Costco’s online market Case-Study

Analyzed by: Morad Moqbel in 01/12/2024

Purpose of case-study: Analyzing sub-categories and their relationship with sales and discounts

Tools that have been used: Power BI

1- Data Source: [www.kaggle.com/datasets/](http://www.kaggle.com/datasets/nelgiriyewithhana/world-educational-data)bhavikjiadara/grocery-store-dataset

Under the license: attribution 4.0 international (CC BY 4.0)

2- Data Validation:

Our data contains 8 columns respectively (All of them are text type):

A- Sub Category: This column categorizes the grocery items into subcategories, providing a detailed classification for easier analysis and organization.

B- Price: Represents the monetary value of the grocery item, indicating its cost or retail price in the specified currency.

C- Discount: Reflects any discounts or promotional offers applicable to the respective grocery item, providing insights into pricing strategies.

D- Rating: Indicates customer satisfaction or product quality based on user ratings, offering a measure of the overall perceived value of the grocery item.

E- Title: Describes the name or title of the grocery item, providing a concise identifier for easy reference and understanding.

F- Currency: Specifies the currency in which the prices are denominated, facilitating proper interpretation and comparison of monetary values.

G- Feature: Includes features or characteristics of the grocery item, offering additional information about its unique attributes or selling points.

H- Product Description: Provides a detailed textual description of the grocery item, offering comprehensive information about its specifications, uses, and other relevant details.

I- Net Price: Calculates the price after the discount (Created by myself)

3- Data Prepare

After we discovered and explored our data, its time to clean and prepare our data in order to perform our analysis.

First we need to figure out wherether or not there are outliers, abnormal values and etc…

The easiest way to discover the errors, its by visuals..

-In our case, there is no need for the “Features,description,rating and title columns”, so we deleted them as well as Currency

-There is no duplicates

-There were 5 rows with no currency, so fix them by replacing their values into $..

-There were 2 rows with ‘.’ values in their “Discount” column, so we deleting them because these 2 rows won’t make a difference

-We changed the “Price” type into decimal ,so we have the ability to perfume some statistical actions

- Since we need the discount column to calculate how much the discount is , we transformed it into decimal

-We removed duplicates from Title column, so our data decreased to 1481 rows..

-We reorder the columns into the order: Title,sub-category-price,discount,currency and rating to keep our data in clear format

4- Data analysis

After we analyzed our data, we figured out the follow:

a- Net Profits = $81.96K

b- There are 16 sub category in our market’s data

c- The most profit category is “Meat and seafood”

d- The most frequently category is “Snaks”

e- There are 35 different types of discounts

5- Conclusion

After we analyzed our data we figured out that the most profit category is Meat & Seafood, and the most category frequently is Snaks...

Although the Snaks category has sold more than any category and had 15 discount offers, it did not generate more revenue than Meat & Seafood category, which sold nealy half of the Snaks category sales..

It is clear that people tend to favor discounts on item that are usually considered expensive. Therefore to archive highe profits, i suggest providing more offers and discounts for Meat & Seafood Category

/-/-/-/-/-/-/-/-/-/-/-/-/-/-/-/-/-/

I hope you have enjoyed in my case-study about costco’s online market and please keep an eye on my github profile to see more case-studies that I am up to create by using a diverse set of analysi tools

Kind Regards

Morad Mobel