

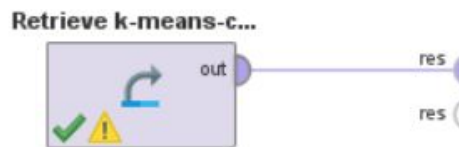
TA1 - Agustín Fernández

Ejercicio 1

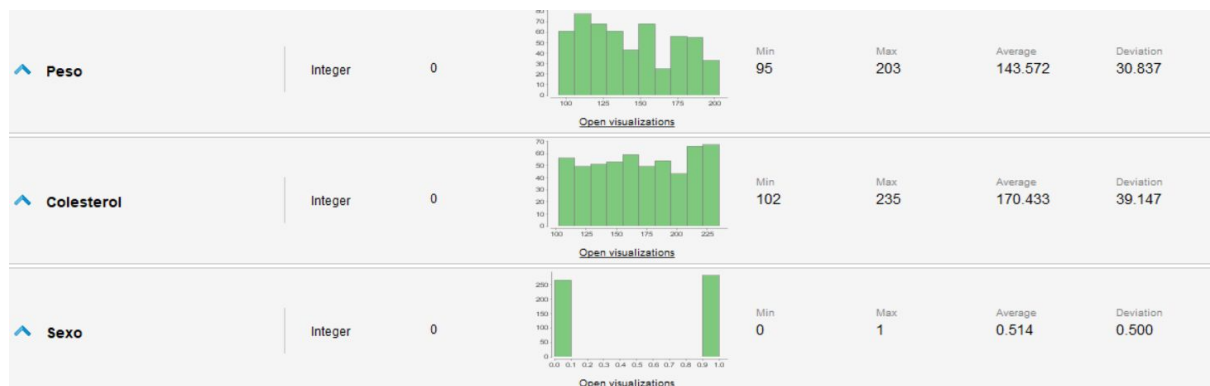
Preparación de los datos

2)

a.



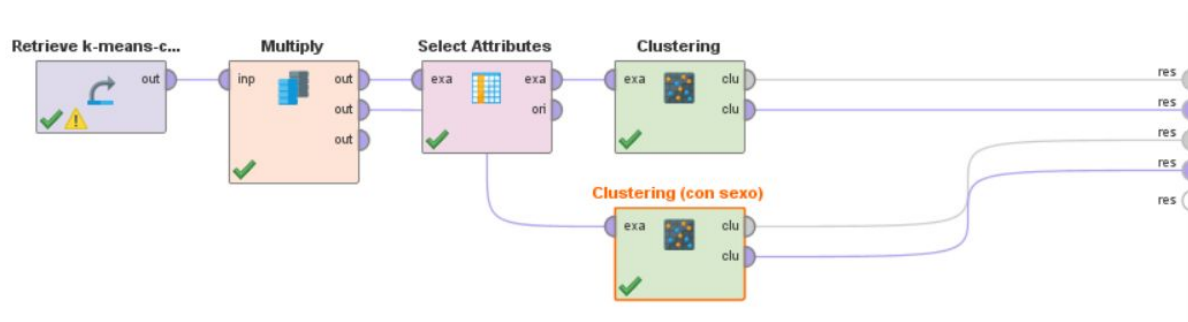
b.



c. No se identificaron outliers que puedan afectar el algoritmo.

Modelado

Sin normalizar



Se aplicó el modelo k-means al dataset incluyendo y sin incluir 'Sexo', y en ambos casos los clusters coincidieron (en cantidad de elementos y en el centroide).

Resultado del clustering:

Cluster Model

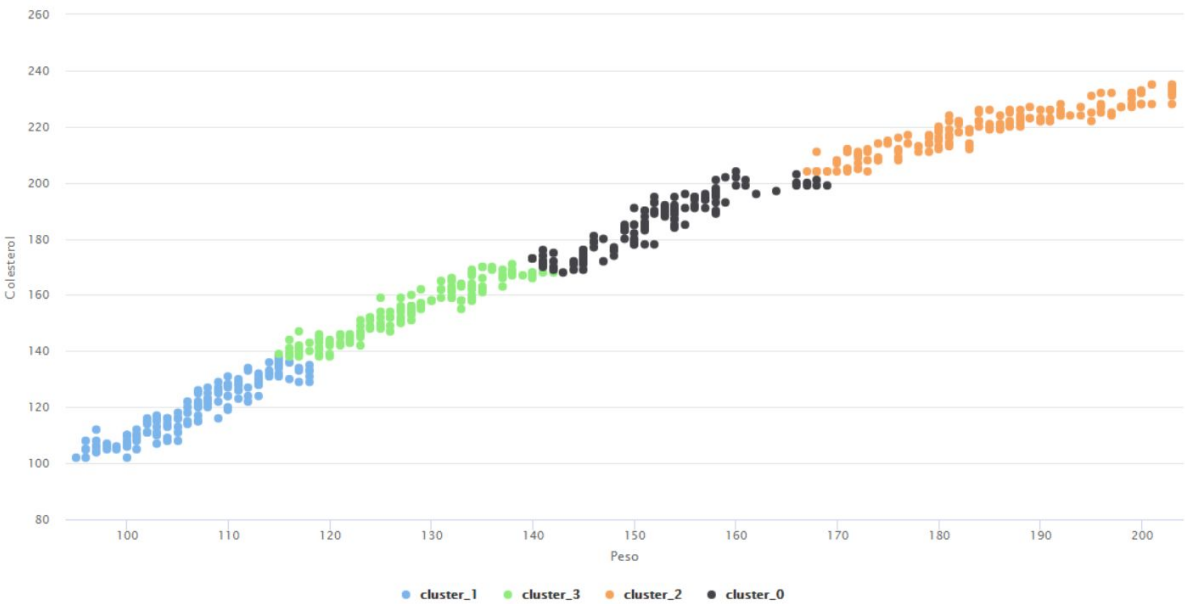
```
Cluster 0: 118 items
Cluster 1: 140 items
Cluster 2: 154 items
Cluster 3: 135 items
Total number of items: 547
```

Tabla de centroides:

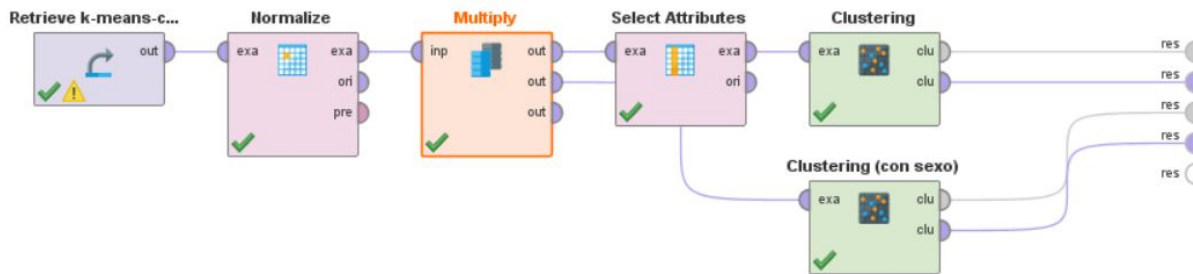
Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Peso	152.093	106.850	184.318	127.726
Colesterol	185.907	119.536	218.916	154.385

Tabla de centroides (con sexo):

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Peso	152.093	106.850	184.318	127.726
Colesterol	185.907	119.536	218.916	154.385
Sexo	0.441	0.543	0.591	0.459



Normalizando los datos



En este caso, la ocurrencia del sexo sí marca un cambio en los clusters:

Sin tener en cuenta el sexo:

Cluster Model

```
Cluster 0: 141 items
Cluster 1: 148 items
Cluster 2: 146 items
Cluster 3: 112 items
Total number of items: 547
```

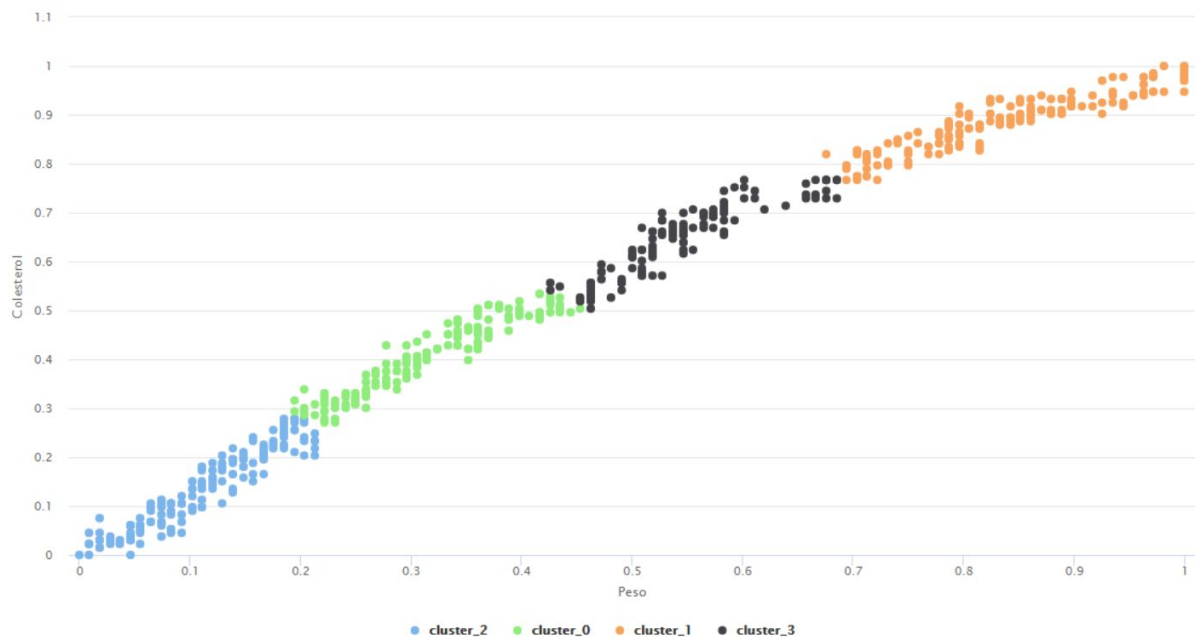
Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Peso	0.318	0.833	0.113	0.547
Colesterol	0.409	0.884	0.138	0.650

Teniendo en cuenta el sexo:

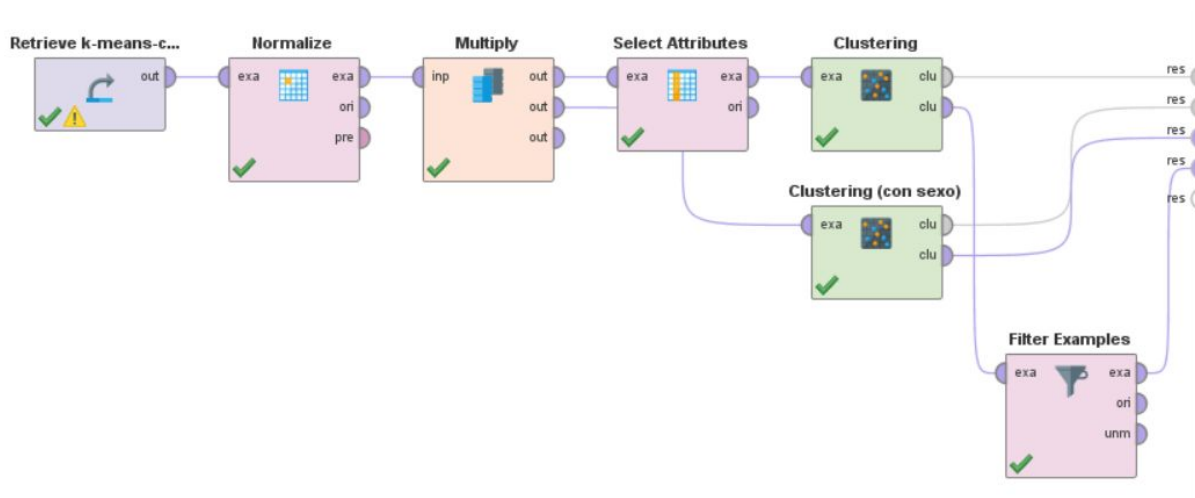
Cluster Model

```
Cluster 0: 141 items
Cluster 1: 157 items
Cluster 2: 109 items
Cluster 3: 140 items
Total number of items: 547
```

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Peso	0.188	0.255	0.726	0.717
Colesterol	0.238	0.320	0.803	0.787
Sexo	1	0	0	1



Despliegue



Las consultas SQL para los demás clusters serían:

```

cluster_0: SELECT First_Name, Last_Name, Policy_Num, Address, Phone_Num
FROM PolicyHolders_view
WHERE Weight >= 140 AND Weight <= 169
AND Cholesterol >= 168 AND Cholesterol <= 204
  
```

```

cluster_1: SELECT First_Name, Last_Name, Policy_Num, Address, Phone_Num
FROM PolicyHolders_view
WHERE Weight <= 118
AND Cholesterol <= 138
  
```

```

cluster_3: SELECT First_Name, Last_Name, Policy_Num, Address, Phone_Num
FROM PolicyHolders_view
  
```

WHERE Weight >= 115 AND Weight <= 142
AND Cholesterol >= 138 AND Cholesterol <= 167