**HBR – CHAMPO CARPETS: IMPROVING BUSINESS TO BUSINESS SALES USING MACHINE LEARNING  
ALGORITHMS**

**Assignment 3.2**

June 18, 2021

Aaron Vo

BIA-678 Big Data Technologies

Stevens Institute of Technology – Summer 2024

Table of Contents

[1 HBR Case STUDY - CHAMPO CARPETS: IMPROVING BUSINESS-TO-BUSINESS SALES USING MACHINE LEARNING ALGORITHMS 3](#_Toc170052588)

[1.1 UVA CASE STUDY QUESTIONS 3](#_Toc170052589)

[1.1.1 With the help of data visualization, provide key insights using exploratory data analysis. (You can use Tableau see instructions below.) 3](#_Toc170052590)

[Given the help of data visualization, there are a few interesting analysis we can gather the following: 3](#_Toc170052591)

[ Even though India was the largest export of rugs, USA is the highest consumer of rugs, with over 56% of customers being from the USA and 21% being from India. The UK customers accumulated to about 9% while Italy was 4th, with just over 3% of the total customer base. 3](#_Toc170052592)

[ In terms of rug quality, we see that TUFTED 60 C was far in a way, the most popular choice as the total number of orders doubled the second highest, which was a variation of the highest, the TUFTED 60C ALL LOOP. However, despite the difference in orders, the most popular quality selection was only 7% of total quality selection, while the second most populator was only 4.5%. This shows that despite the high order count for the TUFTED 60 C quality, the distribution of all carpet quality is close. However, once we group this by country, we see that each country has a specific quality that is most popular. For instance, in the UK, the most popular was the FLATWOVEN 60 carpet quality while the TUFTED 60C and its variation remains out of the top 10 in Belgium. 3](#_Toc170052593)

[ Regarding designs, Plain was the most popular, as compared to the top 20 designs, the plain design was chosen 17.8% while the homer design was 10%. However, since there are over 2000 designs from the given data set, we see that plain only represented 4.3% of the customer base while home was only 2.4%. This is similar to the quality as we see that despite the higher frequency for the top designs, customers have a lot of designs to pick from and it’s showing that they seem to vary. If we apply the popularity in designs to specific country, we also see difference preferences. Italy, the third most popular country for Champoo rugs, prefer Chevron designs, while UK prefers MODASA over Plain as the most popular design. 3](#_Toc170052594)

[ The most popular colors were unsurprisingly grey and blue but the option for multiple colors was the second highest choice. 4](#_Toc170052595)

[ And finally, the most popular shape by far was rectangle 4](#_Toc170052596)

[1.1.2 What kind of analytics and machine learning algorithms can be used by Champo Carpets to solve their problems, and in general, for value creation? 4](#_Toc170052597)

[Champo’s biggest challenge was the low conversion rate from the sample carpets that were being sent to the customers. Since these sample had a high production cost, they produce a challenge to provide customers with the best choice over the different designs while making sure that they’ll enjoy it and buy it. Since the conversion rate was low, this meant that customers often sent back samples with requests for different designs. Accurately predicting and meeting customer preferences through sample designs is crucial. However, this process was challenging and often led to costly iterations to match customer expectations. To do this, we can use a recommender model such as collaborative filtering, content-based filtering, decision trees, random forests, and even K- Means Clustering. 4](#_Toc170052598)

[1.1.3 Develop ML models to help identify features that contribute toward conversion (or non-conversion) of samples sent to customers. (You can use Rapid Miner, R, or Python: We recommend Rapid Miner; see instructions to download below.) 4](#_Toc170052599)

[Check the code attached code in the Jupyter notebook section. 4](#_Toc170052600)

[1.1.4 Discuss the data strategy for building customer segmentation using clustering. What are the benefits Champo Carpets can expect from clustering? 4](#_Toc170052601)

[The data strategy for building customer segmentation using clustering should start with defining the objectives, then data collection, pre processing, analysis, implementation, training and evaluation, actionable insights, and refining. First, we start by clearly outlining the objectives of customer segmentation for Champo Carpets. This may include understanding customer purchasing behavior and using this to enhance the customer experiencing while increasing the value in product offerings. Champo can then collect relevant data from various sources such as sales transactions, customer demographics, online behavior, customer feedback, and social media interactions to align with their objective and ensure that the data pipeline and processes is smooth. Here, the engineer can handle missing values, remove duplicates, and correct inconsistencies to ensure that all of the features can contribute to the data processing. Champo engineers can then apply a chosen clustering algorithm to the preprocessed data, analyze the characteristics of each cluster to interpret their meaning, and provide personalized product recommendations based on customer segments. 5](#_Toc170052602)

[1.1.5 Discuss clustering algorithms that can be used for segmenting Champo Carpets’ customers. 5](#_Toc170052603)

[For segmenting Champo Carpets’ customers, several clustering algorithms can be employed, the one we can use in this situation is K-Means Clustering, which is a clustering algorithm that partitions datasets into K clusters based on distance measurements. This is most useful for Champo’s customers because it is ideal for segmenting customers based on purchasing behavior, where clusters represent groups with similar buying patterns which can be used to help recommend similar products and provide easy samples for higher conversion rates. 5](#_Toc170052604)

[1.1.6 Develop customer segmentation using K-means clustering. Discuss the optimal number of clusters, significant variables, and cluster characteristics. 5](#_Toc170052605)

[Check the code attached code in the Jupyter notebook section. 5](#_Toc170052606)

[1.1.7 What are your final recommendations for Champo Carpets? 6](#_Toc170052607)

[I recommend Champo Carpets to enact a data strategy to help build a successful business between its consumers. There is a lot of data that is available, and using this, Champo Carpets can effectively leverage customer segmentation to enhance and improve customer experiences in order to drive business growth. Implementing a data-driven approach will provide a edge over its competition and foster a healthy system that doesn’t prove to be detrimental to the overall business model 6](#_Toc170052608)

# HBR Case STUDY - CHAMPO CARPETS: IMPROVING BUSINESS-TO-BUSINESS SALES USING MACHINE LEARNING ALGORITHMS

## UVA CASE STUDY QUESTIONS

### With the help of data visualization, provide key insights using exploratory data analysis. (You can use Tableau see instructions below.)

### Given the help of data visualization, there are a few interesting analysis we can gather the following:

### Even though India was the largest export of rugs, USA is the highest consumer of rugs, with over 56% of customers being from the USA and 21% being from India. The UK customers accumulated to about 9% while Italy was 4th, with just over 3% of the total customer base.

### In terms of rug quality, we see that TUFTED 60 C was far in a way, the most popular choice as the total number of orders doubled the second highest, which was a variation of the highest, the TUFTED 60C ALL LOOP. However, despite the difference in orders, the most popular quality selection was only 7% of total quality selection, while the second most populator was only 4.5%. This shows that despite the high order count for the TUFTED 60 C quality, the distribution of all carpet quality is close. However, once we group this by country, we see that each country has a specific quality that is most popular. For instance, in the UK, the most popular was the FLATWOVEN 60 carpet quality while the TUFTED 60C and its variation remains out of the top 10 in Belgium.

### Regarding designs, Plain was the most popular, as compared to the top 20 designs, the plain design was chosen 17.8% while the homer design was 10%. However, since there are over 2000 designs from the given data set, we see that plain only represented 4.3% of the customer base while home was only 2.4%. This is similar to the quality as we see that despite the higher frequency for the top designs, customers have a lot of designs to pick from and it’s showing that they seem to vary. If we apply the popularity in designs to specific country, we also see difference preferences. Italy, the third most popular country for Champoo rugs, prefer Chevron designs, while UK prefers MODASA over Plain as the most popular design.

### The most popular colors were unsurprisingly grey and blue but the option for multiple colors was the second highest choice.

### And finally, the most popular shape by far was rectangle.

### What kind of analytics and machine learning algorithms can be used by Champo Carpets to solve their problems, and in general, for value creation?

### Champo’s biggest challenge was the low conversion rate from the sample carpets that were being sent to the customers. Since these sample had a high production cost, they produce a challenge to provide customers with the best choice over the different designs while making sure that they’ll enjoy it and buy it. Since the conversion rate was low, this meant that customers often sent back samples with requests for different designs. Accurately predicting and meeting customer preferences through sample designs is crucial. However, this process was challenging and often led to costly iterations to match customer expectations. To do this, we can use a recommender model such as collaborative filtering, content-based filtering, decision trees, random forests, and even K- Means Clustering.

### Develop ML models to help identify features that contribute toward conversion (or non-conversion) of samples sent to customers. (You can use Rapid Miner, R, or Python: We recommend Rapid Miner; see instructions to download below.)

### Check the code attached code in the Jupyter notebook section.

### Discuss the data strategy for building customer segmentation using clustering. What are the benefits Champo Carpets can expect from clustering?

### The data strategy for building customer segmentation using clustering should start with defining the objectives, then data collection, pre processing, analysis, implementation, training and evaluation, actionable insights, and refining. First, we start by clearly outlining the objectives of customer segmentation for Champo Carpets. This may include understanding customer purchasing behavior and using this to enhance the customer experiencing while increasing the value in product offerings. Champo can then collect relevant data from various sources such as sales transactions, customer demographics, online behavior, customer feedback, and social media interactions to align with their objective and ensure that the data pipeline and processes is smooth. Here, the engineer can handle missing values, remove duplicates, and correct inconsistencies to ensure that all of the features can contribute to the data processing. Champo engineers can then apply a chosen clustering algorithm to the preprocessed data, analyze the characteristics of each cluster to interpret their meaning, and provide personalized product recommendations based on customer segments.

### Discuss clustering algorithms that can be used for segmenting Champo Carpets’ customers.

### For segmenting Champo Carpets’ customers, several clustering algorithms can be employed, the one we can use in this situation is K-Means Clustering, which is a clustering algorithm that partitions datasets into K clusters based on distance measurements. This is most useful for Champo’s customers because it is ideal for segmenting customers based on purchasing behavior, where clusters represent groups with similar buying patterns which can be used to help recommend similar products and provide easy samples for higher conversion rates.

### Develop customer segmentation using K-means clustering. Discuss the optimal number of clusters, significant variables, and cluster characteristics.

### Check the code attached code in the Jupyter notebook section.

### What are your final recommendations for Champo Carpets?

### I recommend Champo Carpets to enact a data strategy to help build a successful business between its consumers. There is a lot of data that is available, and using this, Champo Carpets can effectively leverage customer segmentation to enhance and improve customer experiences in order to drive business growth. Implementing a data-driven approach will provide a edge over its competition and foster a healthy system that doesn’t prove to be detrimental to the overall business model

References

Goyal, Priyank, et al. Champo Carpets: Improving Business-to-Business Sales Using Machine Learning

Algorithms. Indian Institute of Management Bangalore, 2021. Accessed 23 June 2024.