

La segmentación de clientes y el customer analytics es una herramienta muy poderosa para poder conocer los patrones de comportamiento de nuestros clientes. Una vez sabiendo esto, podemos anticipar necesidades, así como también mantener a nuestros clientes actuales. El poder segmentar con base en comportamientos y características similares, resulta muy conveniente para cualquier organización. A continuación se presenta el siguiente desarrollo del trabajo en donde realizaremos clusters.

## **FASE 1. Análisis exploratorio de datos**

### **Desarrollo de EDA (Exploratory Data Analysis)**

Nuestro Dataset es un archivo en formato cvs, en donde podemos visualizar información general acerca de las transacciones (número de orden, iD de cliente, fecha de la orden, cantidad, tipo de producto, precio, y una columna calculada de ingresos totales. Nuestros tipos de datos son numéricos y de caracteres. Contamos con 4,805 columnas y 9 variables.

Data	
Transactions	4805 obs. of 9 variables
\$ orderId	: num [1:4805] 1 1 1 1 2 2 2 2 3 3 ...
\$ clientId	: num [1:4805] 255 255 255 255 145 145 145 ...
\$ product	: chr [1:4805] "g" "a" "b" "a" ...
\$ gender	: chr [1:4805] "female" "female" "female" ...
\$ orderdate	: chr [1:4805] "03/01/2017" "03/01/2017" "0...
\$ Quantity	: num [1:4805] 3 3 3 1 1 1 1 1 3 2 ...
\$ Price	: num [1:4805] 5 14 22 14 26 26 14 26 12 5 ...
\$ CantidadTotal	: num [1:4805] 15 42 66 14 26 26 14 26 36 1...
\$ IngresoTotal	: num [1:4805] 15 42 66 14 26 26 14 26 36 1...

### **Exploración y visualizaciones de datos**

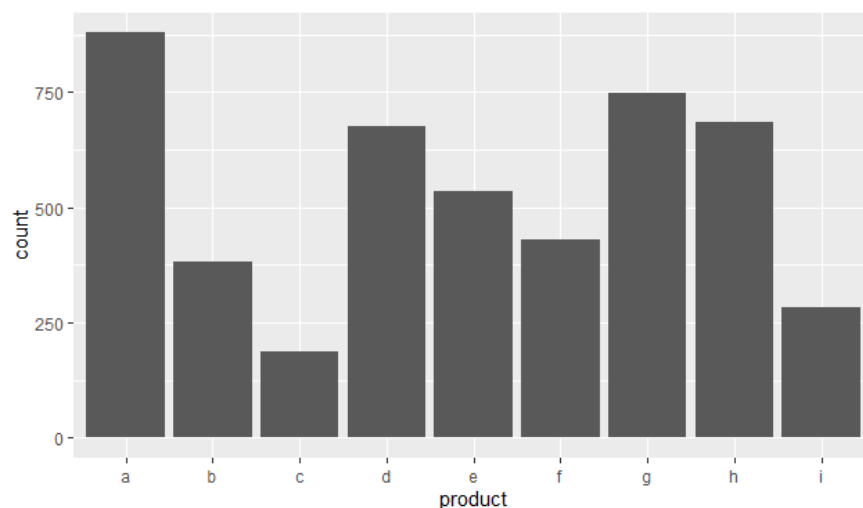
Vamos a ejecutar una función en donde veremos las 10 primeras filas de nuestro dataset.

```
# A tibble: 10 × 9
```

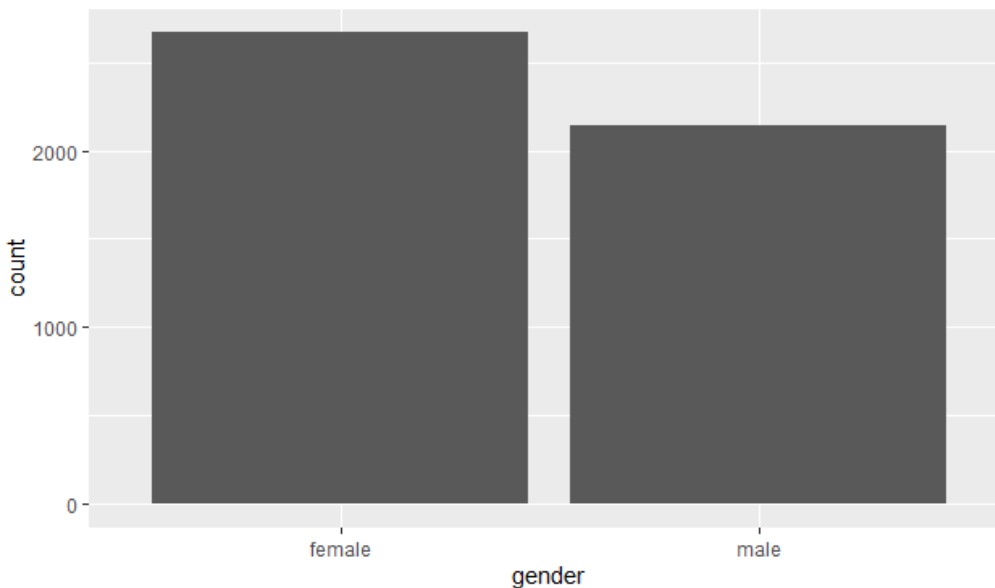
	orderId	clientId	product	gender	orderdate	Quantity	Price	CantidadTotal
	<dbl>	<dbl>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
1	1	255	g	female	03/01/20...	3	5	15
2	1	255	a	female	03/01/20...	3	14	42
3	1	255	b	female	03/01/20...	3	22	66
4	1	255	a	female	03/01/20...	1	14	14
5	2	145	h	male	05/01/20...	1	26	26
6	2	145	h	male	05/01/20...	1	26	26
7	2	145	a	male	05/01/20...	1	14	14
8	2	145	h	male	05/01/20...	1	26	26
9	3	241	d	female	27/02/20...	3	12	36
10	3	241	g	female	27/02/20...	2	5	10

Con ayuda de la librería de ggplot, por medio de una gráfica de barras, exploramos el total de productos vendidos por tipo de producto.

Por lo tanto, obtenemos el siguiente gráfico en donde podemos visualizar que el tipo de producto más vendido fue el **a**.



También, exploramos la cantidad de productos vendidos en función del género del cliente. Por lo que podemos interpretar que el género **femenino** fue el que compró más.



## **FASE 2. Aplicando el algoritmo K-Means sobre el dataset**

Este algoritmo trata de buscar en las observaciones grupos similares con características similares, sin embargo deben ser diferentes a los otros grupos.

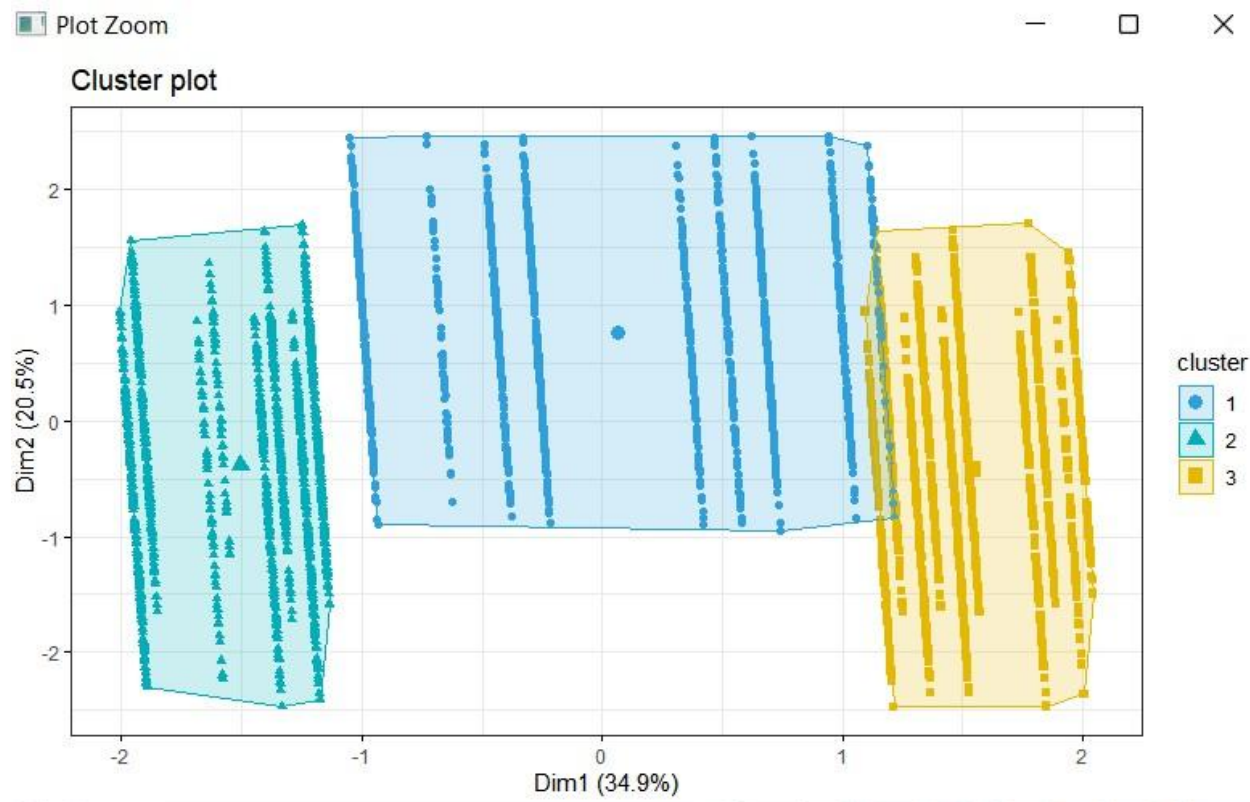
Primero preparamos los datos. Especificaremos que será de 3 clusters. Nos quedaremos con los resultados de variables numéricas. También, realizamos un escalado de datos. Después llamamos al modelo y unimos el tipo de cluster al dataset.

```
Within cluster sum of squares by cluster:
[1] 4400.502 5158.634 4569.498
(between_SS / total_SS = 51.0 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"
[5] "tot.withinss" "betweenss"    "size"         "iter"
[9] "ifault"
```

Posteriormente utilizamos código para poder interpretar nuestros tres tipos de clústeres. En este caso, vamos a hacer una interpretación en función de la **cantidad** de productos comprados, en este caso, nuestra variable es **Quantity**.



## Interpretación

### **Cluster 1**

Fue agrupado en función de la variable (cantidad) que fue menor o igual a 1. Resultando un total de **1664** observaciones.

### **Cluster 2**

Fue agrupado en función de la variable (cantidad) que fue mayor o igual a 2. Resultando un total de **1626** observaciones.

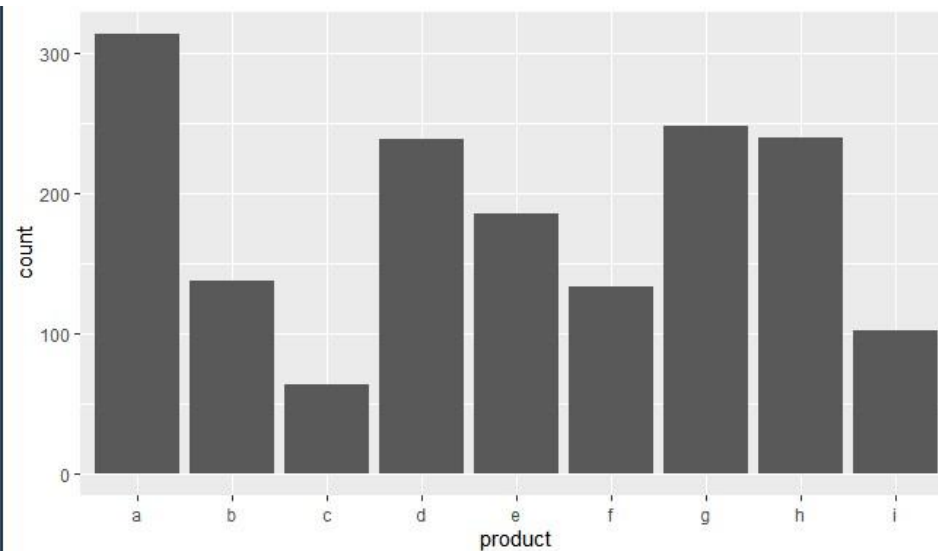
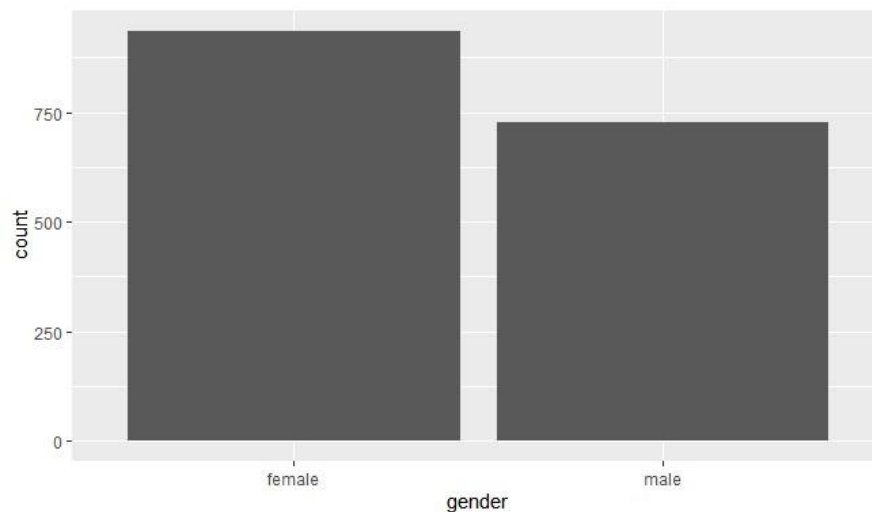
### Cluster 3

Fue agrupado en función de la variable (cantidad) que fue mayor o igual a 2, y que además la cantidad del precio fue mayor a 20. Resultando un total de **1515** observaciones.

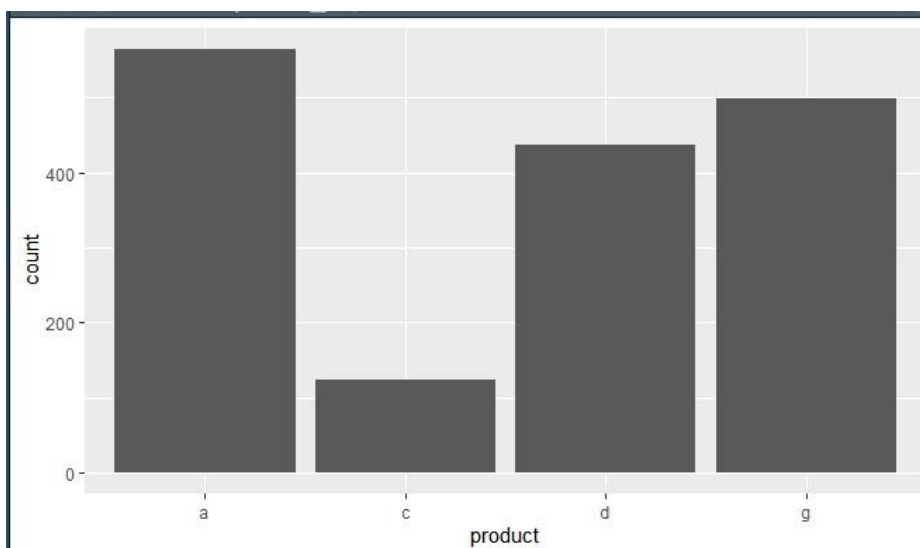
### Conclusiones

Contamos con 3 clusters que fueron interpretados en función de la cantidad de productos que compraron.

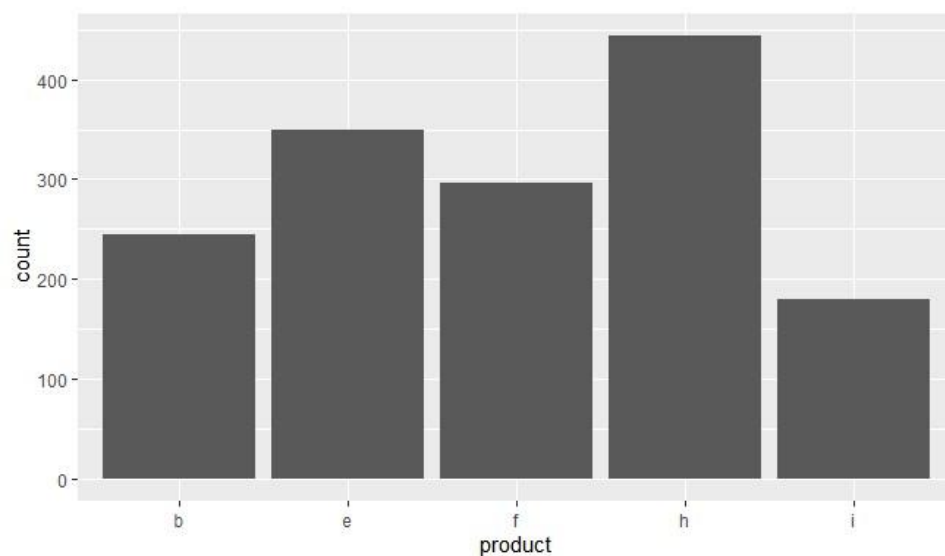
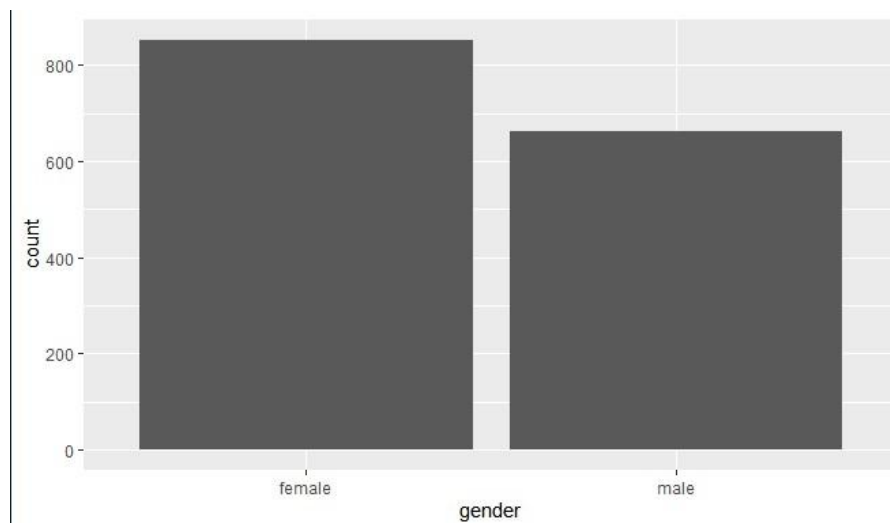
- En el cluster 1, la mayoría de nuestros clientes fueron **mujeres** y el producto más comprado fue el **a**.



- En el cluster 2, la mayoría de nuestros clientes fueron **mujeres** y el producto más comprado fue el **a**



- En el cluster 3, la mayoría de nuestros clientes fueron **mujeres** y el producto más comprado fue el **h**.



Finalmente, podríamos concluir que la empresa debe centrarse en estrategias para poder captar a más clientes del género masculino, así como también enfocarse en mantener una estrategia para no perder a nuestros clientes actuales, especialmente en cuanto al tipo de producto comprado.

Bibliografía: Alboukadel. (2018). K-MEANS CLUSTERING VISUALIZATION IN R: STEP BY STEP GUIDE. 30 /05 /2022, de DataNovia Sitio web: <https://www.datanovia.com/en/blog/k-means-clustering-visualization-in-r-step-by-step-guide/>