Analysis of Facial Tissues Sales

Predictive Analytics with SAS | Project Report | 04/26/2018

Mentored By:

Prof. B.P. Murthi

**Group 1**

Bikramjit Dey

Jing Zhao

Madhumita Jadhav

Mrudula Deore

Surya Thummanapelly

# **Abstract:**

This project was intended to analyze and quantify the relationship between sales of facial tissues and customer demographics, price changes, store-level displays and promotions, store demographics, competitor demographics. The project primarily focuses on identifying ways to enlarge the market share of the second leading brand in the market, “Puffs”.

This report was guided by the following objectives:

1. To identify and segment the customers based on loyalty and revenue contribution
2. To identify potential customers for targeting advertisements and promotions
3. To find out how customer and store demographics affect sales
4. To evaluate the extent of impact on sales by product displays and promotions
5. To identify the effects of closest competitors’ prices and promotions on the sales of “Puffs” brand

The main goal is to enhance the market share gap between “Puffs” and the third leading brand “Scotties”, by improving the sales of the former brand. The project concern is not to find ways to surpass the market leader “Kleenex”, which is difficult if not impossible to achieve considering the limted data and maturity of this market.

# **Descriptive Analytics:**

# **Data Description:**

The data is of majority of the facial tissue brands and products available in the US market which spanned from Jan 1, 2001 to Dec 31, 2001. The provided data can be grouped into following sections:

1. **Household Purchase Data:** Household level weekly purchase data of facial tissues like units bought, bill amount, UPC codes etc.
2. **Store Scanner Data:** Store level weekly scanner data of facial tissue purchases like units sold, UPC codes, Display, Promotions, Featured etc.
3. **Products Data:** Characteristics of various products in the year 2001 like Size, Color, Additives, Type of dispenser, UPC codes etc.
4. **Household Demographics:** Demographics data of households like Family size, Age, Number of children, Race, Income etc.
5. **Store Demographics:** Demographics data of store like location, Estimated ACV, Outlet type etc.

# **Data in Numbers & Visuals:**

Based on our descriptive analysis on overall data, we found following metrics:

* The **Scanner data** and **Household purchase** **data** is week-level, where week starts on Monday. There are total **52** weeks of data.
* The **Scanner data** has data from **1866** stores (Drug stores & Grocery stores) across the US.
* The **Household purchase data** has facial tissues purchase history data of **6078** unique households from **Wisconsin** and **Massachusetts** states in the US.
* **On average,** these 6078 households purchased **1.55 units** of facial tissues **per week** by spending **$2.33.**
* Based on the **Scanner data**, **Puffs** has sold **$** **12,067,731.97** worth of facial tissues, where as the market leader **Kleenex** has done **$ 33,863,872.91** worth of sales and the 3rd leading brand **Scotties** has managed to do **$ 6,982,226.15** sales in the business.

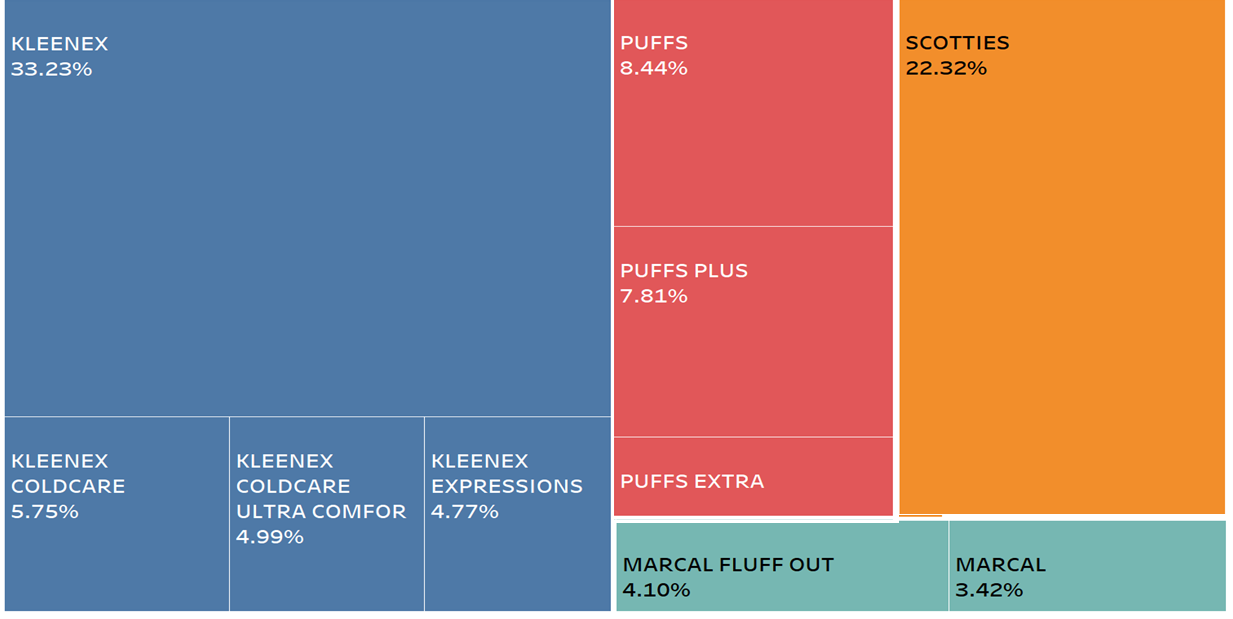
A screenshot of a cell phone

Description generated with very high confidence

**Market Share by Number of units sold - Brands**

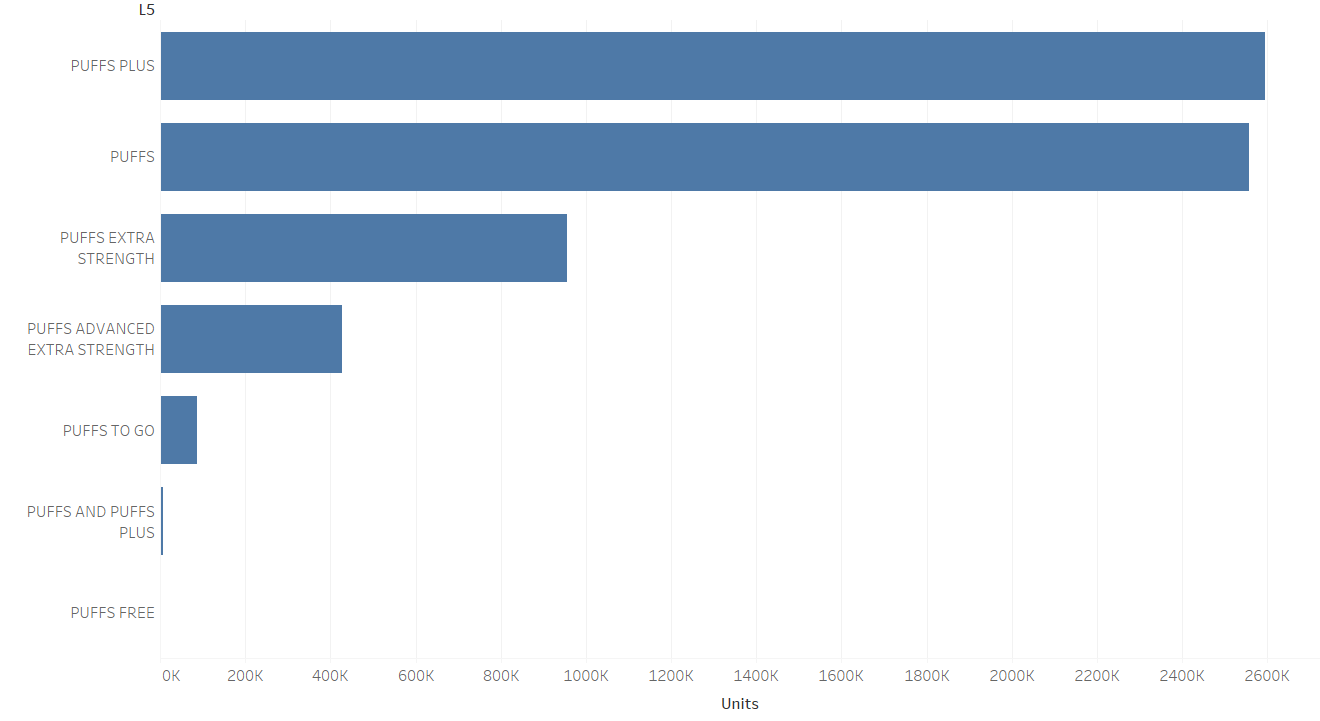
The visualization above shows the market share of major brands of facial tissues by number of units sold. **Kleenex** happens to be the clear leader with **50.34%,** with **Puffs** at second with **17.94%.**

* The visualization below shows the market share of major brands of facial tissues divided into sub-categories for sales by **dollar amount**. Here again, we see a similar trend as above with Kleenex and its sub categories leading followed by Puffs.



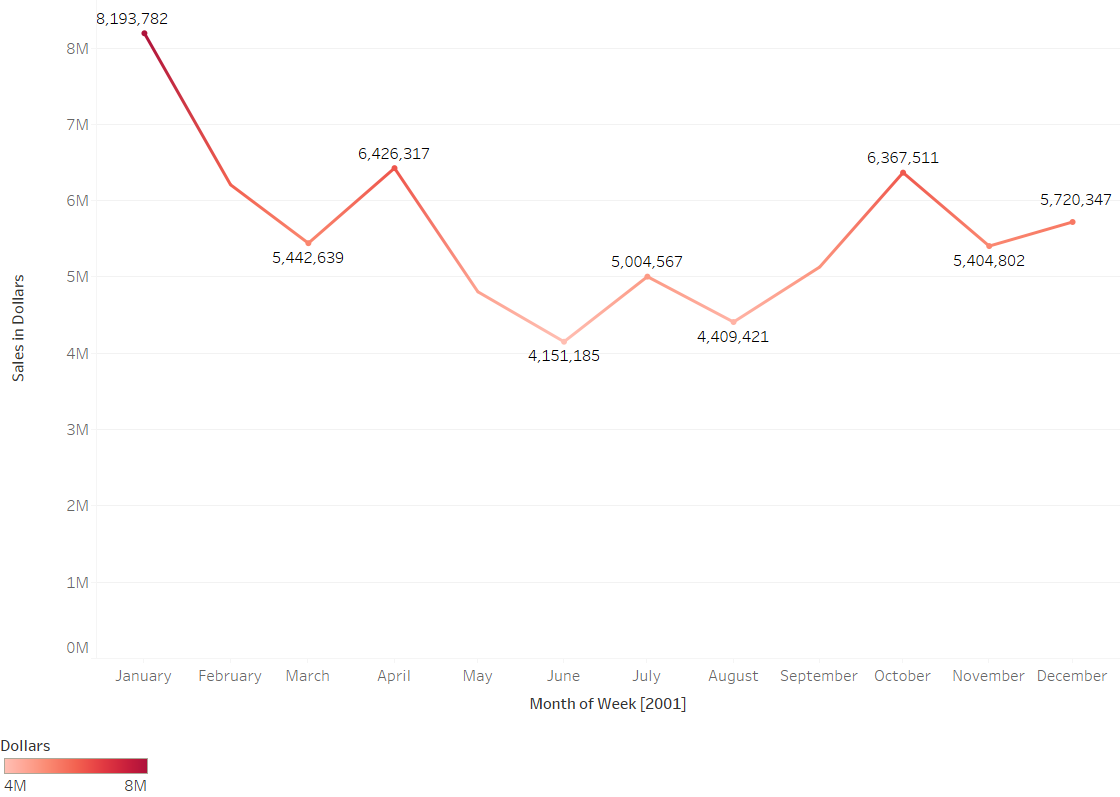
**Market Share by Sales Revenue - Brands**

* **Puffs** sold **7** type of facial tissue products. Their relative share can be observed below:



**Relative Market Share by Number of units sold – Product Types within Puffs**

* **Household purchase data** says, **Puffs** customerson average have a **family size** of **2.47** and **2.8** number of **TV’s.**
* The paper products market happens to be the largest consumer of forest products and one that has seen the largest growth of about 2.1% in the last ten years. It is not seasonal, with consumers using about the same amount of toilet paper, napkins or paper towels every month. But not so with facial tissue: as cold and flu season rears its congested head, monthly sales increase as much as **65%** over the summer months. The visualization below confirms the seasonality with sales figures from 2001.



**Trend in Sales across Months – Cumulative of all brands**

**Correspondence analysis:**

A screenshot of a social media post

Description generated with very high confidence

Correspondence analysis is an exploratory technique used to visualize cross tabulations containing some measure of correspondence between the rows and columns. We conducted a correspondence analysis of the parent company (L4 variable) and family size to generate a perception map to figure out if there is a correspondence between them. As can be seen by the figure above, with a family size of four of less, customers prefer the more expensive brands sold by Kimberly Clark (Kleenex) and P&G (Puffs) while those with bigger family size tend to go for other brands. Price could be a crucial deciding factor here as Kleenex and Puffs happen to be considerably more expensive than the other brands.

1. **Panel Analysis – Scanner Data:**

We analyzed weekly scanner data as a panel data. To achieve this, we have performed data pre-processing and transformations using **Pandas** and **Numpy** libraries in **Python**.

* 1. **Panel Analysis – Data Pre-Processing:**

As Puffs brand has multiple product offerings (Multiple UPC codes), it is essential to calculate weighted attributes for the whole brand within each store and each week of scanner data. After all, we are trying to improve the overall Puff’s brand sales, not just one product type. We have used following calculations to get the weighted characteristics.

*Unit Price = Dollars / Units;*

*Sheets Sold (Number of sheets sold within a store, week and UPC) = Units sold \* Vol\_Eq \* 100;*

*Sheet Price (Price per sheet within each UPC in a store and week) = Unit Price / (Vol\_Eq \* 100) ;*

*Brand Sales (Sum of sheets sold within a store and week across all UPCs under Puffs brand) = data.groupby(store, week, brand).SUM(Sheets Sold);*

*Sales Share (Fraction of sales of a UPC in brand sales within a store, week and Puffs brand) = Sheets Sold / Brand Sales;*

*Share Weighted Price (For each UPC within a store and week) = Sales Share \* Sheet Price;*

*Weighted Average Price (Across all UPCs under Puffs brand within a store and week) = data.groupby(store, week, brand).SUM(Share Weighted Price);*

Similarly, we calculated following metrics with reference to “Sales Share”:

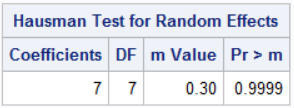
1. Weighted Average Display
2. Weighted Average PR (PR is binary variable to indicate if there is a 5% or more discount on the UPC)
3. Weighted Average F for all levels of “Feature” in the data (F for feature)
4. KL\_Weighted Average Price (KL prefix for Kleenex brand)
5. SC\_Weighted Average Price (SC prefix for Scotties brand)
6. KL\_Weighted Average Display
7. SC\_Weighted Average Display
8. KL\_Weighted Average F
9. SC\_Weighted Average F

Then we took subset of top 49 stores based on the highest **Est\_ACV** (Estimated All Commodity Volume) measurements, which generate 80 – 90% of the revenue for Puffs. Est\_ACV measures the whole volume of sales in a store on average across all products in millions of dollars. Then we filtered out only “Puffs” data.

* 1. **Panel Analysis:**

We performed Fixed effects, Random effects panel models using PROC PANEL on multiple models with various combination of variables. We have improvised the models by removing insignificant variables.

We ran “Hausman Test” to decide between Fixed effects and Random Effects models. The SAS output of the test is below:

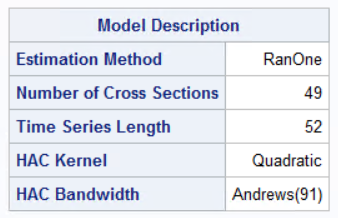


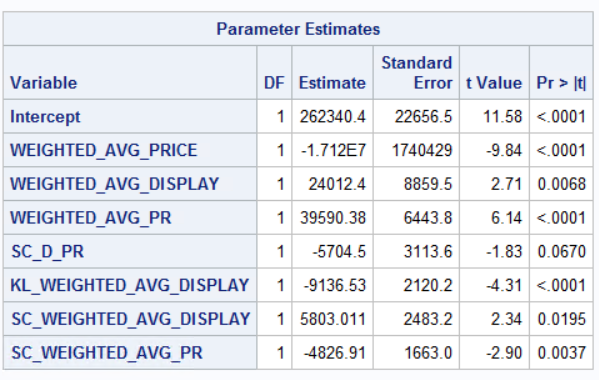
The Hausman test results show that we can’t reject the null hypothesis of “No endogeneity”. So, it is safe to use Random effects model for the data.

Model:

*BRAND\_SALES = WEIGHTED\_AVG\_PRICE + WEIGHTED\_AVG\_DISPLAY + WEIGHTED\_AVG\_PR + SC\_D\_PR + KL\_WEIGHTED\_AVG\_DISPLAY + SC\_WEIGHTED\_AVG\_DISPLAY + SC\_WEIGHTED\_AVG\_PR;*

The output of the RanOne model from SAS is shown below:





**Interpretations:**

1. 0.001 dollars (0.1 cent) increase in price of a sheet is associated with a decrease of 17,120 sales in number of sheets controlling for all other variables (We interpreted for 0.1 cent units because per sheet cost will be too low relative to 1 dollar).
2. Displaying additional 10% of the available product categories under Puffs brand is associated with an increase of 2,401.2 sales in number of sheets controlling for all other variables.
3. Giving 5% or more discount for additional 10% of the available product categories under Puffs brand is associated with an increase of 3,959 sales in number of sheets controlling for all other variables.
4. Displaying and giving discounts for additional 10% of the available product categories under Scotties brand within the store, is associated with a decrease of 570.4 sales in number of sheets for Puffs brand controlling for all other variables. This has a synergistic effect. (This effect is significant at 90% confidence level, but it is insignificant at 95% confidence level)
5. Displaying additional 10% of the available product categories under Kleenex brand is associated with a decrease of 913.6 sales in number of sheets for Puffs brand, controlling for all other variables.
6. Giving 5% or more discount for additional 10% of the available product categories under Scotties brand is associated with a decrease of 48.26 sales in number of sheets controlling for all other variables.
7. **Recommendations:**
8. Puffs lacks in variety of the products, they should introduce the seasonal products like Cold Care and Ultra Comfort like Kleenex. Customers should have more choices to choose what they want. Once customers can’t find the type that they need even just a little bit difference, they will change their mind to other brand, which means Puffs is losing them anyway.
9. Target Promising Customers to convert them into loyal customers by giving them coupons. With coupon, Promising Customers are more likely to buy Puffs other than other brands since there is no difference on them this time. However, they will be used to Puffs for no reason when the next purchase if they think Puffs is good enough.
10. We can build Look-alike Model for Identifying Similar Customers to the prime customers who haven’t purchased. It’s like giving them a chance to try products from Puffs. Sometimes, customers purchase products not because that brand is better, but because they would do as they used to do, which is hard to improve. Look-alike Model breaks this situation.