

# CHAMPO CARPETS

CASO HARVARD

MBA - BUSINESS ANALYTICS

MACARENA HERRERA  
MERCEDES IGLESIAS  
MANUEL LOPEZ  
B. CAROLINA RIVERO



## ÍNDICE

---

**01**

Introducción

**02**

El Problema

**03**

Datos

**04**

Análisis Descriptivo

**05**

Análisis Diagnóstico

**06**

Análisis Predictivo

**07**

Discusión Propuestas

# Introducción

Champo inició su andadura con alfombras tejidas a mano. A lo largo de los años, han ido construyendo su negocio utilizando técnicas de tejido manual y procesamiento a máquina. Con 1,2 millones de pies cuadrados de infraestructura en varias partes de la India como Bhadohi, Panipat y Noida, y con una red de alrededor de 12.500 artesanos.

Con una experiencia de más de 53 años, es uno de los proveedores de marcas como Anthropologie, Target, John Lewis, Urban Outfitters o Joann.

Comprometidos con el medio ambiente, Champo promete intentar reducir su impacto ambiental desde la decisión de los hilos adquiridos hasta el proceso que siguen y las personas que contratan.



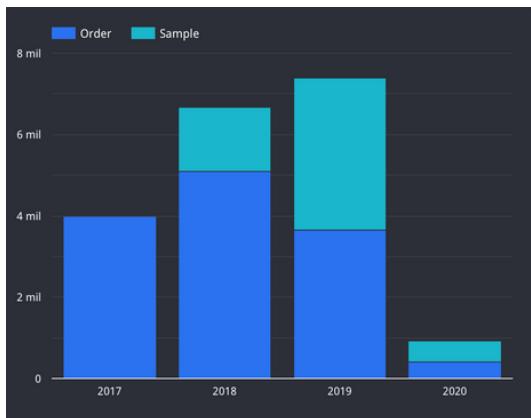
# El Problema

El caso de Champo Carpets, una empresa de fabricación de alfombras en la India, presenta un desafío en el que busca mejorar la selección de muestras de alfombras, utilizar el inventario de manera más eficiente y predecir las conversiones de muestras enviadas a los clientes. El objetivo principal es utilizar datos del sistema ERP para identificar las preferencias de los clientes y desarrollar modelos que les permitan mejorar la selección de diseños de muestras de alfombras. Además, la empresa considera la posibilidad de utilizar los datos de inventario para impulsar la venta de productos específicos, optimizando así la utilización de recursos y aumentando la eficiencia en la producción y venta de alfombras. Para lograr estos objetivos, se plantea realizar un análisis descriptivo de las conversiones por producto, cliente y país, implementar un modelo de Machine Learning para predecir las conversiones y poder recomendar productos a los clientes en base a sus preferencias pasadas y tendencias. Este caso presenta un enfoque innovador en el uso de datos y modelos para mejorar las operaciones y la rentabilidad en la industria de fabricación de alfombras.

Champo Carpets ha enfrentado dificultades significativas en la conversión de muestras de alfombras enviadas a los clientes, lo que ha impactado negativamente en su rentabilidad y en la eficiencia de su producción. La baja tasa de conversión ha generado interrupciones en la programación de la producción, ya que la empresa ha tenido que destinar recursos considerables a la producción de muestras. En 2019, la tasa de conversión de Champo Carpets fue de alrededor del 20%, en comparación con el 35% prevaleciente en la industria, lo que afectó tanto la rentabilidad como la programación de la producción.



# Datos



Para mejorar la relación con los clientes, en 2018 empezaron a mandar muestras.

En el año 2018, Champo Carpets experimenta un aumento de ventas y mandaron algunas muestras. Sin embargo, en 2019 bajaron exponencialmente los pedidos y la cantidad de muestras aumentó, lo que supondría un gran problema

La gráfica del **monto por cliente** representa los ingresos generados por la venta de las alfombras. Esta cantidad tiene en cuenta la cantidad de metros cuadrados vendidos.

Podemos ver que el cliente **TGT** es el que más ha gastado, pero al compararlo con el número de pedidos realizados, este sin embargo no es tan grande, por lo que este cliente compra alfombras de gran tamaño.



Por otro lado, fijándonos en la gráfica de pedidos vs muestras por cliente, hay algo muy llamativo, el cliente registrado como CC tiene muy pocos pedidos, y muchas muestras. Además, es el cliente de India, por lo que nos puede llevar a entender que pueden ser ellos mismos, CC, de Champo Carpets), por lo que sería necesario preguntarles para comprender la manera en la que registran sus datos y ver si es la correcta. Pese a haber identificado eso, una vez se ha mencionado, es importante aclarar que el análisis de aquí en adelante se realizará incluyendo los datos de CC.

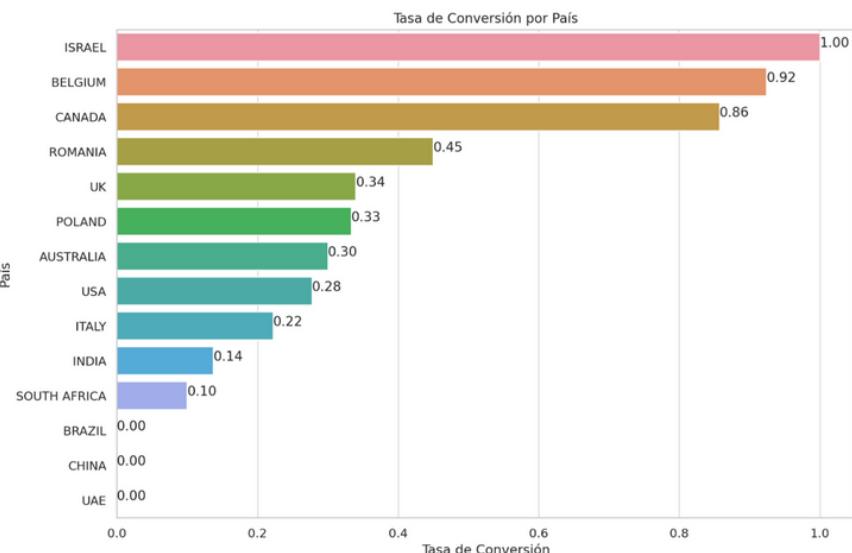
# Análisis Descriptivo

## CLIENTES

Cada cliente tiene asociado un código con el que se le identifica. El total de clientes es 34, de los cuales, menos USA, que tiene 20 clientes, y UK, que tiene 2, los demás solo tienen un cliente.

## CONVERSIÓN POR PAÍS

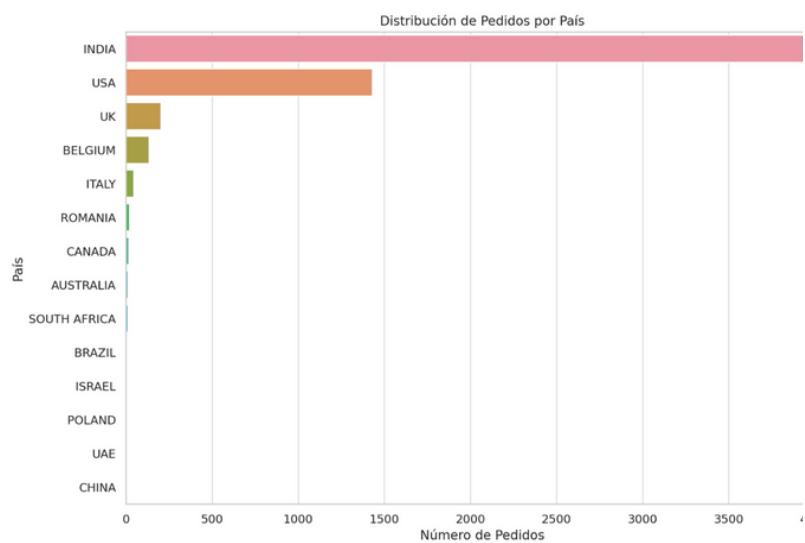
Israel es el país con mayor conversión. Sin embargo, podemos ver que realmente solo ha hecho 6 compras, las cuales en comparación con el total de 11.494, hace que no sea muy significativo. Teniendo en cuenta esto, podemos observar que el país con mayor conversión es Bélgica.



Si nos fijamos en la cantidad de compras, observamos que los países con más compras son USA e India.

Deberíamos hacer un análisis exhaustivo de lo que ocurre en India, ya que su único cliente, CC. En este caso entendemos que podría tratarse de la propia empresa Champo Carpets, y tengan así categorizado como envío de muestras lo que quizás se traten de devoluciones.

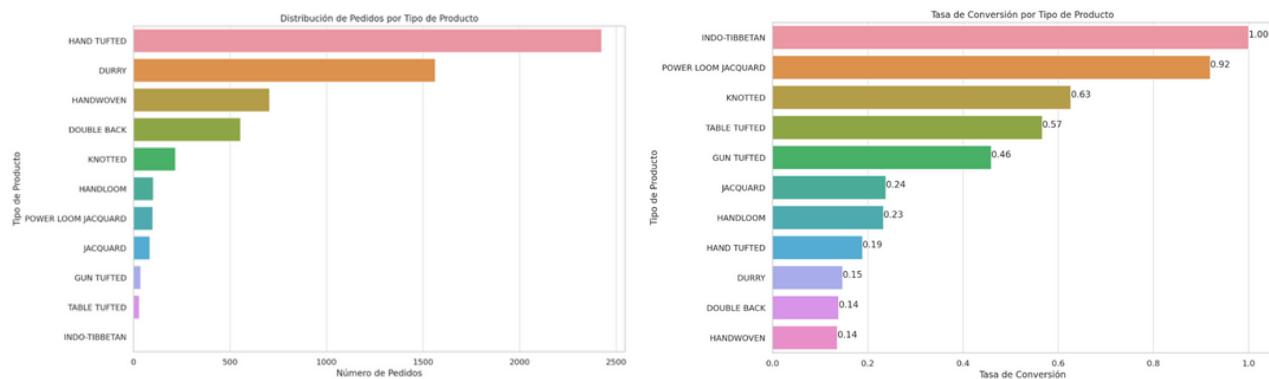
Es por ello que nuestro foco estará puesto en el mercado de USA.



## PRODUCTO

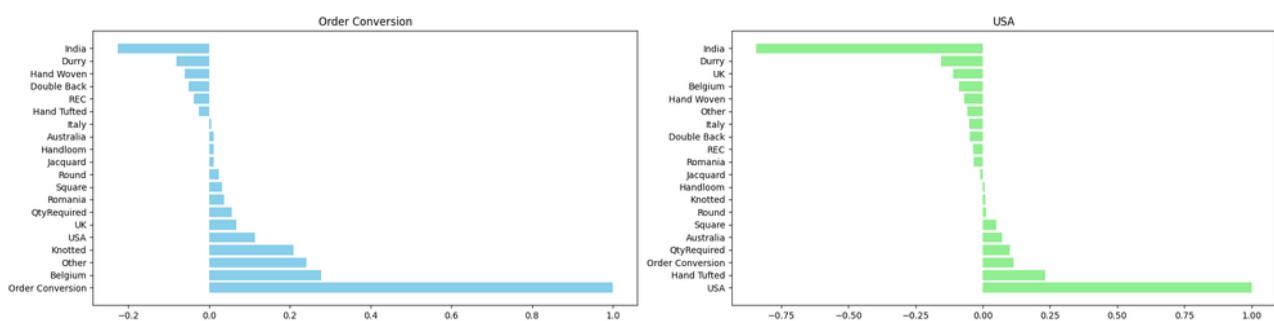
A la hora de analizar una empresa y tomar decisiones, es importante conocer cuáles son los productos con mayor éxito entre sus clientes.

Por un lado, si consideramos las alfombras con mayor número de pedidos han sido, por orden, las alfombras **Hand Tufted**, seguida de las **Durry**, y en tercer lugar **Handwoven**.



# Análisis Diagnóstico

## CORRELACIÓN DE PEARSON

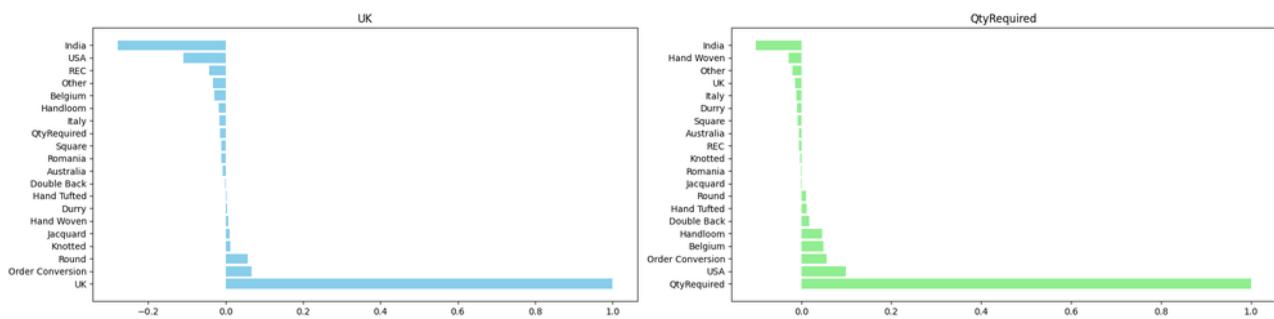


**Gráfico de Order Conversión:**

- **Correlaciones Negativas:** Se observan en India, con los modelos Durry, Hand Woven y Double Back, sugiriendo menores conversiones con estos modelos en India.
- **Correlaciones Positivas:** En Bélgica, con las categorías 'Other', Knotted, USA y UK, indicando mayores conversiones con estos modelos y países.

**Gráfico de Coeficientes de Correlación de USA:**

- Se notan tanto correlaciones negativas como positivas. En USA, hay mayores envíos de modelos Hand Tuften y menores en los modelos Durry y Hand Woven.



**Gráfico de Coeficientes de UK:**

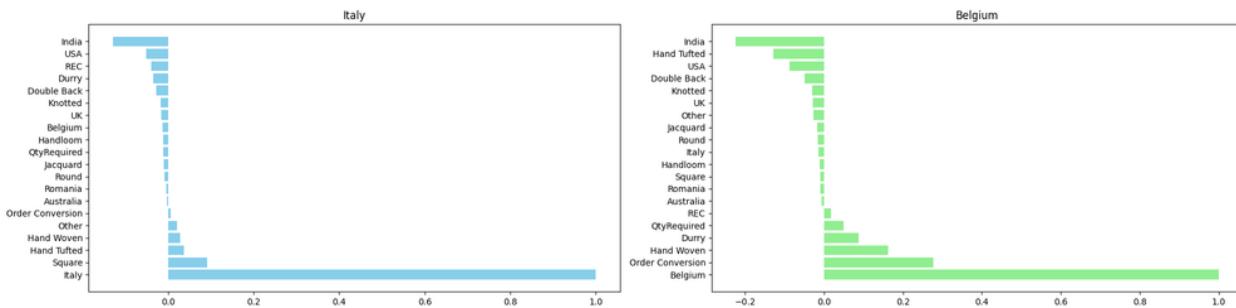
- Correlaciones: Negativas con alfombras rectangulares y positivas con alfombras redondas, sugiriendo una preferencia por estas últimas.

## Coeficientes de QtyRequired:

- USA y Bélgica presentan las mayores cantidades requeridas, especialmente en el modelo Handloom. Por el contrario, el modelo Hand Woven muestra correlaciones negativas.

# Análisis Diagnóstico

## CORRELACIÓN DE PEARSON

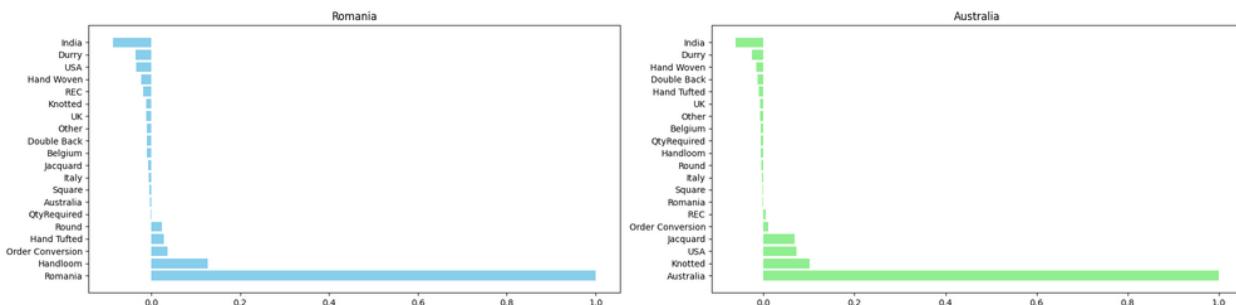


### Coeficientes de Italia:

- Correlaciones negativas con alfombras rectangulares, Durry y Double Back, y positivas con alfombras cuadradas.

### Coeficientes de Bélgica:

- Correlaciones positivas con modelos Hand Woven y Durry, y negativas con Hand Tuften y Double Back.



### Coeficientes de Rumanía:

- Correlaciones negativas con Durry y positivas con Handloom.

### Coeficientes de Australia:

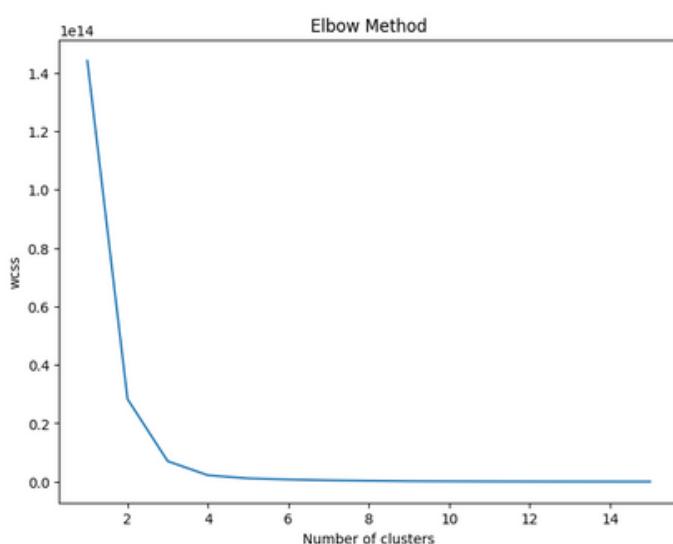
- Alta negatividad con el modelo Durry, mientras que las positivas se inclinan hacia los modelos Knotted y Jaquard.

**Nota:** Se ha decidido omitir el análisis de India debido a la singularidad del único cliente, presumiblemente la misma empresa Champo Carpets.

# Análisis Predictivo

## CLUSTERING & CLASIFICACIÓN

Según los resultados del *clustering*, podemos segmentar a los clientes de *Champo Carpets* en 3 categorías. Esto significa que hay tres grupos de clientes con características distintas en cuanto a sus compras de alfombras. Esta clasificación predictiva de los grupos se ha realizado utilizando el algoritmo **k-means clustering (método del codo)**.



Según los resultados de la gráfica superior (método del codo), el número óptimo de agrupaciones es tres.

Utilizando el algoritmo **k-means clustering**, la empresa puede obtener conocimientos sobre la distribución de sus clientes y ayudará a la empresa a formular mejores estrategias apuntando a diferentes segmentos. A continuación, se muestra el código utilizado para realizar la división de grupos en dicha predicción:

```
numClusters = 3

# Fitting kmeans to the dataset
km = KMeans(n_clusters=numClusters, init='k-means++', max_iter=300, n_init=10,
random_state=0)
y = km.fit_predict(x)

# Number of clients per cluster
print("cluster - clients in cluster")
for c in range(0,numClusters):
    print(c, (y==c).sum())

print('\n Cluster centroids ----- ')
print(pd.DataFrame(km.cluster_centers_,columns=df.columns[1:]).transpose())
print('\n')
```

# Análisis Predictivo

## CLUSTERING & CLASIFICACIÓN

```
# Clients in clusters
for c in range(0,numClusters):
    print('Cluster ',c)
    print( df.values[y==c,0] )
```

Con este código ejecutado se obtiene el siguiente resultado:

Número de clientes  
en cada cluster:

```
cluster - clients in cluster
0 39
1 1
2 5
```

```
Cluster centroids -----
          0           1           2
Sum of QtyRequired 6.323641e+03 1.504500e+04 6.446460e+04
Sum of TotalArea   4.561106e+03 3.763033e+04 7.440661e+04
Sum of Amount      2.076430e+05 1.134105e+07 2.396063e+06
DURRY             2.896949e+03 0.000000e+00 4.133080e+04
HANDLOOM          8.853846e+01 0.000000e+00 9.792000e+02
DOUBLE BACK       2.064359e+02 0.000000e+00 2.061000e+03
JACQUARD          6.017949e+01 0.000000e+00 3.354000e+02
HAND TUFTED       2.616667e+03 1.504500e+04 9.438600e+03
HAND WOVEN        2.869487e+02 0.000000e+00 5.571000e+03
KNOTTED           6.958974e+01 0.000000e+00 2.749000e+03
GUN TUFTED        4.384615e+00 0.000000e+00 3.900000e+01
Powerloom Jacquard -5.684342e-14 0.000000e+00 1.950600e+03
INDO TEBETAN     8.205128e-01 0.000000e+00 -1.110223e-16
```

En la imagen de la izquierda observamos los centroides de cada grupo, es decir, el centro de los grupos (valor de la característica en el centroide de cada grupo).

En la siguiente imagen, se detallan por *cluster* los distintos clientes que pertenecen a cada grupo.

```
Cluster 0
['A-11' 'A-6' 'B-2' 'B-3' 'B-4' 'C-1' 'C-3' 'CC' 'CTS' 'DR' 'E-2' 'F-1'
 'F-6' 'G-1' 'G-4' 'H-1' 'I-2' 'JL' 'K-2' 'K-3' 'L-2' 'L-3' 'L-4' 'L-5'
 'M-2' 'N-1' 'P-4' 'PC' 'PD' 'R-4' 'RC' 'S-2' 'S-3' 'T-2' 'T-4' 'T-5'
 'T-6' 'T-9' 'V-1']
Cluster 1
['TGT']
Cluster 2
['A-9' 'C-2' 'H-2' 'M-1' 'P-5']
```

Beneficios del *clustering*:

- Identificar clientes similares y enviar muestras en consecuencia, para garantizar la conversión de pedidos después de las pruebas de muestra.
- A continuación, se muestra una clasificación de los países que pertenecen a cada *cluster*.

# Análisis Predictivo

## CLUSTERING & CLASIFICACIÓN

Cluster	Country	Clientes
0	USA, China, India, Australia, Canadá, Israel, Sudáfrica, Rumania, Reino Unido, Brasil, Bélgica, Emiratos Árabes Unidos, Líbano, Italia	A-11, A-6, B-2, B-3, B-4, C-1, C-3, CC, CTS, DR, E-2, F-1, F-6, G-1, G-4, H-1, I-2, JL, K-2, K-3, L-2, L-3, L-4, L-5, M-2, N-1, P-4, PC, PD, R-4, RC, S-2, S-3, T-2, T-4, T-5, T-6, T-9, V-1
1	USA	TGT
2	USA	A-9, C-2, H-2, M-1, P-5

### Características de los *clusters*:

El *cluster* 0 tiene 39 clientes, el *cluster* 1 tiene 1 solo cliente y el *cluster* 2 tiene 5 clientes. Esto indica que el *cluster* 0 es el más numeroso en cuanto al número de clientes y el que más contribuye al volumen de ventas de la empresa, mientras que el *cluster* 1 es el más pequeño y el que menos contribuye.

Los *clusters* se diferencian por el tipo de alfombra que compran, que se puede clasificar en diez categorías: "DURRY", "HANDLOOM", "DOUBLE BACK", "JACQUARD", "HAND TUFTED", "HAND WOVEN", "KNOTTED", "GUN TUFTED", "Powerloom Jacquard" y "INDO TEBETAN". Cada *cluster* tiene una preferencia por uno o varios tipos de alfombra, lo que puede reflejar el gusto, el poder adquisitivo o la necesidad de los clientes.

- *Cluster* 0:
  - Un grupo donde los clientes se centran más en la compra focalizada.
  - Este grupo se centra principalmente en dos tipos de alfombra, "DURRY" y "HAND TUFTED", ya que el valor de la característica en el centroide de estos dos grupos es el más elevado.

# Análisis Predictivo

## CLUSTERING & CLASIFICACIÓN

- Sin embargo, el tipo de alfombra del que menos muestras se compran es “Powerloom Jacquard”. Sin duda esto puede estar relacionado con que el cliente que más muestras pide es CC (cliente del *cluster 0*), y por lógica al ser un cliente de la India no es muy común este tipo de alfombras (“Powerloom Jacquard”) en este país, y por ello el poco volumen de muestras compradas de este tipo de alfombra, ya que el cliente con mayor peso de este grupo es CC como bien se ha explicado en el apartado de Datos del informe.
- *Cluster 1:*
  - Solo un cliente (TGT) de EEUU que se centra en un solo tipo de alfombra, en concreto, la de tipo “HAND TUFTED”.
  - El *cluster 1* tiene el mayor valor de ingresos generados por pedido de los tres *clusters* (“Sum of Amount”), con 1.134105e+07 USD, pero esto se debe a que solo hay un cliente en este *cluster*, que se dedica exclusivamente a la compra de muestras de un solo tipo de alfombra. Como ya se mencionó anteriormente, el cliente TGT es el que más ha gastado, pero al compararlo con el número de pedidos realizados (gráfica del apartado de Dato), este sin embargo no es tan grande, por lo que este cliente compra alfombras de gran tamaño.
- *Cluster 2:*
  - Un grupo donde los clientes se centran en la compra más variada.
  - El *cluster 2* tiene el mayor valor de cantidad pedida (“Sum of QtyRequired”) y de área total (“Sum of TotalArea”), lo que indica que los clientes compran una gran variedad y cantidad de alfombras.
  - El tipo de alfombra que más compran con diferencia es el tipo “DURRY”, mientras que, el que menos muestras piden es el tipo “INDO TEBETAN”.

# Análisis Predictivo

## CLUSTERING & CLASIFICACIÓN

### Clasificación

Una vez analizada la predicción realizada por el algoritmo de *clustering*, es necesario hacer una clasificación de los datos entre lo que ha predicho el algoritmo y lo que realmente se obtiene como resultado aplicando el clasificador de regresión logística. Con este clasificador obtenemos la siguiente matriz de confusión:

	^ NC	^ C
NC	2247	68
C	393	183

^ NC: No Convieren (no compran una vez enviadas las muestras)

^ C: Convieren (compran una vez enviadas las muestras)

Para la construcción de este clasificador se han dividido los datos en datos para entrenamiento y datos para testear (50% y 50%), es decir, se ha procedido a entrenar el clasificador con la mitad de los datos, y con la otra mitad a evaluar.

En primer lugar, de las 2891 muestras que se han enviado, 2315 de ellas no convierten (no se compraron), mientras que 576 sí convierten (se compraron). De estas 2315 muestras que no convierten, el algoritmo clasifica las muestras como muestras que no van a convertir a la mayoría, ya que se trata de 2247 muestras predichas por el algoritmo que no convierten de las que realmente no convertían que son 2315, por tanto el algoritmo anda razonablemente bien, con una proporción del 97%.

El objetivo es realmente predecir la conversión de las muestras, por tanto para el caso de las muestras que convierten, el algoritmo no está muy acertado, ya que de 576 muestras que convierten, solo 183 son predichas por el algoritmo como que van a convertir, es decir, solo somos capaces de predecir con el algoritmo esas 183 muestras, porque el resto, las 393 que sí convierten las clasifica como que no se convierten.

Del total de las muestras enviadas (2891 muestras), solo un 20% de las muestras han convertido. Y dentro de la estimación del algoritmo de muestras que convierten, 251, de lo que se había predicho de que se iban a convertir, se ha cumplido el 73%.

# Discusión Propuestas

## DESAFÍO EN LA CONVERSIÓN DE MUESTRAS:

- La baja tasa de conversión en 2019 (20%) en comparación con la industria (35%) señala un desafío crucial en el proceso de selección de muestras.
- La empresa enfrenta interrupciones en la programación de producción debido a la asignación de recursos significativos a la fabricación de muestras con bajos índices de conversión

## SEGMENTACIÓN DE CLIENTES

- El análisis de clustering identifica tres grupos de clientes con preferencias distintas, destacando la necesidad de estrategias personalizadas para cada segmento.
- El mercado de USA lidera, pero la singularidad del cliente CC en India sugiere la necesidad de investigar y adaptar estrategias específicas.

## PRODUCTOS CON MAYOR DEMANDA

- Las alfombras Hand Tufted son las más demandadas, seguidas por Durry y Handwoven.
- Se observan correlaciones positivas y negativas entre productos y países, indicando la necesidad de estrategias adaptadas para cada mercado.

## INVESTIGACIÓN DEL MERCADO INDIO

- Profundizar en el análisis del mercado indio, especialmente con el cliente CC, para comprender la singularidad de los datos y ajustar estrategias.
- Considerar la posibilidad de que las muestras de India puedan estar relacionadas con devoluciones en lugar de nuevos pedidos.

## ENFOQUE EN LA EFICIENCIA PRODUCTIVA

- Mejorar la eficiencia en la producción mediante estrategias basadas en el análisis de datos para optimizar el inventario y minimizar interrupciones en la programación.
- Comprometerse con la mejora continua, adaptándose dinámicamente al mercado y respondiendo proactivamente a los cambios en las preferencias del cliente.

# Discusión Propuestas

## ANÁLISIS DE CLUSTERS

Identificación de tres clusters:

- Cluster 0: Mayoritario, contribuye significativamente al volumen de ventas, con preferencia por "DURRY" y "HAND TUFTED".
- Cluster 1: Único cliente (TGT) enfocado en "HAND TUFTED", con mayor valor por pedido.
- Cluster 2: Compras variadas y en gran cantidad, destacando la preferencia por "DURRY". Aunque el Cluster 2 compra una variedad considerable de alfombras, se destaca la importancia de analizar la eficiencia operativa y los costos asociados con la gestión de una amplia gama de productos. Propondríamos conducir un análisis detallado de los costos asociados con la variedad de productos comprados por el Cluster 2. Identificar oportunidades para optimizar procesos y reducir costos operativos sin comprometer la calidad del servicio.

Se recomienda desarrollar estrategias de marketing personalizadas para cada cluster, adaptando enfoques según las preferencias específicas de cada grupo de clientes.

Dado que el clustering es una técnica que puede evolucionar con el tiempo, se sugiere la implementación de un proceso de evaluación continua para garantizar la relevancia y precisión de las segmentaciones de clientes. Por lo que sería interesante establecer un programa de revisión periódica de la segmentación de clientes utilizando técnicas avanzadas de análisis. Ajustar estrategias según los cambios en el comportamiento del cliente y garantizar que las acciones de la empresa sigan siendo relevantes.

## ANÁLISIS DE RENTABILIDAD POR PRODUCTO

Aunque se identificaron los productos más solicitados, es esencial comprender la rentabilidad asociada a cada tipo de alfombra. ¿Cuáles son los márgenes de beneficio para cada categoría? ¿Existen productos de alto volumen pero baja rentabilidad que podrían requerir ajustes en la estrategia de precios?

Propuesta: Realizar un análisis detallado de la rentabilidad por producto. Esto permitirá ajustar estrategias de precios, promociones y enfoques de marketing para maximizar los beneficios.