**SUPER STORE DATA ANALYSIS:**

**Introduction**

I used the popular superstore data which is the sales record for a retail business located in the United States. It shows the records of sold furniture, office supplies and technological products to a customer set of consumers, home offices and corporate entities across 531 cities in 49 states of the U.S.



## Problem Statement

The store which would be named “The Superstore” for the purpose of this project generated a huge amount of data from sales over the past 4 years which has lied dormant and without use. The owners hope to leverage on it to help drive the growth of the store over the coming years.

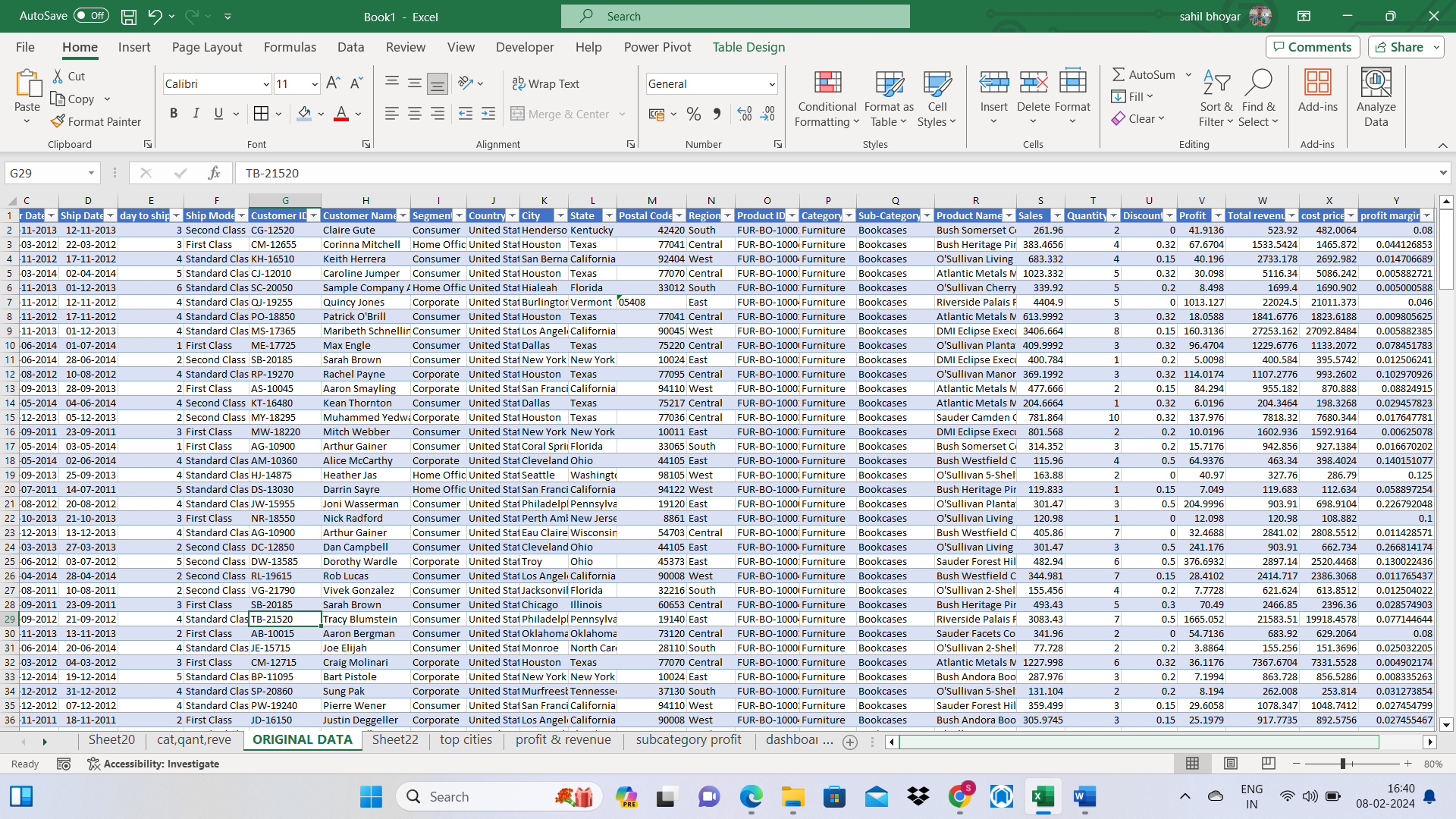
## Basis

The project is to provide in-depth analysis to help the store owners understand their data trends and patterns and provide valuable insights to drive better business decisions. The following criteria would be used for analysis:

* performance of product categories in terms of quantity sold, revenue and profit
* performance of sub-categories
* the preferred ship mode among customers
* the biggest revenue generator among all customer segments
* the regions and cities with revenue generated and profit made
* year by year performance of the store.

## Data Preparation and Exploration

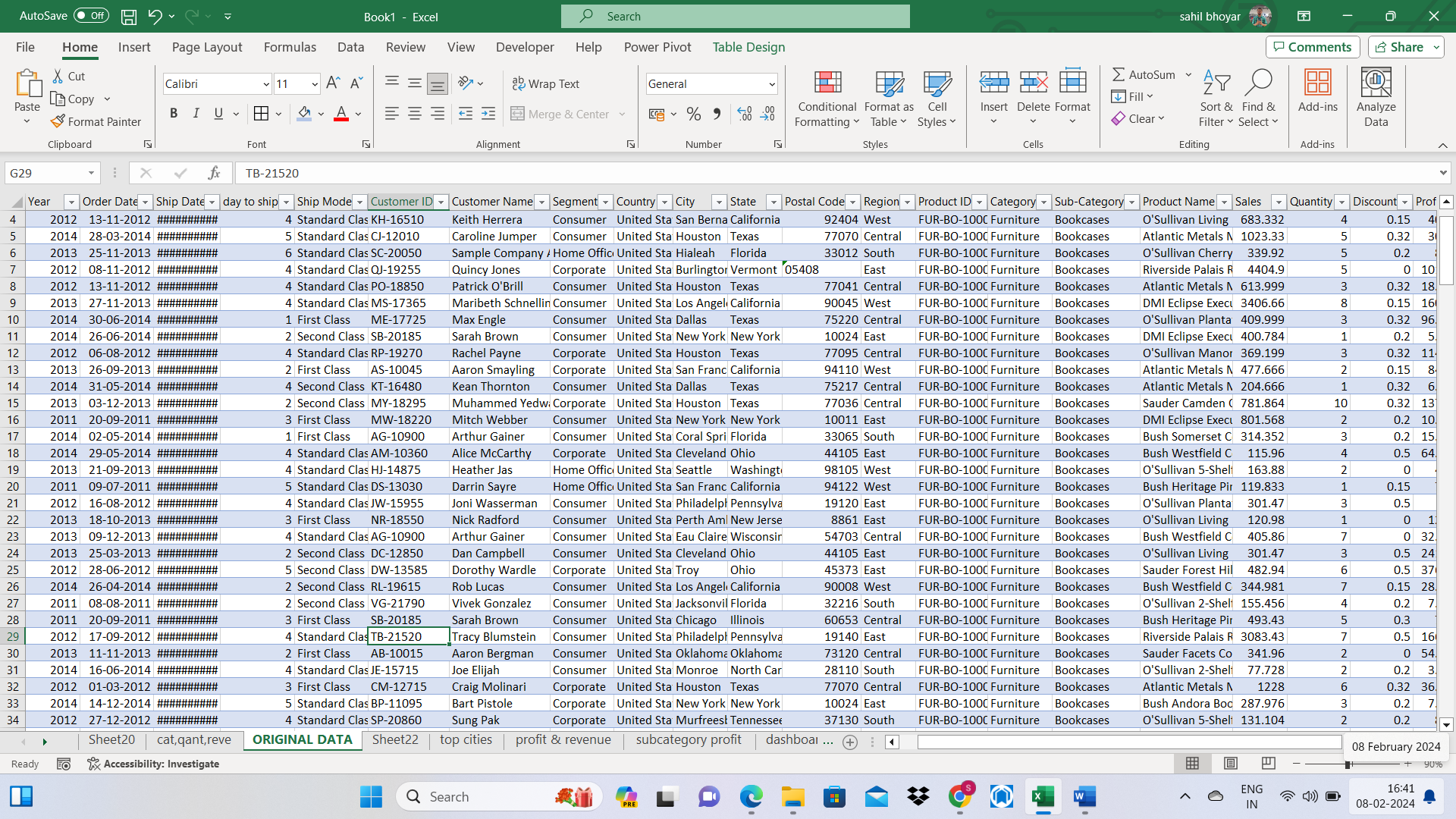
The dataset was downloaded and accessed with Microsoft Excel as it was already in a .xlsx format



Having a look, we can notice a variety of information within the dataset consisting of 21 columns and 9994 rows.

Each row represents an order for a product and it details the order date, ship mode, geographical details, cost and selling price, quantity ordered and discount(in terms of quantity) applied on the said product, amongst others.

## Data Cleaning and Transformation



Common issues such as identifying missing data(empty cells), misspellings, duplicate data were sorted to ensure the integrity of the data.

Columns were formatted with prices set to the currency, order and ship dates columns set to the short date format and others converted to correlate with their corresponding data types.

New fields were created to get the important numbers as the numbers given could not generate the insight needed alone.

***Revenue was calculated by using- Selling Price \* Quantity — (Discounts),***

***Total Cost Price= Revenue- Profit,***

***and profit margin by Profit/Revenue.***

Lastly, the dataset was formatted as a table to pave way for visualization and analysis.

## Analysis

* **Performance of Categories**



As earlier stated, products sold by the store are categorized under

furniture, office supplies and technology. The store owners did not

seem to focus on a category as almost an equal amount was

invested(about $3.5M-$4M) into each category which also yielded

similar amounts of revenue and profit.

Office supplies were by far dominant in terms of quantity sold with

papers and binders being the most popular products in this category.

It also stood out, albeit slightly as the category which generated the

least revenue and profit. mostly due to the fact that it had the cheapest

products across all.

Technology, which had the least quantity sold, generated the highest

revenue and was also the most profitable indicating relatively high

prices to compensate.

|  |  |
| --- | --- |
|  |  |

* **Performance of Sub-categories**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Row Labels** | **Sum of Quantity** | **Sum of cost price** | **Sum of Total revenu** |
| **Technology** | **6939** | **$38,57,402.34** | **$40,80,017.13** |
| Phones | 3289 | $15,69,113.58 | $16,28,690.56 |
| Accessories | 2976 | $8,35,063.83 | $8,78,861.72 |
| Machines | 440 | $8,51,132.58 | $9,14,754.68 |
| Copiers | 234 | $6,02,092.35 | $6,57,710.17 |
| **Furniture** | **8028** | **$37,18,522.85** | **$38,58,846.34** |
| Furnishings | 3563 | $4,27,943.44 | $4,53,984.41 |
| Chairs | 2356 | $16,16,301.12 | $16,62,652.97 |
| Tables | 1241 | $10,97,332.16 | $11,44,430.98 |
| Bookcases | 868 | $5,76,946.13 | $5,97,777.98 |
| **Office Supplies** | **22906** | **$33,11,916.20** | **$35,47,637.52** |
| Binders | 5974 | $9,05,644.38 | $10,12,887.13 |
| Paper | 5178 | $3,55,248.24 | $3,89,301.81 |
| Storage | 3158 | $10,65,034.85 | $10,99,166.28 |
| Art | 3000 | $1,29,433.73 | $1,35,961.51 |
| Appliances | 1729 | $5,11,893.57 | $5,47,290.86 |
| Labels | 1400 | $65,063.50 | $70,609.76 |
| Fasteners | 914 | $16,024.08 | $17,039.99 |
| Envelopes | 906 | $67,194.77 | $74,158.95 |
| Supplies | 647 | $1,96,379.09 | $2,01,221.23 |
| **Grand Total** | **37873** | **$1,08,87,841.39** | **$1,14,86,500.98** |
|  |  |  |  |
|  |  |  |  |

Although office supplies had the most sales and technology was the most profitable category, furniture had a standout sub-category in chairs which singlehandedly generated the most revenue and profit. Only phones from the technology category came close.

* **Customer Segment**

Customer segment was classified into consumers, home offices and corporate for the average person with needs, people with workspaces in their homes and business entities.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Row Labels** | **Sum of Total revenu** | **Sum of Profit** |
| Consumer | $58,18,526.49 | $3,04,010.63 |
| Corporate | $35,15,629.82 | $1,81,553.55 |
| Home Office | $21,52,344.67 | $1,13,095.41 |
| **Grand Total** | **$1,14,86,500.98** | **$5,98,659.59** |
|  |  |  |

The consumer segment was the most profitable and independently got about half of the whole revenue generated by the store. It is followed by corporate segment and home office.

* **Preferred Ship Mode**

The store had 4 systems of shipping ordered products namely first class, same day, second class and standard class.

*To determine delivery time for products, a new column was created by subtracting the order date from the ship date.*

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Row Labels** | **Sum of Quantity** | **Average of day to ship** |
| First Class | 5693 | 2.18 |
| Same Day | 1960 | 0.04 |
| Second Class | 7423 | 3.24 |
| Standard Class | 22797 | 5.01 |
| **Grand Total** | **37873** | **3.96** |

Although it took an average of 5 days to deliver items, with 54% of product shipped, standard class by a large amount is the preferred mode of shipment and perhaps the most economical. It is followed by second class, first class and same day in that order.

Same-day delivery took just like its name, less than a day(the least

time required) to deliver products and had the lowest quantity of

items shipped(around 3% of total). Both first class and second class

had similar delivery times and quantities delivered but with first class

edging slightly by being faster.

* **Regions and Cities**

The store sold items to all regions of the United States, Central, East,

South and West.

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Sum of Profit** | **Sum of Total revenu** |
| West | $1,53,860.37 | $35,94,877.32 |
| East | $1,90,704.00 | $33,78,912.25 |
| Central | $1,52,336.13 | $24,75,274.60 |
| South | $1,01,759.09 | $20,37,436.80 |
| **Grand Total** | **$5,98,659.59** | **$1,14,86,500.98** |

The West and East regions came up as the most profitable regions

while the South made the least profit. This also applies for the revenue

generated.

* **Cities**

New York city generated the highest revenue and profit and this is backed by the fact that it is the most populated city in the United States, and a city with a lot of organizations and corporate entities. It is followed by Los Angeles, the 3rd largest city in terms of population.

All cities in the top 10 had the lowest discount rates for products which also helped boost the profits made from them.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Row Labels** | **Average of Discount** | **Sum of Profit** | **Sum of Total revenu** |
| New York City | 0.056174863 | 69969.0289 | 1263426.211 |
| Los Angeles | 0.074297189 | 32781.4179 | 873153.822 |
| Seattle | 0.064953271 | 29926.4385 | 597587.426 |
| Philadelphia | 0.326815642 | 25343.7148 | 567598.411 |
| San Francisco | 0.066666667 | 20190.9216 | 541391.712 |
| Houston | 0.379681698 | 19417.1851 | 305910.2068 |
| Chicago | 0.383757962 | 15586.6854 | 223934.065 |
| Detroit | 0.006956522 | 13181.7908 | 270414.508 |
| Springfield | 0.142331288 | 12537.143 | 234782.084 |
| Newark | 0.103157895 | 12147.3592 | 114860.433 |
| **Grand Total** | **0.152735063** | **251081.6852** | **4993058.879** |
|  |  |  |  |

The gap between the biggest revenue generating cities and the lowest

was by a lot.

The least profitable cities generated very little revenue and made little

to no profit with most providing losses.

Also, 4 out of 10 least cities were located in Florida, a state in the

south, world renowned for resorts, parks, tourist attractions,

recreation, boating and fishing activities among others. This does not

much fit the demographic of working class people or organizations

that serve as customers for the store and explains the low sales made

from the area. It also explains why the south region generally made

the least revenue out of all regions.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Row Labels** | **Sum of Total revenu** | **Sum of Profit** | **Average of Discount** |
| Jupiter | 1.864 | 0.1548 | 0.2 |
| San Luis Obispo | 7.24 | 1.1946 | 0 |
| Elyria | 1.124 | 1.3984 | 0.7 |
| Springdale | 8.6 | 1.419 | 0 |
| Pensacola | 5.942 | 1.476 | 0.7 |
| Ormond Beach | 7.724 | 1.9656 | 0.7 |
| Keller | 11.8 | 2.1 | 0.2 |
| Port Orange | 7.624 | 2.934 | 0.2 |
| Abilene | 1.984 | 3.7584 | 0.8 |
| Iowa City | 9.99 | 4.4955 | 0 |
| **Grand Total** | **63.892** | **20.8963** | **0.35** |
|  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| * **Yearly Progression**  |  |  |  | | --- | --- | --- | |  |  |  | | **Row Labels** | **Sum of Total revenu** | **Sum of Profit** | | **Home Office** | **$21,52,344.67** | **$1,13,095.41** | | 2011 | $4,62,876.05 | $18,452.92 | | 2012 | $3,80,124.82 | $19,250.08 | | 2013 | $5,25,349.58 | $29,042.54 | | 2014 | $7,83,994.22 | $46,349.88 | | **Corporate** | **$35,15,629.82** | **$1,81,553.55** | | 2011 | $6,46,942.39 | $30,469.50 | | 2012 | $6,48,690.87 | $34,667.25 | | 2013 | $10,53,097.59 | $51,443.56 | | 2014 | $11,66,898.97 | $64,973.24 | | **Consumer** | **$58,18,526.49** | **$3,04,010.63** | | 2011 | $13,58,149.83 | $64,407.10 | | 2012 | $13,55,758.60 | $72,760.06 | | 2013 | $14,68,176.86 | $76,916.26 | | 2014 | $16,36,441.20 | $89,927.21 | | **Grand Total** | **$1,14,86,500.98** | **$5,98,659.59** | |  |  |

The store had a total of of about $14.8M in generated revenue across

all categories over a period of 4 years.

From the chart, it can be deduced that the profit and revenue had somewhat of a steady and arithmetic increase over the years. Also, it can be noticed that sales and revenue generated were usually seasonal over the years with the first quarter of every year usually recording the lowest revenue and sales picking up from the second quarter through to the fourth. A massive drop would then be seen at the beginning of a new year and the cycle repeating itself.

From analyzing the past data, the store is expected to generate a little over $5M in revenue across all segments in the next year while growing at the current pace.

A screenshot of a computer

Description automatically generated