****

Harshil Ajay Bhatt

Description:

|  |
| --- |
| VIP Tires and Services is a tire and automotive services company based in the New England region of the United States. They operate a network of 67 retail locations across New England and the Mid-Atlantic regions that generate $120m revenue annually. They offer a range of services including tire sales, installation, and repairs, as well as automotive repairs and maintenance. The client is soon going to open 2 more stores taking their store count to 69 retail locations. The firm has partnership with manufacturers who reward them with volume bonus on meeting certain thresholds every year. The warehouse is primarily used to stock up these high priority brands.  The firm wants to analyse sales patterns for tires on a store-by-store basis to understand which brands are most popular in each location. By examining sales data alongside external factors such as economic conditions and weather patterns, the firm hopes to identify any trends or correlations that may be influencing customer preferences. Subsequently, these trends and correlations are expected to be leveraged to predict the sales of ‘volume bonus’ brands. These predictions will help the firm in taking informed decision when designing marketing strategies. Ultimately, the goal is to leverage these insights to improve sales and customer satisfaction across all store locations, while also remaining responsive to the unique needs and preferences of each local market. |

Key Stakeholders

|  |  |  |
| --- | --- | --- |
| Name | Title | Project Role |
| Lynn Campbell | Director of Marketing, Advertising & CRM at VIP Tires & Service | Primary stakeholder |
| Blair McGaughey | Director of Tire & Wheel Merchandising | Secondary stakeholder |
| Darren Deschenes | Manager, Technology Development & Support | Primary contact for data related issues |

Potential Timeline:

Additional analyses such as customer affinity, SKU rationalizing

Building Excel based tool to aid in easy assimilation of insights into strategy building

Regression Analysis of Clusters to predict sales of high priority brands

Overlaying demographic variables (Store & tire demographics) data on store clusters

Creation of Analytical Data set, Store Clustering

Exploratory Data Analysis

Literature Review

Understanding data and raising concerns in case of inconsistencies

# **INTRODUCTION**

# VIP tires and Services is an automotive services and tire provider company. The team analysed sales data to analyse the following:

# Effect of external factors on sales of various brands

# Identifying the areas of growth opportunity for each brand in various markets

# Analyse data to increase quotas and annual cash back bonuses

# Moreover, the variables and data were analysed to perform the following tasks:

* Identification of the key variables, target variable (y), explanatory variables (main x) and control variables (control x)
* Identifying the data frequency and the reasons to operate on that frequency
* Exploring the summary statistics for the identified quantitative variables across the key target variables
* Creating a summary graph for the categorical variables
* Exploring the corresponding relationships between key variables through Pearson and Spearman correlations. Further, using a graphical representation to analyse the relationship visually
* Devising a treatment strategy for missing values and stating the reasoning to support the treatment adopted
* Defining outliers, identifying the treatment strategy for outliers, and identifying the corresponding correlations
* Performing clustering analysis
* Creating a summary of the analysis

**BUSINESS QUERIES:**

1. Identifying different groups of stores in the entire network of VIP Tires and Services
2. Analysing the impact of external factors on various groups of stores
3. Identifying underperforming brands across each group in order to enhance brand performance with respect to consumer perspective

**ANALYSIS AND RESEARCH METHODS**

**Key variables, target variable (y), explanatory variables (main x) and control variables (control x)**

Table1: Variable analysis

|  |  |
| --- | --- |
| **Variable** | **Description** |
| **CustomerId** | The ID of the customer |
| **Store#** | The ID of the store |
| **Zipcode** | The zip code of the customer (most likely his residence zip code) |
| **WcYear** | The model year of the car |
| **WcMakeText** | The make of the car (Brand of the vehicle) |
| **WcModelText** | The model of the car |
| **WcEngineText** | Type of engine of the car |
| **Date** | Date of transaction |
| **SkuCode** | Code for Stock Keeping Unit (Unique identifier for a product- can be considered as product ID) |
| **Quantity** | Number of tires sold |
| **Total Price** | Total sale price for the number of tires sold |
| **Price per Tire** | The price of each tire |
| **Discount** | The discount offered for the type of tire |
| **GP$** | The gross profit incurred |

***Key Variables***: **Sales, Tire quantity, Gross Profit and Gross Profit %**

The above stated measures along with number of transactions and number of customers would be used while forming clusters. Apart from these measures, sales mix would also be added in clustering. Other dimensions such as tire brand, vehicle brand would be overlaid on the clusters to identify patterns. Store cluster and month level to predict overall tire quantity based on independent factors. Time-series forecasting.

***Target variable:*** The tire quantity will be selected as the target variable after performing unsupervised learning. The initial part of the project comprises of unsupervised learning analysis, hence there will be no target variable. Further, the algorithm developed will be used to identify store clusters. The latter stages of the project would include regression analysis, for which, the store cluster and month level analysis would be performed to amalgamate **tire quantity** (target variable) into the strategy of volume bonus

***Explanatory Variables:*** The initial analysis based on unsupervised learning would comprise of **tire name, tire type, WcYear, WcMakeText and WcModelText** as the explanatory variables. While performing regression analysis, store cluster and month level variable would be used to perform time series forecasting analysis.

***Control Variable:*** The analysis does not comprise of any control variable since no variables are being kept constant.

**Data frequency**

The dataset currently is at customer-store- date- vehicle type (Model year, vehicle brand, vehicle model and Model engine)-skucode level. The dataset represents the specifics of the tire transactions. For clustering, the dataset would be transformed at store level by considering the recent 104 weeks data. Once, the relevant clusters are identified, the dataset will be modified at store-cluster-month level to build regression/time series models for different clusters. In the next step, the predictions for the next 12 months would be made for each cluster. Eventually, these predictions coupled with volume bonus information would be used to build marketing strategies. The 104 weeks data would enable the team to capture different components of the time series analysis such as trend level, seasonality.

**PRELIMINARY DATA PROCESSING**

A statistical summary of the dataset was obtained which generated the given below result.

Fig1: Statistical summary of the dataset

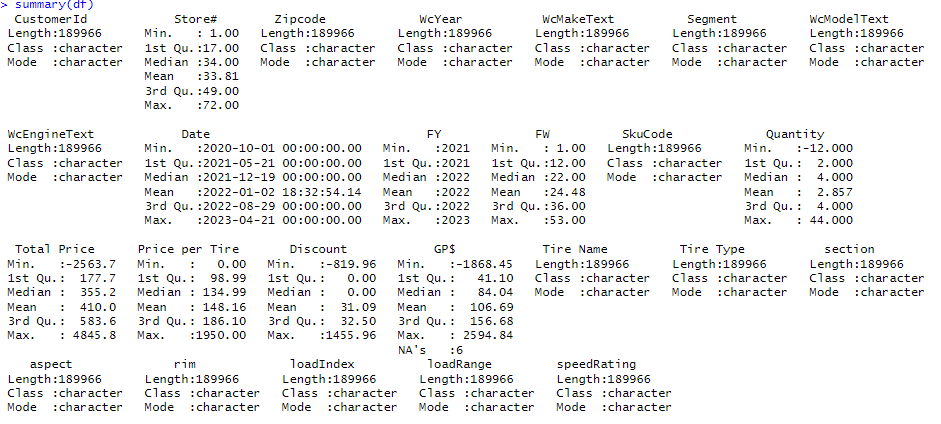


Figure 1 provides the summary of distribution of each variable for the given dataset. There are a total of 189 entries. While examining the dataset, there were negative values in Tire quantity column, indicating the presence of returns.

Fig2: Variance of subset

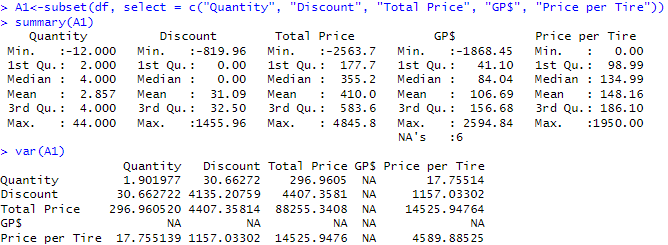
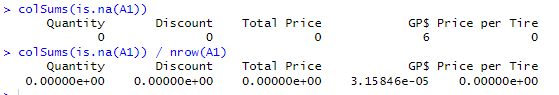


Fig3: Missing values calculation



The total number of missing values was calculated. The result showed that there are no missing values in the measures such as "Quantity", "Discount", "Total Price", and "Price per Tire". During the initial examination, there were few rows where all the measures were 0, hence they were removed. Post identification of the missing values in measures, we looked at the missing rate per dimensions.

Table2: Missing values analysis

Table3: Missing values analysis by Year

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **# of Records with Missing Values for Fiscal Year** | | | **% of Records with Missing Values** | | |
| **Dimension** | **2021** | **2022** | **2023** | **2021** | **2022** | **2023** |
| CustomerId | - | - | - | 0.0% | 0.0% | 0.0% |
| Store# | - | - | - | 0.0% | 0.0% | 0.0% |
| Zipcode | 220 | 785 | 618 | 0.3% | 1.0% | 1.3% |
| WcYear | 14,894 | 15,745 | 7,955 | 22.0% | 20.7% | 17.2% |
| WcMakeText | 14,800 | 15,573 | 7,875 | 21.9% | 20.4% | 17.0% |
| Segment | 14,800 | 15,573 | 7,875 | 21.9% | 20.4% | 17.0% |
| WcModelText | 14,971 | 15,707 | 7,944 | 22.2% | 20.6% | 17.2% |
| WcEngineText | 15,202 | 15,966 | 8,187 | 22.5% | 21.0% | 17.7% |
| Date | - | - | - | 0.0% | 0.0% | 0.0% |
| FY | - | - | - | 0.0% | 0.0% | 0.0% |
| FW | - | - | - | 0.0% | 0.0% | 0.0% |
| SkuCode | - | - | - | 0.0% | 0.0% | 0.0% |
| Tire Name | 1,002 | 596 | 261 | 1.5% | 0.8% | 0.6% |
| Tire Type | 6,776 | 8,411 | 4,843 | 10.0% | 11.0% | 10.5% |
| section | 1,007 | 610 | 443 | 1.5% | 0.8% | 1.0% |
| aspect | 2,174 | 1,602 | 830 | 3.2% | 2.1% | 1.8% |
| rim | 1,007 | 610 | 443 | 1.5% | 0.8% | 1.0% |
| loadIndex | 3,284 | 3,409 | 1,358 | 4.9% | 4.5% | 2.9% |
| loadRange | 6,004 | 6,648 | 4,914 | 8.9% | 8.7% | 10.6% |
| speedRating | 4,802 | 4,723 | 2,199 | 7.1% | 6.2% | 4.8% |

Customer’s vehicle dimensions had higher percentage of missing values when compared to other dimensions. These values accounted for 20% of the transactions where the details of the customer’s vehicle were not captured properly. The variable WcEngineText had the highest number of missing values while Date, FY, FW, SkuCode had missing values. The missing values for customers (20%) was not treated since imputing values would have changed dimensions and the distribution. The tire type had “none” values which were treated as missing values. Invalid zip codes were categorised under missing values.

The missing rate analysis per year highlighted that the data was captures more in the recent years and there was a decline in the missing values.

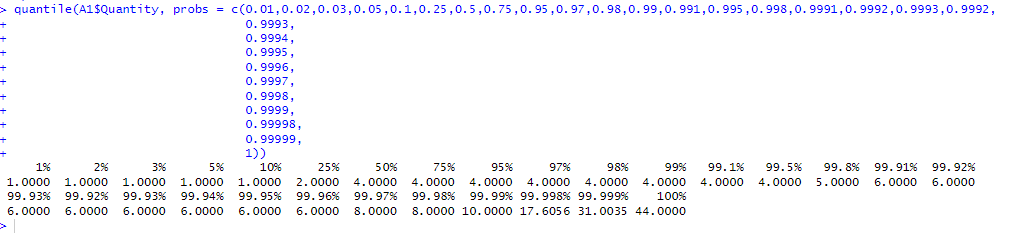
**Outliers’ treatment and corresponding correlations**

Figure 4 describes the percentiles of Tire Quantity variable. The probabilities ranging from 0.01 to 1 were calculated with varying increments. The output showed the values of the quantiles for each of the specified probabilities. The same exercise was done for the other variables, i.e., Discount, Total price and Price Per Tire.

It was observed that outliers were present at the transaction level. There were 16 transactions with tire quantity greater than 10, which were eliminated owing to the unlikely nature of such transactions in real world.

After eliminating the transactions based on tire quantity, we noticed that some of the extreme values which appeared in other measures were also eliminated. For example, transactions with negative tire quantity(returns) which had negative sales were taken care of when we removed them based on tire quantity alone.

Fig4: Quantiles of given probabilities



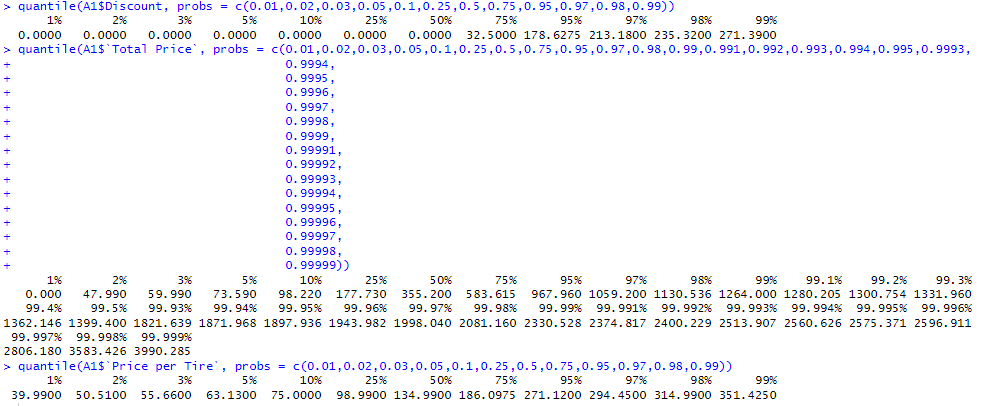
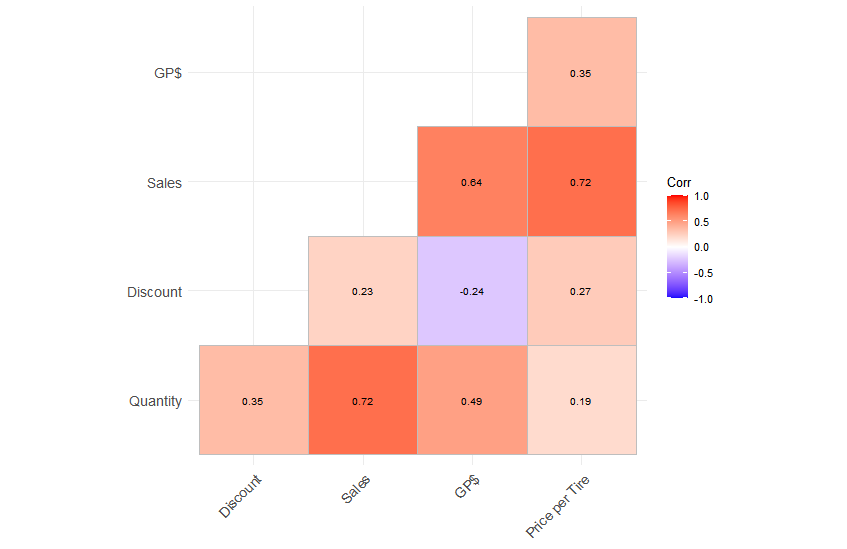


Fig5: Correlation plot

A correlation plot was generated to demonstrate the correlation coefficient for each pair of variables. As expected, there was high positive correlation of 73% between sales and tire quantity.

**Leading tire brand and store analysis**

The total sales generated with 542819 number of tires was $7,78,92,461 and $2,02,65,875 Gross Profit. The brands were ranked based on Sales, Tire Quantity and Gross Profit. The overall Rank was assigned by giving equal weightage to all the three ranks. The Gross Profit% index showed the relative profitability of the brand with respect to overall Gross Profit%. For instance, the brand Goodyear’s GP% was less than the overall GP% of VIP tires which implied that it was not as profitable as the other brands. Goodyear, Michelin, and BF Goodrich had the highest negative GP% index while Multi-Mile, Vanderbilt and Vredestein had the highest positive GP% index. We replicated the similar exercise to identify the top 10 stores. It was observed that store 46 had the highest GP%.

Table4: Brand analysis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Brand** | **Sales Rank** | **Tire Quantity Rank** | **Gross Profit Rank** | **Overall Rank** | **Gross Profit % Index** |
| GOODYEAR | 1 | 2 | 2 | 2 | -32 |
| MULTI-MILE | 2 | 1 | 1 | 1 | 36 |
| SUMITOMO | 3 | 3 | 3 | 3 | 24 |
| MICHELIN | 4 | 7 | 6 | 5 | -43 |
| FALKEN | 5 | 5 | 5 | 4 | -5 |
| BF GOODRICH | 6 | 8 | 9 | 8 | -38 |
| VANDERBILT | 7 | 6 | 4 | 5 | 40 |
| SOLAR | 8 | 4 | 7 | 7 | 29 |
| VREDESTEIN | 9 | 9 | 8 | 9 | 36 |
| KELLY | 10 | 10 | 11 | 10 | -16 |

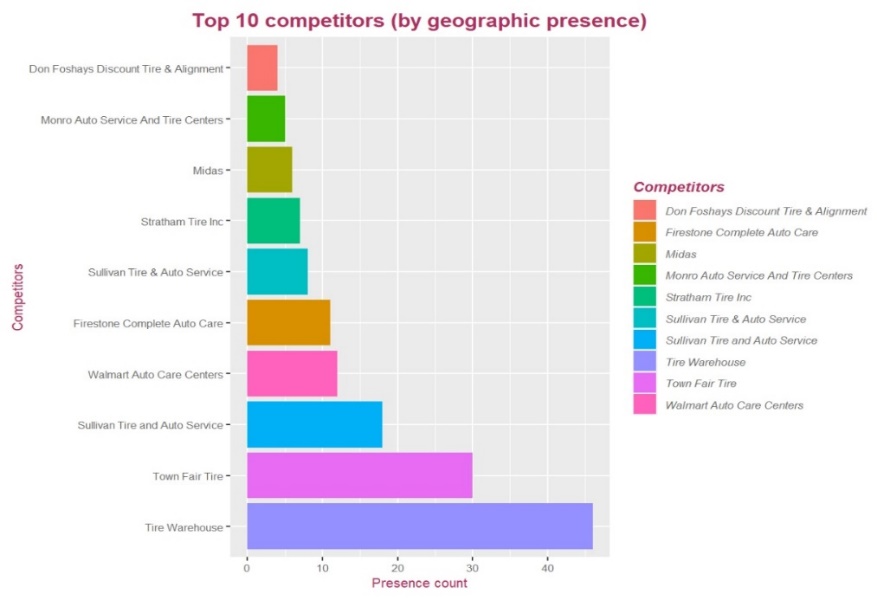
Table5: Store analysis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Store Number** | **Sales Rank** | **Tire Quantity Rank** | **Gross Profit Rank** | **Overall Rank** | **Gross Profit % Index** |
| 28 | 1 | 1 | 2 | 1 | -9 |
| 17 | 2 | 4 | 3 | 3 | -1 |
| 41 | 3 | 3 | 5 | 4 | -9 |
| 46 | 4 | 2 | 1 | 2 | 15 |
| 42 | 5 | 6 | 6 | 5 | -1 |
| 37 | 6 | 7 | 4 | 5 | 8 |
| 15 | 7 | 8 | 10 | 8 | -3 |
| 19 | 8 | 11 | 8 | 9 | 0 |
| 14 | 9 | 5 | 9 | 7 | -1 |
| 8 | 10 | 10 | 11 | 10 | 1 |

**Top competitor analysis**

The graphical representation given below depicts the top 10 competitors across various city/states. Tire warehouse had the highest number of stores across all the cities under consideration while Don Foshays had the least number of stores. On average, the number of competitor stores was between 5 to 20.

Fig6: Competitor analysis

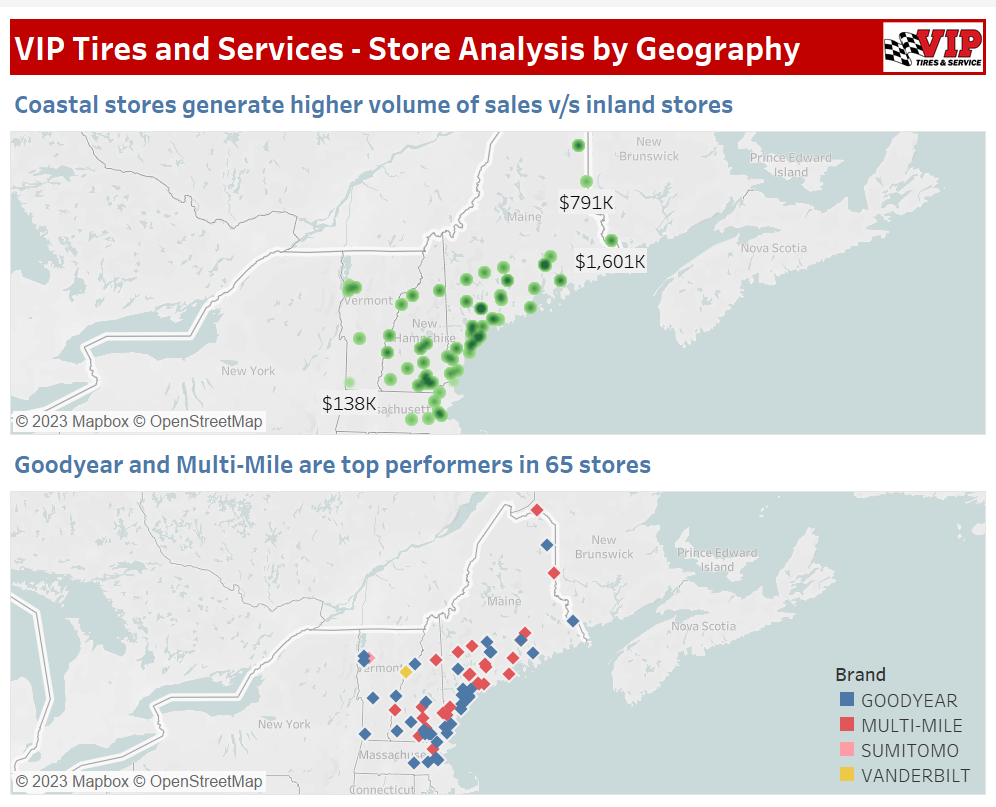


**Store Analysis by Geography**

In general, the coastal stores generated higher volume of sales in comparison to the inland stores. Goodyear and Multi-Mile are the top performers in 65 out of 67 stores. Sumitomo and Vanderbilt topped sales in 1 store each. Interestingly, these 2 stores lie away from other stores where one lies near the border of Vermont and New Hampshire and the other is based out of Vermont.

While analysing the customer base, we noticed that majority of the sales originated from the domicile states of Vermont, New Hampshire, Maine, and Massachusetts. Additionally, most of the customers are based out of eastern side of United States. However, the Sponsors intimated that the zip codes were not reliable and therefore, the inferences might not be accurate.

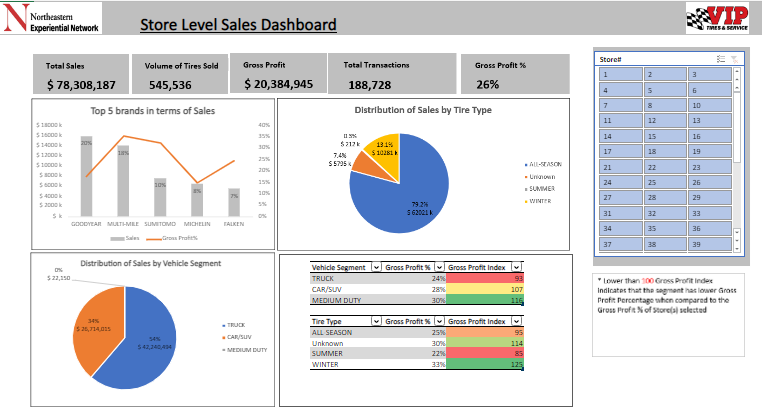
Fig7: Store analysis



**Sales Dashboard**

The team further created a Sales Dashboard in Excel using Pivot tables and charts to understand the store level dynamics on a more fundamental level. Using the dashboard, we are identifying the top 5 brands in terms of sales and their corresponding Gross Profit %. Also, we noticed that higher proportion of retail customer base owns vehicles belonging to Truck segment. As expected, majority of the sales is driven by All-season tires. We also looked at the profitability of different tire types and vehicle segments.

Fig8: Sales Dashboard



**CLUSTERING**

To begin clustering, we rolled up the dataset after eliminating outlier transactions. As stated above, outlier transactions were identified after looking at the percentile values of the tire quantity. The outlier transactions mainly comprised of transactions which had less than 0 tire quantity or above 10 tire quantity. The dataset was grouped by store level with measures such as

1. Total\_Sales
2. Tire\_Quantity
3. Gross\_Profit
4. Total\_Discount
5. COVID\_Sales: Sales generated from Sep 2020 – Dec 2021). The time period selected was based on research which alluded that Covid started to wane by the end of 2021
6. Total\_Transactions
7. Customer\_Count
8. Gross\_Profit\_Percent
9. Sales\_BF GOODRICH
10. Sales\_FALKEN
11. Sales\_GOODYEAR
12. Sales\_KELLY
13. Sales\_MICHELIN
14. Sales\_MULTI-MILE
15. Sales\_SOLAR
16. Sales\_SUMITOMO
17. Sales\_VANDERBILT
18. Sales\_VREDESTEIN

Fields from 9-18 are store level brands sales for the VIP’s top 10 brands. After dataset creation, 2 clustering algorithms namely K-means clustering and Agglomerative Hierarchical clustering were implemented on the scaled dataset. The dataset was scaled to avoid the algorithm from giving more weightage to the variables whose scale is too high. The sections below give a detailed elaboration of the results achieved via the 2 methods.

**K-means clustering**

K-means clustering is an unsupervised algorithm used for partitioning a dataset into distinct groups or clusters based on their similarity. An iterative algorithm aims to minimize the within-cluster sum of squares (WCSS) by assigning data points to the nearest cluster centroid. K-means clustering is one of the easiest methods to implement in any programming language as it is computationally inexpensive. To perform K-means clustering first we need to input different k values repeatedly to find the optimum number of clusters. Post applying the elbow method we found that 4 was the optimal number of clusters as there was no sharp decline in WSS with the increase in clusters.

Fig9: K-Means Clustering

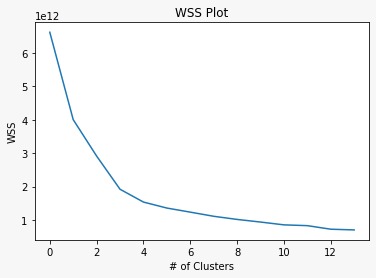
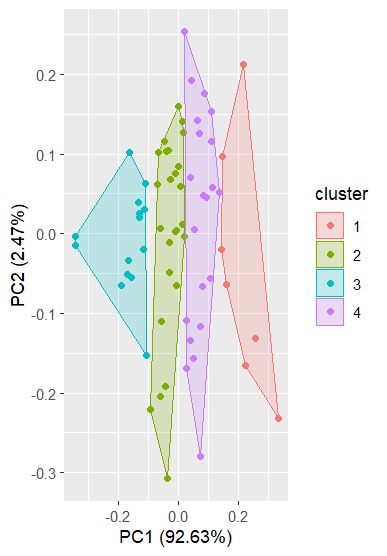


Fig10: After Scaling Clusters



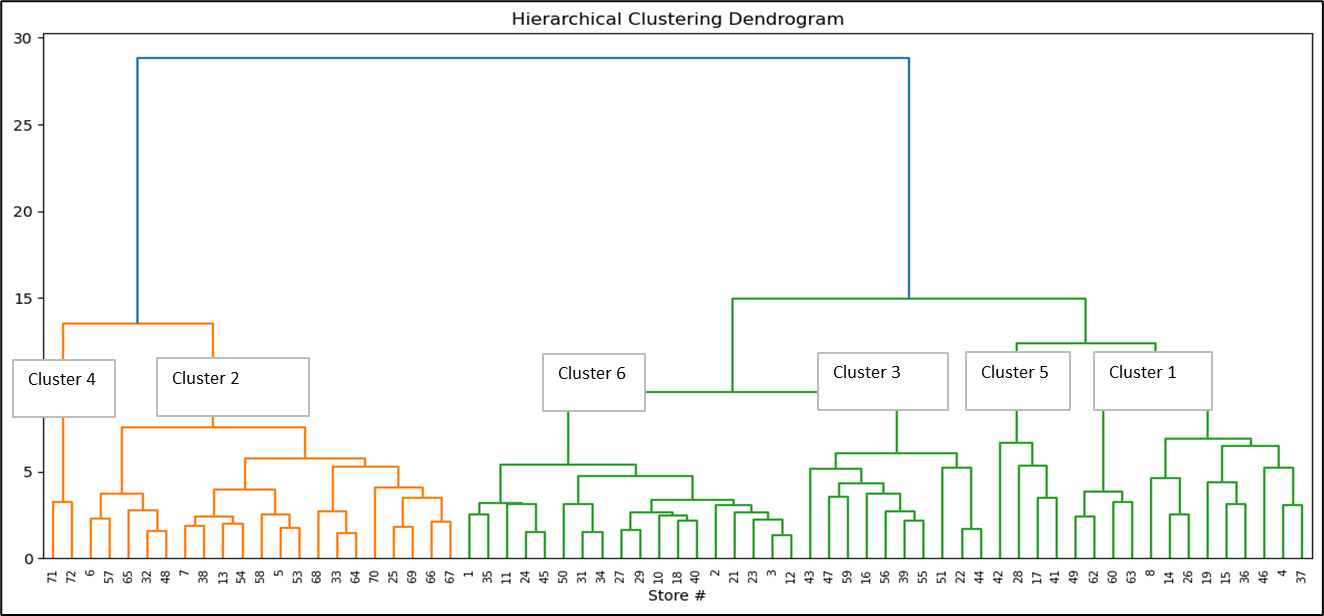
The clustering exercise was implement using R and Python in parallel. Post examining the size of the clusters for k=4, we noticed that the results generated from Python and R were different. This is primarily because of the randomness of the algorithm in selecting the initial centroids. Hence, we shifted out focus to using more robust algorithm -Agglomerative Hierarchical clustering.

**Hierarchical clustering**

Hierarchical clustering [6], [7] is an unsupervised learning method for clustering data points. This algorithm creates clusters by measuring the difference between data. We start by treating each data point as its own cluster. Clusters with the shortest distance between them are then combined to form larger clusters. This step is repeated until a large cluster containing all data points is formed. The Euclidean distance was used to calculate the distance while Ward method was used to combine the clusters.

Given below are the diagrammatic representation of hierarchy of the stores.

Fig11: Hierarchical Clustering



After generating the store cluster labels for each store, we overlaid demographic variables such as:

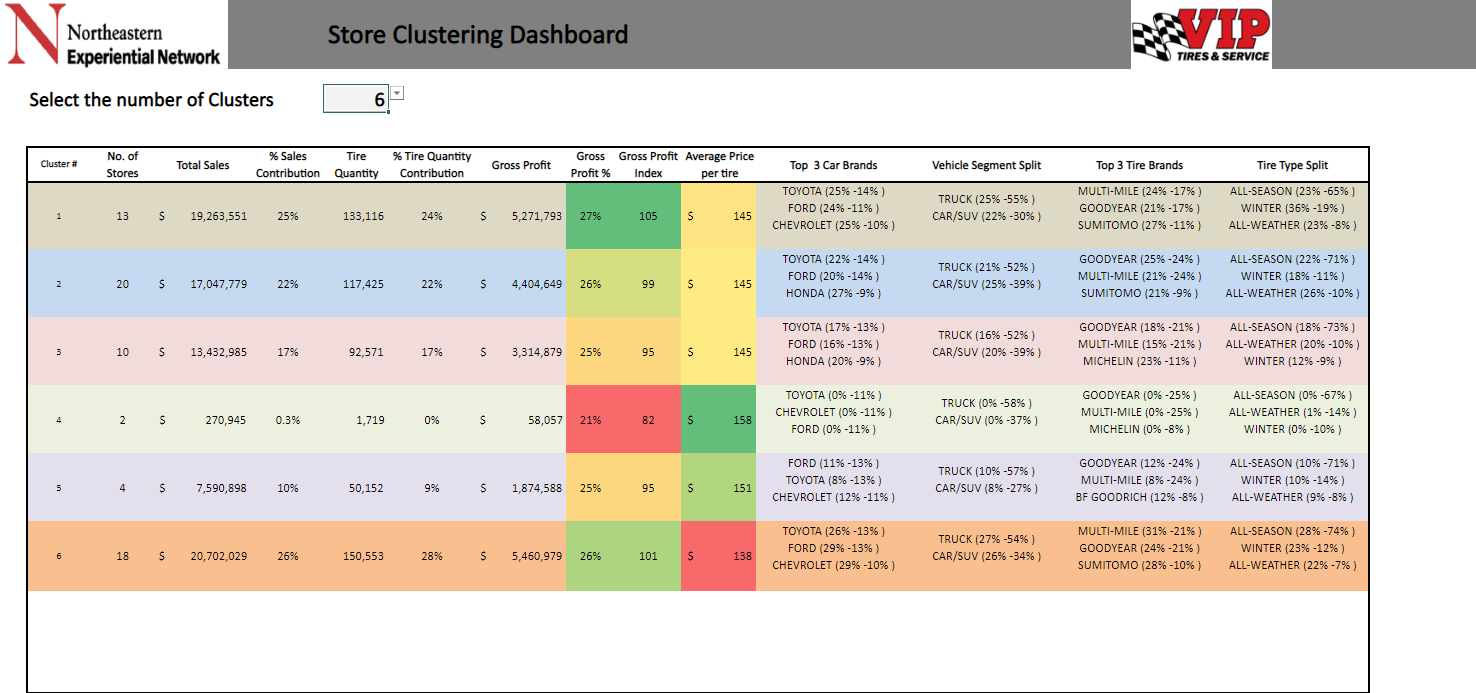
1. Vehicle Manufacturer
2. Vehicle Segment
3. Tire Brand
4. Tire Type

Using clustering output and the demographic output, we built a user-friendly Excel based dashboard to aid the better understanding of the clusters and their features. The user can change the number of clusters from the drop-down menu available to move up/down the hierarchy of the stores and identify their relevant features. Here, we introduced a metric called Gross Profit Index which is calculated using below formula:

Gross Profit Index = Gross Profit of Cluster/ Gross Profit of entire VIP

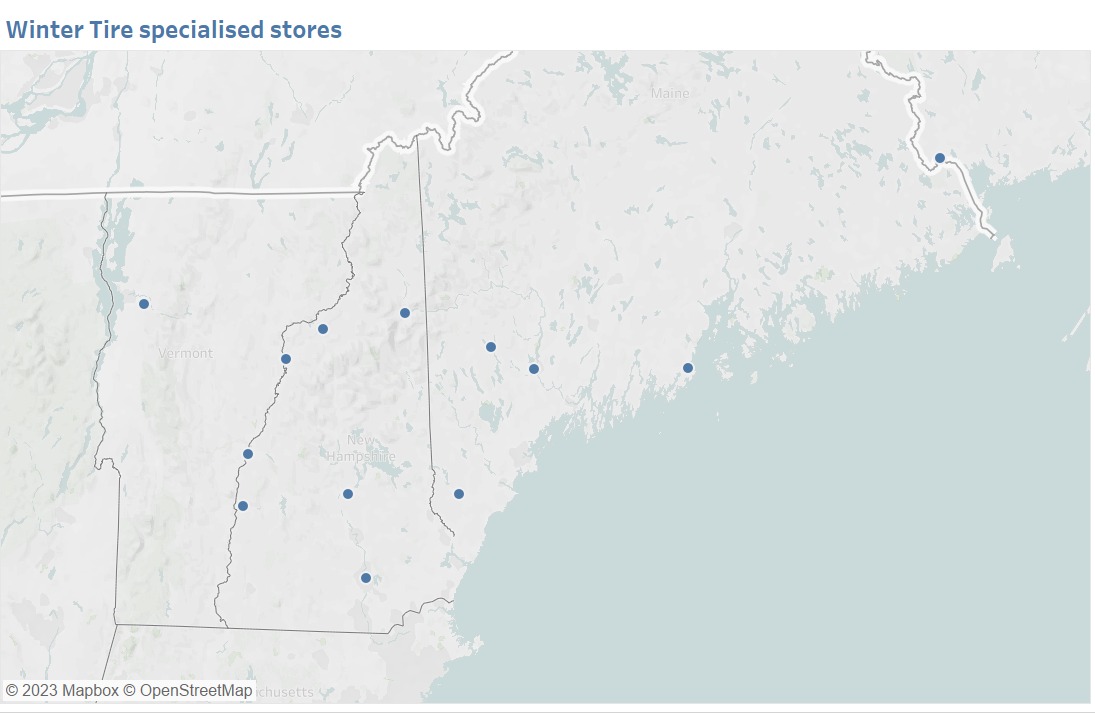
For Gross profit Index, the baseline is 100 and worse performance is lower than 100. If Gross Profit Index is lower than 100, it is interpreted that the corresponding cluster is not generating same gross profit % as the overall network and is slightly underperforming.

Fig12: Store-Clustering Dashboard

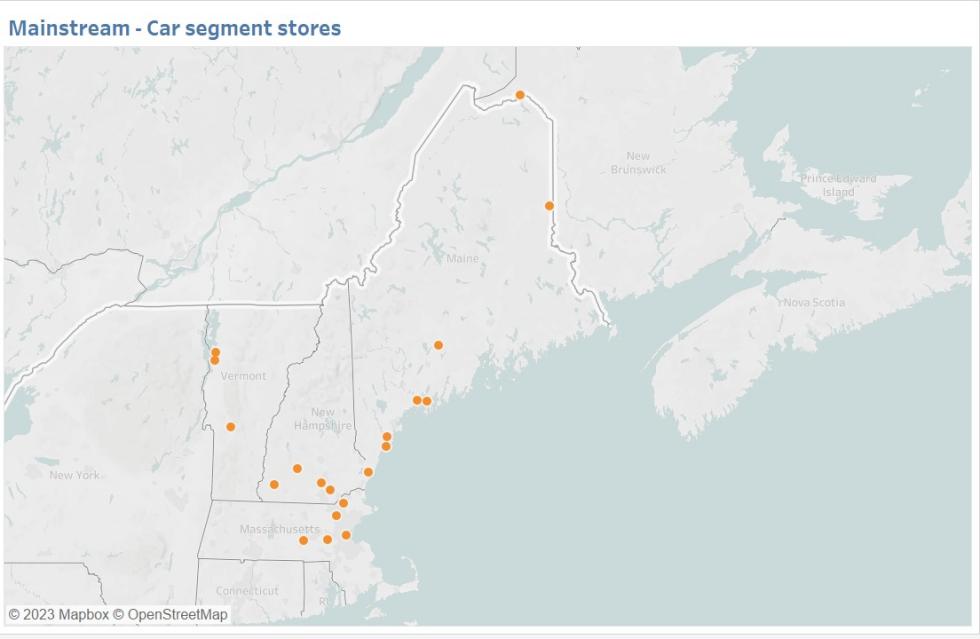


Currently, we have identified 6 as the optimal number of clusters from this exercise. Given below is a detailed explanation of the clusters.

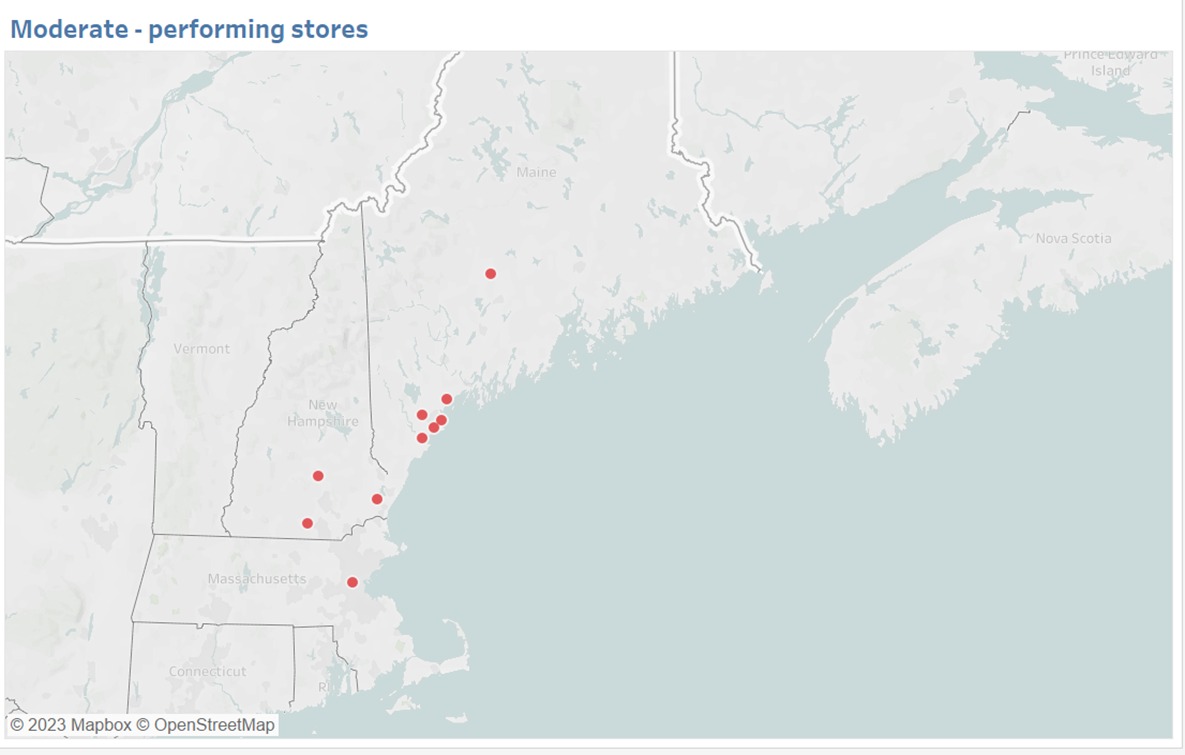
1. **Winter – Tire specialized stores:** This cluster comprised of 13 stores which contributed to 25% of the overall sales. This cluster had higher proportion of sales (19%) incurred from Winter tires when compared to other store clusters (13%). The stores were mostly present at inland locations as expected.



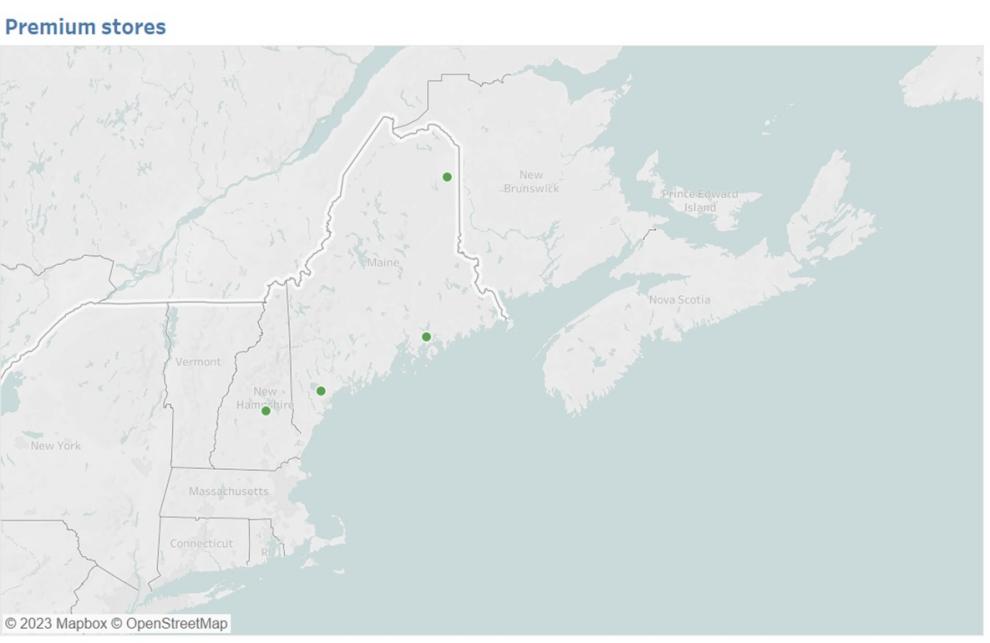
1. **Mainstream – Car segment driven stores:** This cluster is the second largest cluster comprising of 20 stores which contributed to 22% of the overall sales. This cluster has higher proportion of sales (39%) by customers owning Cars/SUVs when compared to other store clusters (33%). The stores are majorly present in New Hampshire and Massachusetts.

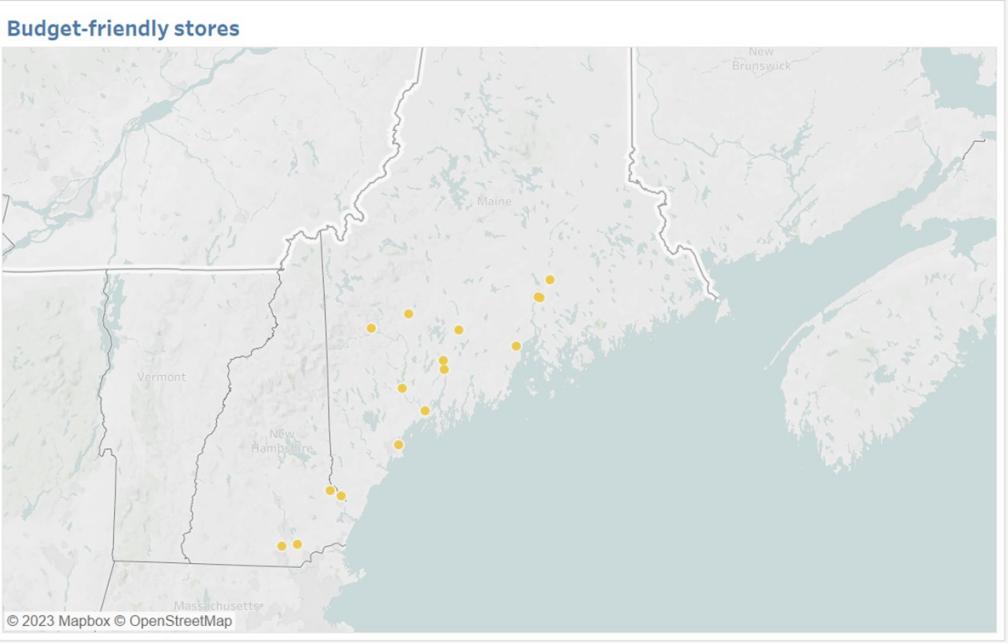


1. **Moderate – Performing stores :** This cluster consist of 10 stores and contributed 17% in overall sales. The average price per tire was $145 which was neither too high nor too low. It comprised of one of the highest numbers of All-season selling tires.



1. **Recently opened stores:** This was the smallest cluster with only 2 stores. Currently, the Gross Profit % of these stores is the lowest despite the high average price per tire of $158. 
2. **Premium customer stores:** This cluster comprised of 4 stores which contributed to 10% of the overall sales. The average price per tire was $151 which was higher than average spend per tire of other clusters except for recently opened stores. 3 of the stores were near state capitals or largest urban agglomerations.



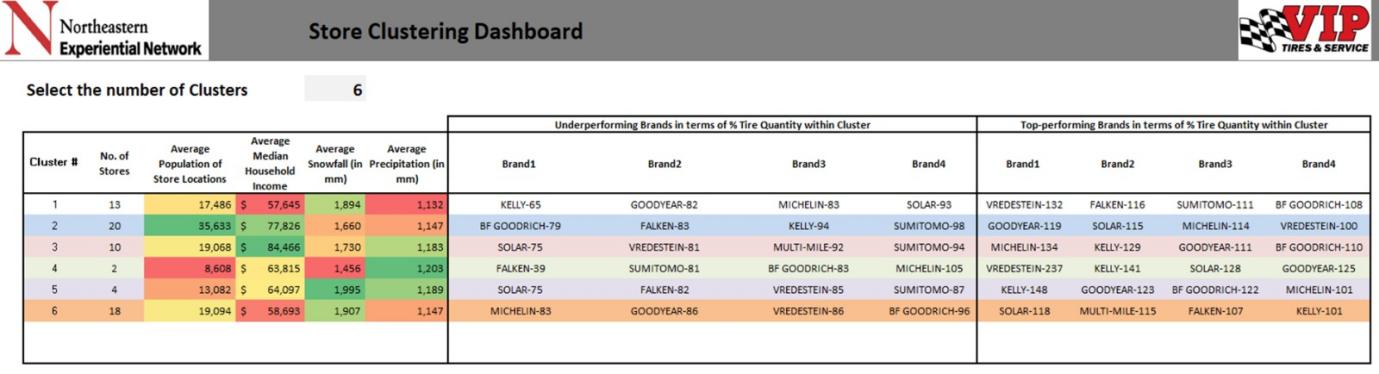
1. **Budget-friendly customer stores:** This cluster is the largest cluster comprising of 18 stores which contributed to 26% of the overall sales. The customers in these stores tend to spend less on a tire as compared to other store clusters. Upon examining the store locations, we noticed that majority of them are from coastal front. 

**EXTERNAL FACTOR ANALYSIS**

For the external analysis for our data, 4 factors were taken into account- Average population of store locations, Average median household Income, Average snowfall(mm) and Average Precipitation(mm). The data had been retrieved from American Community Survey 2017-2021 5-Year Data Release (Population, Household income, Precipitation levels) along with American Community Survey 2010-2014 (Average snowfall data).

The dashboard created consisted of underperforming brands with a relative number given to the brand by dividing the brand’s contribution within the cluster to the brand contribution to the overall quantity. For example, in Cluster 1, Kelly was assigned a number 65 by dividing Kelly’s contribution to Cluster 1 with Kelly’s contribution to the overall quantity.

Fig13: External factors-based cluster analysis

****

1. **Winter – Tire specialized stores:** These stores have lowest average median household income of $57,645. The location of these stores receives 1,894mm of average snowfall, making winter tires preferable for the customers and contributing 19% in sales. Vredestein, Falken, Sumitomo and BF Goodrich are top performing tire brands in these stores. Kelly, Goodyear, Michelin, and Solar are under-performing brands with respect to tire quantity %.
2. **Mainstream – Car segment driven stores:** These stores have highest Average population at store

locations - 35,633 and second Average highest income of $77,826. The top performing brands of this cluster are GoodYear, Solar, Michelin, Vredestein and under-performing brands are BF Goodrich, Falken, Kelly and Sumitomo.

1. **Moderate – Performing stores:** Comprising of 10 stores, this cluster has the highest Average median household income of $84,466.The average tire price at these stores is lowest ($145). The best performing brands here are Michelin, Kelly, GoodYear and BF Goodrich. The under-performing tire brands are Solar, Vredestein, Multi - Mile and Sumitomo.
2. **Recently opened stores:** This cluster comprises of only 2 stores and has the lowest average population of 8,608 nearby the store locations. The average precipitation is highest in the region with 1,203mm.

Although, these stores have lowest population but the average tire prices are the highest $158. The best performing brands here are Vredestein, Kelly, Solar and GoodYear. Under-performing brands are Falken, Sumitomo, BF Goodrich, and Michelin with respect to tire quantity %.

1. **Premium customer stores:** Consisting of 4 stores and an average household income of $64,097, the best performing tire brands in this cluster are Kelly, GoodYear, BF Goodrich and Michelin. Whereas the Under- performing brands are Solar, Falken, Vredestein and Sumitomo.
2. **Budget-friendly customer stores:** This is the largest cluster and has second lowest income of $58,893. The top performing brands are Solar, Multi - Mile, Falken and Kelly. The least performing brands are Michelin, GoodYear, Vredestein and BF Goodrich.

**PREDICTING TIRE QUANTITY FOR 2023 USING SIMPLE INTERPOLATION**

After plotting the tire quantity trends from October 2020 to April 2023 across the six clusters, we observed that some clusters exhibit similar seasonality for tire quantity while having different brand mix internally. For instance, Cluster 1 and Cluster 6 have similar trends but differ in terms of the brands that perform better. In Cluster 1, specialized winter tire stores, Vredestein, Falken, Sumitomo, and BF Goodrich perform well, whereas in Cluster 6 Budget Friendly Stores - Solar, Multi-Mile, Falken, and Kelly are the better-performing brands. All clusters except Cluster 4, the tire quantity was mostly stationary.

Cluster 4 comprised of recently opened stores, resulting in limited available data of the most recent 8 months till April 2023. It is to be noted that even for April, data captured only 20 days. Therefore, we used the overall network ratios for this cluster. Consequently, we decided to make predictions at the cluster level rather than at the overall level. While predicting at the overall level is not flawed, there are no specific trends that can be observed.

Fig14: Cluster wise Tire quantity sales analysis

The chart displays the general trends for tire quantity, with the rise and fall of the trend line relating to the promotional activities. The tire quantities typically reach their peak during October to December, likely due to the sale of winter/all-weather tires, and return to normalcy in the subsequent months. Cluster 4 was excluded from the overall trend analysis since it consisted of recently opened stores. Our focus was primarily on understanding and analyzing the trends of the existing stores.

Fig15: Overall Tire quantity sales analysis

*\*Excluding cluster 4 of recently opened stores*

We conducted an analysis at the cluster level, excluding Cluster 4, to calculate the ratio of yearly tire quantity to the tire quantity in the first quarter for the years 2021 and 2022. Since we did not have complete data for the month of April, we focused on the first quarter only. Using this ratio, we multiplied it by the tire quantity of the first three months of this year (2023) to predict the overall tire quantity for the year 2023 for the corresponding cluster.

Table6: Overall Tire quantity prediction

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | |  |
|  | **Cluster #** | | | | | |
| **Year** | **1** | **2** | **3** | **5** | **6** | **Grand Total** |
| 2021 | 5.5 | 5.4 | 4.9 | 5.7 | 5.2 | 5.3 |
| 2022 | 5.8 | 5.5 | 5.0 | 5.7 | 5.2 | 5.4 |
| **Average** | 5.6 | 5.4 | 4.9 | 5.7 | 5.2 | 5.4 |

|  |  |
| --- | --- |
| **Year** | 2023 |
| **Month** | (Multiple Items) |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Cluster #** | | | | | | |
|  | **1** | **2** | **3** | **4** | **5** | **6** | **Grand Total** |
| **Sum of Quantity** | **8,268** | **9,158** | **6,999** | **755** | **3,354** | **10,467** | **39,001** |
| **2023 Prediction** | **46,671** | **49,844** | **34,609** | **4,046** | **19,152** | **54,693** | **209,007** |

Subsequently we calculated the average of this ratio for every cluster as demonstrated below. We then utilized this average ratio to predict the overall tire quantity for 2023 by multiplying the tire quantity of the first three months of 2023 with the previously obtained ratio. However, for Cluster 4, we used the overall ratio of 5.4 for the prediction. It is to be noted that even for the overall ratio, we used all clusters except Cluster 4.

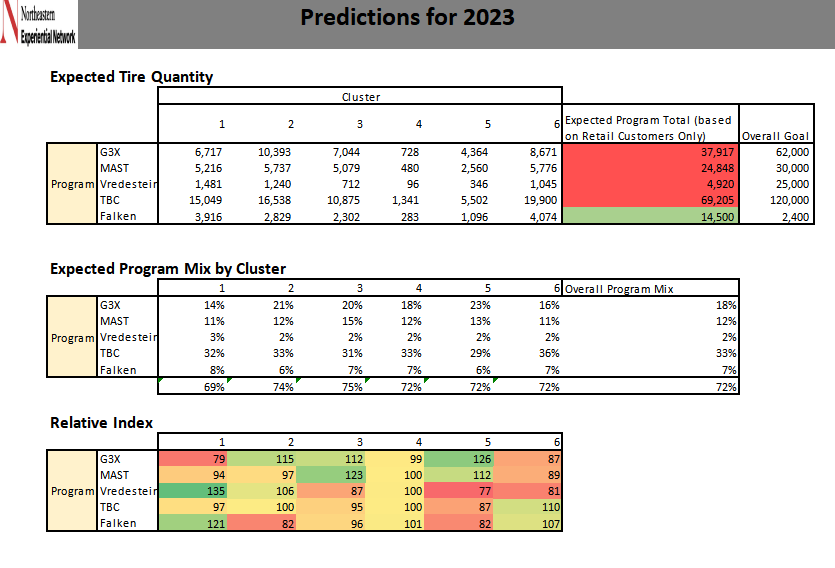
Table7: Program contribution to sales

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **2021 and 2022 tire quantity by program** | | | | | | | |
| **Cluster #** | **1** | **2** | **3** | **4** | **5** | **6** | **Total** |
| **G3X** | 14,785 | 19,405 | 14,499 | - | 8,707 | 18,369 | 75,765 |
| **MAST** | 11,482 | 10,711 | 10,454 | - | 5,109 | 12,235 | 49,991 |
| **Vredestein** | 3,259 | 2,316 | 1,466 | - | 691 | 2,214 | 9,946 |
| **TBC** | 33,125 | 30,878 | 22,384 | - | 10,979 | 42,157 | 139,523 |
| **Falken** | 8,619 | 5,282 | 4,739 | - | 2,187 | 8,630 | 29,457 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **2021 and 2022 tire quantity by program** | | | | | | | |
| **Cluster #** | **1** | **2** | **3** | **4** | **5** | **6** | **Total** |
| **G3X** | 14% | 21% | 20% | - | 23% | 16% | 18% |
| **MAST** | 11% | 12% | 15% | - | 13% | 11% | 12% |
| **Vredestein** | 3% | 2% | 2% | - | 2% | 2% | 2% |
| **TBC** | 32% | 33% | 31% | - | 29% | 36% | 33% |
| **Falken** | 8% | 6% | 7% | - | 6% | 7% | 7% |

In the presented table, we examined the tire quantity at the volume bonus program level for different clusters by considering tire quantities of years 2021 and 2022. Cluster 4 was excluded from the table due to incomplete information. During the analysis, we calculated and observed the distribution of tires sold under various volume bonus programs and determined their proportional mix.

Fig16: 2023 predictions



Next, we predicted tire quantity sold in 2023 and historical proportional mix of volume bonus program at cluster level to predict the tire quantity. The expected program mix % at a cluster level was calculated followed by relative indexing to identify programs performance across each cluster. The relative index was calculated by dividing the program mix of the cluster by the program mix at the overall level. For example, for Cluster 1, the G3X Expected Program mix by Cluster for Cluster 1 was 14% and the overall program mix for the same cluster was 18%. By dividing the former percentage (14%) by the latter (18%), the relative index was obtained. This would be used to suggest boost of programs across each cluster. Moreover, the expected program goal was calculated for FY2023 by considering only retail customers. However, the overall goal was comprised of other customers as well. The table given below lists the brands under each program.

Table8: Brand list under each program

|  |  |
| --- | --- |
| **Brand Name** | **Program** |
| GOODYEAR | G3X |
| DUNLOP | G3X |
| KELLY | G3X |
| BF GOODRICH | MAST |
| MICHELIN | MAST |
| UNIROYAL | MAST |
| VREDESTEIN | Vredestein |
| SUMITOMO | TBC |
| MULTI-MILE | TBC |
| various | TBC |
| FALKEN | Falken |

By analysing the table and the tire quantities sold, it was noticed that GoodYear performed better than Kelly and Dunlop under the same program. To boost consumer spending and increase the tire quantities sold, A/B testing can be performed to identify clusters wherein the consumers preferred economical tires and branded tires respectively. This would enable the Sponsors to promote certain brands within each program according to the spending trend of the consumer.

**CONSUMER PERSPECTIVE**

The following table shows the factors influencing consumer perspective. It includes services, warranty, product quality, consumer awareness, advanced tech (technological advancements in tires), merchandising policy (brand support), adjustment policy (how helpful the brand is) and line coverage (types of tires covered under the brand). Michelin has been rated highest of 9.2 and 9.4 in terms of product quality and consumer awareness. GoodYear tops the list for advanced technology in their tires. Michelin has highest merchandising policy of 7.9. Multi - Mile is considered best when it comes to adjustment policy and line coverage with the ratings of 8.5 and 8.8 respectively. The benefits offered by these brands can be studied in order to understand the preference of these brands as compared to other brands.

Table9: Analysing consumer analysis of top brands

****

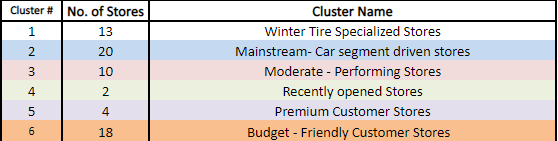
**ALTERNATIVE APPROACHES**

We tried seasonal ARIMA model to predict tire quantities for the remaining 2023. They were pretty much closer to the simple interpolation and therefore, we went forward with the method discussed above.

**CONCLUSION**

* Identified the invalid values for each dimension. For example, invalid zip codes are treated as missing values. Similarly, the WcYear values such as 0 and 1900 are treated as missing values as these are meaningless.
* Similarly, in transactions where either tire was purchased over the counter loosely for trailer or customer did not share vehicle type information, the vehicle related dimensions are not populated properly. Hence, the missing rate for these dimensions is as high as 20%. However, we would not remove these records from our analysis. While analysis, we will find top vehicle brands for cluster based by not including ‘missing’ values.
* For measures, we did not notice any missing values. However, there were close to 6 rows where each measure was 0. Hence, we will remove them. Additionally, we decided to remove weird transactions- transactions where tire quantity > 10 or < 1.
* Sales and tire quantity have decent correlation of 72%.
* Some of the top revenue generating tire brands have lower gross profit % when compared to the overall gross profit % of the business. These brands include Goodyear, Michelin, and BF Goodrich.
* Store #46 is one of the top performing stores as it features in top 6 stores in terms of sales and has high gross profit % index.
* 6 clusters obtained through Hierarchical clustering

Fig19: Store-Cluster Categories



* The external factors and consumer perspective helped to identify the characteristics of clusters and consumer inclination towards certain brands respectively. The external factors included Average population of store locations, Average median household Income, Average snowfall(mm) and Average Precipitation(mm). The consumer perspectives included services, warranty, product quality, consumer awareness, advanced tech (technological advancements in tires), merchandising policy (brand support), adjustment policy(how helpful the brand is) and line coverage(types of tires covered under the brand).
* The 2023 predictions for tire quantities sold under each program across each cluster was calculated and compared with the overall goal. The expected program mix % at a cluster level was calculated followed by relative indexing to identify programs performance across each cluster. This would be used to suggest boost of programs across each cluster. Moreover, the expected program goal was calculated for FY2023 by considering only retail customers. However, the overall goal was comprised of other customers as well.
* To boost consumer spending and increase the tire quantities sold, A/B testing can be performed to identify clusters wherein the consumers preferred economical tires and branded tires respectively. This would enable the Sponsors to promote certain brands within each program according to the spending trend of the consumer.
* The relative index of G3X program was the least in Cluster 1, Falken in Cluster 2, Vredestein in Cluster 3, TBC in Cluster 5, and MAST in Cluster 6. This showcased that these volume bonus programs can be boosted in these clusters according to the consumer spending trend and subsequently, devising appropriate marketing strategies. The above analysis could not be done for Cluster 4 as the segment seems to be inclined towards expensive tires could be because of effect of heavy discounts offered in these stores which in a way trigger decoy pricing effect.
* Lastly, the next steps were identified as exploring other forecasting methods to improve accuracy of tire quantity prediction and suggesting marketing strategies based on consumer analysis to boost underperforming brands across each cluster.

**REFERENCE**

1. MDPI | MDPI Journal List. (n.d.). <https://www.mdpi.com/about/journals>
2. Agarwal, K., Jain, P., & Rajnayak, M. A. (2019). Comparative Analysis of Store Clustering Techniques in the Retail Industry. In DATA (pp. 65-73)
3. Pattanaik, P. P., & Balu, V. (2021b). Plus size tire: Effect on the performance of the vehicle. Materials Today: Proceedings. <https://doi.org/10.1016/j.matpr.2021.03.431>
4. Guest. (2023, March 7). How to Identify and Promote Underperforming Products [+Tips]. Extensiv. <https://www.extensiv.com/blog/underperforming-products>
5. An assessment of tire-buying among millennial consumers. (n.d.). Retrieved April 27, 2023, from <https://ideaexchange.uakron.edu/cgi/viewcontent.cgi?article=1578&context=honors_research_projects>
6. Keita, Z. (2023, January 19). An introduction to hierarchical clustering in Python. DataCamp. <https://www.datacamp.com/tutorial/introduction-hierarchical-clustering-python>
7. Machine learning - hierarchical clustering. Python Machine Learning - Hierarchical Clustering. (n.d.). <https://www.w3schools.com/python/python_ml_hierarchial_clustering.asp>
8. 8th December 2022. American Community Survey 2017-2021 5-Year Data Release. [www.census.gov.](http://www.census.gov.) <https://www.census.gov/newsroom/press-kits/2022/acs-5-year.html>