

Customer Analysis Dataset

Grupo 5 Richard Douglas G. – Trabajo Final
Curso Big Data Expert



Customer Personality



Descripción del Dataset

- Información general
- Características
- Datos que contiene

Dataset Customer Analysis : compuesto por características que describen a los consumidores, información como edad, ingresos, comportamiento de compra y canal de compra.



```
# Validando los tipos de datos en el dataset
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 29 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   ID                                    2240 non-null   int64
1   Year_Birth                           2240 non-null   int64
2   Education                             2240 non-null   object
3   Marital_Status                       2240 non-null   object
4   Income                               2216 non-null   float64
5   Kidhome                              2240 non-null   int64
6   Teenhome                             2240 non-null   int64
7   Dt_Customer                          2240 non-null   object
8   Recency                              2240 non-null   int64
9   MntWines                             2240 non-null   int64
10  MntFruits                             2240 non-null   int64
11  MntMeatProducts                       2240 non-null   int64
12  MntFishProducts                       2240 non-null   int64
13  MntSweetProducts                      2240 non-null   int64
14  MntGoldProds                         2240 non-null   int64
15  NumDealsPurchases                     2240 non-null   int64
16  NumWebPurchases                       2240 non-null   int64
17  NumCatalogPurchases                  2240 non-null   int64
18  NumStorePurchases                    2240 non-null   int64
19  NumWebVisitsMonth                    2240 non-null   int64
20  AcceptedCmp3                         2240 non-null   int64
21  AcceptedCmp4                         2240 non-null   int64
22  AcceptedCmp5                         2240 non-null   int64
23  AcceptedCmp1                         2240 non-null   int64
24  AcceptedCmp2                         2240 non-null   int64
25  Complain                             2240 non-null   int64
26  Z_CostContact                        2240 non-null   int64
27  Z_Revenue                            2240 non-null   int64
28  Response                             2240 non-null   int64
dtypes: float64(1), int64(25), object(3)
memory usage: 507.6+ KB
```



Información referente a las personas o clientes

- ID: identificador , codigo de cliente, valor único
- Year_Birth: el año de nacimiento de los clientes.
- Education: Nivel educativo de los clientes
- Marital_Status: CEstado civil de los clientes
- Income: Ingreso de los clientes expresado en forma anual
- Kidhome: Numeros de niños por hogar
- Teenhome: Numero de adolescentes por hogar
- Dt_Customer: Fecha en que el cliente se vinculó con la empresa
- Recency: Numero de días desde la última compra
- Complain: Indicador si el cliente se a quejado en lso ultimos años. 1 indica que sí. o indica que no.

Información sobre los productos de consumo

- MntWines: El monto que los clientes gastan en vinos o bebidas alcoholicas en los ultimos 2 años.
- MntFruits: El monto que los clientes gastan en frutas en los ultimos 2 años
- MntMeatProducts: El monto que los clientes gastan en productos carnicos en los ultimos 2 años
- MntFishProducts: El monto que los clientes gastan en pescado en los ultimos 2 años
- MntSweetProducts: El monto que los clientes gastan en dulces o golocinas en los ultimos 2 años
- MntGoldProds: El monto que los clientes gastan en productos de oro, en los ultimos dos años.



Información sobre características promocionales en el proceso de compra

- NumDealsPurchases: Number of purchases made with a discount.
- AcceptedCmp1: 1 indica si el cliente aceptó la oferta en la primer campaña, 0 que no.
- AcceptedCmp2: 1 indica si el cliente aceptó la oferta en la segunda campaña , 0 que no.
- AcceptedCmp3: 1 indica si el cliente aceptó la oferta en la tercer campaña , 0 que no.
- AcceptedCmp4: 1 indica si el cliente aceptó la oferta en la cuarta campaña , 0 que no.
- AcceptedCmp5: 1 indica si el cliente aceptó la oferta en la quinta campaña , 0 que no.
- Response: 1 ndica si el cliente aceptó la oferta en la ultima campaña, 0 que no.

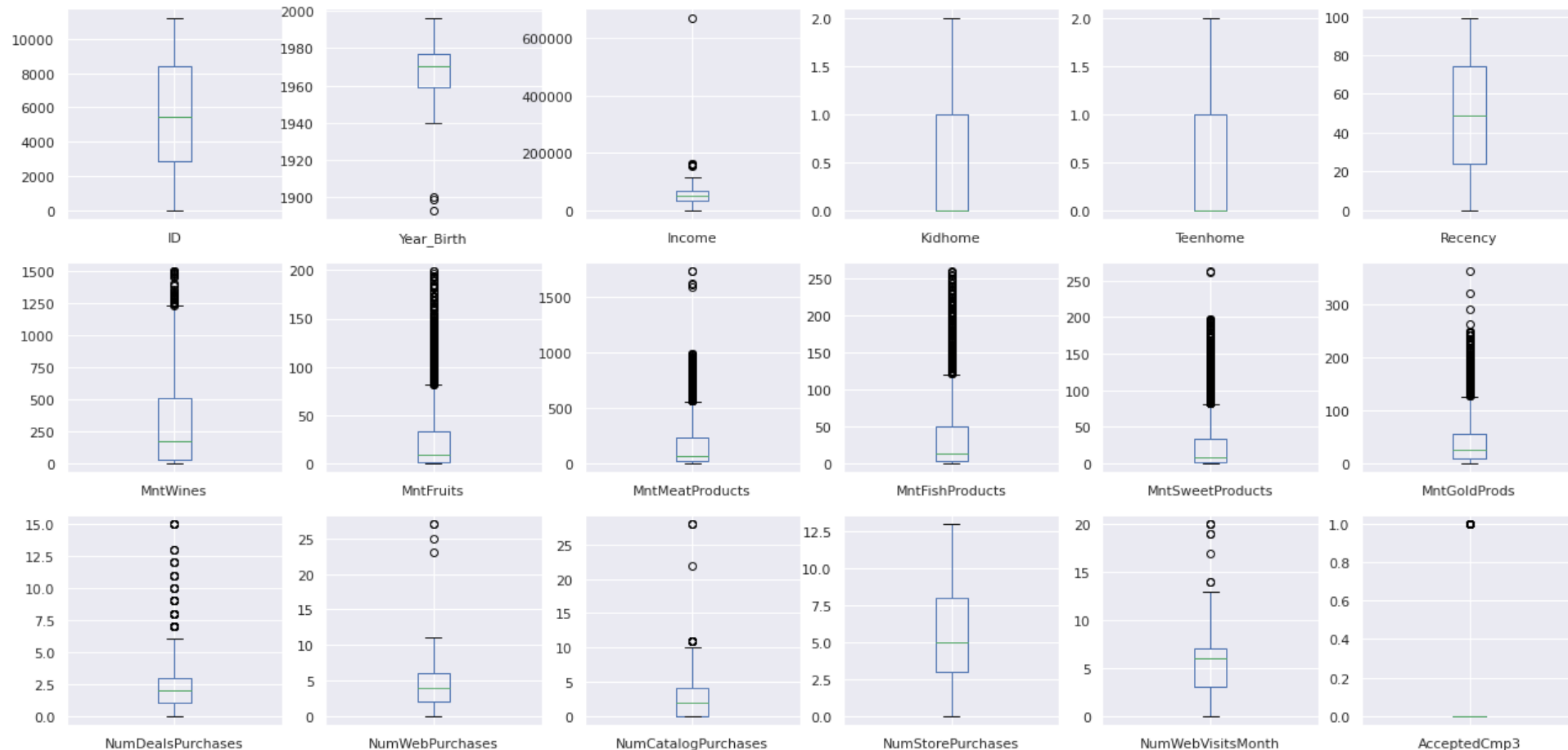
Medios de compra o canal de compra, en que el cliente elige para el proceso de compra

- NumWebPurchases: Cantidad de compras realizadas por medio de la pagina web.
- NumCatalogPurchases: Cantidad de compras realizadas por medio del catalogo.
- NumStorePurchases: Cantidad de compras realizadas directamente en las tiendas.
- NumWebVisitsMonth: Cantidad de visitas a la pagina de la empresa en el ultimo mes.

Dataset Customer Analysis : compuesto por características que describen a los consumidores, información como edad, ingresos, comportamiento de compra y canal de compra.



```
data.plot(kind='box',subplots=True, layout=(6,6), sharex=False, sharey=False)
sns.set(rc={'figure.figsize':(22,22)})
plt.show() # Representación de las características como vienen en el daset original, sin tratamiento de los outliers - atípicos
```



EDA



Exploración de los datos



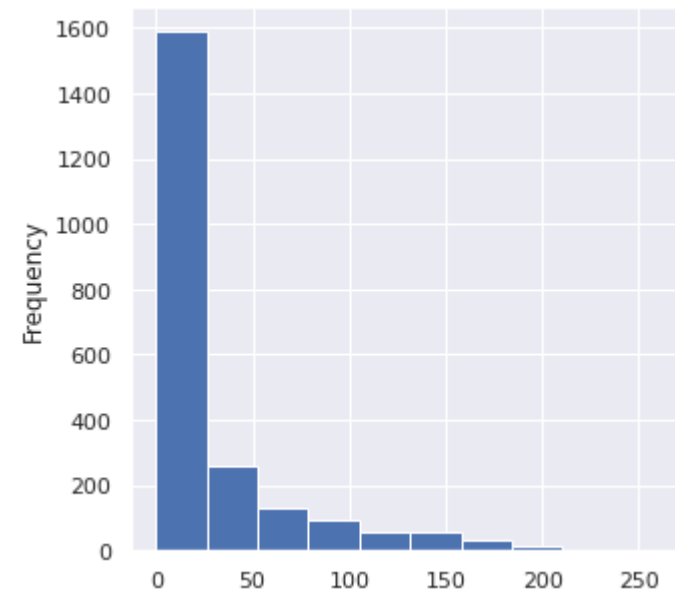
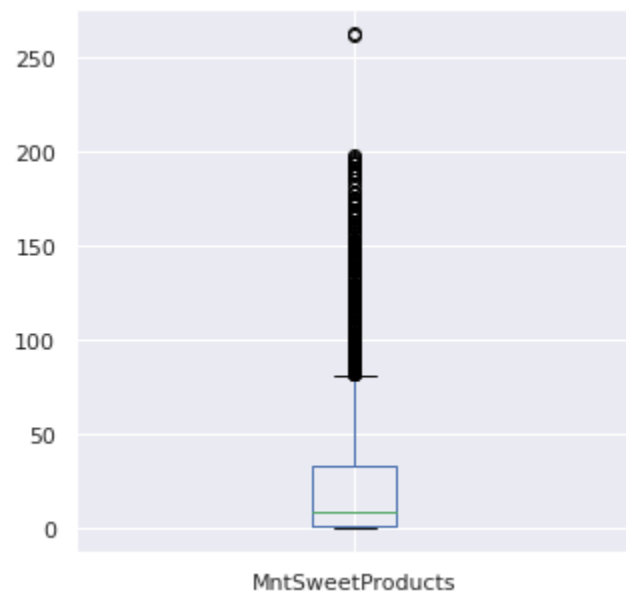
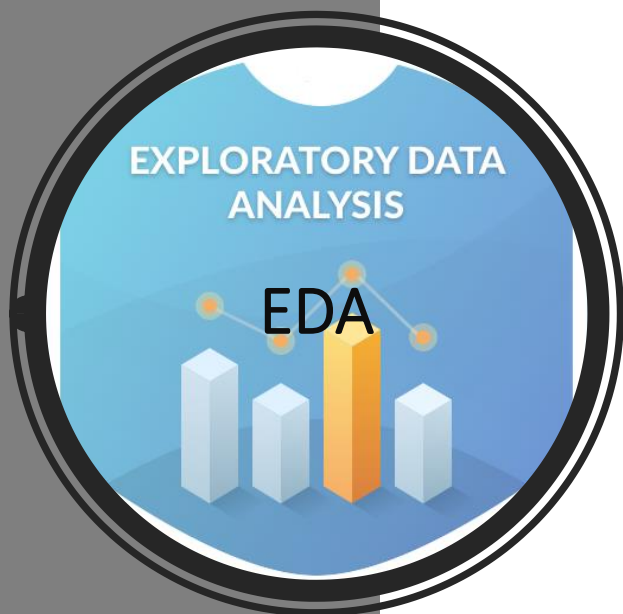
Extracción de
información

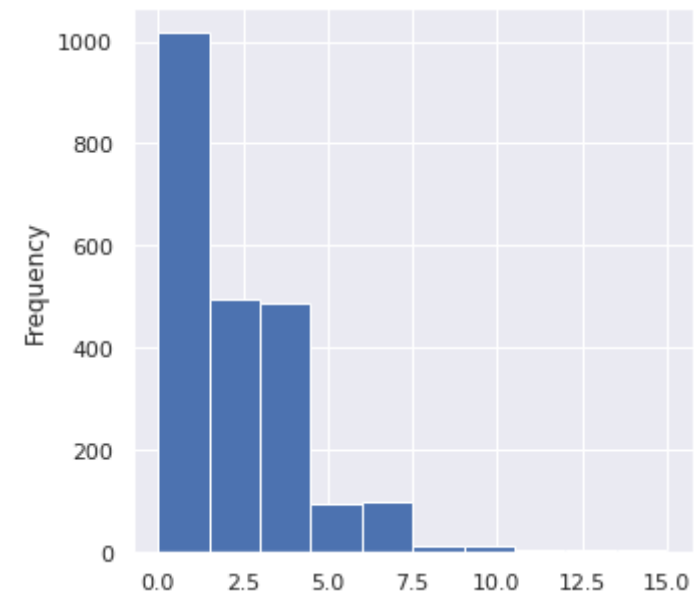
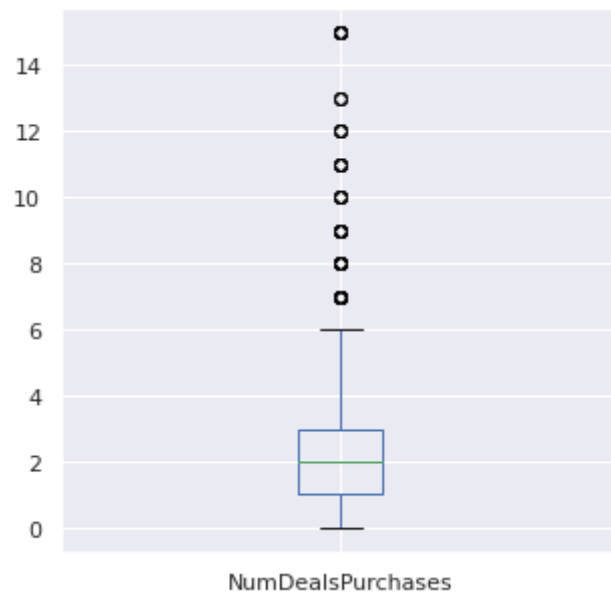
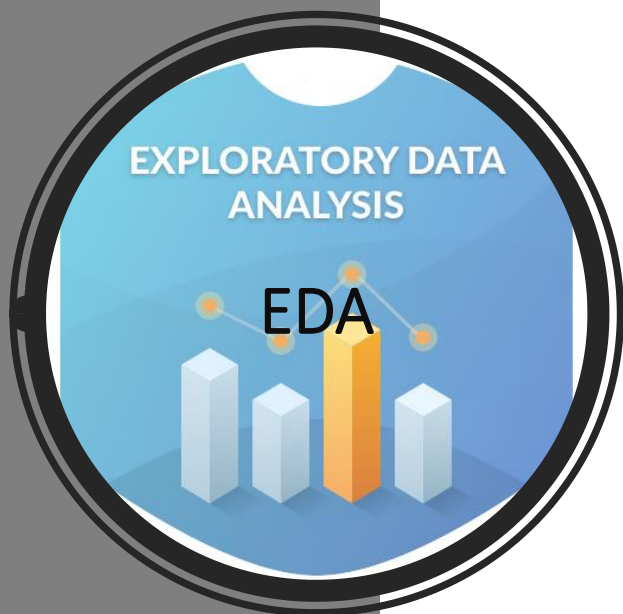


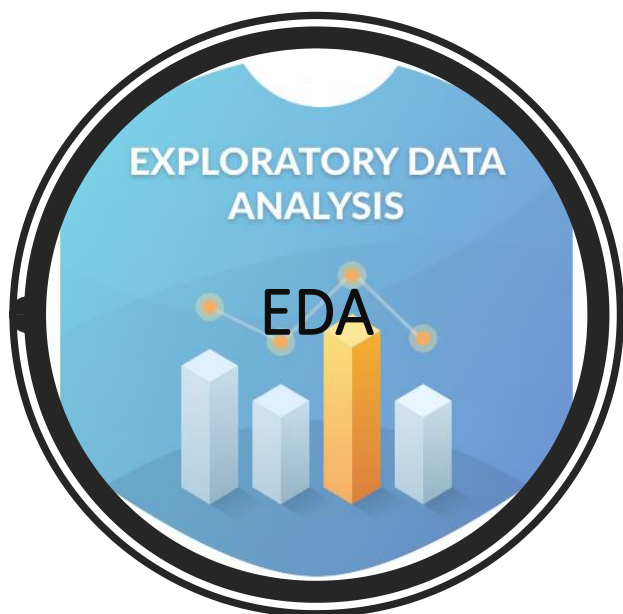
Medidas de
tendencia central



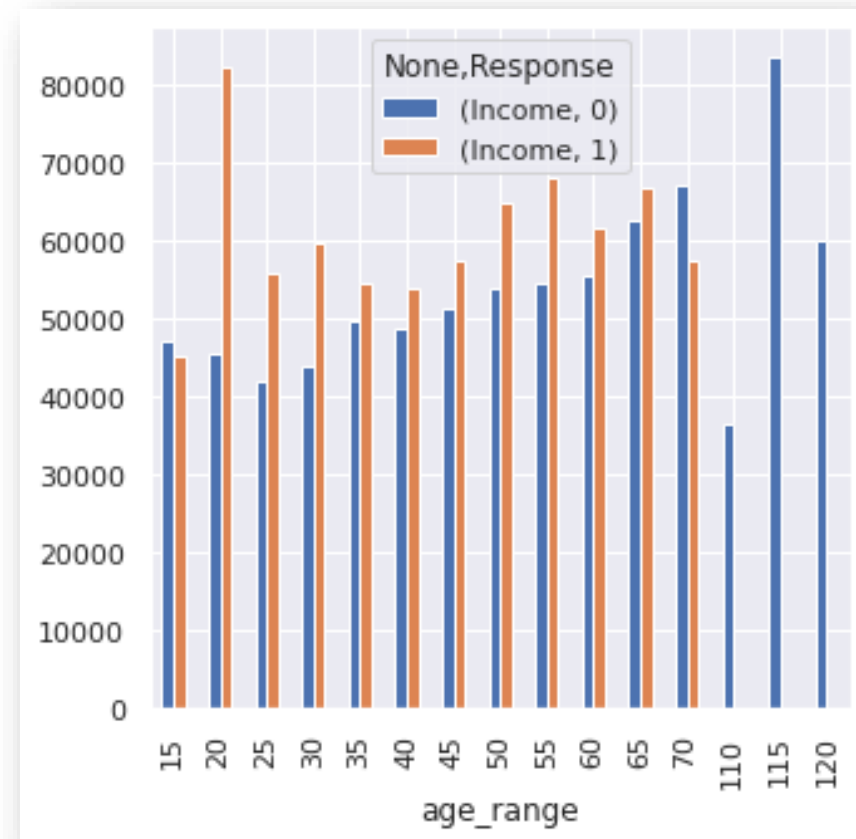
Gráficos

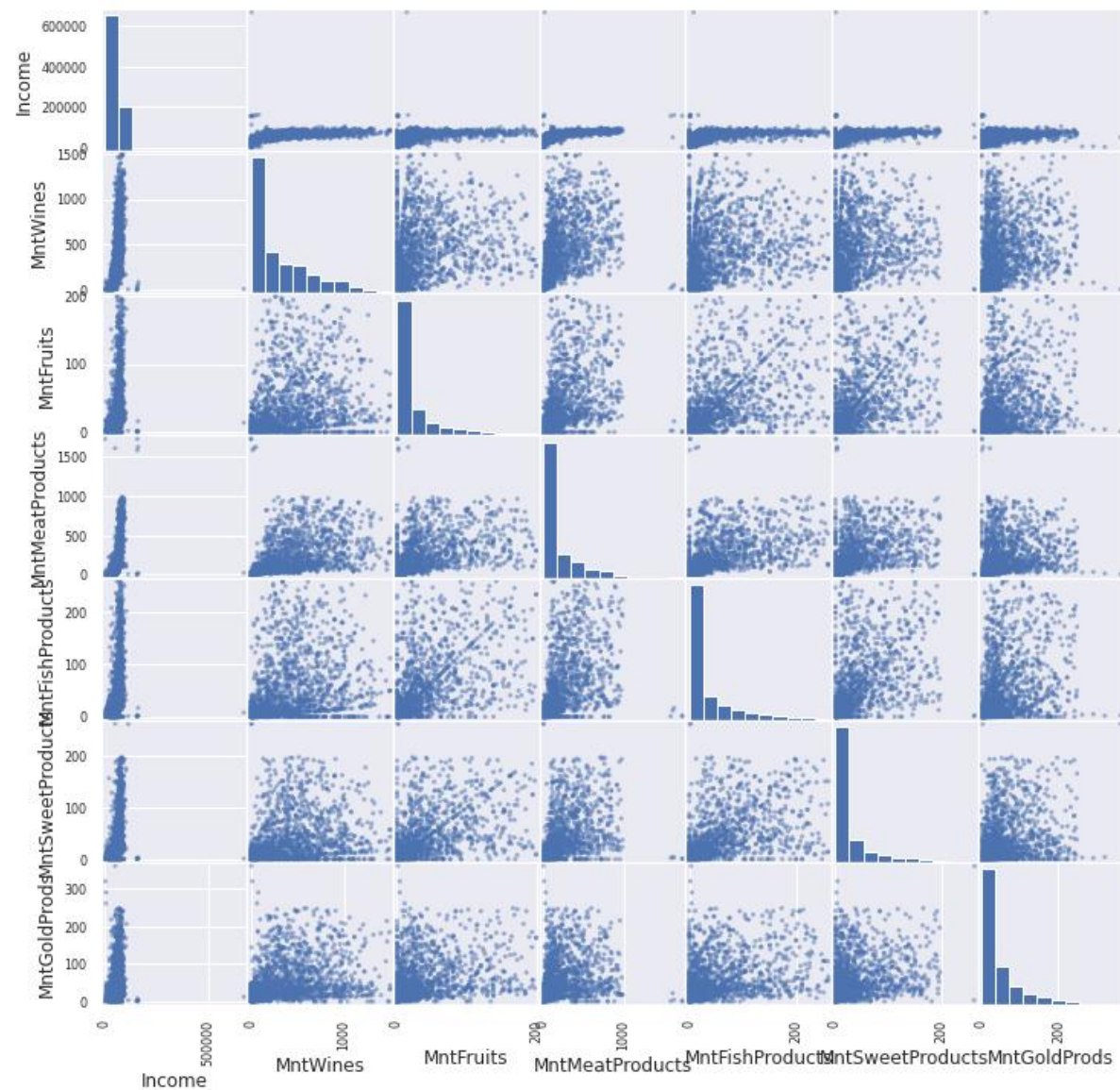
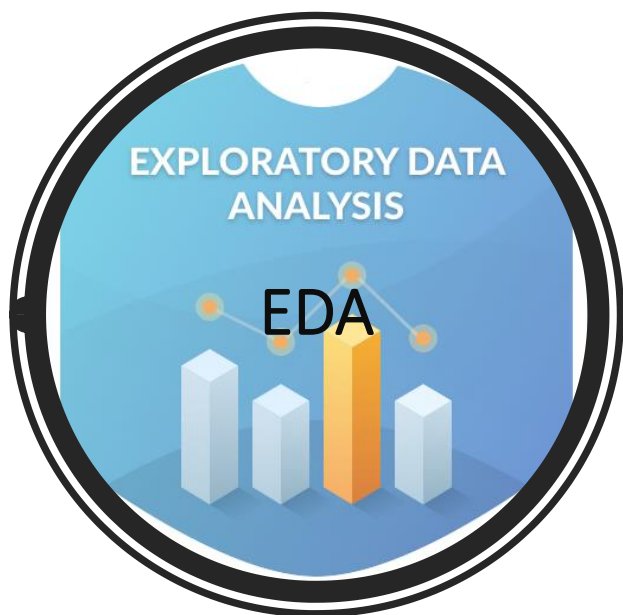






Response	Income	
	0	1
age_range		
15	47169.80	45378.50
20	45447.38	82508.42
25	41942.80	55807.64
30	44090.53	59777.49
35	49649.64	54639.90
40	48883.83	53977.36
45	51511.21	57509.11
50	54076.08	65049.43
55	54758.11	68336.91
60	55579.22	61810.50
65	62803.75	66809.77
70	67080.67	57466.62
110	36640.00	0.00
115	83532.00	0.00
120	60182.00	0.00





Cliente ideal



No temenos información
de género



Edad

Mínimo
18

Máximo
121

Promedio
45

Estado Civil

Top
Casados

Segundo
Union Libre

Tercero
Solteros

Ingresos

Mínimo
\$1730

Máximo
\$666666

Promedio
\$51381

Educación

Top
Graduation

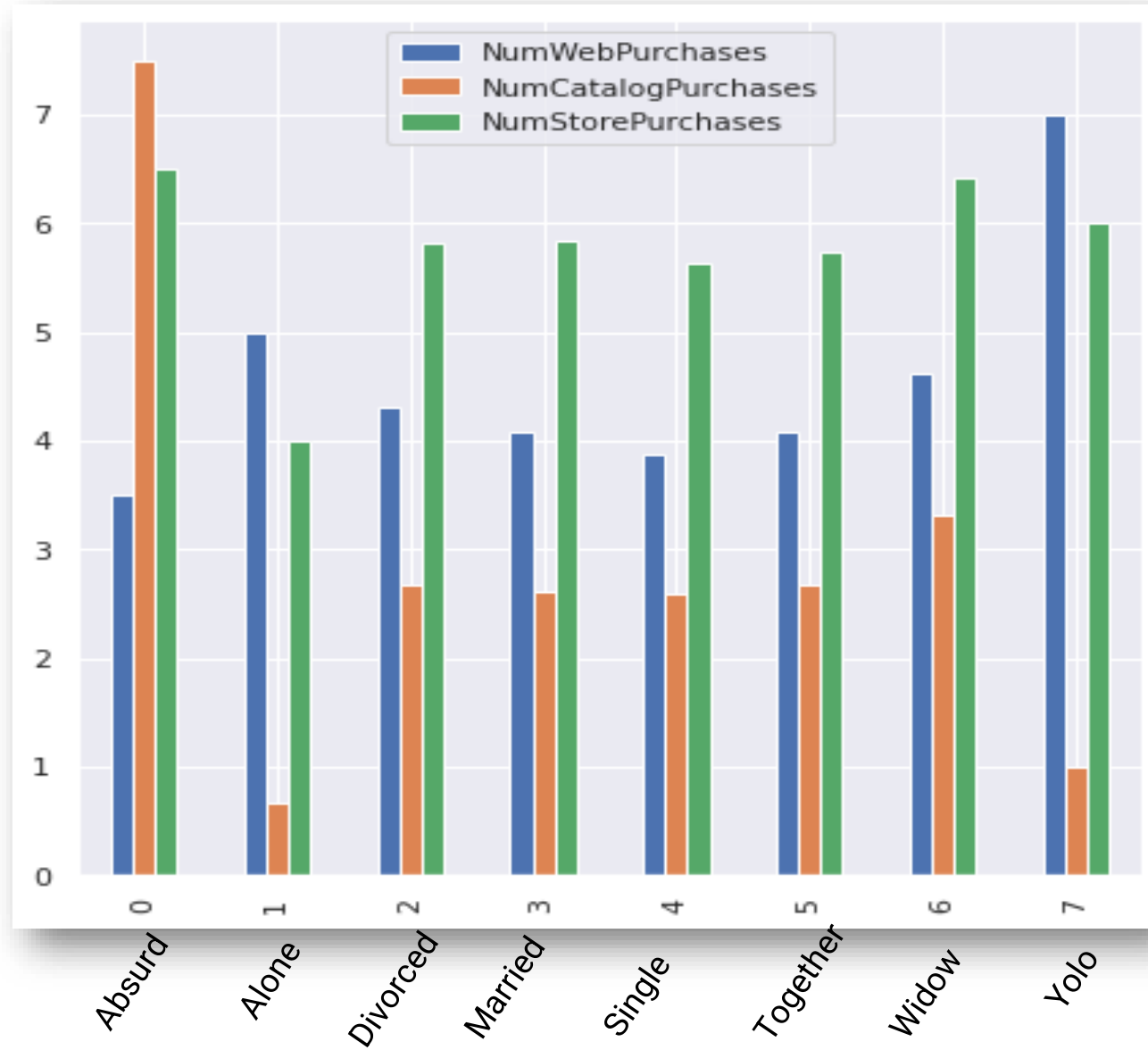
Segundo
PhD

Tercer
Master

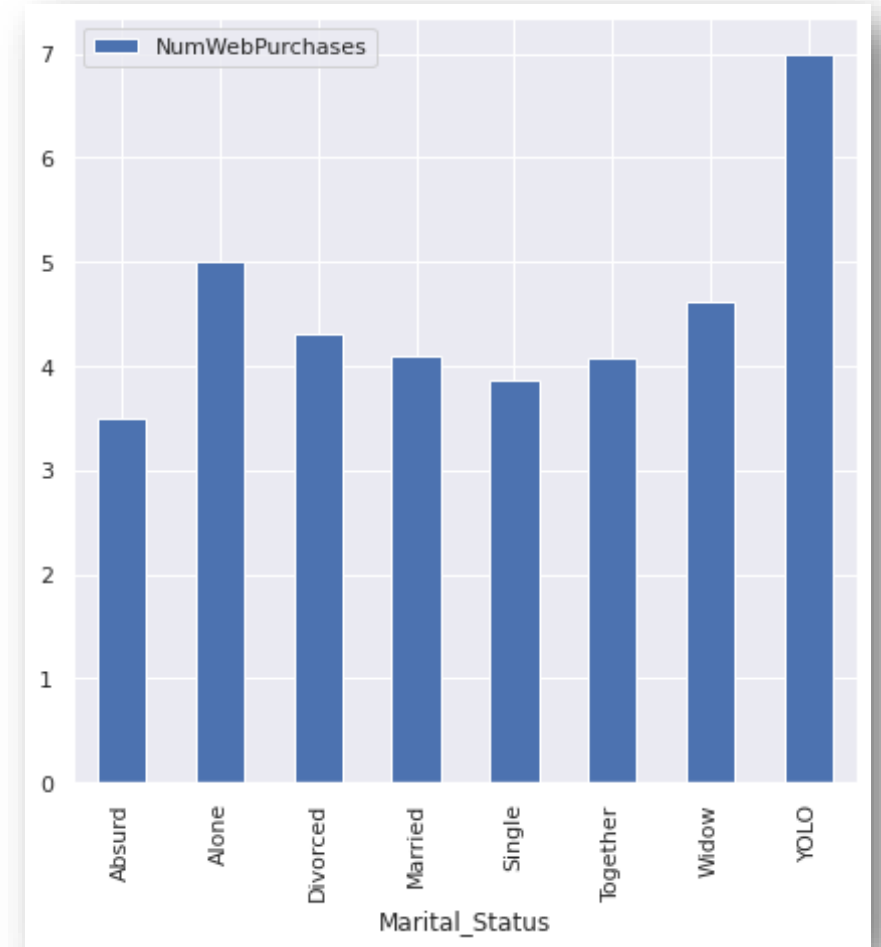
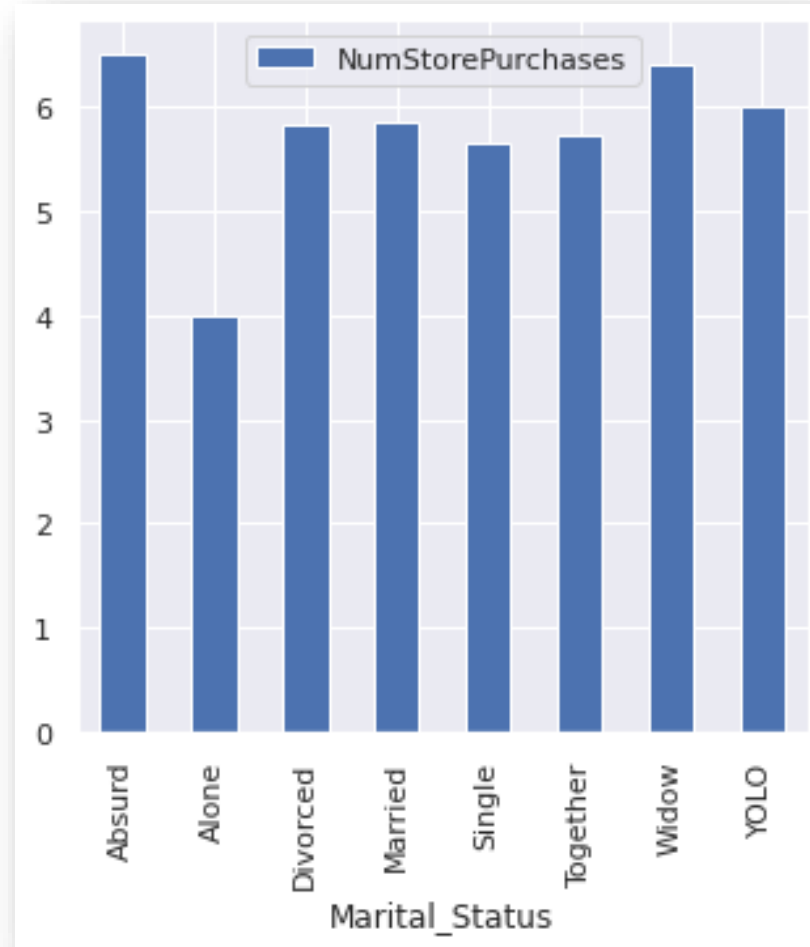
Otros Aspectos de los Clientes



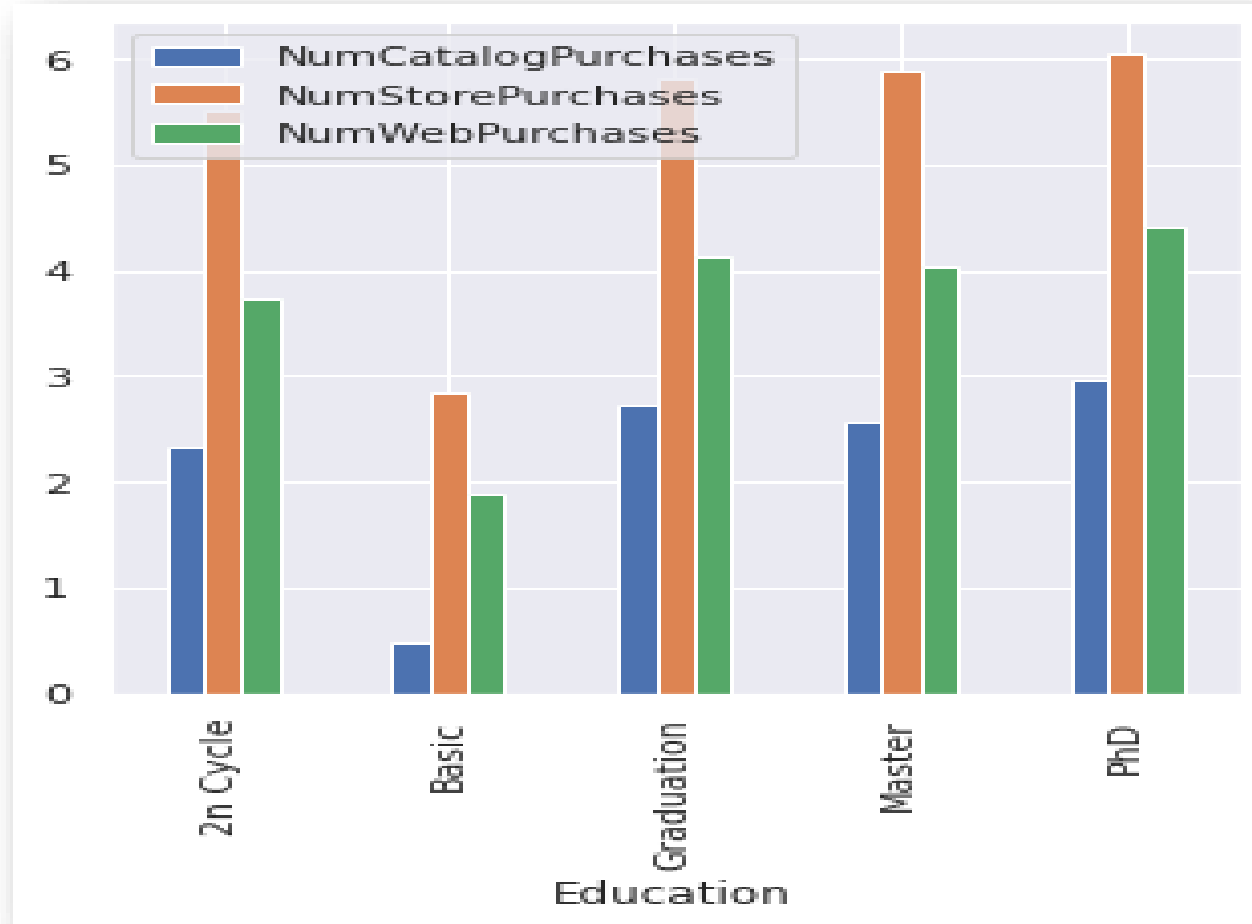
Comportamiento de Compra Según Canal de Venta



Comportamiento de Compra Según Canal de Venta

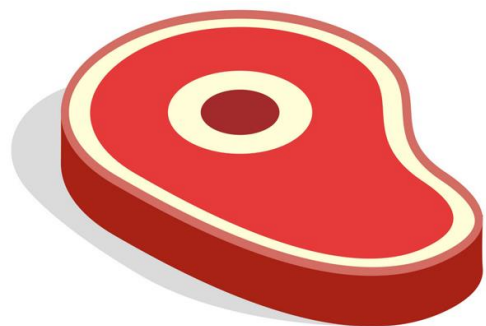


Comportamiento de Compra Según Canal de Venta



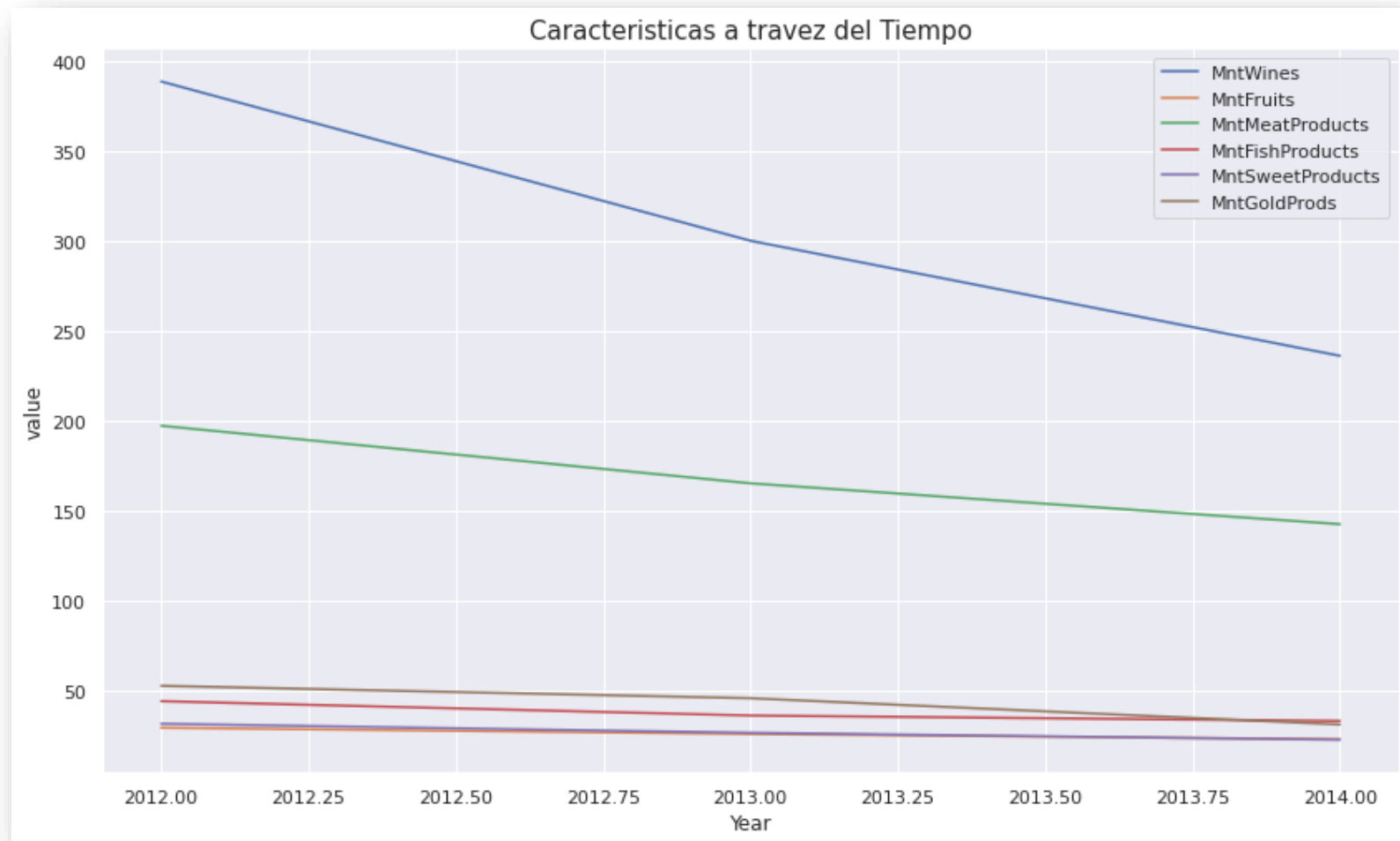


Comportamiento de Compra



Comportamiento Compra Rango Edad Promedio

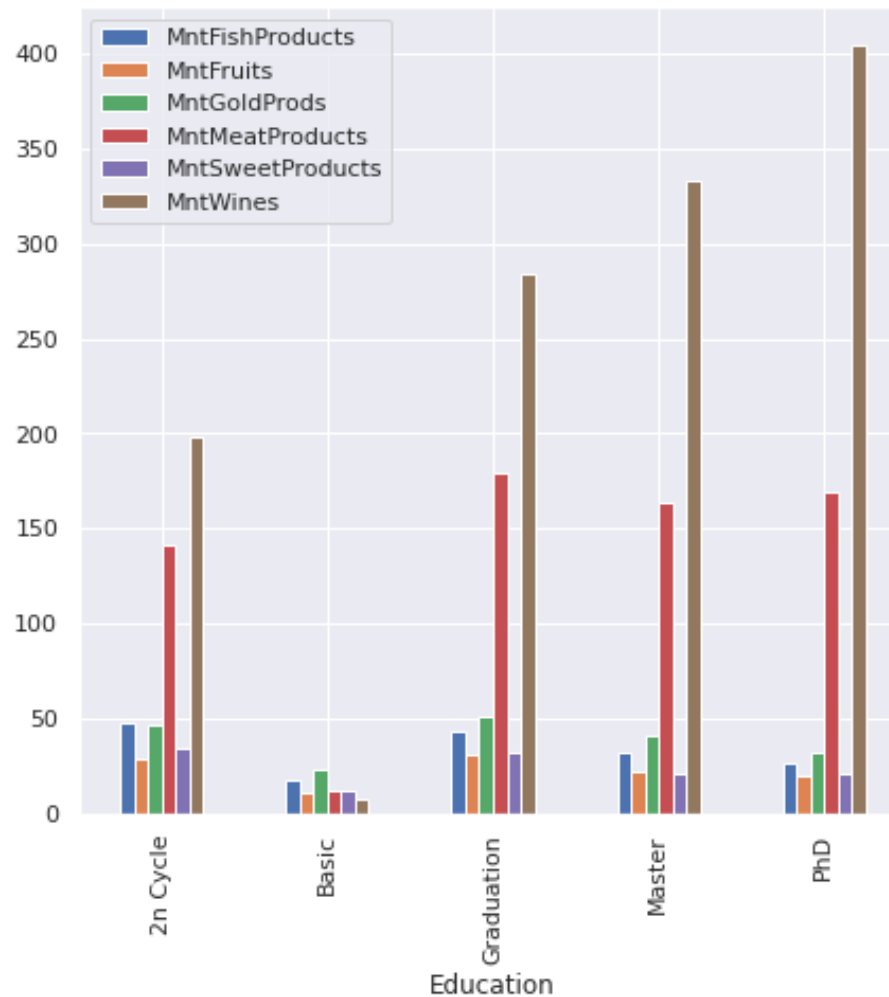
	age_range	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds
0	15	203.428571	16.714286	233.000000	38.571429	41.857143	29.857143
1	20	349.907407	41.537037	298.537037	55.092593	34.592593	62.944444
2	25	228.656250	27.631250	173.037500	35.118750	29.875000	38.037500
3	30	240.384236	26.369458	160.068966	38.423645	25.990148	40.931034
4	35	244.310734	24.121469	141.573446	35.641243	25.598870	41.144068
5	40	268.948187	22.673575	148.500000	31.349741	24.626943	37.059585
6	45	306.375862	24.893103	150.658621	32.182759	24.351724	44.155172
7	50	372.620370	28.194444	163.643519	38.851852	26.310185	48.560185
8	55	356.326693	29.011952	185.063745	42.613546	29.095618	47.043825
9	60	344.602871	23.516746	166.330144	39.215311	26.693780	54.602871
10	65	476.967033	34.032967	252.637363	49.978022	39.362637	49.538462
11	70	614.812500	46.187500	298.812500	87.000000	34.937500	28.937500



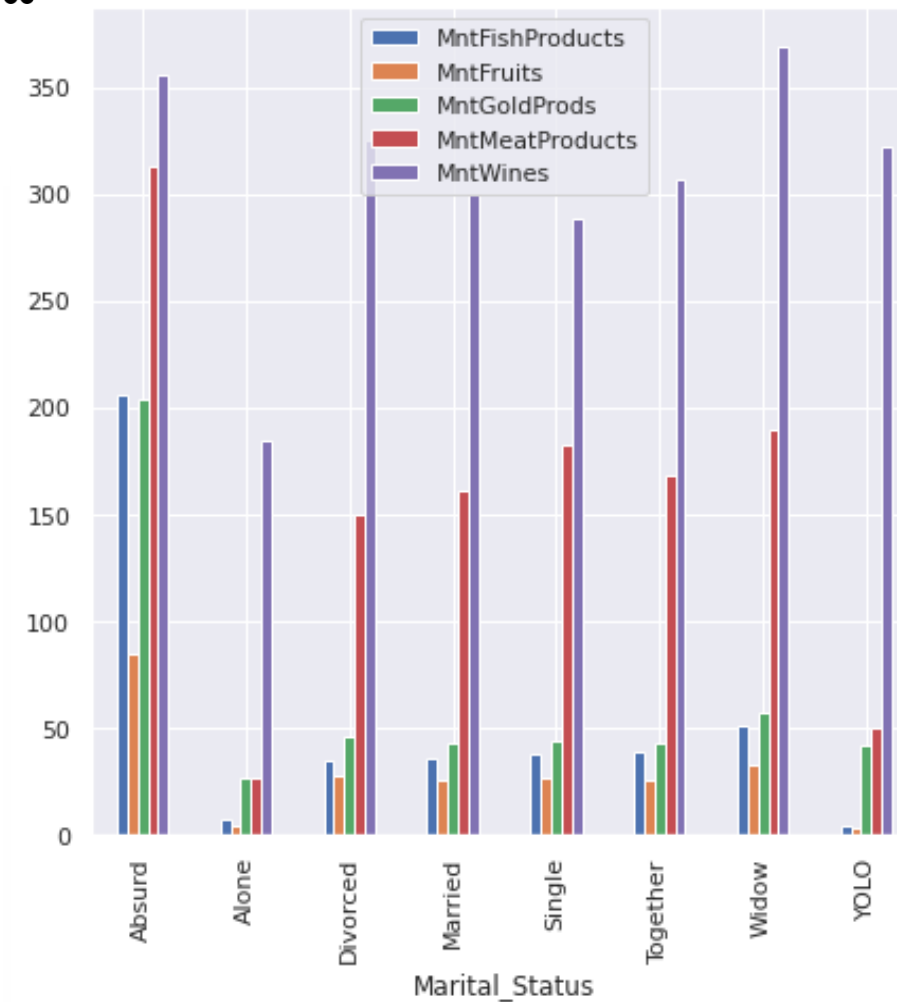
El nivel de compra de los productos viene a la baja

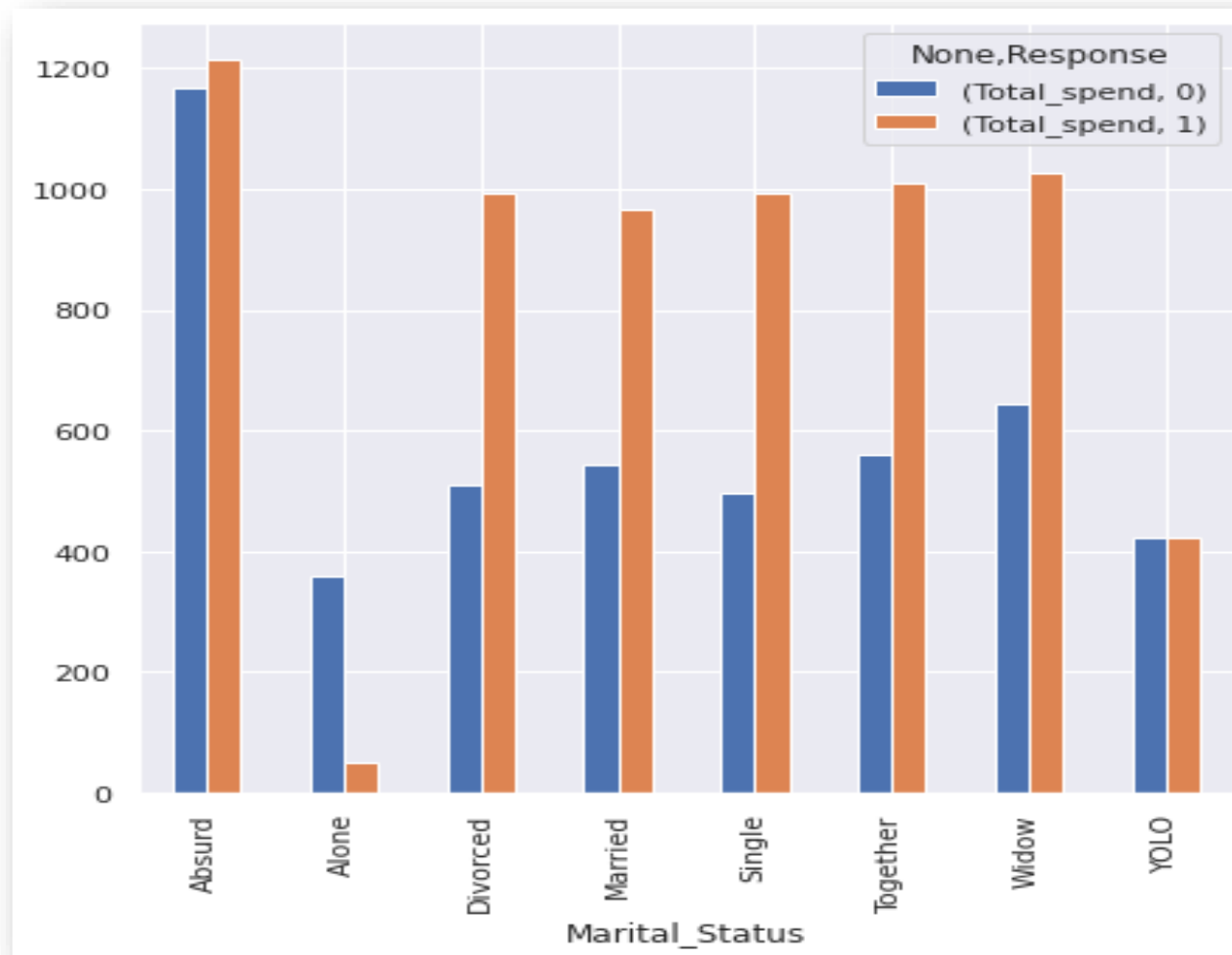


Comportamiento de Compra según Nivel Educativo



Comportamiento de Compra según Estatus Marital





Comportamiento de Compra según la respuesta a las campañas, nivel de ingreso y estado civil de los clientes.



	MntFishProducts	MntFruits	MntGoldProds	MntMeatProducts	MntSweetProducts	MntWines
Year						
2012	44.47	29.72	53.01	197.52	31.95	388.80
2013	36.51	26.29	46.14	165.54	26.90	300.32
2014	33.52	23.30	31.54	142.84	23.06	236.39

Las líneas en las que las personas gastan más por año:

1. Bebidas alcohólicas
2. Productos cárnicos

Anova

Influencia del Respuesta campana sobre la vistia web

F_onewayResult(statistic=0.035569030632842884, pvalue=0.850425363665631)

Influencia del Respuesta campana sobre compra de frutas

F_onewayResult(statistic=35.690773060613886, pvalue=2.6831159908772832e-09)

Influencia del Respuesta campana sobre Las Ultima compra

F_onewayResult(statistic=91.73834607514921, pvalue=2.5203765627580955e-21)

Se observa que el nivel de Respuesta positiva o no a la campaña influye en las visitas a las paginas web. pv 0.85 \geq 0.05

Así como la Repuesta positiva o no a la campaña influye en el consumo o compras de los clientes. pv 2.68 \geq 0.05

De igual forma la Repuesta positiva o no a la campaña influye en los días que tiene una persona de su ultima compra (influye en que vuelva a comprar o compre más seguido) pv 2.52 \geq 0.05

Anova

Datos e Información sobre la Influencia de la Publicidad y las Decisiones de Compra de los Clientes.

Con respecto a esto se procede a citar algunos aspectos de la influencia de la publicidad y como esta promueve el consumo:

Una de las funciones de la publicidad es **Persuadir**, se crea preferencia por la marca, se debe convencer a los consumidores de que compren en el momento, como también que acepten visitas por el equipo de ventas.

La función de la publicidad no es vender: la publicidad es responsable de crear condiciones para la venta de productos. Para ello, transmite la idea al segmento de mercado objetivo, esperando que pueda coincidir con la idea y tenerla en cuenta a la hora de comprar.

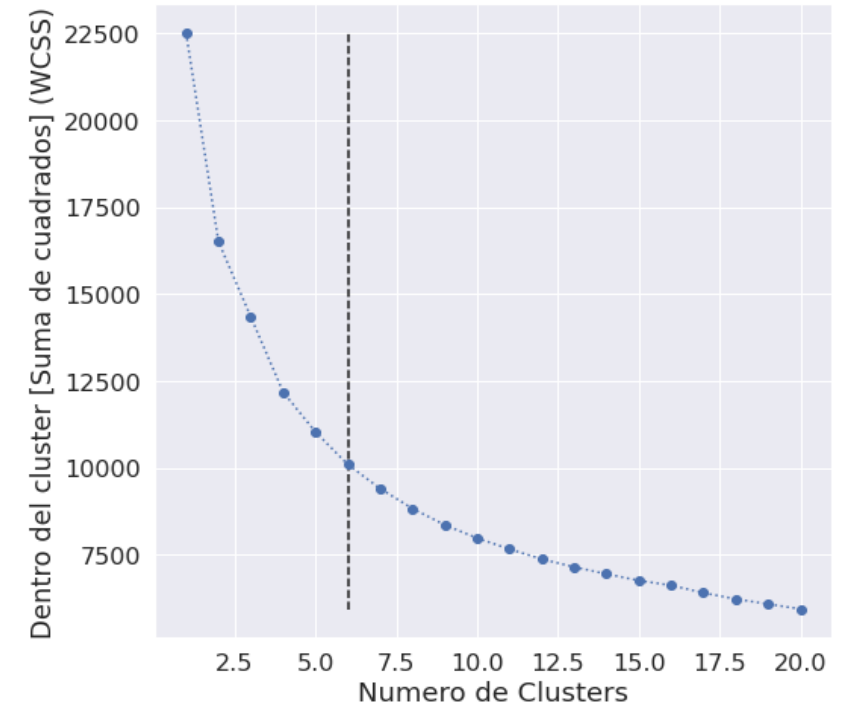
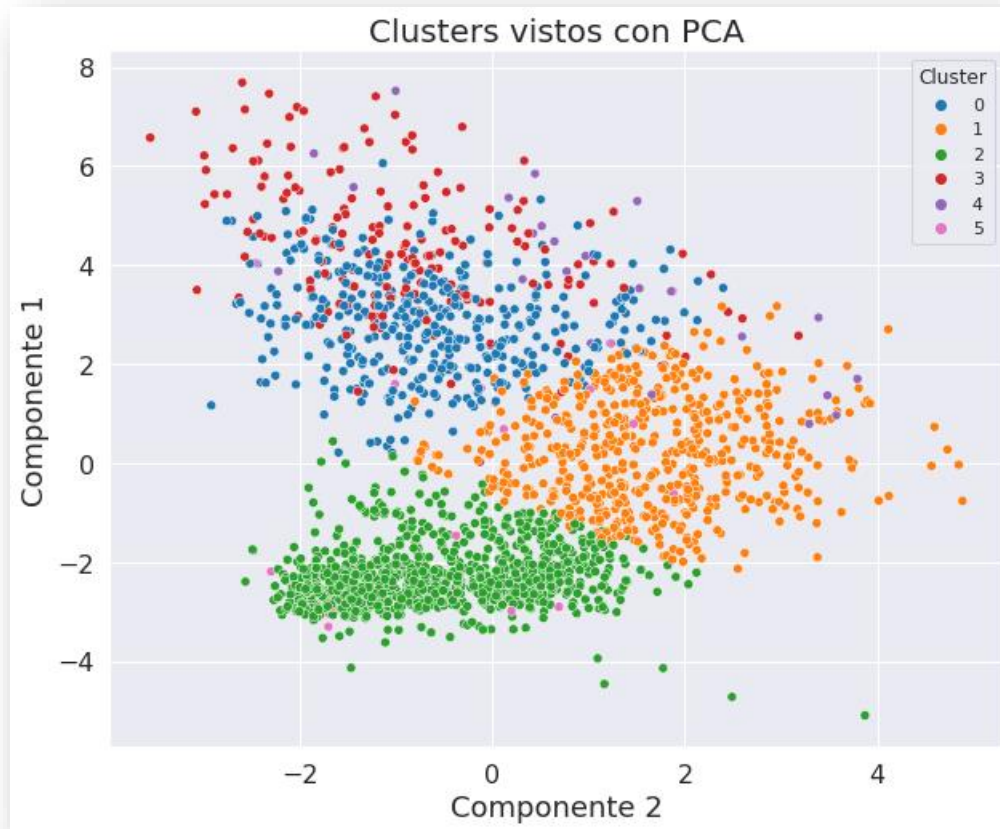
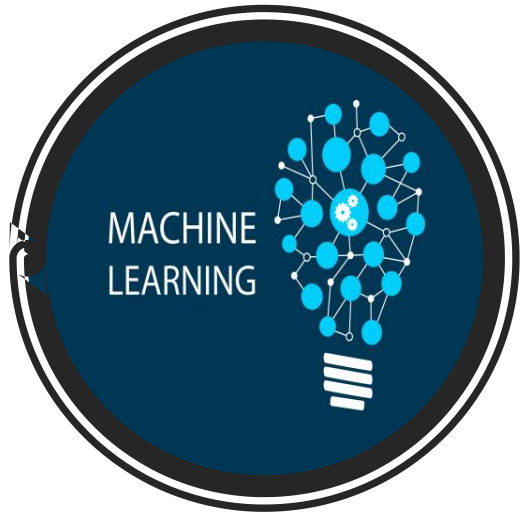
han identificado diferentes factores que afectan la intención de compra entre los cuales se encuentran: la publicidad, el brand equity y el brand engagement. La publicidad actúa como un divulgador de información o de ideas con la intención de cambiar el comportamiento de un grupo de personas para que piense o actúe de determinada manera, también actúa como un diferenciador de una marca frente a la competencia.

1. Influencia de la publicidad en el comportamiento de compra de los estudiantes de mercadeo de la Extensión Universitaria de Aguadulce. vol. 5, núm. 1, 2020 . <http://portal.amelica.org/ameli/jatsRepo/212/2121146005/html/index.html>

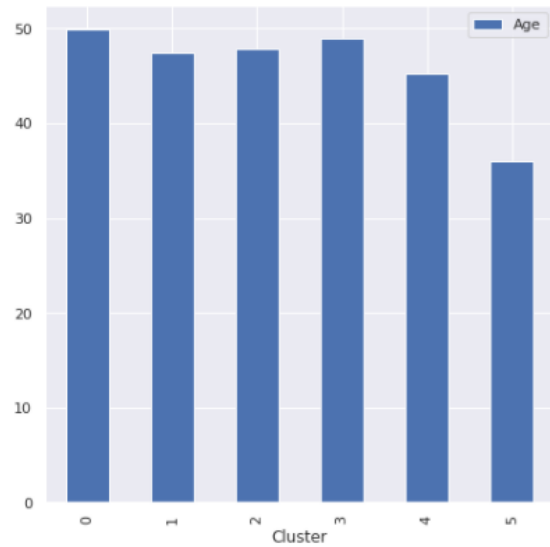
2. Análisis del efecto de la publicidad en la intención de compra y el papel que en esa relación juegan el brand engagement y el brand equity.2018

https://repository.icesi.edu.co/biblioteca_digital/bitstream/10906/84330/1/TG02186.pdf

Modelo No Supervisado



```
round(df_seg_pca_kmeans.pivot_table('Age', index= 'Cluster', aggfunc= 'mean',fill_value=0),2).plot(kind = 'bar')
sns.set(rc={'figure.figsize':(7,7)})
plt.show()
```

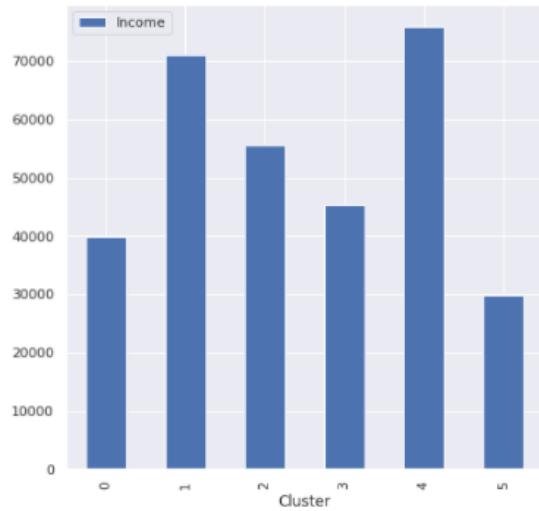


```
round(df_seg_pca_kmeans.pivot_table('Age', index= 'Cluster', aggfunc= 'mean',fill_value=0),2)
```

Age	
Cluster	
0	49.93
1	47.44
2	47.88
3	48.90
4	45.19
5	35.96

- Del gráfico anterior se puede ver que las edades entre los primeros 4 clúster son muy similares, mientras que los últimos dos clúster tienen a los clientes más jóvenes

```
round(df_seg_pca_kmeans.pivot_table('Income', index= 'Cluster', aggfunc= 'mean',fill_value=0),2).plot(kind = 'bar')
sns.set(rc={'figure.figsize':(7,7)})
plt.show()
```

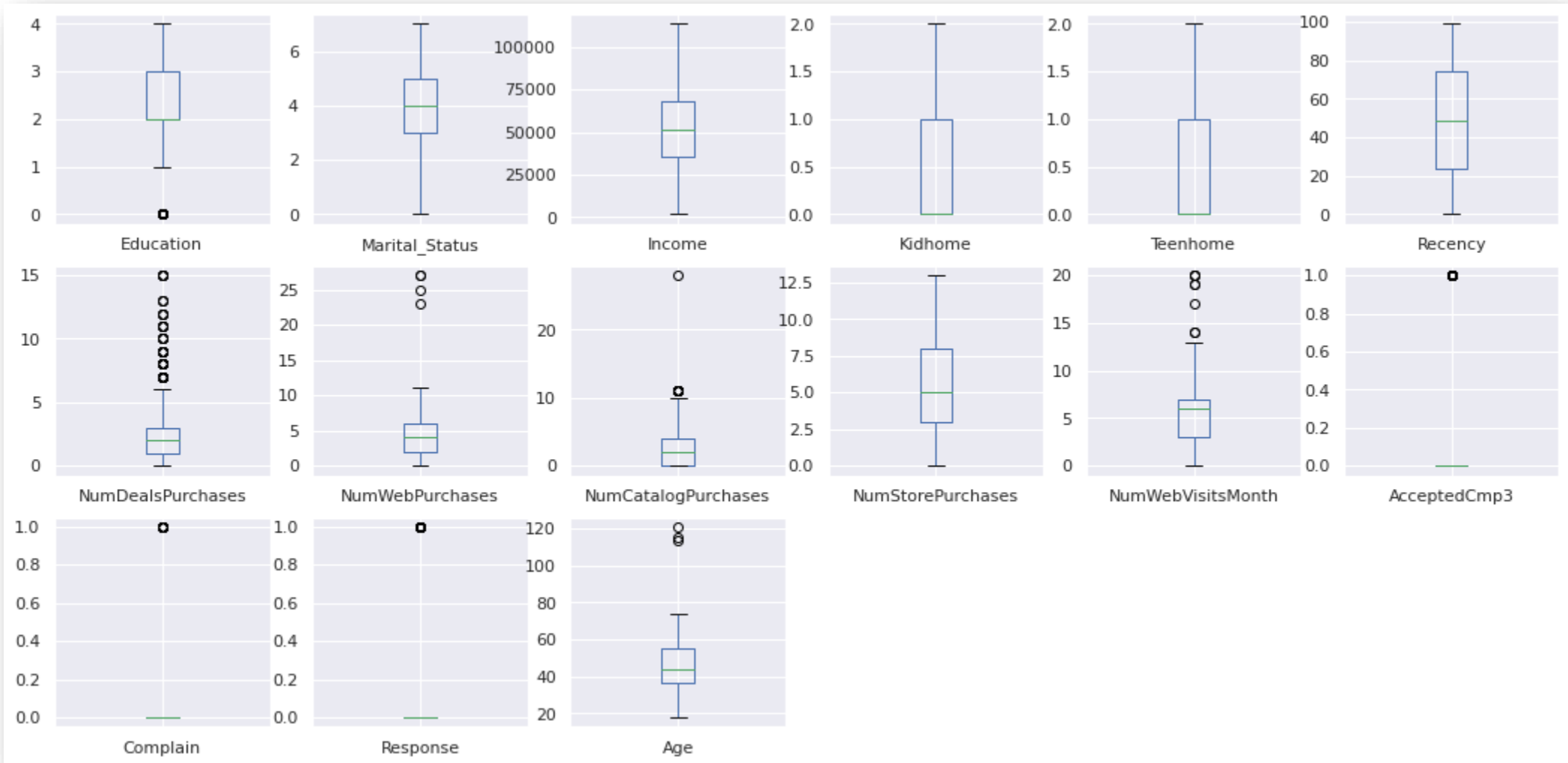


```
round(df_seg_pca_kmeans.pivot_table('Income', index= 'Cluster', aggfunc= 'mean',fill_value=0),2)
```

Income	
Cluster	
0	39826.32
1	70963.72
2	55645.15
3	45242.29
4	75879.98
5	29844.13

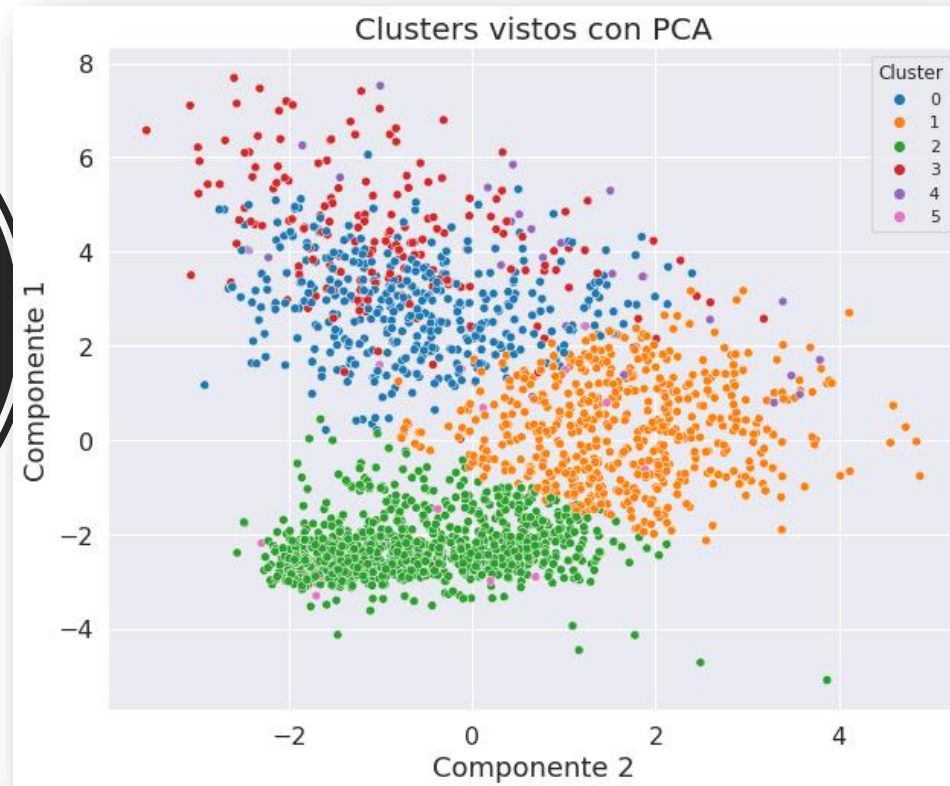
Según se puede ver el clúster 1 y 4 poseen los clientes de mayor ingreso

Características Seleccionadas

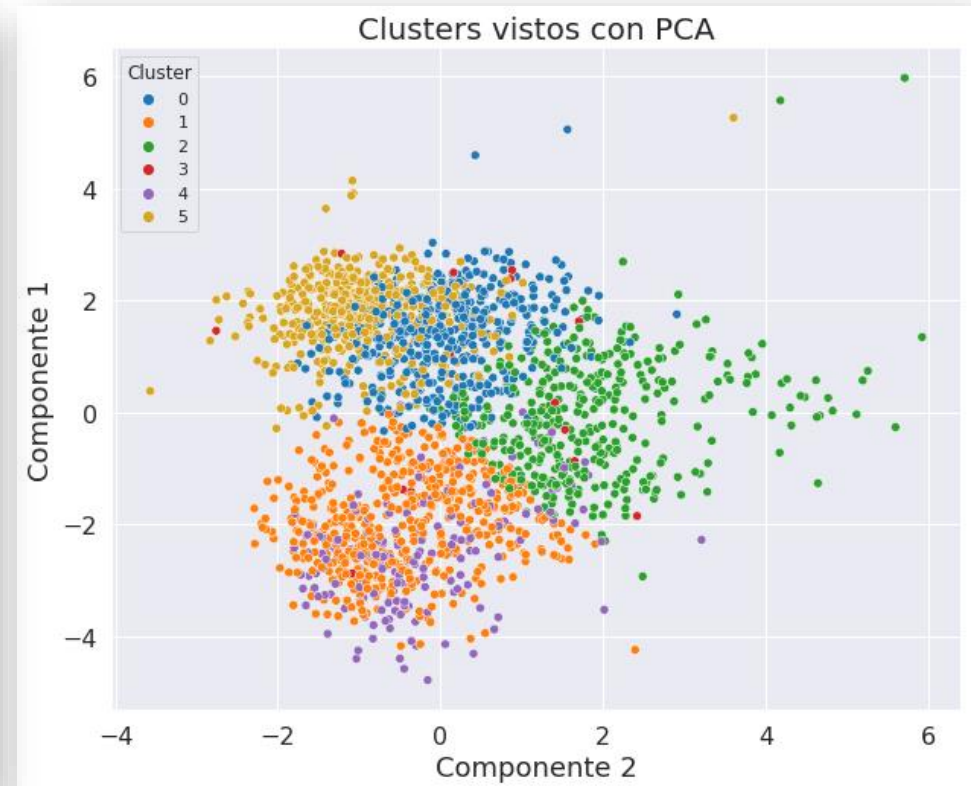


Modelo No Supervisado

Total de Características



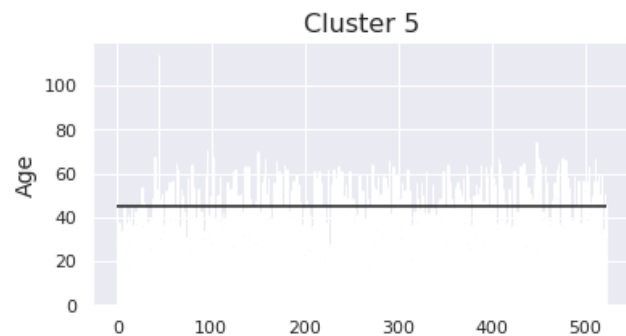
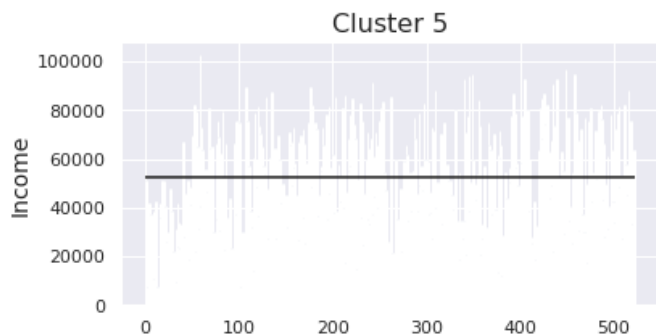
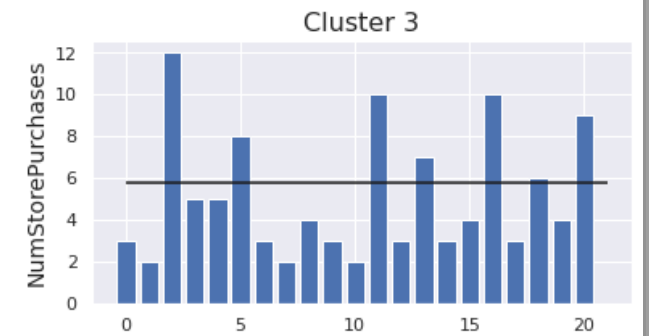
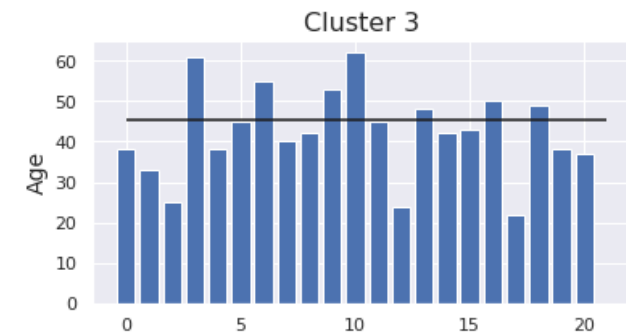
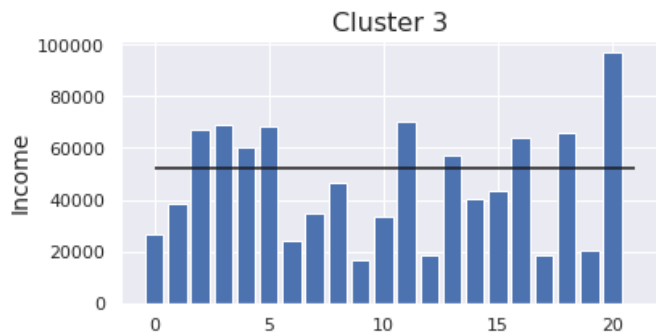
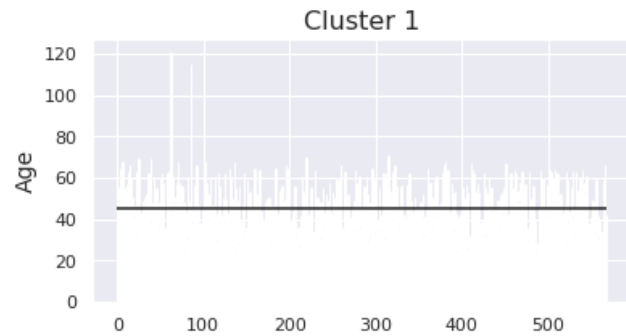
Características Seleccionadas y Tratamiento Outliers



MACHINE
LEARNING



Visualizando Características en los Cluster



Prueba de Modelo Supervisado

```
y_pred = classifier.predict(X_test)
```

```
reaccion_positiva = classifier.predict([[2,5,89000,2,2,12,1,1,42]]) # predicción de reacción o no a la campaña apublicitaria  
print(reaccion_positiva)
```

```
[0.]
```

```
reaccion_positiva = classifier.predict([[3,4,59000,2,0,38,0,0,45]]) # predicción de reacción o no a la campaña apublicitaria  
print(reaccion_positiva)
```

```
[0.]
```

```
reaccion_positiva = classifier.predict([[4,4,100000,2,8,44,1,0,28]]) # predicción de reacción o no a la campaña apublicitaria  
print(reaccion_positiva)
```

```
[0.]
```

```
reaccion_positiva = classifier.predict([[3,2,75000,2,1,8,1,1,28]]) # predicción de reacción o no a la campaña apublicitaria  
print(reaccion_positiva)
```

```
[1.]
```

```
data5.iloc[:,[0,1,2,3,4,5,11,12,13,14]]
```

```
Education  Marital_Status  Income  Kidhome  Teenhome  Recency  AcceptedCmp3  Complain  Response  Age
```

Conclusiones

Cómo son los Clientes?

En términos de Edad, Ingreso, Estado Civil, Nucleo Familiar Comportamiento de Compra

Entre menor se los días de la ultima compra, mejor respuesta tendrá el cliente a la campaña.

Existe relación entre comprar un producto junto con otro, además aquellos que tienen una mayor cantidad de hijos tienden a realizar compras con descuentos.

La cantidad promedio de hijos por hogar corresponde alrededor de un hijo, la edad promedio de los clientes es de 45 años, la edad que más se repite más veces en los clientes es 38.

El ingreso promedio de los clientes es de \$52 237

El monto promedio que los clientes gastan en las compras totales es de \$ 605.79

Los clientes prefieren en su mayoría realizar las compras en las tiendas o establecimientos físicos, seguido de las compras por la web.

Aquellos que van más a las tiendas corresponde a los que se encuentran casados o en una relación.

Según se observó en nivel de compras totales con el pasar de los años va en descenso.

Por medio de la aplicación de un método supervisado, se puede determinar si un cliente puede o no reaccionar positivamente a las campañas.

Además la creación de Cluster de los clientes, nos permite determinar por rangos de edad o niveles de ingreso algún tipo de conducta de compra, que permita formular una campaña publicitaria por el medio de compra preferido del cliente (tiendas y Web)