

Intro

Currently, Tantalizing Textiles operates with one store in each state within the United States of America. With expectations for growth, upper management wants to get a head start and begin to plan how they can save shipping costs and engage better with customers in their physical stores. As of today, all shipped orders made on Tantalizing Textiles' website are packaged up and sent out to customers from their lone shipment warehouse in the greater Los Angeles area.

They have come up with the idea of expanding one store within each of the four regions (Northeast, Midwest, South, and West) of the country that will serve as flagship stores to offer a more enhanced in-store shopping experience, in addition to serving as each region's inventory warehouse from which items get shipped to customers that have chosen to have their purchases delivered to them.

DISCLAIMER

Before digging into the analysis I completed of the data, it is vital to highlight how I've interpreted this "dummy dataset" which I obtained publicly via Kaggle thanks to [insert Author name]. The dataset is not representative of truthful purchases made by customers and is simply intended for personal project exploration. The analysis I have conducted and presented in this project by no means takes into account the multitude of factors that characterizes the United States of America today. Simply put, my analysis draws only from the data contained in the dataset I initially downloaded.

Assumptions Made

In the early phases of exploring the dataset, there were various columns that required deeper thought in terms of how I was going to approach my future analysis. Firstly, it is important to state that the initial dataset contains 3900 records of different customer_id. As there are only 3900 records in the dataset, this means that the dataset only shows one purchase for each customer. For purposes of this project, I have made the assumption that this is a database of the 3900 unique customers the company has in its history of operation and that each row reflects characteristics of each customer as well as the most recent purchase made by each of them. Therefore, it is assumed that there is a wealth of transaction history that predates the snapshot that this dataset presents.

Another area in the dataset in which I had to make a decision on how to interpret what was being presented is between the two columns "previous_purchases" and "frequency_of_purchases". The problem lying here was that these two columns didn't line up logically. For purpose of my analysis, I removed the "previous_purchases" column and kept the column "frequency_of_purchases" which I have assumed to mean the average rate (Weekly, Fortnightly, Monthly, Quarterly, and Annually) by which each customer purchases an item from Tantalizing Textiles based on their purchase history.

An additional point to make in regard to how I interpreted the data is under the "shipping_type" column. All unique values in this column pertain to items being shipped except for "Store Pickup". For purposes of my analysis I have used "Store Pickup" to mean customers that physically came into the store, chose a product, and then bought the product upon leaving. Therefore, "Store Pickup" refers to purchases that were not made using the company's online website, whereas the rest of the purchases were.

Data Cleaning

There was a very minor amount of cleaning that needed to be done to the dataset, prior to me embarking on my exploration and analysis. Within the column "frequency_of_purchases", There were two sets of values that I grouped together due to them having the same thing. Thus, I replaced every instance of "Bi-Weekly" with "Fortnightly", and replaced every instance of "Every 3 Months" with "Quarterly".

Analysis

In regards to upper management approaching me with the initial data exploration of expanding a select number of physical Tantalizing Textiles stores, my first step was to understand which columns from the dataset would be most relevant. Tantalizing Textiles seeks to not only expand four specific stores in terms of enhancing the in-store shopper experience, but also to serve as inventory warehouses and shipping headquarters for each region.

By assisting in the preliminary analysis to justify the selection of four different states to house flagship stores, it was essential for me to identify a core set of data points which reflected aspects of the in-store and purchasing-habits shopping experience. The main columns that I chose to analyze were “subscription”, “frequency_of_purchases”, and “shipping_type”. My final decision on where each region’s flagship store should be located takes into account the various factors that I outline below.

Subscribers as “Investors”

When it comes to in-store shopping, I immediately thought it was important to query and then visualize the areas of the country based on where subscribers were located. To me, a customer that is identified as a subscriber is one who has a personal attachment to the company and expresses an ongoing connection and interest with the company’s culture, growth, products and services. Thus, those states in which there are a high number of subscribers are areas that are emblematic of a hearth waiting to spread in terms of interest, popularity, and loyalty towards a company. The process of one subscriber bringing a friend who has no awareness of the company to the physical store does a tremendous amount for the company. For this metric, I took the count of subscribers for each state and divided it by the count of customers in each state, then converting that value into a percentage.

Building Habits with Frequency

Similarly to subscribers of the company being representative of those that are loyal followers of the brand, there is no better metric of the satisfaction of the brand than the frequent association with that brand. For this reason, I chose to include a visualization displaying the percentage of customers from each state that have a purchase frequency of at least one item per month. For purposes of limitation of space on my dashboard as well as being interested in states with the most customers, I chose to only analyze states that have a customer count of 80 and higher. I achieved this by converting the string values within the “frequency_of_purchases” column of “Weekly”, “Fortnightly”, and “Monthly” into values of 1 for each customer (row in the dataset) and summing them together and then dividing by the total amount of customers in order to get a percentage for each state. Choosing to only include states with a customer count of 80 and above also serves to display which states house the most customers.

A frequent shopper of a brand is one who had a positive initial experience, and has chosen to continue associating themselves with and supporting the brand through consumption. This connection that one has with the company, shown through the repeated ordering of products certainly comes with its advantages when the flagship store is located close to the frequent shopper, due to shorter shipping costs and delivery times, in addition to close physical proximity to a brand they love.

Sharing Time Together

It only makes sense that if the company is looking to increase the size and in-shopper experience of a select number of stores, that those stores are ones which see a significant number of customers that last purchased an item in store. It can be inferred that if a particular customer in the database had purchased an item in their state’s physical store, it is more likely than not that they’ll enact their shopping habit in the store again in the future. This is especially true considering that the store will be an upgraded and enhanced version which would only draw the customer in even more. Despite many consumers only

choosing to continuously purchase products from a brand because of the material product, there are an abundance of brands that supplement their customers' attachment to products based on very strategically designed physical spaces in stores in addition to quality of customer support, and other marketing techniques. These pull factors are detrimental to the flagship stores that Tantalizing Textiles wants to open in each region.

Income by State

Additionally, I incorporated a third chart showing total income generated by state based on each frequent customer's last purchase. Despite my other points of analysis supporting a claim for each region's flagship store location, company revenue (only income being defined in the dataset) is a defining factor to the operation of a business, especially when forecasting future success when looking to expand. This visualization is intended to be thought of in tandem with the other ones. For instance, if a state secures a high income from customers who make at least one purchase per month, then future forecasts of sales would expect to increase if a flagship store is positioned closer to shoppers from that specific state.

Proposal

Northeast

Upon observing the data both visually and from initial querying, it is clear that the Northeast region is not characteristic of the company's primary region in terms of popularity and customer loyalty. Whilst having multiple states in which over 20% of customers obtained their last item in the physical store, only three states fall within the top 20 nationally by subscriber percentage. When considering that the development of flagship stores will appeal strongly to those already familiar with the company, I recommend that **Delaware** is the state in the Northeast that would best house the region's flagship store.

Delaware impressively displays the 7th highest subscriber percentage count of every state in the nation, with 32.56% of the state's 86 customers. In addition, nearly 57% of its customers are characterized as frequent customers. Furthermore, the above two metrics are highlighted more when one sees that Delaware showed the 4th highest income in the region based on frequent customers' last purchases. It is important to recognize that Delaware shows a national middle-of-the-pack percentage for customers having made their last purchase in-store, but with frequency of purchases among customers and percentage of subscribers being high, I anticipate that there is significant potential for growth of current customers and new ones coming into the physical store, specifically if it is an updated one with more pull factors. Delaware's high subscriber percentage is its strongest asset in forecasting future success for Tantalizing Textiles in the company as I expect spoken word about the company to be influential in attracting current and new customers with the opening of a flagship store.

Midwest

When digging into customer representation of the Midwest, it is clear to see that this region is of strong importance for the company. Of the factors explored in my analysis, there were several clear front-runners that the company's customer database showed as the most optimal places for location of the region's flagship store with Iowa, Illinois, and Indiana being noteworthy. However, when taking into account each visualization and the importance of having the flagship store located centrally in the region, I came to the decision of **Illinois** as being the best site for the Midwest.

Although Illinois (ranked #41) doesn't possess the strongest subscriber percentage among its customers, it stands on top as the state with the highest percentage of most frequent customers with 64.13% of them making at least one purchase a month. Additionally, Illinois ranks the highest in the country at 26.09% of its customers making their most recent purchase in-store. Furthermore, Illinois' significance in monetary terms is seen with the state accruing the 2nd highest income of most recent purchases by customers throughout the United States thus solidifying its case for the location of the region's flagship store. Despite Illinois ranking in the bottom 10 for subscriber percentage, it is clear that the state dominates the nation in other defining facets as the state and region would benefit enormously with the arrival of a flagship store.

South

Whereas the data pertaining to the Midwest region displayed a pretty clear winner in the proposal for where a flagship store would be best placed, the South served as a region that required more thought. Boasting four states in the top 10 nationally on the basis of subscriber percentage, in addition to the prominence of southern states having high frequent customer percentages, it is clear that customers in the South engage strongly with Tantalizing Textiles. The map portrays Oklahoma as a particularly prime location, but by taking into account all three visualizations in addition to thinking geographically-central for importance of shipping purposes throughout the region, **Alabama** positions itself as a particularly promising state for the South's flagship store.

In my exploration of the data, Alabama proves as the state that featured most prominently based on the defining factors in my analysis. Alabama sets itself atop the region by ranking 4th in frequent customer percentage and ranking 1st in total income accrued from frequent customers' most recent orders. Alabama's subscriber count isn't the highest in the region (ranked #33 nationally), but its percentage of customers who purchased their most recent item in-store (21.35%) is third highest in the region. Of all regions across the United States, Alabama's central position in the South serves as its strongest asset. As previously mentioned, states within the region have consistently high values of frequent customer percentage meaning that as the shipment hub for the region, Alabama would be a prime location for meeting the high demand of Tantalizing Textiles within the South.

West

Of each region in the country, the West proved to be the most difficult one to pinpoint the best optimal location for a flagship store. As seen in the dashboard, Nevada and California both rank quite high by subscriber percentage, frequent customer percentage, and total income accrued from frequent customers' most recent purchases, but these two states possess quite low percentages for customers that last made in-store purchases. When taking into account subscriber percentage and percentage of customers purchasing on at least a monthly basis, Colorado looks quite promising. However, Colorado doesn't feature positively in terms of total income accrued by the state, with Colorado in fact ranking 12th out of 13 in the West when I explored beyond the top 5 by region.

Seeing that the regionally central states of Nevada and Utah aren't ones within the region whose data insights prove to be the most optimal location for a flagship store, **New Mexico** stood alone as the West's best option for flagship store given its relevance in the various viewpoints of my analysis. With the highest subscriber percentage by state in the whole country being South Carolina at 35.53%, New Mexico's value of 27.16% is notable. Additionally, New Mexico has the second highest percentage of frequent shoppers among customers in the West with 56.79% of customers purchasing at least one item per month. Looking deeper into their loyal frequent customers, New Mexico also recorded the 5th highest income based on most recently purchased items within the West. The one metric that New Mexico slightly lacks is the percentage of most recent purchases being made in store by their customers, but like Delaware I foresee

this figure increasing with an enhanced in-store shopping experience emerging with the development of New Mexico's flagship store.