variables

Row ID	<b>S</b> Gender	<b>I</b> Age	<b>I</b> Annual ...	<b>I</b> Spend...	<b>S</b> Cluster
3	Female	20	16	6	cluster_1
5	Female	31	17	40	cluster_1
7	Female	35	18	6	cluster_1
13	Female	58	20	15	cluster_1
17	Female	35	21	35	cluster_1
23	Female	46	25	5	cluster_1
25	Female	54	28	14	cluster_1
27	Female	45	28	32	cluster_1
29	Female	40	29	31	cluster_1
35	Female	49	33	14	cluster_1
37	Female	42	34	17	cluster_1
39	Female	36	37	26	cluster_1
41	Female	65	38	35	cluster_1
45	Female	49	39	28	cluster_1

Como se observa en la imagen en este cluster se tienen 14 filas en las que el atributo género tiene la clase mujer.

▲ Filtered - 0:32 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 9 Spec - Columns: 5 Properties Flow Variables

Row ID	<b>S</b> Gender	<b>I</b> Age	<b>I</b> Annual ...	<b>I</b> Spend...	<b>S</b> Cluster
1	Male	19	15	39	cluster_1
9	Male	64	19	3	cluster_1
11	Male	67	19	14	cluster_1
15	Male	37	20	13	cluster_1
19	Male	52	23	29	cluster_1
21	Male	35	24	35	cluster_1
31	Male	60	30	4	cluster_1
33	Male	53	33	4	cluster_1
43	Male	48	39	36	cluster_1

Como se observa en la imagen en este cluster se tienen 9 filas en las que el atributo género tiene la clase hombre.

La siguiente imagen muestra el conjunto de datos generado por el cluster\_2

Filtered - 0:23 - Rule-based Row Filter (Filtramos cluster\_2)

File Edit Hilite Navigation View

Table "default" - Rows: 80 Spec - Columns: 5 Properties Flow Variables

Row ID	S Gender	I Age	I Annual ...	I Spend...	S Cluster
44	Female	31	39	61	cluster_2
47	Female	50	40	55	cluster_2
48	Female	27	40	47	cluster_2
49	Female	29	40	42	cluster_2
50	Female	31	40	42	cluster_2
51	Female	49	42	52	cluster_2
52	Male	33	42	60	cluster_2
53	Female	31	43	54	cluster_2
54	Male	59	43	60	cluster_2
55	Female	50	43	45	cluster_2
56	Male	47	43	41	cluster_2
57	Female	51	44	50	cluster_2
58	Male	69	44	46	cluster_2
59	Female	27	46	51	cluster_2
60	Male	53	46	46	cluster_2
61	Male	70	46	56	cluster_2
62	Male	19	46	55	cluster_2
63	Female	67	47	52	cluster_2
64	Female	54	47	59	cluster_2
65	Male	63	48	51	cluster_2
66	Male	18	48	59	cluster_2
67	Female	43	48	50	cluster_2
68	Female	68	48	48	cluster_2
69	Male	19	48	59	cluster_2
70	Female	32	48	47	cluster_2

En la imagen anterior podemos observar que contiene 80 elementos, y cada una de las filas contiene los datos de género, edad, ingresos anuales y puntuación de gastos de cada cliente

A continuación se muestran las estadísticas del cluster\_2

▲ Statistics Table - 0:24 - Statistics (estadísticas de)

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 16 Properties Flow Variables

Row ID	S Column	D Min	D Max	D Mean	D Std. de...	D Variance
Age	Age	18	70	42.938	16.43	269.933
Annual Incom...	Annual Inco...	39	76	55.087	8.845	78.233
Spending Sco...	Spending Sc...	35	61	49.712	6.333	40.106

En las estadísticas podemos observar que podemos observar que:

La **edad mínima** de las personas dentro de este cluster es **18** años.

La **edad máxima** de las personas dentro de este cluster es **70** años.

El **ingreso anual mínimo** de las personas de este cluster es **39**.

El **ingreso anual máximo** de las personas de este cluster es **76**.

La **puntuación mínima de gastos** de las personas de este cluster es **35**.

La **puntuación máxima de gastos** de las personas de este cluster es **61**.

La **media de edad** en este cluster es **42.93** años.

La **media de ingresos anuales** en este cluster es **55.08**.

La **media de puntuación de gastos** en este cluster es de **49.71** puntos.

La **desviación estándar de la edad** de las personas en este cluster es de **16.43**.

La **desviación estándar de los ingresos anuales** de las personas en este cluster es de **8.84**.

La **desviación estándar de la puntuación de gastos** de las personas del cluster es de **6.33**.

La **varianza de la edad** de las personas en este cluster es de **269.93**.

La **varianza de los ingresos anuales** de las personas en este cluster es de **78.23**.

La **varianza de la puntuación de gastos** de las personas en este cluster es de **40.10**.

Por último para el cluster\_2 se muestra la cantidad de hombres y mujeres que hay

Filtered - 0:34 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 47 Spec - Columns: 5 Properties Flow Variables

Row ID	S Gender	I Age	I Annual ...	I Spend...	S Cluster
44	Female	31	39	61	cluster_2
47	Female	50	40	55	cluster_2
48	Female	27	40	47	cluster_2
49	Female	29	40	42	cluster_2
50	Female	31	40	42	cluster_2
51	Female	49	42	52	cluster_2
53	Female	31	43	54	cluster_2
55	Female	50	43	45	cluster_2
57	Female	51	44	50	cluster_2
59	Female	27	46	51	cluster_2
63	Female	67	47	52	cluster_2
64	Female	54	47	59	cluster_2
67	Female	43	48	50	cluster_2
68	Female	68	48	48	cluster_2
70	Female	32	48	47	cluster_2
72	Female	47	49	42	cluster_2
73	Female	60	50	49	cluster_2
74	Female	60	50	56	cluster_2
77	Female	45	54	53	cluster_2
79	Female	23	54	52	cluster_2
80	Female	49	54	42	cluster_2
84	Female	46	54	44	cluster_2
85	Female	21	54	57	cluster_2
87	Female	55	57	58	cluster_2
88	Female	22	57	55	cluster_2

Como se observa en la imagen en este cluster se tienen 47 filas en las que el atributo género tiene la clase mujer.

Filtered - 0:35 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 33 Spec - Columns: 5 Properties Flow Variables

Row ID	S Gender	I Age	I Annual ...	I Spend...	S Cluster
52	Male	33	42	60	cluster_2
54	Male	59	43	60	cluster_2
56	Male	47	43	41	cluster_2
58	Male	69	44	46	cluster_2
60	Male	53	46	46	cluster_2
61	Male	70	46	56	cluster_2
62	Male	19	46	55	cluster_2
65	Male	63	48	51	cluster_2
66	Male	18	48	59	cluster_2
69	Male	19	48	59	cluster_2
71	Male	70	49	55	cluster_2
75	Male	59	54	47	cluster_2
76	Male	26	54	54	cluster_2
78	Male	40	54	48	cluster_2
81	Male	57	54	51	cluster_2
82	Male	38	54	55	cluster_2
83	Male	67	54	41	cluster_2
86	Male	48	54	46	cluster_2
92	Male	18	59	41	cluster_2
93	Male	48	60	49	cluster_2
96	Male	24	60	52	cluster_2
99	Male	48	61	42	cluster_2
100	Male	20	61	49	cluster_2
103	Male	67	62	59	cluster_2
104	Male	26	62	55	cluster_2

Como se observa en la imagen en este cluster se tienen 33 filas en las que el atributo género tiene la clase hombre.



La siguiente imagen muestra el conjunto de datos generado por el cluster\_3

Filtered - 0:25 - Rule-based Row Filter (Filtramos cluster\_3)

File Edit Hilite Navigation View

Table "default" - Rows: 36 Spec - Columns: 5 Properties Flow Variables

Row ID	S Gender	I Age	I Annual ...	I Spend...	S Cluster
125	Female	23	70	29	cluster_3
129	Male	59	71	11	cluster_3
131	Male	47	71	9	cluster_3
133	Female	25	72	34	cluster_3
135	Male	20	73	5	cluster_3
137	Female	44	73	7	cluster_3
139	Male	19	74	10	cluster_3
141	Female	57	75	5	cluster_3
145	Male	25	77	12	cluster_3
147	Male	48	77	36	cluster_3
149	Female	34	78	22	cluster_3
151	Male	43	78	17	cluster_3
153	Female	44	78	20	cluster_3
155	Female	47	78	16	cluster_3
157	Male	37	78	1	cluster_3
159	Male	34	78	1	cluster_3
161	Female	56	79	35	cluster_3
163	Male	19	81	5	cluster_3
165	Male	50	85	26	cluster_3
167	Male	42	86	20	cluster_3
169	Female	36	87	27	cluster_3
171	Male	40	87	13	cluster_3
173	Male	36	87	10	cluster_3
175	Female	52	88	13	cluster_3
177	Male	58	88	15	cluster_3

En la imagen anterior podemos observar que contiene 36 elementos, y cada una de las filas contiene los datos de género, edad, ingresos anuales y puntuación de gastos de cada cliente

A continuación se muestran las estadísticas del cluster\_3

Statistics Table - 0:26 - Statistics (estadísticas de)

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 16 Properties Flow Variables

Row ID	S Column	D Min	D Max	D Mean	D Std. de...	D Variance
Age	Age	19	59	40.667	11.497	132.171
Annual Incom...	Annual Inco...	70	137	87.75	16.387	268.536
Spending Sco...	Spending Sc...	1	39	17.583	10.205	104.136

En las estadísticas podemos observar que podemos observar que:

**La edad mínima** de las personas dentro de este cluster es **19** años.

**La edad máxima** de las personas dentro de este cluster es **59** años.

**El ingreso anual mínimo** de las personas de este cluster es **70**.

**El ingreso anual máximo** de las personas de este cluster es **137**.

**La puntuación mínima de gastos** de las personas de este cluster es **1**.

**La puntuación máxima de gastos** de las personas de este cluster es **39**.

**La media de edad** en este cluster es **40.66** años.

**La media de ingresos anuales** en este cluster es **87.75**.

**La media de puntuación de gastos** en este cluster es de **17.58** puntos.

**La desviación estándar de la edad** de las personas en este cluster es de **11.49**.

**La desviación estándar de los ingresos anuales** de las personas en este cluster es de **16.38**.

**La desviación estándar de la puntuación de gastos** de las personas del cluster es de **10.20**.

**La varianza de la edad** de las personas en este cluster es de **132.17**.

**La varianza de los ingresos anuales** de las personas en este cluster es de **268.63**.

**La varianza de la puntuación de gastos** de las personas en este cluster es de **104.13**.

Por último para el cluster\_3 se muestra la cantidad de hombres y mujeres que hay

▲ Filtered - 0:36 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 17 Spec - Columns: 5 Properties Flow Variables

Row ID	<b>S</b> Gender	<b>I</b> Age	<b>I</b> Annual ...	<b>I</b> Spendi...	<b>S</b> Cluster
125	Female	23	70	29	cluster_3
133	Female	25	72	34	cluster_3
137	Female	44	73	7	cluster_3
141	Female	57	75	5	cluster_3
149	Female	34	78	22	cluster_3
153	Female	44	78	20	cluster_3
155	Female	47	78	16	cluster_3
161	Female	56	79	35	cluster_3
169	Female	36	87	27	cluster_3
175	Female	52	88	13	cluster_3
181	Female	37	97	32	cluster_3
185	Female	41	99	39	cluster_3
187	Female	54	101	24	cluster_3
189	Female	41	103	17	cluster_3
191	Female	34	103	23	cluster_3
195	Female	47	120	16	cluster_3
197	Female	45	126	28	cluster_3

Como se observa en la imagen en este cluster se tienen 17 filas en las que el atributo género tiene la clase mujer.

▲ Filtered - 0:37 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 19 Spec - Columns: 5 Properties Flow Variables

Row ID	<b>S</b> Gender	<b>I</b> Age	<b>I</b> Annual ...	<b>I</b> Spendi...	<b>S</b> Cluster
129	Male	59	71	11	cluster_3
131	Male	47	71	9	cluster_3
135	Male	20	73	5	cluster_3
139	Male	19	74	10	cluster_3
145	Male	25	77	12	cluster_3
147	Male	48	77	36	cluster_3
151	Male	43	78	17	cluster_3
157	Male	37	78	1	cluster_3
159	Male	34	78	1	cluster_3
163	Male	19	81	5	cluster_3
165	Male	50	85	26	cluster_3
167	Male	42	86	20	cluster_3
171	Male	40	87	13	cluster_3
173	Male	36	87	10	cluster_3
177	Male	58	88	15	cluster_3
179	Male	59	93	14	cluster_3
183	Male	46	98	15	cluster_3
193	Male	33	113	8	cluster_3
199	Male	32	137	18	cluster_3

Como se observa en la imagen en este cluster se tienen 19 filas en las que el atributo género tiene la clase hombre.

La siguiente imagen muestra el conjunto de datos generado por el cluster\_4

▲ Filtered - 0:27 - Rule-based Row Filter (Filtramos cluster\_4)

File Edit Hilite Navigation View

Table "default" - Rows: 39 Spec - Columns: 5 Properties Flow Variables

Row ID	S Gender	I Age	I Annual ...	I Spend...	S Cluster
124	Male	39	69	91	cluster_4
126	Female	31	70	77	cluster_4
128	Male	40	71	95	cluster_4
130	Male	38	71	75	cluster_4
132	Male	39	71	75	cluster_4
134	Female	31	72	71	cluster_4
136	Female	29	73	88	cluster_4
138	Male	32	73	73	cluster_4
140	Female	35	74	72	cluster_4
142	Male	32	75	93	cluster_4
144	Female	32	76	87	cluster_4
146	Male	28	77	97	cluster_4
148	Female	32	77	74	cluster_4
150	Male	34	78	90	cluster_4
152	Male	39	78	88	cluster_4
154	Female	38	78	76	cluster_4
156	Female	27	78	89	cluster_4
158	Female	30	78	78	cluster_4
160	Female	30	78	73	cluster_4
162	Female	29	79	83	cluster_4
164	Female	31	81	93	cluster_4
166	Female	36	85	75	cluster_4
168	Female	33	86	95	cluster_4
170	Male	32	87	63	cluster_4
172	Male	28	87	75	cluster_4

En la imagen anterior podemos observar que contiene 39 elementos, y cada una de las filas contiene los datos de género, edad, ingresos anuales y puntuación de gastos de cada cliente

A continuación se muestran las estadísticas del cluster\_4

Statistics Table - 0:28 - Statistics (estadísticas de)

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 16 Properties Flow Variables

Row ID	S Column	D Min	D Max	D Mean	D Std. de...	D Variance
Age	Age	27	40	32.692	3.729	13.903
Annual Incom...	Annual Inco...	69	137	86.538	16.312	266.097
Spending Sco...	Spending Sc...	63	97	82.128	9.364	87.694

En las estadísticas podemos observar que podemos observar que:

**La edad mínima** de las personas dentro de este cluster es **27** años.

**La edad máxima** de las personas dentro de este cluster es **40** años.

**El ingreso anual mínimo** de las personas de este cluster es **69**.

**El ingreso anual máximo** de las personas de este cluster es **137**.

**La puntuación mínima de gastos** de las personas de este cluster es **63**.

**La puntuación máxima de gastos** de las personas de este cluster es **97**.

**La media de edad** en este cluster es **32.69** años.

**La media de ingresos anuales** en este cluster es **86.53**.

**La media de puntuación de gastos** en este cluster es de **82.12** puntos.

**La desviación estándar de la edad** de las personas en este cluster es de **3.72**.

**La desviación estándar de los ingresos anuales** de las personas en este cluster es de **16.31**.

**La desviación estándar de la puntuación de gastos** de las personas del cluster es de **9.36**.

**La varianza de la edad** de las personas en este cluster es de **13.90**.

**La varianza de los ingresos anuales** de las personas en este cluster es de **266.09**.

**La varianza de la puntuación de gastos** de las personas en este cluster es de **87.69**.

Por último para el cluster\_4 se muestra la cantidad de hombres y mujeres que hay

Filtered - 0:38 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 21 Spec - Columns: 5 Properties Flow Variables

Row ID	\$ Gender	I Age	I Annual ...	I Spendi...	\$ Cluster
126	Female	31	70	77	cluster_4
134	Female	31	72	71	cluster_4
136	Female	29	73	88	cluster_4
140	Female	35	74	72	cluster_4
144	Female	32	76	87	cluster_4
148	Female	32	77	74	cluster_4
154	Female	38	78	76	cluster_4
156	Female	27	78	89	cluster_4
158	Female	30	78	78	cluster_4
160	Female	30	78	73	cluster_4
162	Female	29	79	83	cluster_4
164	Female	31	81	93	cluster_4
166	Female	36	85	75	cluster_4
168	Female	33	86	95	cluster_4
176	Female	30	88	86	cluster_4
182	Female	32	97	86	cluster_4
184	Female	29	98	88	cluster_4
190	Female	36	103	85	cluster_4
192	Female	32	103	69	cluster_4
194	Female	38	113	91	cluster_4
196	Female	35	120	79	cluster_4

Como se observa en la imagen en este cluster se tienen 21 filas en las que el atributo género tiene la clase mujer.

Filtered - 0:39 - Row Filter (Cantidad de)

File Edit Hilite Navigation View

Table "default" - Rows: 18

Spec - Columns: 5

Properties

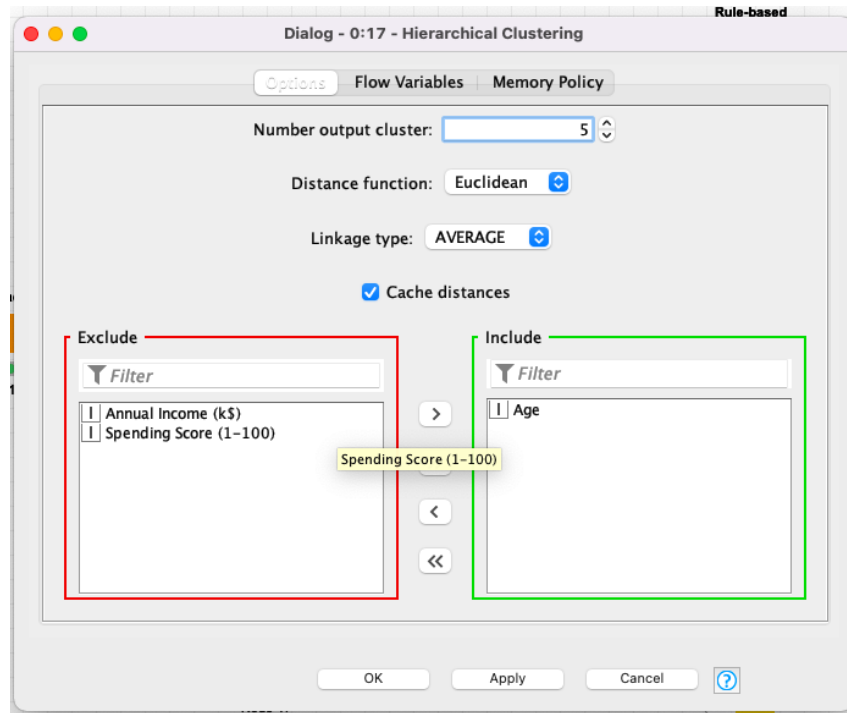
Flow Variables

Row ID	<b>S</b> Gender	<b>I</b> Age	<b>I</b> Annual ...	<b>I</b> Spend...	<b>S</b> Cluster
124	Male	39	69	91	cluster_4
128	Male	40	71	95	cluster_4
130	Male	38	71	75	cluster_4
132	Male	39	71	75	cluster_4
138	Male	32	73	73	cluster_4
142	Male	32	75	93	cluster_4
146	Male	28	77	97	cluster_4
150	Male	34	78	90	cluster_4
152	Male	39	78	88	cluster_4
170	Male	32	87	63	cluster_4
172	Male	28	87	75	cluster_4
174	Male	36	87	92	cluster_4
178	Male	27	88	69	cluster_4
180	Male	35	93	90	cluster_4
186	Male	30	99	97	cluster_4
188	Male	28	101	68	cluster_4
198	Male	32	126	74	cluster_4
200	Male	30	137	83	cluster_4

Como se observa en la imagen en este cluster se tienen 18 filas en las que el atributo género tiene la clase hombre.

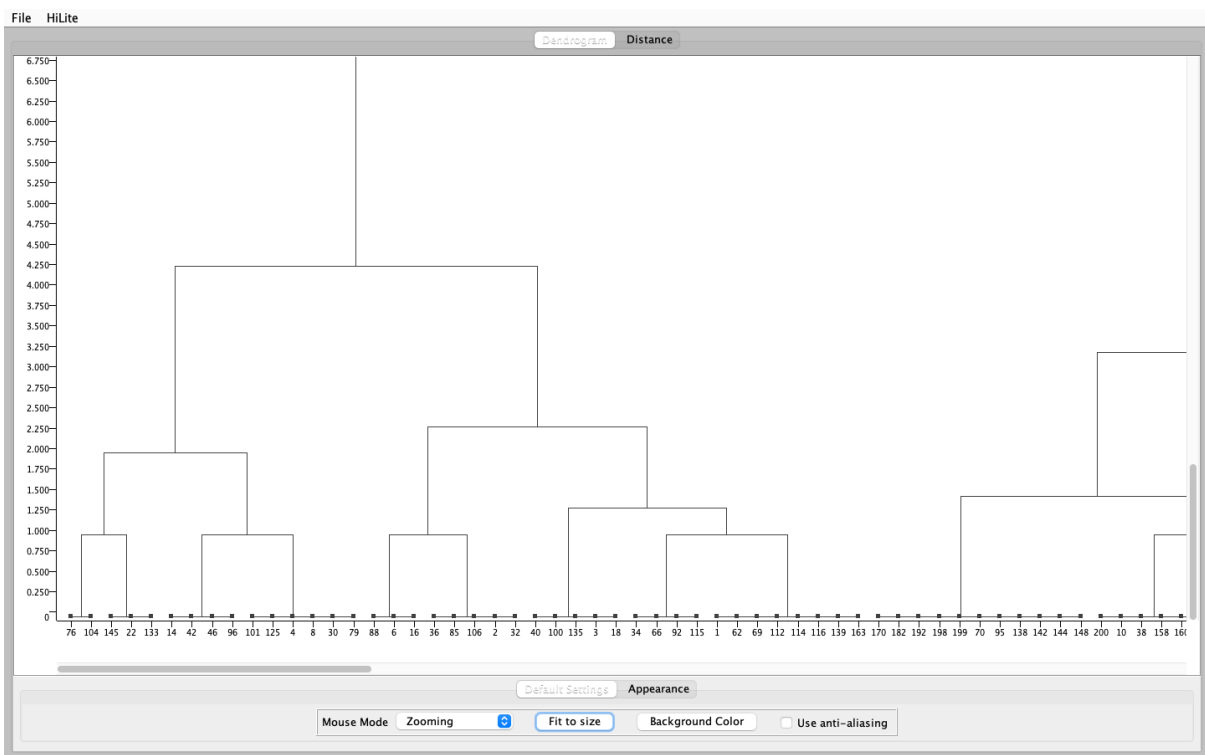
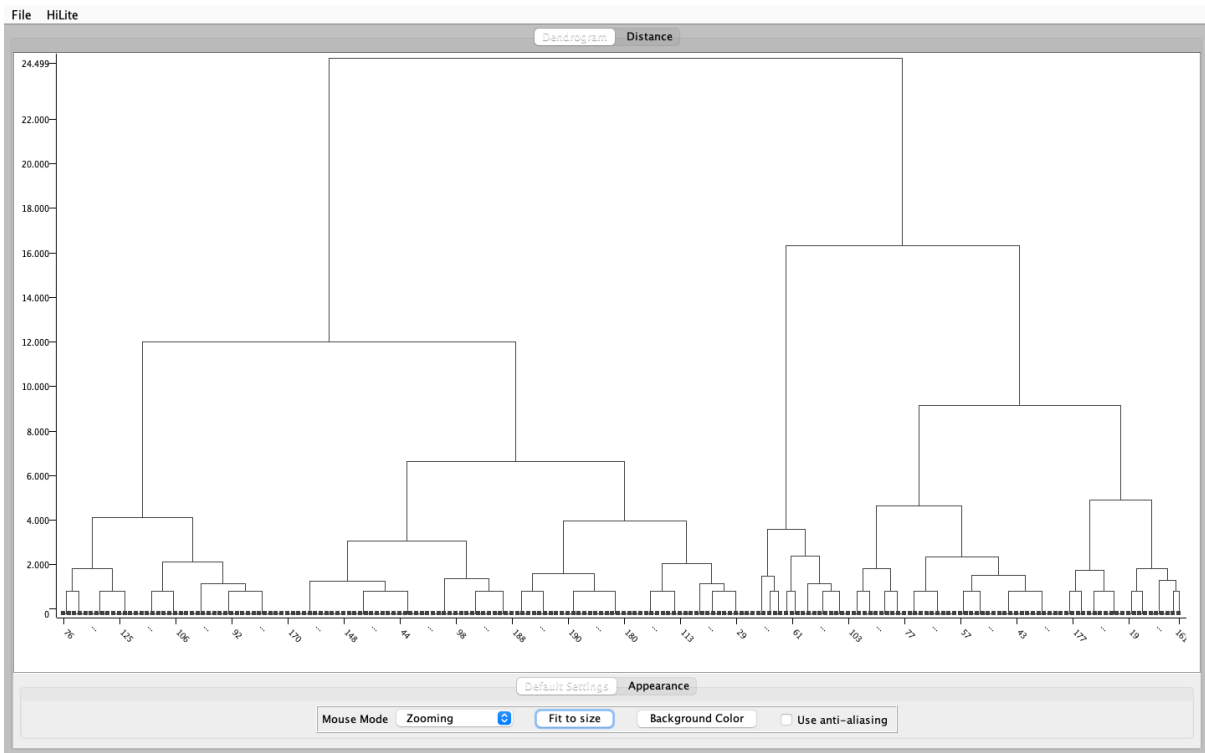
# Hierarchical Clustering (Jerárquico)

Para el Clustering Jeararquico tomamos el mismo archivo “Mall\_Customers”, pero tomamos solo el valor de la edad.



El módulo Hierarchical Clustering permite visualizar las agrupaciones de los clientes por edad con un Denograma:





Por distancia

