DATA EXPLORATION & PRE – PROCESSING:

Checking the columns in the data frame

```
In [17]: champ_df.columns

Out[17]: Index(['OrderType', 'OrderCategory', 'CustomerCode', 'CountryName',
                'CustomerOrderNo', 'Custorderdate', 'UnitName', 'QtyRequired',
                'TotalArea', 'Amount', 'ITEM_NAME', 'QualityName', 'DesignName',
                'ColorName', 'ShapeName', 'AreaFt'],
               dtype='object')
```

Checking the data types information of columns present in data frame:

```
RangeIndex: 18955 entries, 0 to 18954
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   OrderType        18955 non-null  object
 1   OrderCategory    18955 non-null  object
 2   CustomerCode     18955 non-null  object
 3   CountryName      18955 non-null  object
 4   CustomerOrderNo  18946 non-null  object
 5   Custorderdate    18955 non-null  datetime64[ns]
 6   UnitName         18955 non-null  object
 7   QtyRequired      18955 non-null  int64
 8   TotalArea        18955 non-null  float64
 9   Amount           18955 non-null  float64
 10  ITEM_NAME        18955 non-null  object
 11  QualityName      18955 non-null  object
 12  DesignName       18955 non-null  object
 13  ColorName        18955 non-null  object
 14  ShapeName        18955 non-null  object
 15  AreaFt           18955 non-null  float64
```

We can observe that [QtyRequired , TotalArea, Amount and Areaft] are the numerical columns In the data frame with others are categorical variables except Date as Datetime data type.

Checking missing(null) values in the columns if any –

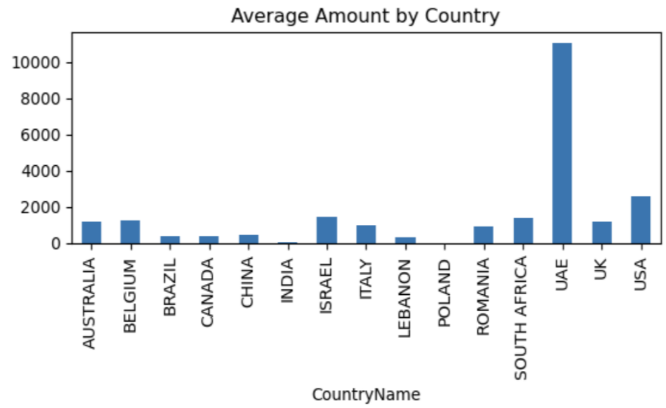
```
# Check for missing values
print(champ_df.isnull().sum())

Sum of QtyRequired    0
Sum of TotalArea      0
Sum of Amount         0
DURRY                 0
HANDLOOM              0
DOUBLE BACK           0
JACQUARD              0
HAND TUFTED           0
HAND WOVEN            0
KNOTTED               0
GUN TUFTED            0
Powerloom Jacquard    0
INDO TEBETAN          0
dtype: int64
```

VISULAIZING DATA :

Average Amount by Country

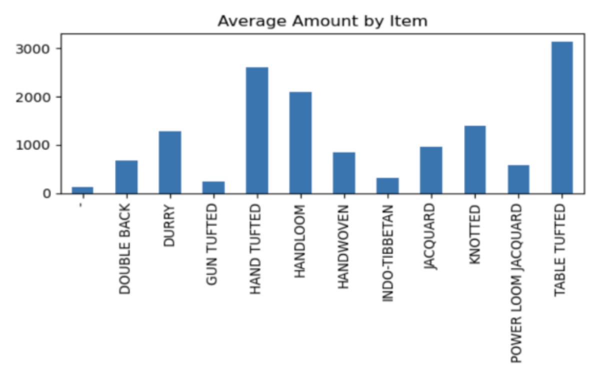
```
Average Amount by Country:
 CountryName
AUSTRALIA        1147.713389
BELGIUM          1233.501186
BRAZIL            362.892521
CANADA            406.893031
CHINA             429.654467
INDIA              35.688996
ISRAEL           1427.406267
ITALY             944.796724
LEBANON           337.754345
POLAND              0.000000
ROMANIA           935.583439
SOUTH AFRICA     1387.850957
UAE             11058.500000
UK               1160.219143
USA              2548.769442
Name: Amount, dtype: float64
```



We can see that UAE has more average amount of average sales worth ~$11058 followed by USA ~$2548 and Israel ~$1427.

AVERGAE AMOUNT BY ITEM

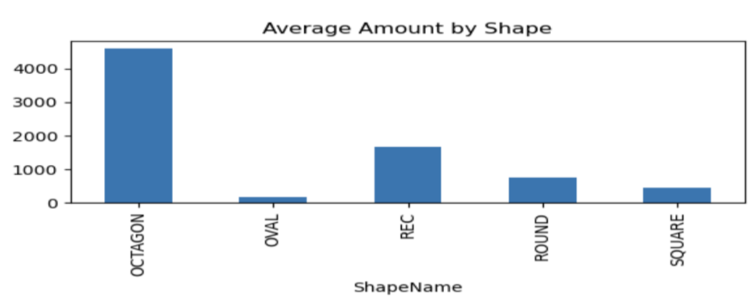
```
Average Amount by Item:
 ITEM_NAME
-                        125.850000
DOUBLE BACK              677.279032
DURRY                   1286.138784
GUN TUFTED               237.268791
HAND TUFTED             2608.156636
HANDLOOM                2092.241130
HANDWOVEN                854.331671
INDO-TIBBETAN            324.650909
JACQUARD                 967.648071
KNOTTED                 1394.853380
POWER LOOM JACQUARD      585.411458
TABLE TUFTED            3149.028571
Name: Amount, dtype: float64
```



As per above details and bar graph, we can observe that Table Tufted is the has the average highest sales by item ~$3149 followed by Hand Tufted $2608.15 and Handloom $2092.24

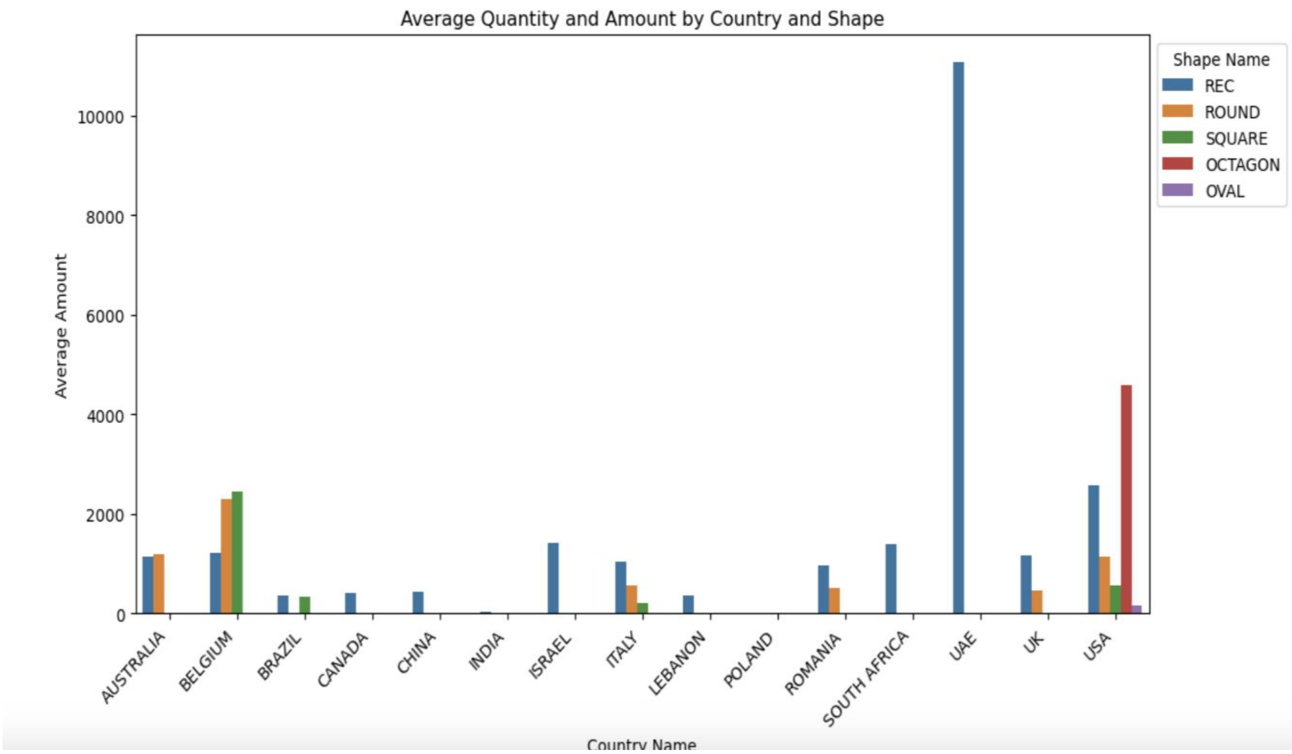
AVERAGE AMOUNT BY SHAPE

```
Average Amount by Shape:
 ShapeName
OCTAGON    4590.000000
OVAL        162.000000
REC        1679.501061
ROUND       766.589575
SQUARE      443.827014
Name: Amount, dtype: float64
```



As per above details and the graph , we can observe that highest average sales by shape is on OCTAGON $4590 followed by Rectangle $1679.50 and Round $766.59

AVERAGE QUANTITY AND AMOUNT BY COUNTRY AND SHAPE:



	Amount				
ShapeName	OCTAGON	OVAL	REC	ROUND	SQUARE
CountryName					
AUSTRALIA	NaN	NaN	1145.072412	1199.828667	NaN
BELGIUM	NaN	NaN	1223.722450	2298.505000	2457.600000
BRAZIL	NaN	NaN	363.279420	NaN	342.000000
CANADA	NaN	NaN	406.893031	NaN	NaN
CHINA	NaN	NaN	437.064370	7.290000	NaN
INDIA	NaN	NaN	36.054239	0.000000	0.571429
ISRAEL	NaN	NaN	1427.406267	NaN	NaN
ITALY	NaN	NaN	1043.940223	573.663608	201.896875
LEBANON	NaN	NaN	356.433270	7.760000	NaN
POLAND	NaN	NaN	0.000000	NaN	NaN
ROMANIA	NaN	NaN	966.011032	518.430968	NaN
SOUTH AFRICA	NaN	NaN	1387.850957	NaN	NaN
UAE	NaN	NaN	11058.500000	NaN	NaN
UK	NaN	NaN	1169.410938	461.642727	NaN
USA	4590.0	162.0	2577.261557	1153.073242	560.835444

	QtyRequired				
ShapeName	OCTAGON	OVAL	REC	ROUND	SQUARE
CountryName					
AUSTRALIA	NaN	NaN	10.442568	6.133333	NaN
BELGIUM	NaN	NaN	34.795918	11.000000	6.000000
BRAZIL	NaN	NaN	2.697531	NaN	2.000000
CANADA	NaN	NaN	2.094077	NaN	NaN
CHINA	NaN	NaN	14.017544	1.000000	NaN
INDIA	NaN	NaN	2.741021	2.771429	1.714286
ISRAEL	NaN	NaN	127.083333	NaN	NaN
ITALY	NaN	NaN	10.285714	5.567010	1.000000
LEBANON	NaN	NaN	10.490566	4.888889	NaN
POLAND	NaN	NaN	1.000000	NaN	NaN
ROMANIA	NaN	NaN	17.240000	6.806452	NaN
SOUTH AFRICA	NaN	NaN	9.840426	NaN	NaN
UAE	NaN	NaN	195.500000	NaN	NaN
UK	NaN	NaN	36.624402	41.318182	NaN
USA	51.0	1.0	41.525125	356.820000	8.044444

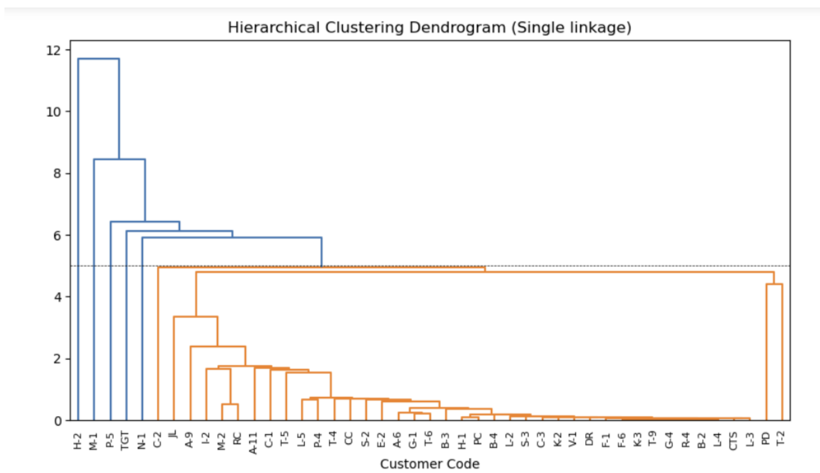
The above graph shows the holistic information about the average quantity sold in the country classified by shape. As per above graph we can observe that Belgium has dominant preference for round shapes with highest average amount of $2299 with moderate order quantity of 11. This suggest that market here has good opportunity of business with round shapes. Similarly, USA has dominant preference of rectangle shape with highest average amount of $2577 with order quantity of 41.53, with notable preference for ROUND and SQUARE shapes as well as an potential opportunity for these businesses.

Conversely, China and India has limited in interest in ROUND shapes , with a lower average amount of $ 7.29 and $0.00 needs new innovations and tailored strategies for capturing market base in significant way.

For clustering, we have moved the RowLabels which is customer code to the index to compare how each variable changes for different types of customers. We then normalized all the columns in the customer data.

Various Hierarchical methods have been used to cluster the customer groups.

Single Linkage Method:

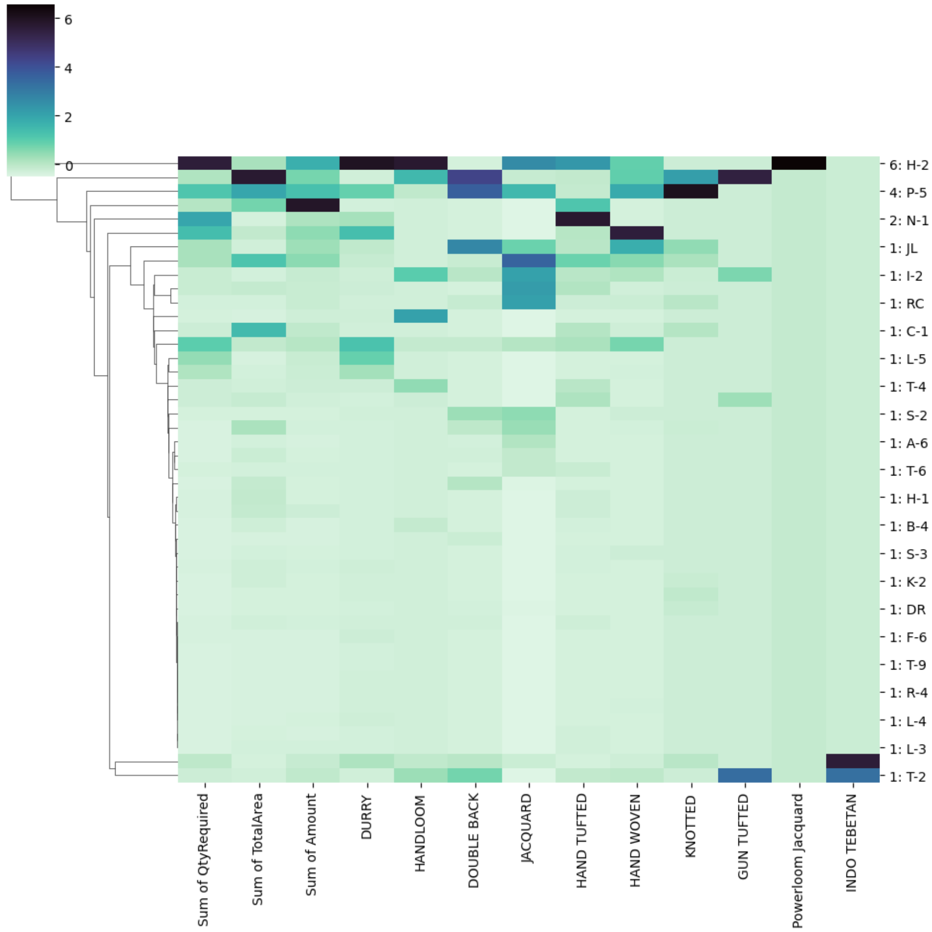


From the above Dendrogram graph, which was designed on single linkage method, we can see that there's a line in mid of the graph, where the threshold for cluster development was considered as 5. From this single linkage dendrogram, 6 clusters were identified and each of these customer categories has been mapped in one of these 6 clusters as follows

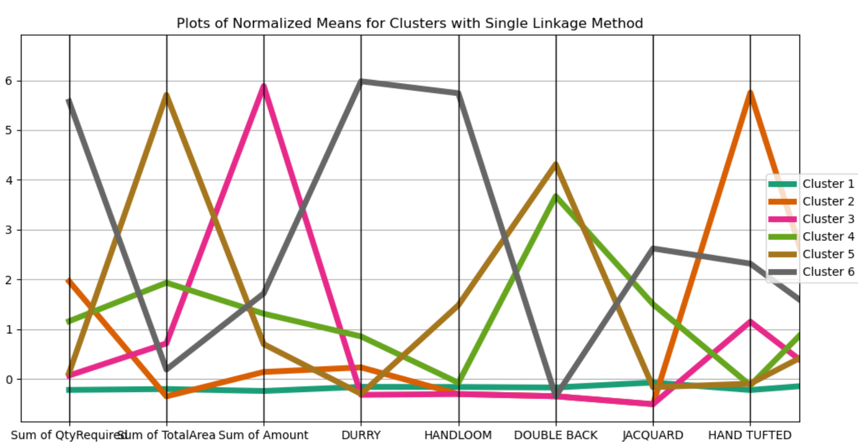
```
1 :  A-11, A-6, A-9, B-2, B-3, B-4, C-1, C-2, C-3, CC, CTS, DR, E-2, F-1, F-6, G-1, G-4, H-1, I-2, JL, K-2, K-3, L-2, L-3, L
-4, L-5, M-2, P-4, PC, PD, R-4, RC, S-2, S-3, T-2, T-4, T-5, T-6, T-9, V-1
2 :  N-1
3 :  TGT
4 :  P-5
5 :  M-1
6 :  H-2
```

We can see that clusters – 2,3,4,5,6 are individual single clusters having only unique customer segment.

A heat map was developed using this singular linkage and from the below heatmap it's seen that each customer group has been mapped with it's cluster number and how each variable influences these clusters.So, as per single linkage method, most of the customer groups from Cluster 1 doesn't have interest in buying the carpets except some customer groups such as JL, I-2, RC who are more inclined to try Jacquard carpets from Champo.

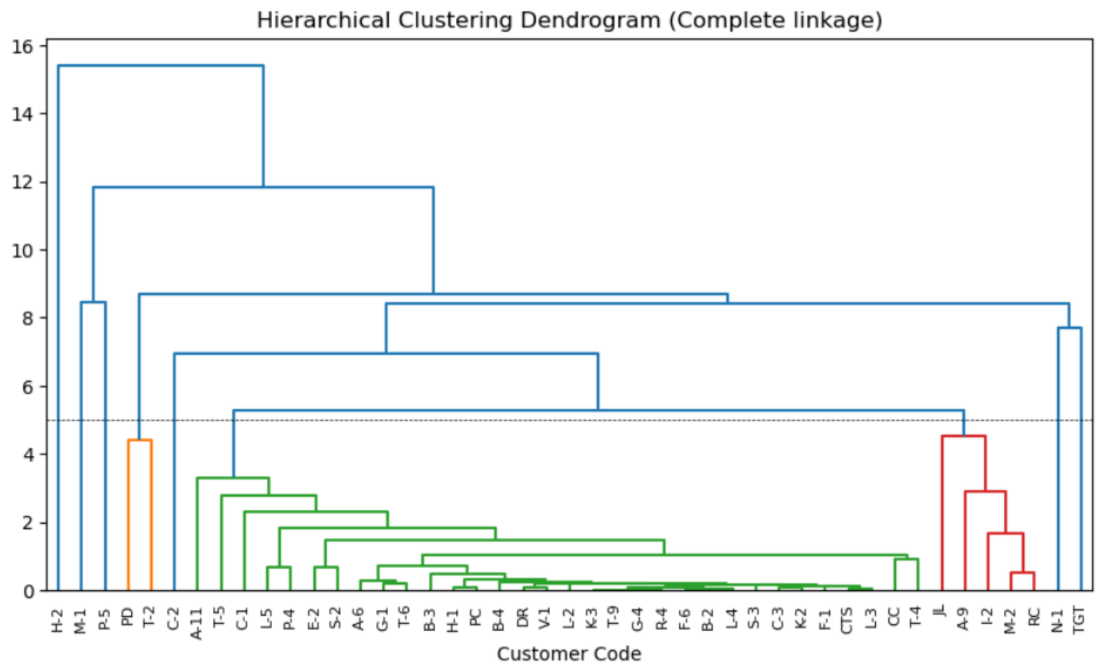


Customers in Cluster 2 are buying Hand tufted and Hand-woven carpets which are expensive comparative to others, hence getting more revenue. So, champo can send their samples of these carpet segments to cluster 2 ,4 and 6 for higher rate of customer churning. Cluster 6 customers are interested in trying all the type of carpets, so sending the samples to this cluster will result in bringing more orders to Champo.



From the above graph we can see that Cluster 1 has the lowest interest in champo carpets as their numbers are low in each of the variable. Cluster 3 & 4 are buying Double back carpets and cluster is more inclined to hand-tufted so sending the samples of various dyes, designs of that carpets to these clusters can result in better customer conversion.

Later, another cluster model using **Complete linkage method** was developed



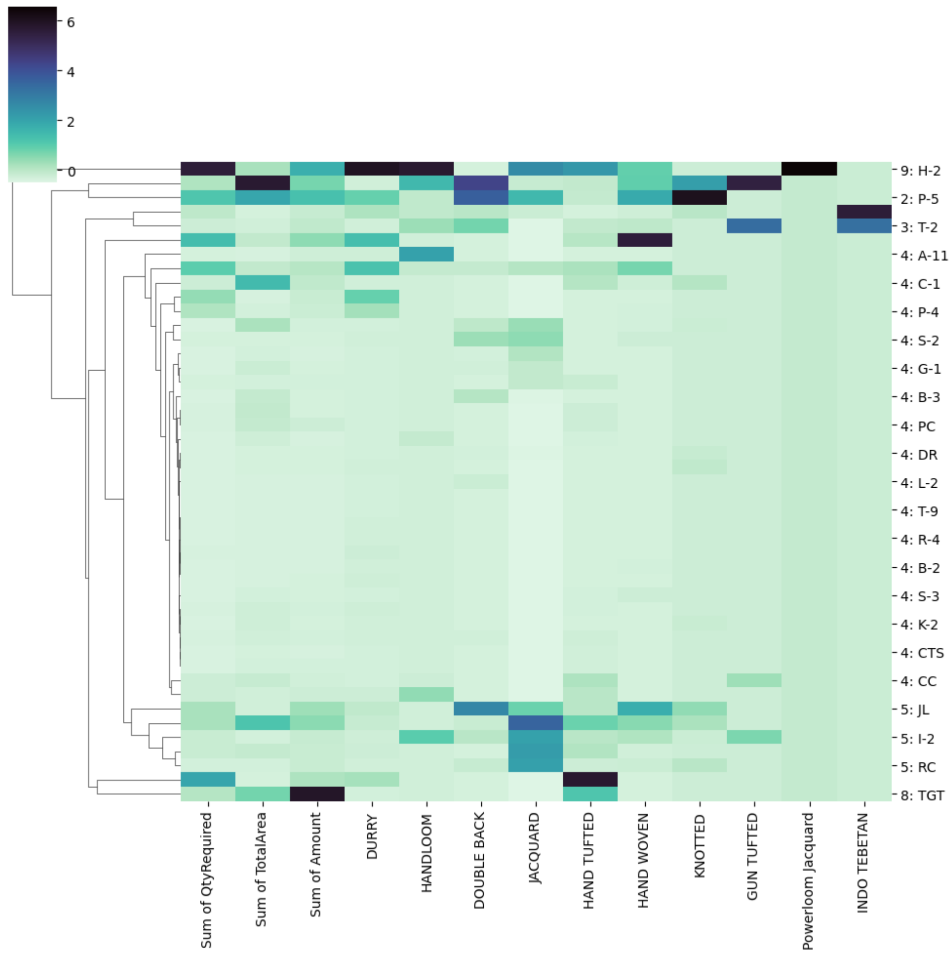
In this model, we have considered the same cluster development threshold of 5 and upon which 9 new clusters were formed as follows

```
1 :  M-1
2 :  P-5
3 :  PD, T-2
4 :  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, K-2, K-3, L-2, L-3, L-4, L-5, P-4, PC, R
-4, S-2, S-3, T-4, T-5, T-6, T-9, V-1
5 :  A-9, I-2, JL, M-2, RC
6 :  C-2
7 :  N-1
8 :  TGT
9 :  H-2
```

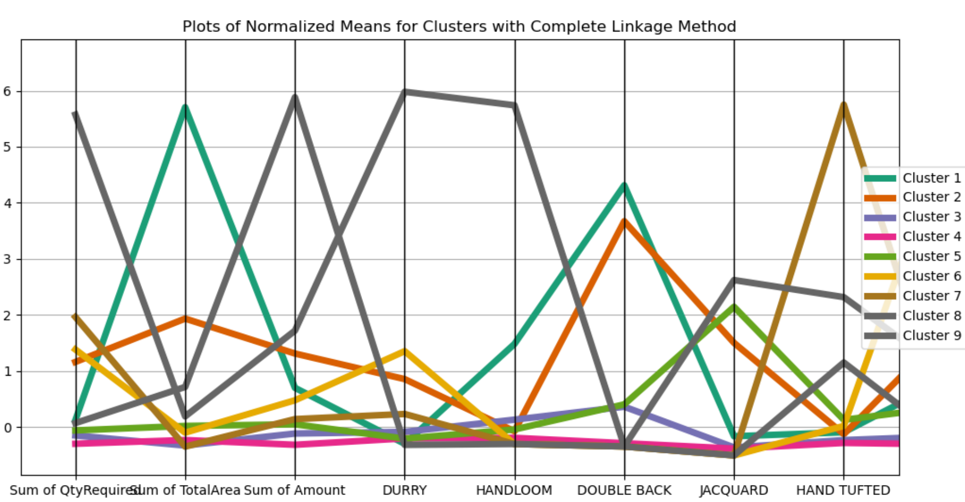
From this complete method cluster model, 9 new clusters were formed with most of the customer groups falling under cluster 4 and the rest of clusters have either single customer group or double. Cluster 9 is ordering more quantities of handloom, durry and powerloom jacquard. Hence new samples of these carpets with new designs and colors can be sent to cluster 9 as they are interested in buying these from Champo.

Cluster 8 has only customer group TGT who has been buying Hand tufted carpets of more area with Champo boosting their revenue. Hence Champo should share more such samples of various colors, designs, textures of Hand tufted carpets to cluster 9 customers.

Customers in Cluster 3 are more likely to be from Indo Tibetan region hence sharing gun tufted carpets to this cluster instead of jacquard or durry carpets can make more sales out of this conversion.



Apart from cluster 9, other clusters are not interested in the durry and Handloom carpets from Chapmo, hence sharing of these samples will result in more operations and supply chain costs compared to the revenue that can be generated from it. Cluster 7 is ordering more of the hand tufted carpets and hence sending these samples will help Champo sales.



Similarly, clusters were made using Average Linking, Centroid and Ward methods. Where the Euclidean distance thresholds were different and different clusters were formed.

Champo can cluster their customers using another Machine learning models to segment them based on their carpet choice and another product types.

Champo can cluster their customers using K-means clustering, where the columns are normalized, and we can obtain centroids. It also gives us how similar each cluster is with-in and Champo can decide to concentrate on which cluster to send sample for each carpet type. Also, Champo can find optimal number of clusters that can be formed with their customers.

Decision trees can also be used to segment the customers based on their previous orders. Using these trees, we can classify customers into various tiers or interest groups. Models will determine the attributes which are useful for segmentation. Based on the tree formed, we can make rules and classify the customer as whether he's a handcraft liking person or which color does he likes. Which makes it easy for Champo to send the sample which suits the customer preferences.

Neural networks play a crucial role in clustering by utilizing customer order data as input. This information is introduced into the input layer and traverses through multiple layers containing numerous neurons designed to identify patterns. During an iterative training procedure, the connections between neurons are continually adjusted to capture intricate relationships. Ultimately, the output layer classifies customers into specific segments, such as those interested in handcrafted carpets, durries, knotted items, and so forth.

By considering the initial EDA and the clusters, Champo has huge market in UAE and least in Poland, so Champo must promote their carpets in Poland and send more samples to both Poland and UAE to acquire more customers. They can also introduce the other shaped carpets to UAE. From the cluster models made, the Customer group H-2 is interested in buying various types of carpets from Champo, so sending various samples to that customer segment without limiting certain types can result in more retention and sales. There is a particular cluster in every model which has many customer groups, this cluster customers are less likely to purchase the products from Champo, so Champo can share some samples to these customers to acquire their attention and provide any offers for the initial customer acquisition.

Customer group T-2 is the key purchaser for Indo Tibetan carpets, hence while bringing a new carpets in this segment instead of sending samples for every customer, Champo can concentrate on this cluster itself to lower manufacturing costs and operations cost for samples.

A specific customer group, TGT is inclined in buying the expensive carpets – Hand tufted and they were buying it in bulk. Hence Champa can directly connect with this customer group if they are bringing in a new variety in hand tufted carpets.

Champo should send samples which are costly in terms of manufacturing and the materials used to the restricted groups who have already bought their carpets and re-bought in bulk as they are most likely to be the prospective customers for new products and Champo should market their products in specific demographic and send samples which are not costly in manufacturing to the prospective customers from these demographic as it is not sure whether the customers in this region will purchase the product or not.