

GROUP 5 PROJECT

Brand Assigned: Hills Brothers



Team Members

Alankrita Roy

Amritha Rekha

Taniya Dhar

Rahul Bhatt

Rushabh Shah

Introduction to the dataset:

The dataset provided to us was the scanner data of grocery stores which had details about the purchase of the coffee. Following are the information provided by the datasets.

- Customer Demographic data
- Stores location details
- Details of different coffee product of different brands
- Panel data giving the detail about the store from which the data was purchased, the type of outlet it is.
- Grocery store data giving details of product UPC, whether the product was displayed or featured or was there a price reduction.

Goal of the Project:

The goal of the project was to analyze the given information and present the insights along with recommendations to improve sales of the coffee brand called “Hills Brothers”. The interpretations have been drawn by running various models taught in class on the provided data.

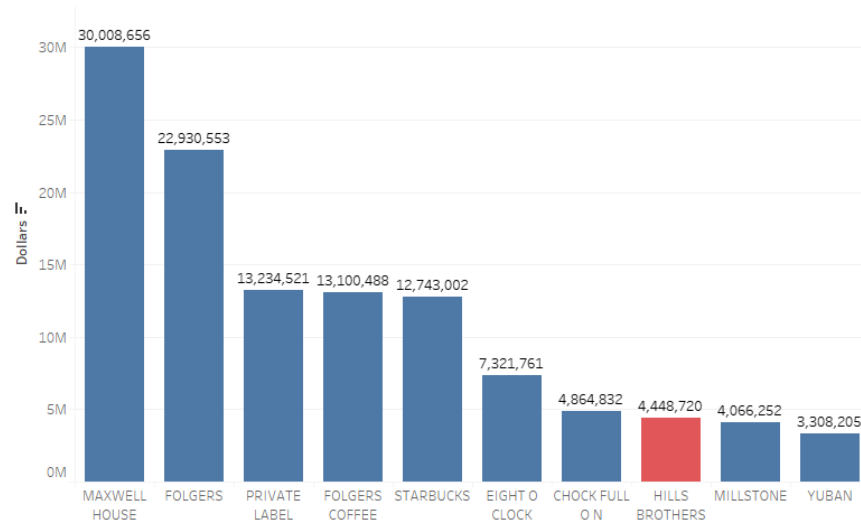
Objective of the project:

1. Analyze the effect of various promotion methods using Panel regression.
2. Analyze the effect of change in sales to the change in price for the brand itself (self-price elasticity) and based on the competitor’s price (cross-price elasticity) and analyze the effect of brand specific and customer specific variables on the choice of Hills brother brand.
3. Forecasting the sales of Hills Brothers brand for the successive 20 weeks by using Time Series Analysis Method.

EXPLANATORY DATA ANALYSIS

Top Brands in terms of Dollar Sales:

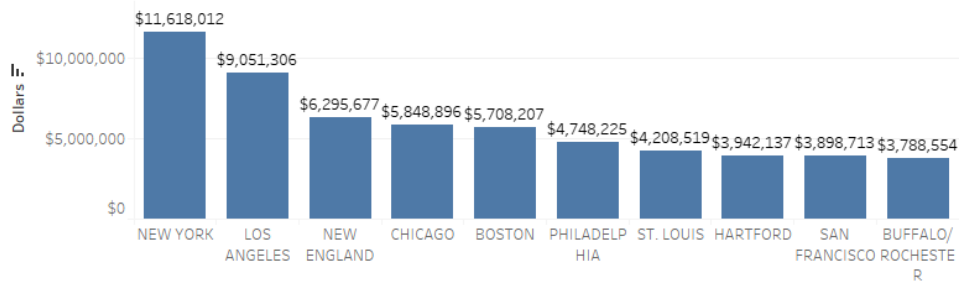
Top Brands in terms of Dollar Sales



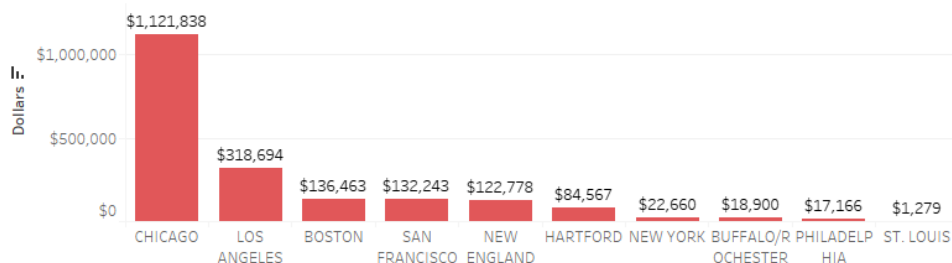
After running proc means, on the sum of sales we get the top 10 coffee brands. As we can see the top brand is Maxwell House and “Hills Brothers” is ranked 8th in terms of sales.

Top regions in Terms of Dollar Sales:

Top regions in Terms of Sales for all the brands



Top regions in Terms of Sales for Hills Brothers



Based on the Top Regions for the entire dataset, we can observe that, New York has the maximum sales of coffee in the United States.

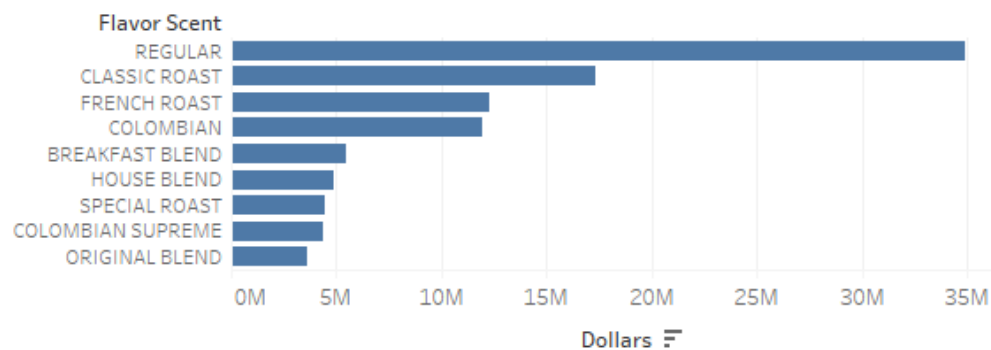
Drilling down further to the Hills Brothers Brand, we observed that Chicago has the maximum sales then followed by Los Angeles and the Boston.

Recommendations:

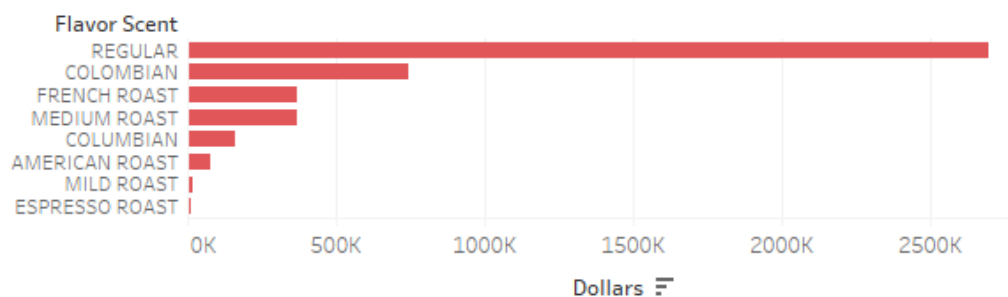
- New York City has the most coffee shops, coffee houses, cafes, and pretty much any coffee-related store imaginable than any other city in the nation. As sales of Hills Brothers coffee is very less in NY, they should probably open new stores in NY and include other strategies like adding donuts to the menu etc. to entice more customers in NY

Most Preferred Flavors by Customers:

Top Flavor/Scent in terms of Sales for all the Brands

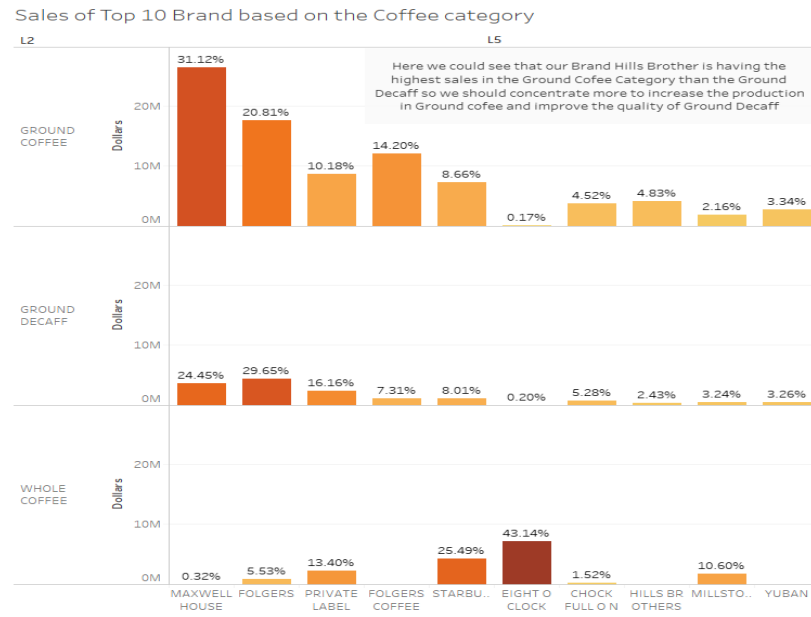


Top Flavor/Scent in terms of Sales for Hills Brothers



Based on the entire dataset, the most preferred Flavor/Scent is "Regular". Similarly, even Hills Brothers Brand also has "Regular" as the most preferred flavor or scent.

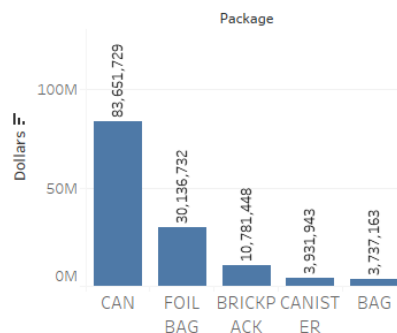
Top 10 Brand based on the Coffee Category in terms of Sales:



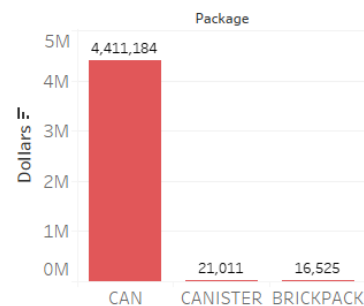
Here we could see that our Brand Hills Brothers is having the highest sales in the Ground Coffee Category than the Ground Decaf so we should concentrate more to increase the production in Ground coffee and improve the quality of Ground Decaf.

Most Preferred Packages in Terms of Sales:

Top 5 Packages in terms of Sales for all the Brands



Packages in terms of Sales for Hills Brothers



As we can see from the visualization above for all brands, the most preferred packaging is Can type followed by Foil Bag. For Hills Brothers, it seems like the coffee packaging is most preferred in can type.

Recommendation:

- We recommend that Hills Brothers should start selling coffee using "Foil Bag" package to increase its market share. Hills Brother should reassess their packing strategy and use Foil bags as it ensures that the beverages stay fresh and retain their aroma.

PART I: PANEL REGRESSION

The data provided to us has information about different purchases made by consumers. The dataset also contains information about various promotions being deployed in order to increase the sales of the coffee brand. Hence, using panel regression is an appropriate method, to find the most effective method of Marketing from: Types of Features and Display and Price Reduction.

The PANEL Procedure

Wansbeek and Kapteyn Variance Components (RanTwo)

Dependent Variable: TotalSales

Model Description	
Estimation Method	RanTwo
Number of Cross Sections	1101
Time Series Length	52

Fit Statistics			
SSE	713511719.1	DFE	43157
MSE	16532.9314	Root MSE	128.5804
R-Square	0.1541		

Variance Component Estimates	
Variance Component for Cross Sections	18512.51
Variance Component for Time Series	262.0484
Variance Component for Error	16585.02

Hausman Test for Random Effects				
Coefficients	DF	m	Value	Pr > m
	7	7	93.89	< .0001

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	101.0142	8.4588	11.94	< .0001
AvgPrice	1	-138.849	28.9308	-4.80	< .0001
AvgDisplay	1	233.5768	5.9185	39.47	< .0001
AvgFeat_coupon	1	827.9735	15.3134	54.07	< .0001
AvgFeat_large	1	201.2477	5.3668	37.50	< .0001
AvgFeat_medium	1	71.46891	4.6557	15.35	< .0001
AvgFeat_small	1	78.39123	31.0470	2.52	0.0116
AvgPriceRed	1	7.197949	2.5245	2.85	0.0044

- As seen in the above screenshot, the Hausman test has the p-value which is less than 0.05. Hence, null hypothesis is rejected, and we conclude that the fixed effect is the best model.

The SAS System			
The PANEL Procedure			
Fixed Two-Way Estimates			
Dependent Variable: TotalSales			
Model Description			
Estimation Method		FixTwo	
Number of Cross Sections		1101	
Time Series Length		52	

Fit Statistics			
SSE	696670250.5	DFE	42006
MSE	16585.0176	Root MSE	128.7828
R-Square	0.5965		

F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
1151	42006	40.83	<.0001

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Pr > t	Label
Intercept	1	-165.517	25.7126	-6.44	<.0001	Intercept
AvgPrice	1	-113.827	30.5833	-3.72	0.0002	
AvgDisplay	1	234.202	5.9672	39.25	<.0001	
AvgFeat_coupon	1	825.3466	15.3721	53.69	<.0001	
AvgFeat_large	1	201.8837	5.4173	37.27	<.0001	
AvgFeat_medium	1	71.99188	4.6937	15.34	<.0001	
AvgFeat_small	1	51.50484	31.1905	1.65	0.0987	
AvgPriceRed	1	8.648385	2.5701	3.36	0.0008	

- After running fixed effect model, we see that Coupon is the best way, since it has the maximum impact on sales. This shows that, in order to increase sales, coupon is the most effective mode of marketing.
- Another noticeable fact is that price reduction also increases the sales significantly. This is obvious since as the price is reduced the sales increase for the brand.
- According to the output, we can see that all types of marketing communications like feature, display and price reduction are significant. It influences the total sales of the coffee.
- The stores which have coupon policies on coffee have \$825.35 more sales than the stores with no coupon in that week for that store.
- The sale of product with large advertisement is \$201.88 more as compared to the products with no advertisement in that week for that store.
- The sale of product with medium advertisement is \$72 more as compared to the product with no advertisement in that week for that store.
- The product with small advertisement is \$51.5 more as compared to the product with no advertisement in that week for that store.
- From above interpretations, we can conclude that any form of Featuring whether it is large, medium or small has incremental effect on sales.

- In the dataset, Coupon refers to A+, Large refers to A, Medium refers to B and Small refers to C.

Order of Significance of Feature on Sales is as follows:

Coupon > Large>Medium>Small

- Looking at the result of Display we see that, if the coffee product was displayed in the store then the sale in that week for that store will increase by \$234.20 as compared to when the coffee product was not on display.
- Similarly, if there was a price reduction on coffee then the sale of coffee goes up by \$8.64 compared to no price reduction in that week for that store.

PART II: BINARY LOGIT

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Cable_TV 0 vs 1	2.909	1.532	5.522
Family_Size_1 0 vs 1	2.048	1.164	3.606
Family_Size_2 0 vs 1	1.007	0.878	1.155
Feature_largead 0 vs 1	0.890	0.578	1.372
Feature_mediumad 0 vs 1	0.728	0.492	1.077
Feature_coupon 0 vs 1	1.900	1.136	3.179
Feature_smallad 0 vs 1	0.393	0.292	0.529
HH_Income 1 vs 2	0.874	0.765	0.998
MAJOR_DISPLAY 0 vs 1	1.881	1.110	3.187
MALE_SMOKE 0 vs 1	1.788	1.556	2.054
MINOR_DISPLAY 0 vs 1	0.491	0.317	0.761
NOT_Employed 0 vs 1	1.550	1.133	2.122
No_Children 0 vs 1	0.596	0.501	0.710
Panel_YES 0 vs 1	0.186	0.165	0.210
PriceReduction 0 vs 1	0.579	0.503	0.666
VOL_EQ	2.223	2.074	2.382
DOLLARS	0.854	0.839	0.869

- **Model Fit statistics:** As the AIC, SC and $-2\log L$ is decreasing when adding covariates, it signifies that our model is having the significant covariates that helps us to predict our Y variable.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	12804.071	10409.333
SC	12812.323	10557.872
-2 Log L	12802.071	10373.333

Prediction Metric: Percent Concordant is 82.6. So, 82.6% of times model correctly predicted that phat is higher in Hill Brothers brand than in the Top 6 brand

Frequency Percent Row Pct Col Pct	Table of Brand by P_final			
	Brand	P_final		Total
		no	yes	
0	3800117	38	3800155	
		95.65	0.00	95.65
		100.00	0.00	
		95.65	57.58	
1	172745	28	172773	
		4.35	0.00	4.35
		99.98	0.02	
		4.35	42.42	
Total	3972862	66	3972928	
		100.00	0.00	100.00

The percentage of accurate prediction is 95.6%. $(3800117 + 28 / 3972928)$

1. Odds of a cable-TV user being Hills Brothers consumer is approximately thrice (2.90 times) more than that of a non-cable TV user.
2. Odds of a small family (family members < 2) being Hills Brothers consumer is approximately twice (2.048 times) more than that of a large family (family members >2).
3. Odds of hills brother coffee consumptions using a rebate or coupon is twice (1.9 times) more than buying them via other features (small, large advertisement).
4. Odds of hills brother coffee consumption by higher income people (more than 35,000 USD) is 0.874 times more than lower income people (less than 35,000 USD).
5. Odds of attracting customers to buy Hills Brothers coffee by putting by a major display is 1.88 times more than when not keeping a display.
6. Odds of attracting customers to but Hills Brothers coffee by putting a minor display is 0.49 times more than when not keeping a display.

7. Odds of a male smoker being a Hills Brothers consumer is 1.788 times more than a male non-smoker.
8. Odds of buying Hills Brothers coffee is 0.579 times more when there is a price reduction.
9. Odds of buying Hills Brothers coffee being an unemployed is 1.55 times more than being a full-time worker.

Recommendation based on above interpretations:

1. We are recommending Hills Brothers brand to show more advertisements, so that the cable tv users are likely to see the advertisement and the advertisement will have an impact on them, which in turn will influence the customers to buy the Hills Brothers Coffee brand.
2. Hills Brothers coffee brand should target small families in order to increase their sales, as the small families are likely to buy the coffee & the large families might have some other grocery priorities.
3. Hills Brothers coffee brand must distribute more coupons to attract more customers. With more loyalty and rewards program, they will not lose money but will instead gain more customers
4. Hills Brothers should target higher income people as they are more likely to spend money on coffee
5. One of the best methods to campaign customers is by putting up banners displaying the quality, characteristics and offers on the product.
6. People who are addicted to tobacco are likely to be addicted to coffee. This fact can be used in shelf management in stores. If Hills Brothers coffee products are placed next to tobacco products then that will result in increase of sales.
7. Seasonal price reductions will be a better option to increase the market share of the Hills Brothers coffee brand, so try to include price reduction.
8. If there is panel for hills brother coffee brand, they can share their approaches and suggestions to improve coffee quality.
9. People who are unemployed are more likely to drink Hills Brothers coffee than full-time workers. As majority of the hills brother's coffee production is that of ground coffee which is mostly consumed by people at home. So, Hills Brothers should increase the production of tea bags and single serve pods.

Price Elasticity Calculation:

Average Price(per Unit), Display, Feature

The MEANS Procedure

Brand	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
CHOCK FU	150919	AvgPPU	150919	0.2586890	0.0754767	0.0408333	1.2825640
		Display	150919	0.0450043	0.2884895	0	2.0000000
EIGHT O	307216	AvgPPU	307216	0.3168152	0.1914487	0.0207692	4.9074497
		Display	307216	0.0577737	0.3257103	0	2.0000000
FOLGERS	1185591	AvgPPU	1185591	0.3655002	0.3500271	0.0251282	2.0633333
		Display	1185591	0.0513676	0.3016201	0	2.0000000
HILLS BR	172773	AvgPPU	172773	0.2271379	0.0653051	0.0228205	0.6453846
		Display	172773	0.0625329	0.3298028	0	2.0000000
MAXWELL	994586	AvgPPU	994586	0.3399930	0.2823405	0.0253846	2.9966667
		Display	994586	0.0546981	0.3095207	0	2.0000000
OTHERS	1975431	AvgPPU	1973056	0.5236544	0.2625274	0.000714286	14.4300000
		Display	1973056	0.0266277	0.2071048	0	2.0000000
PRIVATE	725573	AvgPPU	725573	0.3652469	0.2416732	0.0286937	4.2218921
		Display	725573	0.0624582	0.3276333	0	2.0000000
STARBUCK	436270	AvgPPU	436270	0.6684828	0.0797450	0.0808333	3.1949776
		Display	436270	0.0167648	0.1611910	0	2.0000000

Market Share of the Brands

	Sum_of_Sales
	ColPctSum
Brand	
CHOCK FU	3.44
EIGHT O	5.17
FOLGERS	26.78
HILLS BR	3.11
MAXWELL	21.34
OTHERS	21.60
PRIVATE	9.45
STARBUCK	9.11
All	100.00

Brand	Share	N	Avg PPOZ	Average Display	Prob (MS))	Own price elasticity	Cross Price elasticity
FOLGERS	26.78	1185591	0.37	0.051	0.268	-0.04	0.016
MAXWELL	21.34	994586	0.34	0.055	0.213	-0.04	0.012
PRIVATE	9.45	725573	0.37	0.062	0.095	-0.05	0.006
STARBUCK	9.11	436270	0.67	0.017	0.091	-0.10	0.010
EIGHT O	5.17	307216	0.32	0.058	0.052	-0.05	0.003
CHOCK FU	3.44	150919	0.26	0.045	0.034	-0.04	0.001
HILLS BR	3.11	172773	0.23	0.063	0.031	-0.04	0.001
OTHERS	21.60	1975431	0.52	0.027	0.216	-0.07	0.018

As price is a brand specific variable, the coefficient of price of all the brands will be the same while the price of the brands will be different. Here, the coefficient of price is -0.16.

Own Price Elasticity:

If Hill Brothers were to lower prices by 1%, its market share would increase by 0.04%, all else remaining the same.

Cross Price Elasticity:

If Hill Brothers were to lower price by 1%, Folgers's market share would decrease by 0.001%, all else remaining the same.

If Hill Brothers were to lower price by 1%, Maxwell's market share would decrease by 0.001%, all else remaining the same.

PART III: TIME SERIES ANALYSIS

Using Customer Panel Data we ran a model to forecast sales for the next successive 20 weeks. We calculated the total sales for each week from the data we have which is of 52 weeks. On running a Time Series Forecast Model we predicted the sales for 20 weeks of the next year.

The output was forecasted with a 95% confidence interval so a manager can predict the sales for the successive weeks along with some cushion to expect a surge.

Descriptive Analysis on Time Series Sales Data on entire dataset:

This is the pattern for sales along the duration of a year. We can spot increase sales in 3 weeks.

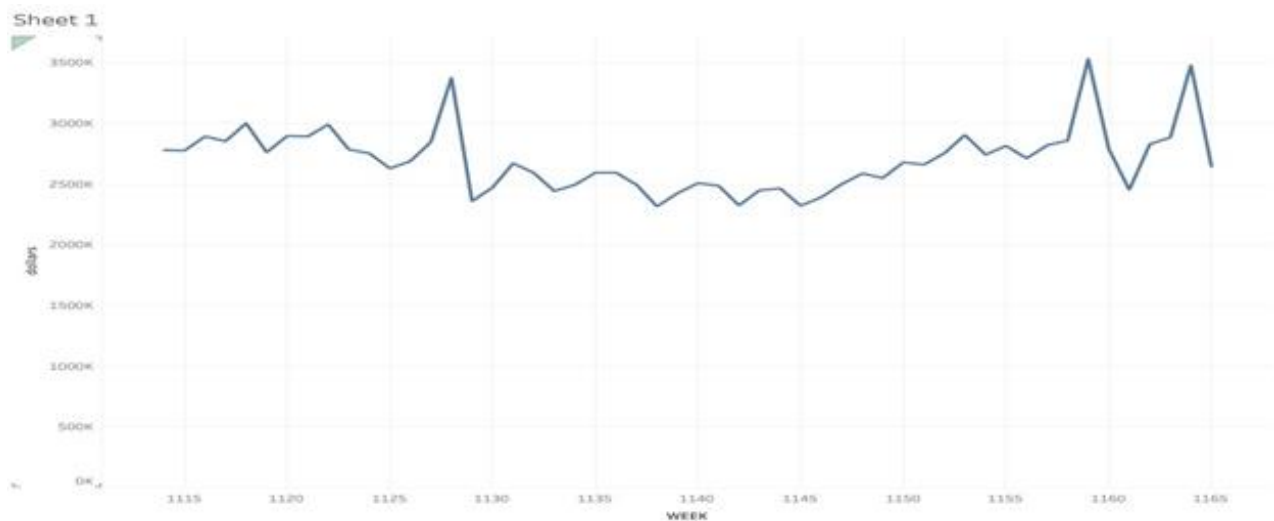
Observing the graph and its pikes we could infer that the highest sales were between the week

1.Week 1128-Week 1114= Week 14 that is the Month of April

2.Week 1159-Week 1114= Week 45 that is the month of November

3.Week 1164-Week 1114= Week 50 that is the month of December

We can infer that during the months of Holidays that is Easter, Thanksgiving, Christmas are the times when the sales see the highest peak which is justified by the Black Friday Sale, Easter weekend as well as the Christmas Holidays.



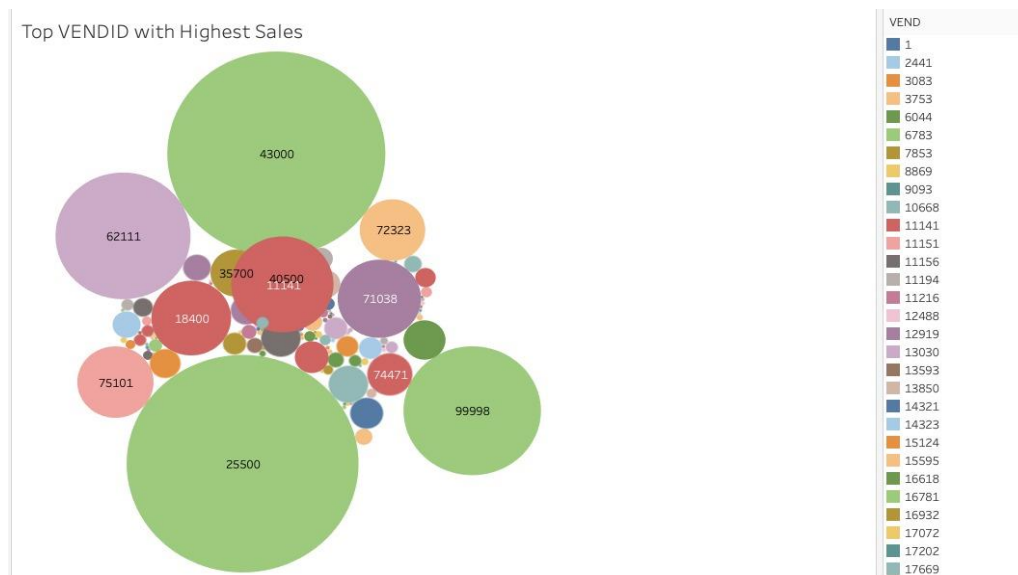
VENDID with Highest Sales:

In the vend data set we found the items that had been vended and bought the maximum number of times. We calculated the total revenue generated by each item during the entire year.

As we can see the VENDID 25500 has the maximum sales of \$37,684,852 and VENDID 43000 also has high sales of \$33,316,852.

Recommendation:

A manager can know which items to pitch for and in which area should he maximize his marketing. He can also know which item to stock more as they have increased sales. He can optimize his wholesale order of products accordingly.



Forecasting Sales of Hills Brothers:

After looking at the sales trend on the time series plot for the entire dataset, we drilled down to forecast next 52 weeks sales for “Hills Brothers” brand and came up with some interesting insights as follows:

- Highest Sales is forecasted in the second week which is January 7th to January 13th. As we can see the sales in second week is \$91,971.
- Lowest Sales is forecasted in the first week which is December 31st to January 6th. Here the sales are of \$78,970.
- Using tableau features, we filtered out top 10 sales and bottom 10 sales weeks in the forecasted 52 weeks.

- Hence, we can conclude that existing sales pattern and forecast sales pattern are similar. There is seasonality in the data as the sales are high in beginning of year and then towards the end of year.
- The bars in blue shows top 10 sales and bars in orange shows bottom 10 sales.

Top 10 and Bottom 10 Forecast Sales

