



ANÁLISIS DE REBAJAS

“EL CORTE INGLÉS”

ANÁLISIS EXPLORATORIO DE DATOS
DOBLE GRADO CIENCIA DE DATOS E
INGENIERÍA DE ORGANIZACIÓN INDUSTRIAL

UPV



ÓSCAR ANTONINO MANRIQUE
MIGUEL GIL JIMÉNEZ
DANIEL RIVAS GARCÍA

1. OBJETIVO.....	2
2. HIPÓTESIS INICIALES.....	2
3. OBTENCIÓN DE LOS DATOS.....	2
4. ANÁLISIS EXPLORATORIO DE DATOS.....	3
4.1. ANÁLISIS UNIVARIANTE	3
4.2. CONCLUSIÓN DEL ANÁLISIS UNIVARIANTE	9
4.3. ANÁLISIS MULTIVARIANTE.....	10
5. CONCLUSIONES	19

Contenido

1. OBJETIVO

El objetivo de este trabajo es realizar un Análisis Exploratorio de Datos sobre una fuente de datos escogida como parte de la asignatura de Análisis Exploratorio de Datos. En este caso, nos hemos decidido por analizar los productos que se encuentran en descuento de El Corte Inglés.

Para realizar este análisis vamos a hacer uso de herramientas como Statgraphics para el análisis de datos y de Python para la recolección y limpieza de datos junto a la extracción de algunas gráficas más complejas programadas con matplotlib. El objetivo que nos planteamos es plantearnos una hipótesis sobre los productos en descuento de esta tienda online, conseguir adquirir los datos mediante técnicas de webscrapping, analizarlos y confirmar o desmentir nuestras hipótesis inicialmente propuestas haciendo una conclusión del análisis realizado.

2. HIPÓTESIS INICIALES

Consideramos que este análisis puede ser de interés, ya que inicialmente nos hemos planteado algunas hipótesis como que la mayor parte de los productos que se encuentren en descuento en la web de El Corte Inglés irán dirigidos a mujeres, ya que creemos que compran mucho más frecuentemente que los hombres. Por ello clientes mucho más potenciales que los hombres y realizan campañas más dirigidas a este público. Como consecuencia de ello, suponemos que además de tener un mayor descuento, tendrán una mayor variedad de productos.

Por otro lado, creemos que los hombres tendrán menor presencia en subsecciones como la Ropa, pero mayor presencia en Accesorios y que los productos con un mayor precio inicial, tendrán un mayor descuento. Esta hipótesis se debe a que suponemos que los productos con mayor precio inicial tendrán un mayor margen de beneficio, por lo que se podrán permitir mayores porcentajes de descuentos.

Creemos que existirán grandes diferencias entre los productos de Hombre y de Mujer, en cuanto a la cantidad, la presencia en las subsecciones y los descuentos. Por eso nuestro análisis inicialmente se centrará en comparar estas dos grandes categorías. Es por ello que consideramos necesario realizar inicialmente un análisis univariante de las variables que obtengamos, para poder redefinir correctamente nuestros objetivos e hipótesis y así pasar finalmente a realizar un análisis multivariante donde podamos analizar todas estas relaciones de una forma precisa.

3. OBTENCIÓN DE LOS DATOS

Los datos se han obtenido de la página oficial de El Corte Inglés en concreto de la subsección “[Descuentos](#)” del día 8 de noviembre del 2023. Para ello se ha hecho uso de técnicas de web scraping mediante Python. En concreto se han usado las librerías selenium y bs4. En este [link](#) se puede acceder al repositorio de Github en el que se

encuentra el código fuente para ejecutar el webscraping. Ejecutando el archivo “webscraping_main.ipynb” se descargará un .csv para cada subsección de la Sección de Descuentos tanto de hombre como de mujer. Posteriormente, ejecutando el archivo “data_union.ipynb” se unirán los diferentes .csv con el fin de crear un único archivo con todos los datos. Para más información sobre el proceso de la adquisición de datos, puede visitar el readme del repositorio de GitHub.

Tras realizar este proceso, se obtuvieron finalmente 4497 productos diferentes y para cada producto se especifican las siguientes variables:

- 1) Nombre del producto
- 2) Marca
- 3) Precio inicial
- 4) Descuento
- 5) Precio descuento
- 6) Sección
- 7) Subsección

El proceso de Feature Engineering se realizó en el propio proceso de obtención de datos, de forma que al analizar los datos posteriormente estos se encontrasen ya limpios. El webscraping permitió obtener todos los datos sin ningún nulo, dado que fueron sacados de la propia página de El Corte Inglés donde para cada producto encontramos los datos pertenecientes a cada variable.

4. ANÁLISIS EXPLORATORIO DE DATOS

4.1. ANÁLISIS UNIVARIANTE

En el primer paso realizado en este Análisis Exploratorio de Datos sobre los productos en descuento de El Corte Inglés se proceden a estudiar las diferentes variables de forma individual para así tener un primer contacto con nuestros datos y buscar posibles anomalías, errores y datos llamativos.

ANÁLISIS UNIVARIANTE: NOMBRE DEL PRODUCTO	
Tipo variable	Cualitativa nominal
Descripción	Nombre del producto con el que se anuncia en la web de El Corte Inglés
Cantidad de nulos	0
Recuento	4497
Categorías diferentes	3650

Esta variable no la consideramos de gran importancia, ya que, aunque no se trata de un id único por producto, es simplemente un dato informativo del nombre del producto que no aporta información. Además, observamos que hay 4497 productos

en total y 3650 productos con nombres diferentes, por lo que no nos servirá para poder encontrar relaciones o agrupar los productos.

ANÁLISIS UNIVARIANTE: MARCA	
Tipo variable	Cualitativa nominal
Descripción	Nombre de la empresa fabricante del producto
Cantidad de nulos	0
Recuento	4497
Categorías diferentes	99

Consideramos que la variable marca es de gran importancia ya que tenemos la hipótesis de que algunas marcas harán un descuento mucho superior a las demás. Sin embargo, creemos que un gráfico para representar todas las marcas no es relevante dado el alto número de categorías (99).

Si analizamos la tabla de frecuencias de la variable Marca, cabe destacar que la marca “Singularu” tiene 694 productos, “Fiorella Rubino” tiene 421 productos y “Tous” 240. Mientras que por otro lado hay 14 marcas que solo tienen un producto, por lo que estas marcas podrían considerarse residuales.

ANÁLISIS UNIVARIANTE: PRECIO INICIAL	
Tipo variable	Cuantitativa continua
Descripción	Precio en euros del producto con anterioridad al descuento
Cantidad de nulos	0
Recuento	4497
Media	231,037
Mediana	64,9
Media recortada 5%	93,6073
Media winsorizada	114,696
Varianza	742687
Desviación estándar	861,793
Variación	373,011%
Desviación media absoluta	0,767961
Mínimo	5,99
Máximo	12999,0
Rango	12993,0
Cuartil inferior	37,99
Cuartil Superior	110,0
Rango intercuartílico	72,01
Sesgo estandarizado	228,105
Curtosis	83,0447
Curtosis estandarizada	1136,7 6

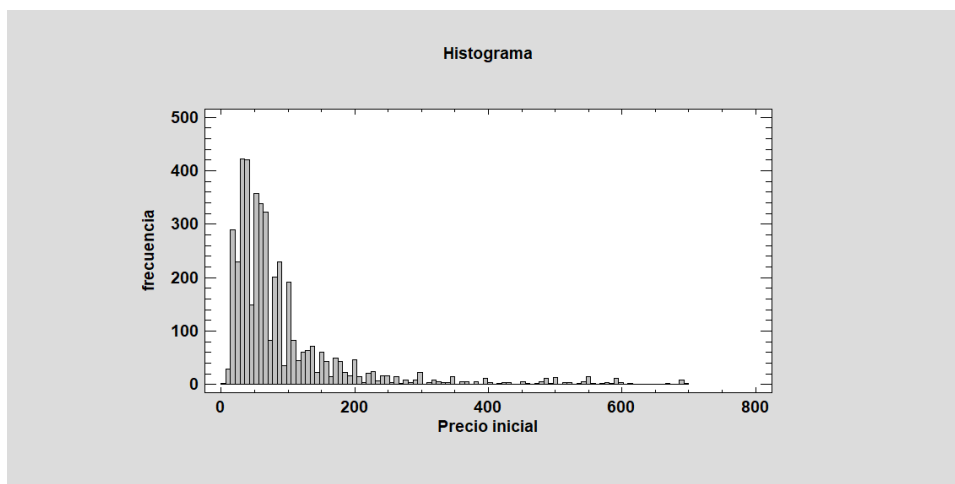
En la Gráfica 1 se han recortado los valores a partir del 800, dado que el máximo es 12999 aunque no es muy representativo dado que el cuartil superior es igual a 110.

Es por ello que podemos afirmar que existen valores atípicos y en consecuencia la media (231,037) no es una buena medida de centralización, por ello podemos fijarnos mejor en la mediana (64,9). El cuartil inferior del precio es 37,99 y el cuartil superior 110, por lo que el 50% de los productos tiene un precio inicial entre ambos valores.

Si analizamos la dispersión de esta variable observamos que la desviación típica es 861,793 aunque en este caso deberíamos fijarnos en un parámetro robusto como la desviación media absoluta que toma el valor de 0,767961.

Por último, si nos centramos en la forma podemos afirmar que tiene una asimetría a la derecha debido a todos estos valores atípicos y que los datos siguen una forma leptocúrtica debido a una curtosis de 83,0447.

De este análisis univariante podemos concluir que la mayoría de los productos tienen un precio entre 37,99 y 100, aunque existen productos con un precio mucho más elevado, aunque en mucha menor cantidad, dado que no todas las personas se lo pueden permitir.

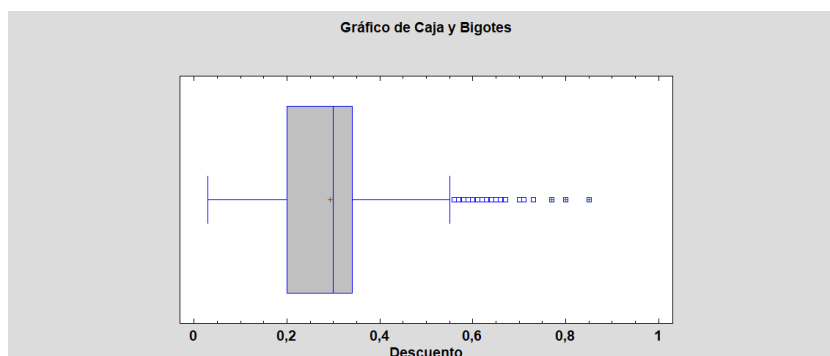


Gráfica 1: Histograma del Precio inicial

ANÁLISIS UNIVARIANTE: DESCUENTO	
Tipo variable	Cuantitativa continua
Descripción	Porcentaje de descuento aplicado al precio inicial
Cantidad de nulos	0
Recuento	4497
Media	0,293845
Mediana	0,3
Media recortada 5%	0,287057
Media winsorizada	0,293351
Varianza	0,0188924
Desviación estándar	0,13745

Variación	46,7763%
Desviación media abs.	0,101448
Mínimo	0,03
Máximo	0,85
Rango	0,82
Cuartil inferior	0,2
Cuartil Superior	0,34
Rango intercuartílico	0,14
Sesgo estandarizado	15,7482
Curtosis estand.	0,6462

En la Gráfica 2 observamos que hay bastantes datos anómalos, por lo que en este caso como parámetro de centralización tomamos la mediana (0,3) por su robustez, aunque tiene un valor similar a la media. Observamos que el 50% de los datos se encuentran distribuidos entre 0,2 y 0,34 (rango intercuartílico 0,14). Como parámetro de dispersión tomamos la desviación media absoluta que toma un valor de 0,101448. Por otro lado, debemos conocer el mínimo y el máximo de esta variable, ya que representan el descuento máximo realizado (85%) y el descuento mínimo de un producto (3%). Finalmente, concluimos que lo importante de este análisis ha sido conocer que la mayoría de los descuentos se encuentran en torno al 30%, aunque existe una asimetría a la derecha que representa a los productos que cuentan con un gran descuento, que son la minoría de los casos.



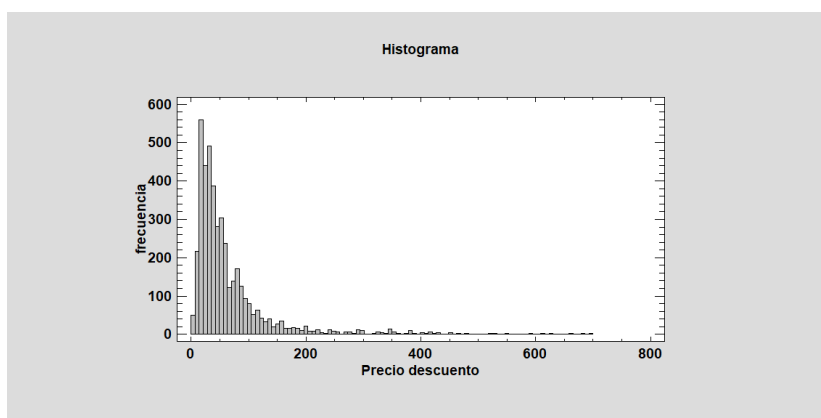
Gráfica 2: Caja y Bigotes del Descuento

ANÁLISIS UNIVARIANTE: PRECIO DESCUENTO	
Tipo variable	Cuantitativa continua
Descripción	Precio en euros del producto tras aplicar el descuento
Cantidad de nulos	0
Recuento	4497
Media	165,978
Mediana	46,9
Media recortada 5%	64,7831
Media winsorizada	79,0648
Varianza	404564
Desviación estándar	636,054
Variación	383,215%

Desviación media absoluta	207,467
Mínimo	4,75
Máximo	9999,0
Rango	9994,25
Cuartil inferior	26,99
Cuartil Superior	83,3
Rango intercuartílico	56,31
Sesgo estandarizado	228,07
Curtosis estandarizada	1117,91

Debemos tener en cuenta que el Precio de Descuento es una variable dependiente de otras dos variables, ya que es igual al producto del Precio inicial y del Descuento.

Como era de esperar, estos datos siguen una distribución similar a la del Precio Inicial. Podemos observar que la media y la mediana disminuyen respecto a la del Precio Inicial y que el rango intercuartílico pasa de ser 72,01 a 56,31 tras aplicar el descuento. Esto nos lleva a concluir que los datos se cuentan menos dispersos. En concreto, si nos fijamos en la mediana, esta pasa de ser 64,9 a 46,9 por lo que se puede concluir que los productos sufren un descuento mediano de un 27,73%, valor que es muy similar a la mediana de la variable Descuento (30%).



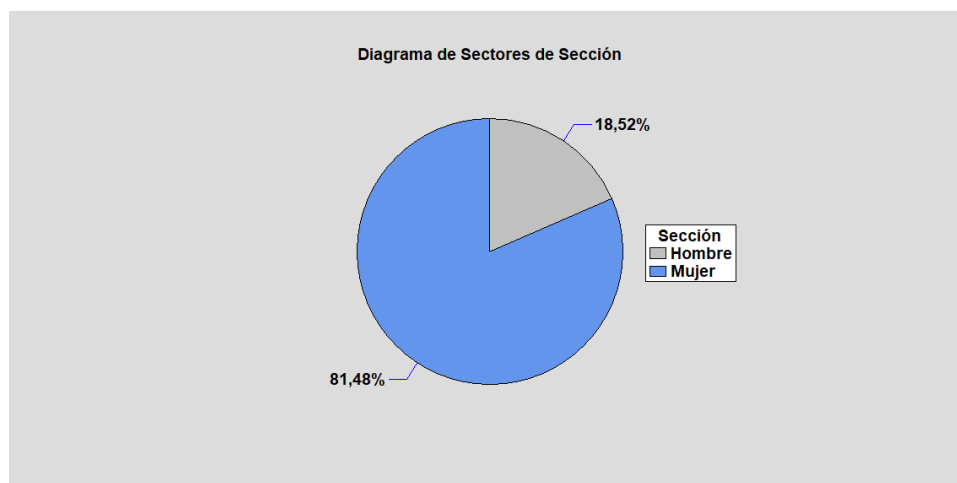
Gráfica 3: Histograma del Precio de Descuento

ANÁLISIS UNIVARIANTE: SECCIÓN	
Tipo variable	Cualitativa nominal
Descripción	Sección a la que pertenece el producto. En este caso solo se ha realizado webscraping sobre las secciones de Hombre y Mujer.
Cantidad de nulos	0
Recuento	4497
Categorías diferentes	2 (Hombre y Mujer)

Tal y como observamos en la Tabla de Frecuencias (Tabla 1), la cartera de productos en descuento de El Corte Inglés se compone en su mayoría de productos de mujeres.

Entrando más en detalle podemos reconocer que el 18,52% de los productos van orientados al género masculino, mientras que el 81,48% de ellos son para las mujeres.

Analizando estos datos se podría tomar la hipótesis de que las mujeres consumen un número mayor de productos que los hombres, por lo que El Corte Inglés está más preparado para este conjunto de clientes potencial ofreciéndole una gama más amplia de productos.



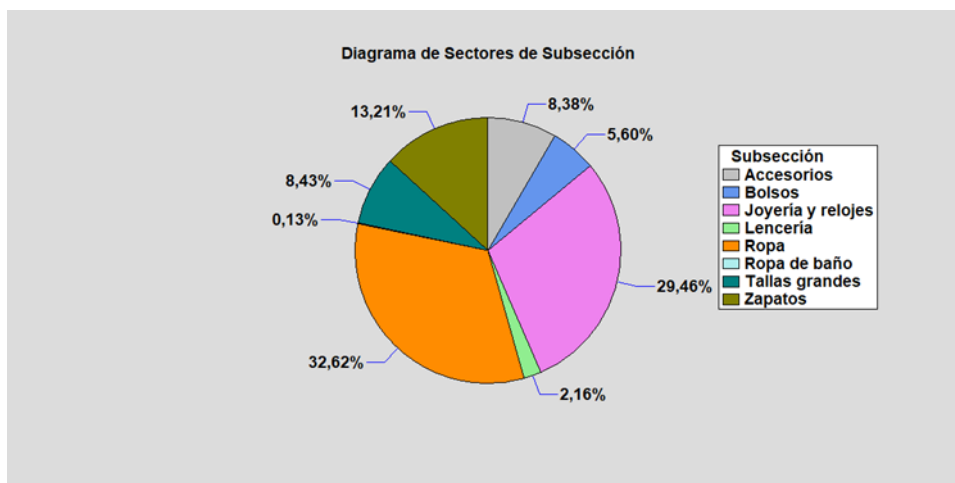
Gráfica 4: Diagrama de Sectores de la Sección

Clase	Valor	Frecuencia	Frecuencia relativa	Frecuencia acumulada	Frecuencia relativa acumulada
1	Hombre	833	0,1852	833	0,1852
2	Mujer	3664	0,8148	4497	1,0000

Tabla 1: Tabla de Frecuencia para Sección

ANÁLISIS UNIVARIANTE: SUBSECCION	
Tipo variable	Cualitativa nominal
Descripción	Dentro de la sección de Hombre y Mujer, el tipo de producto o subsección a la que pertenece
Cantidad de nulos	0
Recuento	4497
Categorías diferentes	8

Tal y como podemos observar en la Gráfica 5 y la Tabla 2, la subsección con mayor frecuencia es la de Ropa tal y como era de esperar. Por otro lado, llama la atención que la segunda subsección es Joyería y relojes, lo que inicialmente podría deberse a esa diferencia entre la cantidad de productos de Mujer y de Hombre. En el extremo opuesto llama la atención que solo hay 6 productos pertenecientes a la subsección Ropa de baño, lo que podría considerarse una subcategoría residual, ya que supone únicamente un 0,13% de los productos.



Gráfica 5: Diagrama de Sectores de la Subsección

Clase	Valor	Frecuencia	Frecuencia relativa	Frecuencia acumulada	Frecuencia relativa acumulada
1	Accesorios	377	0,0838	377	0,0838
2	Bolsos	252	0,0560	629	0,1399
3	Joyería y relojes	1325	0,2946	1954	0,4345
4	Lencería	97	0,0216	2051	0,4561
5	Ropa	1467	0,3262	3518	0,7823
6	Ropa de baño	6	0,0013	3524	0,7836
7	Tallas grandes	379	0,0843	3903	0,8679
8	Zapatos	594	0,1321	4497	1,0000

Tabla 2: Tabla de Frecuencia para Subsección

4.2. CONCLUSIÓN DEL ANÁLISIS UNIVARIANTE

Además de las conclusiones obtenidas para cada variable individualmente, creemos que es importante volver a tomar una perspectiva general sobre nuestros datos, teniendo ahora en cuenta la información obtenida en el análisis univariante. De esta forma, podemos volver a analizar los objetivos de nuestro trabajo y enfocar el análisis multivariante de la mejor forma posible.

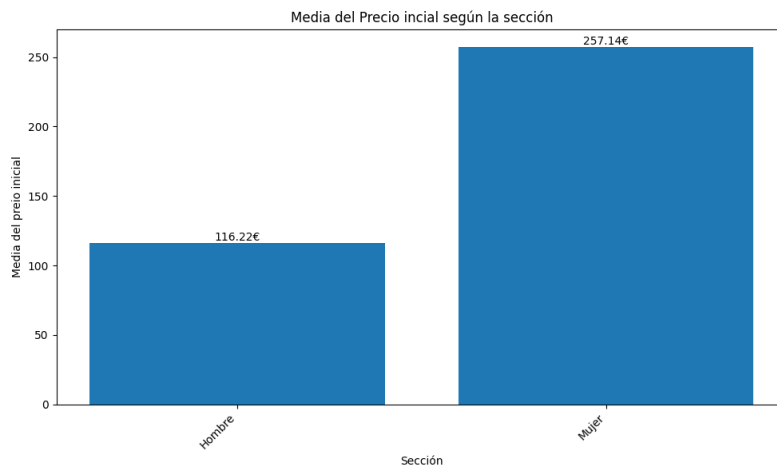
Ya que hemos observado que hay marcas que tienen mucha cantidad de productos y otras que tienen muy pocos, por lo que sería de interés analizar si las marcas que tienen muchos productos dan un menor descuento que aquellos que tienen un menor número de productos en descuento. Por otro lado, también hemos comprobado que el precio de los productos varía bastante. Por ello, tendremos que profundizar para intentar encontrar relaciones entre el precio y el descuento, ya que consideramos que aquellos productos con un mayor precio se podrán permitir hacer mayores descuentos que aquellos productos con un precio inferior, dado que los de mayor valor tendrán normalmente un mayor margen de beneficio. También hemos podido confirmar que la mayoría de los productos van dirigidos a Mujeres (análisis univariante de la sección)

tal y como suponíamos inicialmente. Es por ello que creemos que sería momento de comprobar si existe una relación entre la Sección y la Subsección, ya que los hombres tienen muy pocos productos, pero se podría dar el caso de que predominen sobre la sección Mujer en una de las Subsecciones.

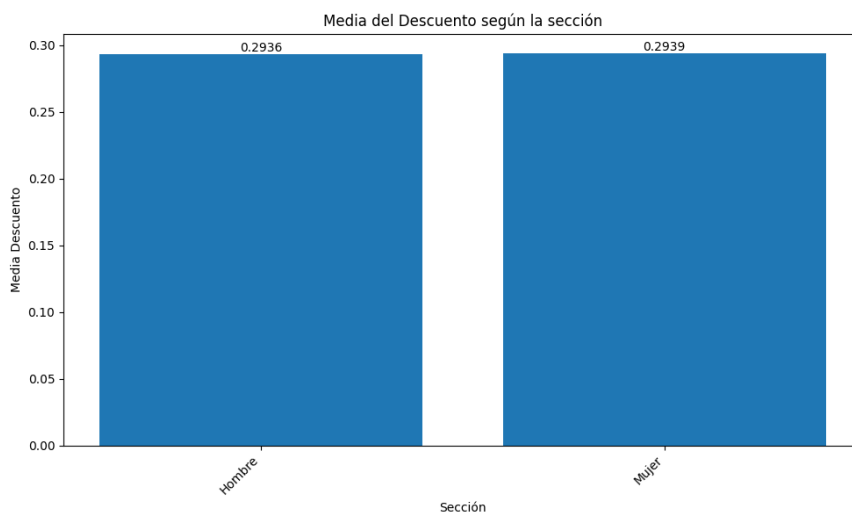
4.3. ANÁLISIS MULTIVARIANTE

4.3.1 PRECIO INICIAL Y DESCUENTO RESPECTO A SECCIÓN

Teniendo en cuenta nuestra hipótesis inicial de que las campañas de descuento de El Corte Inglés deben estar más enfocadas a mujeres que a hombres, procedemos a comparar el precio medio y el descuento medio de todos los productos según la Sección, de este modo podremos confirmar o desmentir nuestra hipótesis.



Gráfica 6: Diagrama de Barras del precio medio respecto de la sección



Gráfica 7: Diagrama de Barras del descuento medio respecto de la sección

Los gráficos muestran que, a pesar de que los productos de las mujeres son de media más caros que los de los hombres, el descuento medio (0,29) se mantiene casi exactamente igual en ambos casos, siendo contrario a la hipótesis inicial. Al no tratarse de un parámetro robusto, se ha relaizado el mismo cálculo con la mediana y se ha obtenido exactamente el mismo resultado. Esto es un valor orientativo ya que se trabajan con datos bastante generales, por lo que el resultado puede ser variable dependiendo del caso. Sin embargo, aunque a ambos se les aplica el mismo descuento, seguimos manteniendo nuestra hipótesis inicial, ya que el número de productos de mujeres es muy superior al de hombres. Los productos orientados a mujeres suponen el 81,5% del total de productos, como pudimos confirmar en el análisis univariante de la Sección.

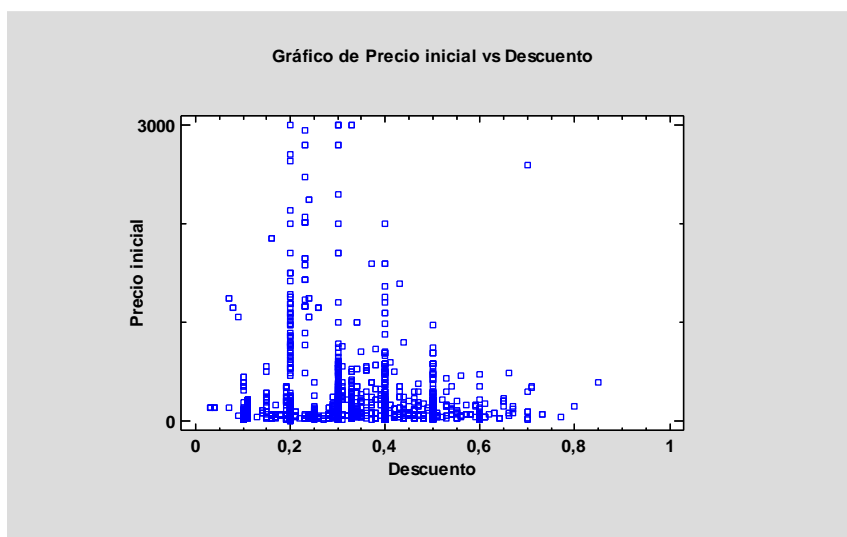
4.3.2 DESCUENTO SEGÚN PRECIO INICIAL

	Precio inicial	Descuento
Recuento	4497	4497
Promedio	231,037	0,293845
Mediana	64,9	0,3
Moda	59,95	0,3
Varianza	742687,	0,0188924
Desviación Estándar	861,793	0,13745
Coefficiente de Variación	373,011%	46,7763%
Desviación Media Absoluta	286,176	0,101448
DAM	30,1	0,1
Mínimo	5,99	0,03
Máximo	12999,0	0,85
Rango	12993,0	0,82
Rango Intercuartílico	72,01	0,14
Sesgo Estandarizado	228,105	15,7482
Curtosis Estandarizada	1136,76	0,6462

Tabla 3: Análisis del Precio inicial Frente al Descuento

Hemos hecho un análisis multivariante respecto las variables precio y descuento. Queríamos hacerlo para obtener una relación entre ambas. En un principio teníamos la hipótesis de que cuanto menor fuera el precio inicial menor sería el descuento, ya que si aplicamos un gran descuento a un producto barato el precio se reduce mucho. Tras hacer el análisis hemos observado que nuestra hipótesis era errónea. Al observar el gráfico de dos dimensiones X-Y, vemos que los productos que más descuento tienen son los que menor precio inicial tienen. Hay una tendencia inversa respecto al precio inicial y el descuento.

También cabe destacar que hay una gran acumulación de productos en los descuentos con números enteros múltiplos de 10 (como el 20%, 30% 40% 50% o 60%). Esto es para llamar más la atención del consumidor al ser números tan “redondos”.



Gráfica 8: Scatterplot del precio inicial frente al descuento

Realizando el coeficiente de relación de Pearson nos da $-2.23 \cdot 10^{-4}$. Esto quiere decir que tiene una ligera relación inversa. Nosotros pensábamos que habría más relación entre ambas variables aunque los datos nos dicen lo contrario.

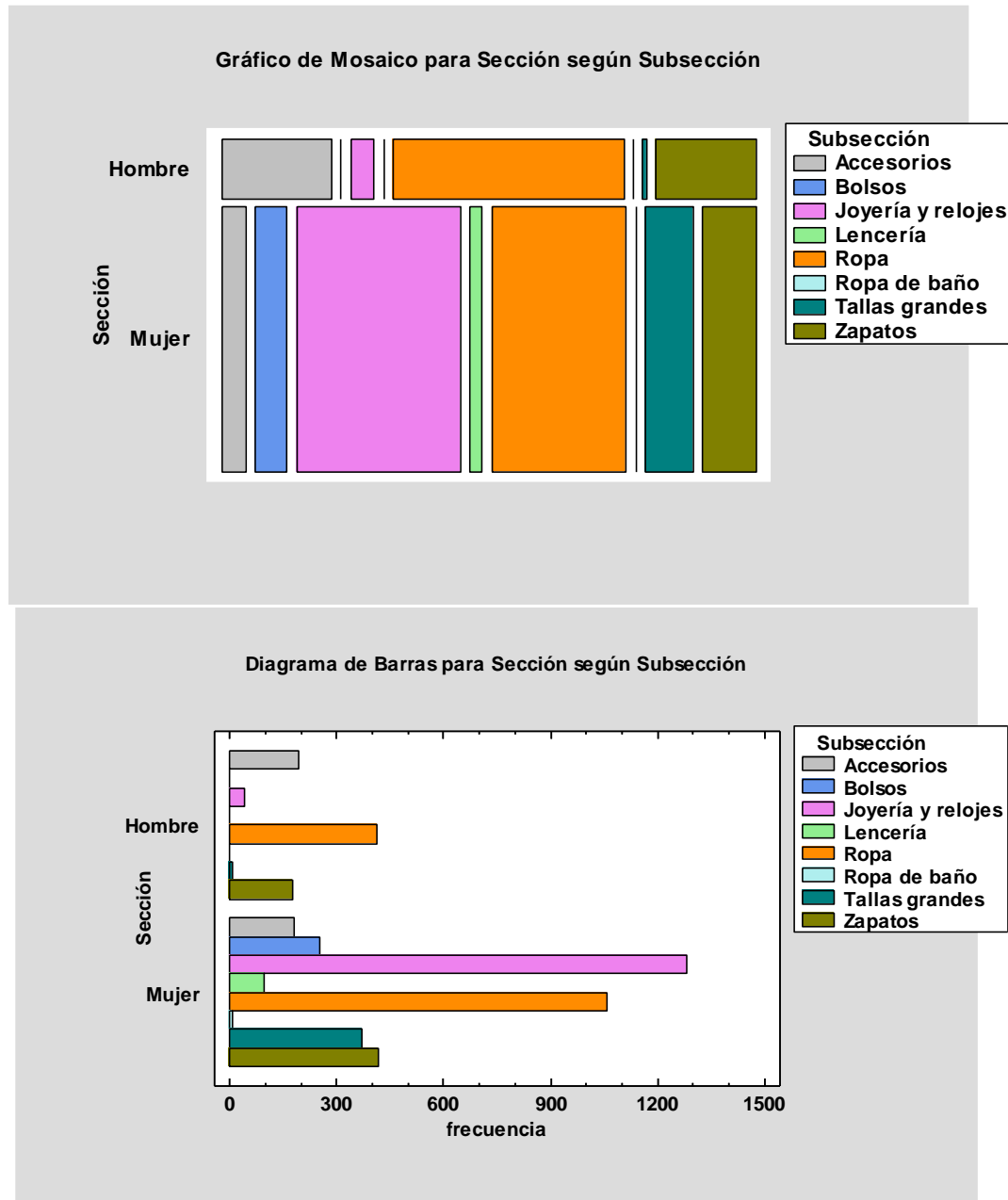
3.3.3 ANÁLISIS DE LA SUBSECCIÓN RESPECTO A LA SECCIÓN

	Accesorios	Bolsos	Joyería y relojes	Lencería	Ropa	Ropa de baño	Tallas grandes	Zapatos	Total por Fila
Hombre	195	0	42	0	411	0	7	178	833
	4,34%	0,00%	0,93%	0,00%	9,14%	0,00%	0,16%	3,96%	18,52%
Mujer	182	252	1283	97	1056	6	372	416	3664
	4,05%	5,60%	28,53 %	2,16%	23,48 %	0,13%	8,27%	9,25%	81,48%
Total por Columna	377	252	1325	97	1467	6	379	594	4497
	8,38%	5,60%	29,46 %	2,16%	32,62 %	0,13%	8,43%	13,21%	100,00%

Tabla 4: Tabla de frecuencias de las subsecciones agrupadas por secciones

Hemos hecho un análisis multivariante de las variables cualitativas sección y subsección. Respecto a este análisis queríamos observar cuantas subsecciones había respecto a cada sección de género o en que secciones se realizaban más rebajas en función del género. Es un análisis muy amplio donde se pueden sacar muchas conclusiones. En un principio teníamos la hipótesis de que había muchas más subsecciones de mujer que de hombre. Una vez hecho, lo hemos ratificado. En los hombres destacan solo tres subsecciones: accesorios, ropa y zapatos. En las mujeres

tenemos mucha más variedad en cuanto secciones y los productos están más repartidos pudiendo destacar hasta seis subsecciones: accesorios, bolsos, joyería y relojes, ropa, tallas grandes y zapatos. A través de esto nos damos cuenta de que las mujeres tienden a comprar muchos más productos de diferentes tipos que los hombres. Las subsecciones de bolsos, lencería y ropa de baño no existen para hombre.

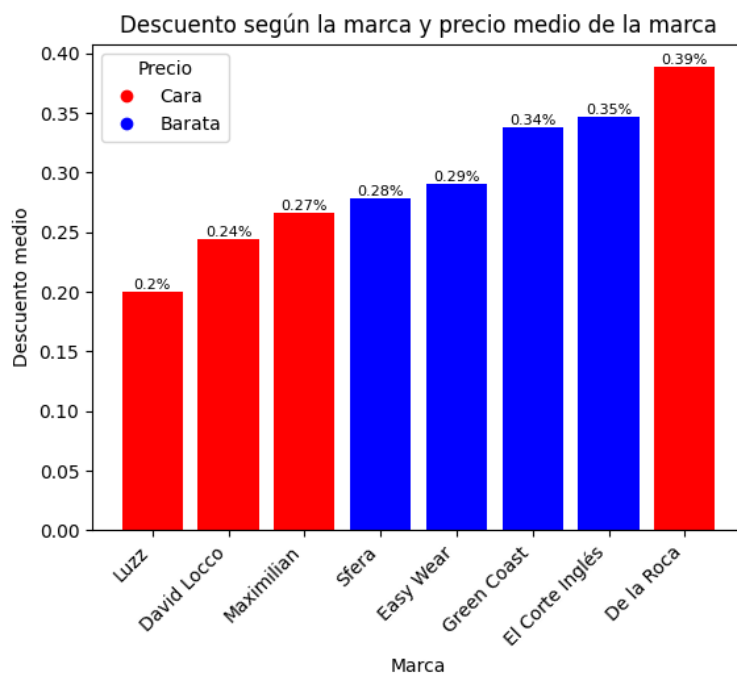


Gráficas 9 y 10: Descripción del contenido

Respecto a las tres subsecciones de hombre, hay más productos dentro de estas en mujer, pero en relación y la cantidad en hombres es mucho más grande.

3.3.4 ANÁLISIS DEL DESCUENTO RESPECTO A LA MARCA

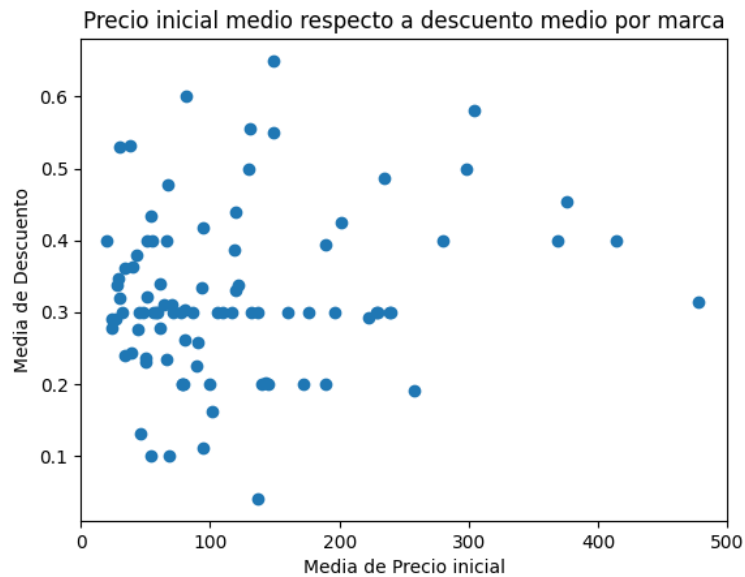
Se han estudiado en conjunto las variables de la marca del producto y su precio de descuento con el fin de encontrar una relación. Al contar con un número tan elevado de marcas, se ha optado por la reducción (para este análisis exclusivamente) a las 4 con los productos más caros y a las 4 con los más baratos. En este caso, contaríamos con las marcas *Maximilian*, *David Locco*, *De la Roca*, *Luzz*, *Sfera*, *El Corte Inglés*, *Easy Wear* y *Green Coast*.



Gráfica 11: Descuento medio de las marcas con mayor y menor media en Precio Inicial (cata y barata)

Se han obtenido los descuentos medios de las marcas mencionadas anteriormente y, como se puede observar en el anterior gráfico, es mayoritario que las marcas con precios más elevados cuenten con un porcentaje medio de descuentos menor al de las marcas más baratas. Esto puede mostrar el enfoque de las marcas más accesibles a tratar de vender los máximos productos a las personas con menor nivel económico, quienes conforman la mayoría de la sociedad, teniendo por ende más beneficio.

Sin embargo, al aumentar el número de datos con los cuales se forma el gráfico se puede observar que no existe ninguna relación, siendo el coeficiente de correlación 0,05. Esto se ha obtenido con datos orientativos, es decir, con el precio medio y descuento medio de todos los productos agrupados por las marcas estudiadas.

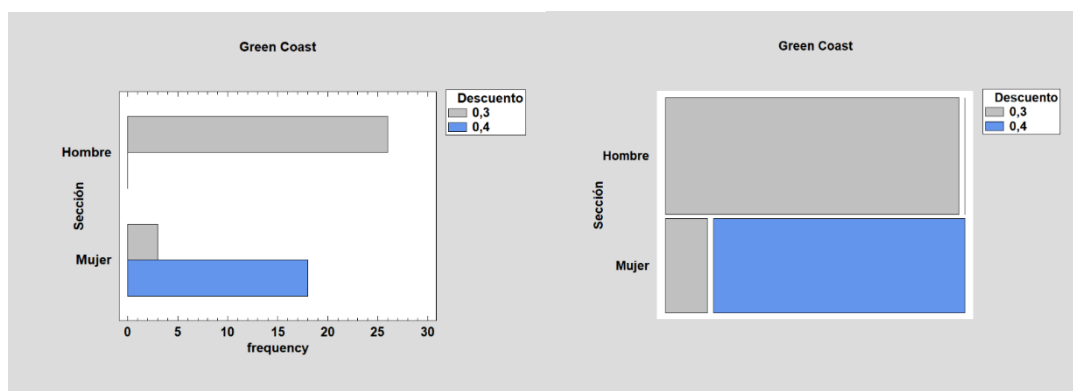


Gráfica 12: Scatterplot del descuento medio respecto al precio inicial medio por Marca

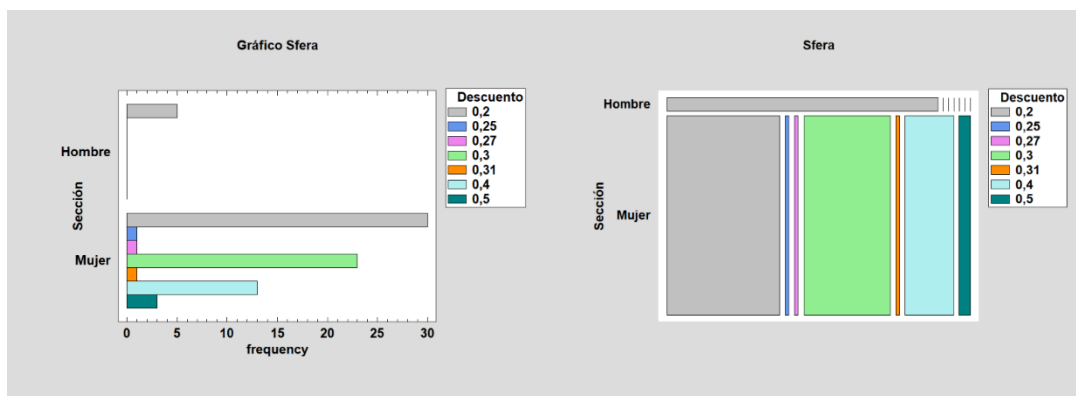
3.3.5 DESCUENTO RESPECTO A LA MARCA Y SECCIÓN

Para el siguiente análisis multivariante se ha analizado la posible relación entre las variables de marca, descuento y género con el fin de identificar si, dependiendo de la marca de los productos y del precio de estos, se enfoca la venta más hacia los hombres o las mujeres. De nuevo, y como con el caso anterior, al contar con una cantidad tan grande de datos se ha optado a la reducción de marcas a las 4 con productos más caros y a las 4 con los más baratos, siendo estas *Maximilian*, *David Locco*, *De la Roca*, *Luzz*, *Sfera*, *El Corte Inglés* y *Green Coast*.

Con el conjunto de datos se han obtenido tan solo dos gráficos que reflejen la relación a estudiar:

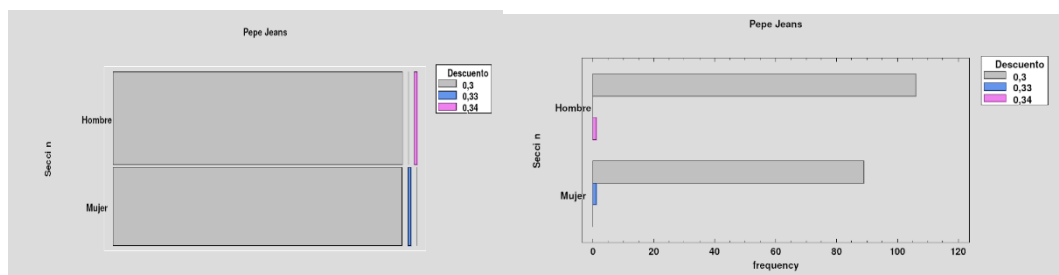


Gráficas 13 y 14: Descripción del contenido



Gráficas 15 y 16: Descripción del contenido

Como se puede observar en ambos gráficos, o bien se hacen más descuentos para los productos de mujer o bien los descuentos son más grandes. Respecto al resto de marcas, tan solo hacen productos para mujer, luego de esto podemos obtener dos conclusiones; hay una mayoría de marcas que se enfocan tan solo en productos para mujeres, y que, en relación también con el análisis anterior, los descuentos se centran en las marcas más accesibles y, a su vez, en las mujeres.

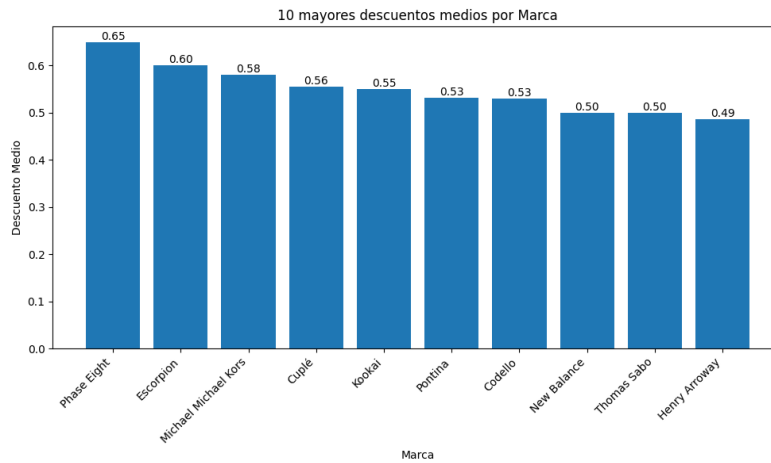


Gráficas 17 y 18: Scatterplot. Número de productos frente al descuento medio por marca

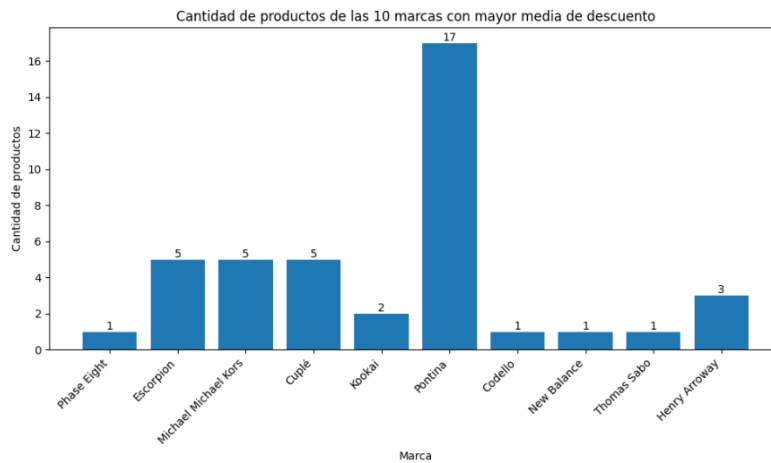
Sin embargo, al estudiar una muestra tan pequeña del conjunto de datos, también se pueden encontrar excepciones, como es el caso de la marca *Pepe Jeans*, en la cual hay un descuento mayoritario para los hombres. Por ello, sería necesario un estudio más extenso de todas las marcas para poder obtener conclusiones válidas.

3.3.6 DESCUENTO RESPECTO A LA CANTIDAD DE PRODUCTOS DE UNA MARCA

El último análisis multivariante realizado trataba de comprobar la relación entre los productos que tiene una marca y los descuentos que da esa marca. Ya que tenemos una cantidad elevada de productos diferentes y de descuentos, decidimos inicialmente buscar relaciones entre la cantidad de productos por marca, la media de descuentos por marca. Ya que seguimos teniendo una elevada cantidad de marcas diferentes, representamos a continuación gráficas de las 10 marcas con mayor descuento medio y las 10 marcas con menor descuento medio.

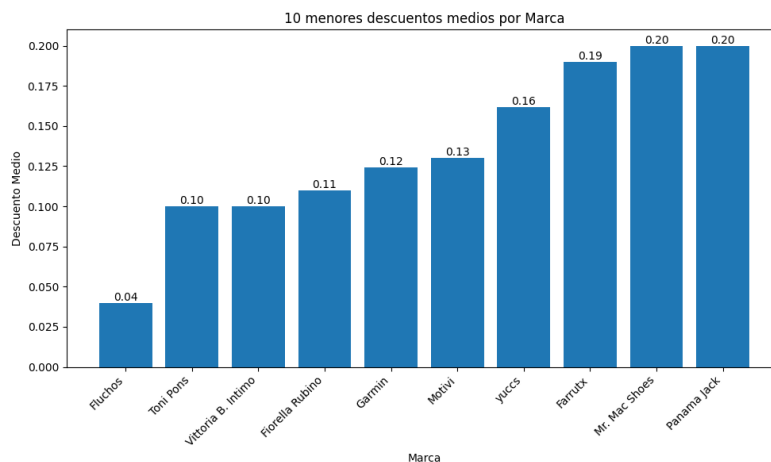


Gráfica 19: Marcas con mayor descuento medio en sus productos

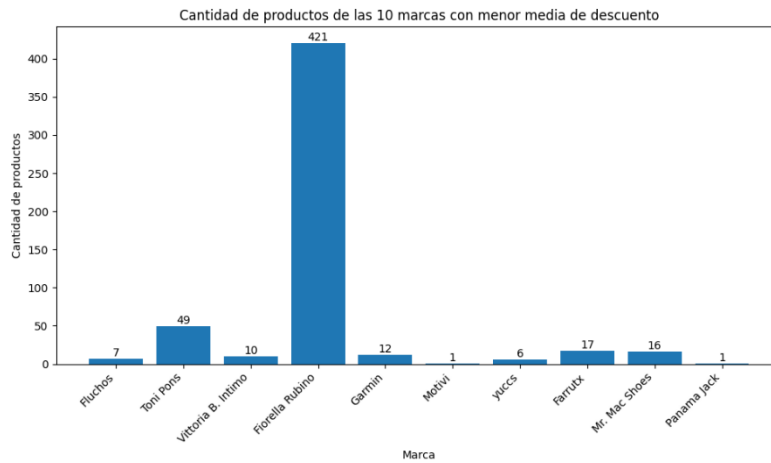


Gráfica 20: Cantidad de productos disponibles, de las 10 marcas con mayor descuento medio (Gráfica 14)

Analizando las gráficas 19 y 20, parece que las marcas que dan descuentos muy elevados son aquellas que tienen una cantidad de productos en descuento muy baja.



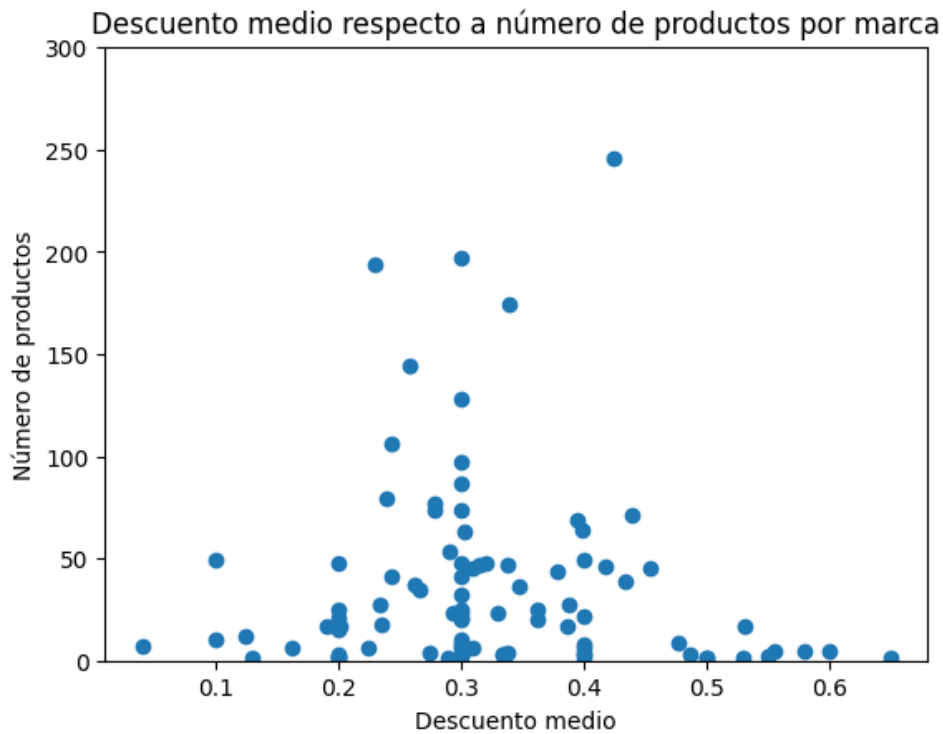
Gráfica 21: Marcas con menor descuento medio en sus productos



Gráfica 22: Cantidad de productos disponibles, de las 10 marcas con menor descuento medio (Gráfica 16)

Analizando las gráficas 21 y 22, parece que las marcas que dan descuentos muy bajos son aquellas que tienen una cantidad de productos en descuento más elevada que los casos anteriores. Se podría plantear la hipótesis de que el descuento medio de una marca es inversamente proporcional a la cantidad de productos que tiene esta marca en descuento. Sin embargo, consideramos que el análisis realizado hasta ahora no es suficiente para poder confirmar esta hipótesis.

Es por ello que procedemos a mostrar un Scatterplot en que se ve reflejada la relación entre el descuento medio de una marca y la cantidad de productos que tiene esta marca (Gráfica 23).



Gráfica 23: Scatterplot. Número de productos frente al descuento medio por marca

Finalmente, analizando esta gráfica observamos que no se puede afirmar que el número de productos y el descuento medio de una marca sean linealmente dependientes. Además, extrayendo el coeficiente de correlación de -0.10 por lo que confirmamos que no existe una correlación entre ambas variables.

5. CONCLUSIONES

Tras haber realizado un análisis univariante y multivariante de nuestros datos y haber representado la relación entre variables procedemos ahora a concluir el trabajo comprobando las hipótesis iniciales que habíamos planteado al inicio del trabajo.

Considerábamos al principio la posibilidad de las mujeres compren mucho más frecuentemente que los hombres, por lo que las campañas de descuento estarían dirigidas a este público objetivo. Se ha podido comprobar en el análisis univariante que la mayoría de los productos que se encuentran en descuento están dirigidos a mujeres, aunque en el análisis multivariante encontrábamos que el descuento que se le aplica de media a los productos de hombres y mujeres es similar. Por ello, confirmamos nuestra hipótesis inicial sobre la diferencia entre la cantidad de productos de mujeres y de hombres, aunque debemos tener en cuenta que el descuento aplicado a ambos géneros es el mismo.

Por otro lado, pensábamos inicialmente que los hombres tendrían una menor presencia en la subsección de Ropa y mayor en la de Accesorios. Tenemos que ser conscientes que al tener las mujeres una cantidad muy superior de productos que los hombres, en todas las secciones encontraremos mayor cantidad de productos de mujeres que de hombres. Por ello se debe analizar la cantidad de productos presentes en cada sección de forma relativa a la sección. De esta forma se ha comprobado en las gráficas 9 y 10, que la cantidad relativa de Ropa para la sección Mujer es mayor que la cantidad relativa de Ropa para la sección Hombre y que pasa exactamente lo contrario con la Subsección de Accesorios, donde predomina de forma relativa sobre los productos dirigidos a hombres.

Seguidamente, otra de las hipótesis iniciales era que a mayor el precio de un producto, mayor sería su descuento. Sin embargo, como se puede observar en el gráfico 8 se muestra una ligera relación inversa, es decir, que hay varios casos en los cuales el precio inicial era considerablemente bajo respecto al resto de productos y aun así su descuento era mayor que los productos que contaban con un precio inicial mucho más elevado, y lo mismo sucedía de forma contraria.

Además, cabe resaltar que, en el último análisis multivariable, el cual relaciona la cantidad de productos de una marca respecto a su descuento medio, y concretamente el gráfico 23, muestra que nuestra hipótesis inicial sobre la relación entre la cantidad de productos que tiene una marca y los descuentos que esta se puede permitir no es correcta.

Finalmente, haciendo este trabajo hemos conseguido afianzar los conocimientos adquiridos en la asignatura y aplicarlos a un caso práctico real en el que hemos podido sacar conclusiones analizando los datos de una empresa real.