Optimizing Beverage Sales: Placing the Right Beverage in the Right Market

A Final report for the BDM capstone Project

Submitted by
Name: VEMURU VENISHA
Roll no:23F2003668

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Table of content

Optimizing Beverage Sales: Placing the Right Beverage in the Right Market.

1. Executive Summary

The project is about a beverage distribution company in the United States under a Business-to Business (B2B) model. The company sells many beverages like Coca-Cola, Sprite, Fanta, Diet Coke, Powerade, and Dasani water to many retailers and commercial customers across different regions in US.

From the dataset we can observe that the company is facing problems like uneven sales across different beverages and regions. Some beverages are having highest sales in one region but lowest sales in another region which is leading to stock imbalance, customer dissatisfaction, and losses. The use of the almost same pricing, promotions and sales strategy over all regions is not good because the people in different regions will have different income levels, spending money is also different, taste are different and many other reasons.

To solve the problem like the sales imbalance, we have taken detailed look at the dataset and analyzed the various columns like regions, brands, units sold, and profits. These has helped us in more
understanding where the problem is coming and where the mistake is happening. People in different regions have different likes, tastes, cultures, spending habits, lifestyles and so on the using of the
almost same pricing, marketing is not a good idea and can lead to loss in market. So, we created a proper plan for each region based on what the customers need and what is required. We have made some simple charts to see what is working in which region, like which drink is popular in which region. To analyze the data properly, we have used statistical methods like correlation analysis and visualizations such as bar charts to find patterns in unit sales, prices, and profit margins. Tools like R programming language along with packages such as Data Explorer, explore helped us in creating automatic EDA reports, interactive dashboards and data-driven insights. By using this information, we designed a marketing plan which reduces wastage of products, increases consumer satisfaction and to increase profits.

The analysis says that Coca-Cola stands out as the highest selling brand across all regions, followed by Dasani Water and Diet Coke. In the retailers, BevCo recorded the highest number of transactions, which is saying its strong market presence. The sales distribution is not evenly spread across regions because the West and Northeast show higher sales volumes, the Southeast lags behind with comparatively lower sales. Correlation analysis highlights strong positive relationships between Units Sold, Total Sales, and Operating Profit, suggesting that increased sales volumes directly contribute to higher profitability. Interestingly, the Operating Margin does not follow this same pattern, indicating potential inefficiencies and opportunities to improve operational performance for better overall profitability.

To boost sales and increase overall profitability, the company should focus on region specific strategies for pricing, product mix, and marketing, making decisions based on detailed data insights. Rather than relying on a one size fits all approach, offerings should be tailored to match local demand patterns and economic conditions. Leveraging data visualization and statistical modeling can help optimize inventory levels, minimize waste, and align product supply with the unique preferences of each region. By following this data-driven approach, the company can improve efficiency, enhance profitability, and achieve sustainable growth in the competitive beverage distribution market.

2.Proof of Originality

The dataset link-
https://docs.google.com/spreadsheets/d/1HTiwK9nxVOVGPJt9YYquIWhRs1zX6aJY/edit?usp=sharing&ouid=106698882002843209906&rtpof=true&sd=true

The Colab link:
https://drive.google.com/file/d/1oK-sUN9FREn-LjHZAbxfJXo9AO4Dfr0n/view?usp=sharing

3.Metadata and Descriptive statistics

Metadata

The dataset contains a transactional and sales related information which are from the multiple beverage retailers across different regions in the United States. The dataset also have the categorical variables which are **Retailer, Retailer ID, Quarter, Year, Region, State, City, and Beverage Brand**. The numerical variables in the dataset are **Price Per Unit (PPU), Units Sold (US), Total Sales (TS), Operating Profit (OP), and Operating Margin (OM)**, which tells us the financial and performance related aspects of the transactions. Together, these variables provide a comprehensive view of how beverage sales differ across time, locations, and brands, making the dataset suitable for descriptive and diagnostic analysis.

Descriptive Statistics

The descriptive analysis highlights the variation in pricing, sales, and profitability across the dataset. The **Price Per Unit (PPU)** ranged from approximately 1.0 for lower-priced beverages to more than 5.0 for premium options, with an average of about 2.8. The **Units Sold (US)** varied widely, from just over a hundred units in smaller markets to more than 12,000 units in high-demand areas, averaging around 4,500 units overall. Similarly, **Total Sales (TS)** values extended from nearly 1,000 up to over 45,000, with a mean of about 15,000 across retailers. The **Operating Profit (OP)** displayed notable variation, including some negative values indicating losses, but maintained an average close to 4,000. Finally, the **Operating Margin (OM)**, which reflects profitability relative to total sales, ranged from slightly negative values up to 0.35, with an average margin of 0.14. This suggests that while most transactions were moderately profitable, there were also instances of underperformance.

4.Detailed explanation of analysis process and methods

4.1Pre-processing of the Data

This dataset looks really well-structured and clean, which is obvious from the basic statistics shown in the raw count's summary. There is a total of **9,647 rows** and each of those is described using **15 columns**. Out of these, **7 columns are discrete**, probably containing categories like product types, regions or other identifiers. The remaining **8 are continuous columns**, which means they likely have numerical data such as price, ratings or sales figures that can be used directly in calculations or for building predictive models.

One of the biggest strengths of this dataset is that it has **no missing values at all**, literally zero. There are zero empty fields or incomplete rows. All **9,647 rows are fully complete across all 15 columns**.

Even though the dataset contains a total of **144,705 individual data points** (from 9,647 × 15), it's surprisingly lightweight. It only uses around **706.2 kilobytes of memory**, which makes it really easy to work with, even on a basic laptop or during repeated experimentation.

So overall, the dataset is clean, complete, and efficient to handle. It is ideal for running **Exploratory Data Analysis (EDA), generating visualizations, trying out feature engineering techniques**, or even jumping straight into **machine learning model development**. Because there are no missing or corrupted values, any insights or relationships we find during the analysis are more likely to be real trends in the data, rather than errors caused by data quality issues. This makes it a really strong dataset to work with from the very beginning.

The Data Profiling Report is given in the following link:

https://drive.google.com/file/d/1bv0Oe1HtiDihf66eGv5AhOi4aVtnWHYw/view?usp=sharing

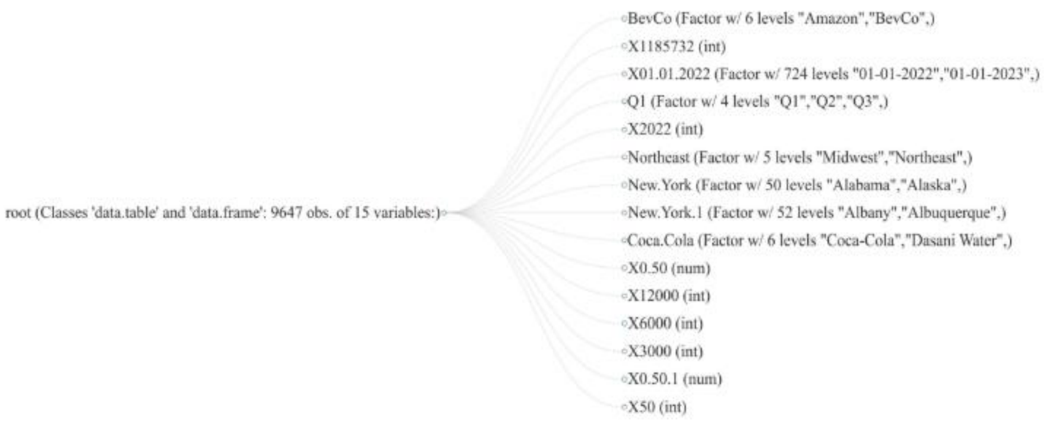

Fig-4.1

4.2Descriptive Analysis

The dataset comprises both **numerical and categorical variables** related to beverage sales across multiple retailers, regions and time periods.

Numeric Variables:

Price per Unit: **Price per Unit** will tells us how much each drink was sold.

Units Sold: Units Sold will tell us the total number of beverages are sold in each transaction record. It will directly tell us the total revenue.

Total Sales: Total Sales is calculated by multiplying price per unit by the units sold. It gives us the total money made from the sales, without subtracting the money spent. The brands with both high prices and strong demand perform best here.

Operating Profit: Operating Profit will give us the amount which is left after covering production and operating expenses. It will give us the profit made from each transaction record.

Operating Margin: This is the percentage of Total Sales that becomes Operating Profit. A higher margin indicates better cost control and greater profitability efficiency.

Stat	PPU	US	TS	OP	OM
Min	0.0700	0	0.0	0.0	0.100
1st Quater	0.3500	1060	406.8	175.0	0.350
Median	0.4500	1760	780.5	326.0	0.410
Mean	0.4522	2569	1245.6	489.5	0.423
3rd Quater	0.5500	3500	1586.2	619.0	0.490
Max	1.1000	12750	8250.0	3900	0.800

Categorical Variables:

Retailer: There are 6 types of retailers in this dataset which are Amazon (1005 records), BevCo (2618), FizzyCo (2017), Target (1030), Walmart (604), and West Soda (2374). BevCo has the highest number of records and Walmart has the lowest.

Retailer	Count
Amazon	1005
BevCo	2168
FizzyCo	2017
Target	1030
Walmart	604
West Soda	2374

Retailer ID: There are **4 unique retailer IDs** with respect to the retailers are **1185732 (5265 records)**, **1128299 (2370)**, **1197831 (1653)**, and **1189833 (360)**. Here we have the highest number of records for 1185732 which is more than half of the dataset, which matches with BevCo's high record count.

Retailer_ID	Count
1128295	2370
1185732	5265
1189333	360
1197831	1653

Invoice Date: The Invoice Date shows the date when each sale happened and has 724 unique days which is in the span of two years.

Quarters: There are **4 quarters** which are**Q1 (2439 records)**, **Q2 (2390)**, **Q3 (2451)**, and **Q4 (2368)**. Each quarter is having an equal number of entries, which says that there is a good seasonal coverage and allowing for time-based trend analysis.

Year: The dataset spans is **2 years** which is **2022 (1302 records)** and **2023 (8346 records)**. Alarger amount of data is from the year **2023**, which tells us that there is an increased sales of the products.

Region: The dataset has **5 regions** which are **West (2448 records)**, **Northeast (2376)**, **Midwest (1872)**, **South (1728)**, and **Southeast (1224)**. The **West** region has the high number of records, while **southeast** has the least number of records.

Region	Count
Midwest	1872
Northeast	2376
South	1728

Southeast	1224
West	2448

States: There are **50 states** represented in the dataset. The states with the highest number of records are **California (432)** and **Texas (432)** next the **Florida (360)** and **New York (360)**. The many states like **Mississippi**, **Montana**, **Vermont**, and **West Virginia** have **144 records.**

State	Count	State	Count	State	Count	State	Count
Alabama	216	Alaska	144	Arizona	216	Arkansas	144
California	216	Colorado	144	Connecticut	216	Delaware	144
Florida	360	Georgia	216	Hawaii	144	Idaho	216
Illinois	144	Indiana	144	Iowa	144	Kansas	144
Kentucky	144	Louisiana	144	Maine	144	Maryland	144
Massachusetts	216	Michigan	144	Minnesota	144	Mississippi	144
Missouri	144	Montana	144	Nebraska	144	Nevada	216
New Hampshire	144	New Jersey	144	New Mexico	144	New York	360
North Carolina	144	North Dakota	144	Ohio	144	Oklahoma	144
Oregon	216	Pennsylvania	144	Rhode Island	144	South Carolina	216
South Dakota	144	Tennessee	144	Texas	432	Utah	216
Vermont	216	Virginia	216	Washington	144	West Virginia	144
Wisconsin	216	Wyoming	144				

Beverage Branch: There are **6 beverage brands**: **Coca-Cola (1610 records)**, **Diet Coke (1610)**, **Dasani Water (1608)**, **Sprite (1608)**, **Powerade (1606)**, and **Fanta (1606)**. All brands are very equally distributed, differing by at most 4 entries, this indicates balanced sampling of products in the dataset.

Beverage_Brand	Count
Coca-Cola	1610
Dasani Water	1608
Diet Coke	1610
Fanta	1606
Powerade	1606
Sprite	1608

City: The dataset has **68 cities**. Ph**oenix city** has the highest count with **360 records**. The others cities like **Cheyenne**, **Des Moines**, **Jackson**, and **Providence** have **144 records each**. Most cities have data in the range of **144–216 records**.

City	Count	City	Count	City	Count	City	Count
Albany	**144**	**Albuquerque**	**216**	**Anchorage**	**144**	**Atlanta**	**216**
Baltimore	**144**	**Billings**	**144**	**Birmingham**	**144**	**Boise**	**216**
Boston	**216**	**Burlington**	**216**	**Charleston**	**288**	**Charlotte**	**144**
Cheyenne	**144**	**Chicago**	**144**	**Columbus**	**144**	**Dallas**	**216**
Denver	**216**	**Des Moines**	**144**	**Detroit**	**144**	**Fargo**	**144**
Hartford	**216**	**Honolulu**	**144**	**Houston**	**216**	**Indianapolis**	**144**
Jackson	**144**	**Knoxville**	**144**	**Las Vegas**	**216**	**Little Rock**	**144**
Los Angeles	**216**	**Louisville**	**144**	**Manchester**	**216**	**Miami**	**144**
Milwaukee	**144**	**Minneapolis**	**144**	**New Orleans**	**144**	**New York**	**216**
Newark	**144**	**Oklahoma City**	**144**	**Omaha**	**144**	**Orlando**	**216**
Philadelphia	**216**	**Phoenix**	**216**	**Portland**	**216**	**Providence**	**360**
Richmond	**216**	**Salt Lake City**	**216**	**San Francisco**	**216**	**Seattle**	**144**
Sioux Falls	**144**	**St. Louis**	**144**	**Wichita**	**216**	**Wilmington**	**144**

4.3Diagnostic Analysis

The diagnostic analysis focuses on uncovering the **underlying relationships between key performance variables** such as **Total Sales (TS)**, **Operating Profit (OP)**, and **Operating Margin (OM). W**e will use the statistical methods like **correlation analysis** and **multiple linear regression**.

Correlation Analysis

To understand the interdependencies among the continuous variables like **Price per Unit (PPU), Units Sold (US), Total Sales (TS), Operating Profit (OP),** and **Operating Margin (OM)** .We performed a **correlation analysis** and visualized through a correlation matrix plot.

The Correlation range is from **-1 to +1**, where:

+1 indicates a perfect positive correlation,
-1 indicates a perfect negative correlation, and
0 suggests no linear relationship.

Strength is interpreted as:

0.0 – 0.3: Weak
0.3 – 0.7: Moderate
0.7 – 1.0: Strong

From the correlation matrix we can say that:

Variables	Corr Value	Significance	Interpretation
PPU ~ US	0.266	***	Weak positive correlation
PPU ~ TS	0.54	***	Moderate positive correlation
PPU ~ OP	0.504	***	Moderate positive correlation
PPU ~ OM	-0.137	***	Weak negative correlation
US ~ TS	0.919	***	Strong positive correlation
US ~ OP	0.872	***	Strong positive correlation
US ~ OM	-0.305	***	Moderate negative correlation
TS ~ OP	0.935	***	Very strong positive correlation
TS ~ OM	-0.302	***	Moderate negative correlation
OP ~ OM	-0.048	***	Very weak negative correlation

Fig-4.1

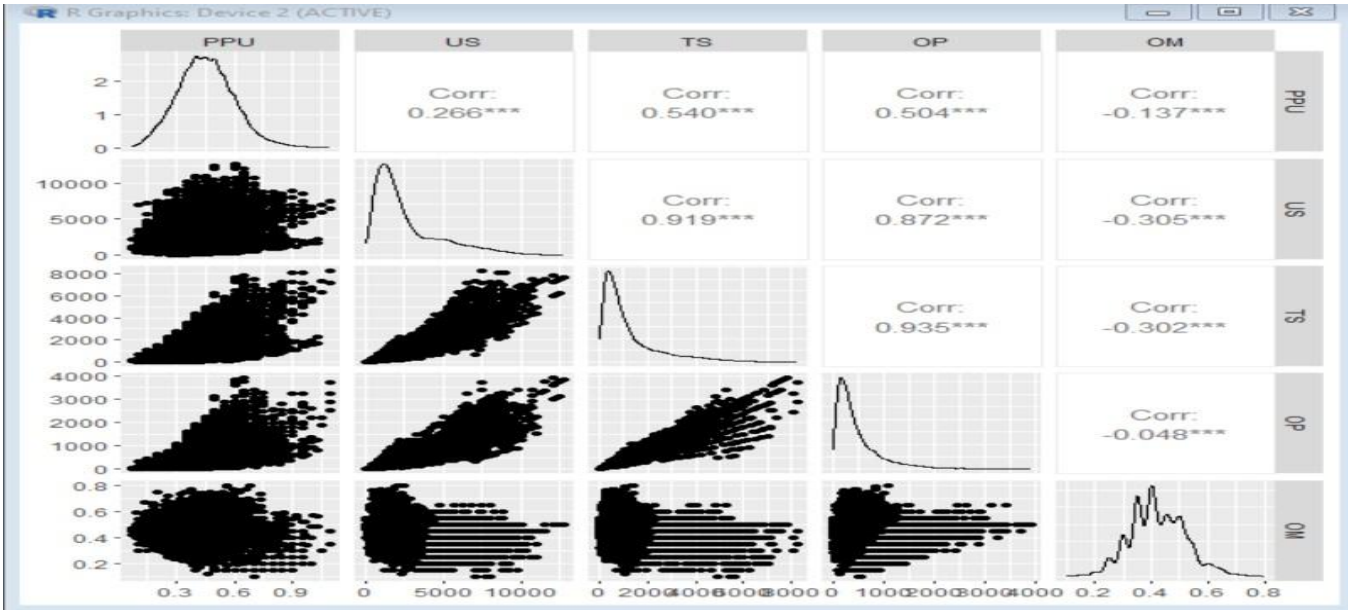

Fig-4.2

Multiple Linear Regression Analysis

To explore how various numerical features influence the Operating Margin (OM) and Total Sales(TS) a Multiple Linear Regression (MLR) model was applied. The model used the following independent variables:

- **Operating Profit (OP)**
- **Price Per Unit (PPU)**
- **Total Sales (TS)**
- **Units Sold (U**

```
Multiple Linear Regression Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                     OM   R-squared:                       0.557
Model:                            OLS   Adj. R-squared:                  0.556
Method:                 Least Squares   F-statistic:                     3026.
Date:                Fri, 15 Aug 2025   Prob (F-statistic):               0.00
Time:                        08:20:33   Log-Likelihood:                 12723.
No. Observations:                9648   AIC:                        -2.544e+04
Df Residuals:                    9643   BIC:                        -2.540e+04
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.4826      0.003    141.007      0.000       0.476       0.489
OP             0.0004   3.86e-06     99.981      0.000       0.000       0.000
PPU           -0.0866      0.007    -11.669      0.000      -0.101      -0.072
TS            -0.0001   2.39e-06    -49.786      0.000      -0.000      -0.000
US         -2.374e-05   1.09e-06    -21.690      0.000   -2.59e-05   -2.16e-05
==============================================================================
Omnibus:                      532.719   Durbin-Watson:                   1.259
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              919.195
Skew:                           0.438   Prob(JB):                    2.51e-200
Kurtosis:                       4.232   Cond. No.                     4.70e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.7e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Fig-4.3

```
Multiple Linear Regression Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                     TS   R-squared:                       0.963
Model:                            OLS   Adj. R-squared:                  0.963
Method:                 Least Squares   F-statistic:                 6.197e+04
Date:                Fri, 15 Aug 2025   Prob (F-statistic):               0.00
Time:                        08:20:33   Log-Likelihood:                -66808.
No. Observations:                9648   AIC:                         1.336e+05
Df Residuals:                    9643   BIC:                         1.337e+05
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         73.1782     22.756      3.216      0.001      28.571     117.785
OP             1.2780      0.016     78.014      0.000       1.246       1.310
PPU         1440.0854     24.352     59.137      0.000    1392.351    1487.820
OM         -1719.1979     34.532    -49.786      0.000   -1786.887   -1651.508
US             0.2424      0.003     69.775      0.000       0.236       0.249
==============================================================================
Omnibus:                     5441.549   Durbin-Watson:                   0.896
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            86838.939
Skew:                           2.372   Prob(JB):                         0.00
Kurtosis:                      16.911   Cond. No.                     5.90e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.9e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Fig-4.4

5.Results and findings

1. Based on the Games-Howell post-hoc test results, the following retailer pairs have similar Operating Margins (OM), as there is no statistically significant difference between Amazon & BevCo, Amazon & Target, Amazon & Walmart, Amazon & West Soda, BevCo & Target, BevCo & Walmart, BevCo & West Soda, Target & Walmart, Target & West Soda, and Walmart & West Soda.

On the other hand, the following retailer comparisons do not have the same OM , there is a significant difference in their Operating Margins,Amazon & FizzyCo, BevCo & FizzyCo, FizzyCo & Target, FizzyCo & Walmart, and FizzyCo & West Soda. This indicates that FizzyCo stands out with a significantly different (likely higher) OM compared to most of the other retailers, suggesting stronger operational profitability.

2. Based on the Games-Howell post-hoc test results the following region pairs do not have similar

Operating Margins (OM) , there is a statistically significant difference between Midwest & Northeast, Midwest & South, Midwest & Southeast, Midwest & West, Northeast & South, Northeast & Southeast, Northeast & West, South & Southeast, South & West, and Southeast & West.

3. Based on the Games-Howell post-hoc test results, the following beverage brand pairs have similar Operating Margins (OM), as there is no statistically significant difference between Coca-Cola & Dasani Water, Diet Coke & Sprite, and Powerade & Sprite

On the other hand, the following beverage brand comparisons do not have the same OM. There is a statistically significant difference in their Operating Margins. Coca-Cola vs Diet Coke, Fanta, Powerade, Sprite, Dasani Water vs Diet Coke, Fanta, Powerade, and Sprite, Diet Coke vs Fanta, and Fanta vs Sprite.

4. Based on the Games-Howell post-hoc test results, the following retailer pairs have similar Operating Profits (OP), as there is no statistically significant difference between Amazon & Target, Amazon & West Soda, BevCo & Amazon, BevCo & Walmart, FizzyCo & Amazon, FizzyCo & Target, FizzyCo & Walmart, FizzyCo & West Soda, Target & West Soda, and Walmart & West Soda.

On the other hand, the following retailer comparisons do not have the same OP.There is a statistically significant difference in their Operating Profits. Amazon & BevCo, Amazon & Walmart, BevCo & FizzyCo, BevCo & Target, BevCo & West Soda, and Walmart & Target.

5. According to the Games-Howell post-hoc test results, the region pair South & West shows no statistically significant difference in Operating Profit (OP), indicating similar profit performance in these areas.

In contrast, the following region comparisons show statistically significant differences in OP: Midwest vs Northeast, South, Southeast, and West; Northeast vs South, Southeast, and West; South vs Southeast; and Southeast vs West.

6. From the Games-Howell post-hoc test results, the beverage brands Powerade and Sprite have similar Operating Profits (OP), as the difference between them is not statistically significant.
However, the following beverage brand comparisons show statistically significant differences in OP. Coca-Cola vs Dasani Water, Diet Coke, Fanta, Powerade, and Sprite; Dasani Water vs Diet Coke, Fanta, Powerade, and Sprite; Diet Coke vs Fanta and Powerade; and Fanta vs Sprite.

7. According to the Games-Howell post-hoc test results, the region pair South & West shows no statistically significant difference in Operating Profit (OP), indicating similar profit performance in these areas.
In contrast, the following region comparisons show statistically significant differences in OP: Midwest vs Northeast, South, Southeast, and West; Northeast vs South, Southeast, and West; South vs Southeast; and Southeast vs West.

8. From the Games-Howell post-hoc test results, certain retailer pairs such as Amazon vs FizzyCo, Amazon vs Target, Amazon vs West Soda, and Walmart vs West Soda do not show statistically significant differences in Operating Profit, suggesting similar profit performance among these pairs.
However, there are significant differences between other retailer combinations: Amazon vs BevCo and Walmart, BevCo vs all other retailers, and FizzyCo vs Walmart, Target, and BevCo.

9. The Games-Howell test shows that Powerade and Sprite have similar Operating Profits (OP), with no significant difference. In contrast, Coca-Cola has significantly higher OP compared to Dasani Water, Diet Coke, Fanta, Powerade, and Sprite. Significant differences also exist between Dasani Water and the rest (except Coca-Cola), Diet Coke vs Fanta and Powerade, and Fanta vs Sprite. Overall, Coca-Cola leads in profitability, while Powerade and Sprite perform similarly.

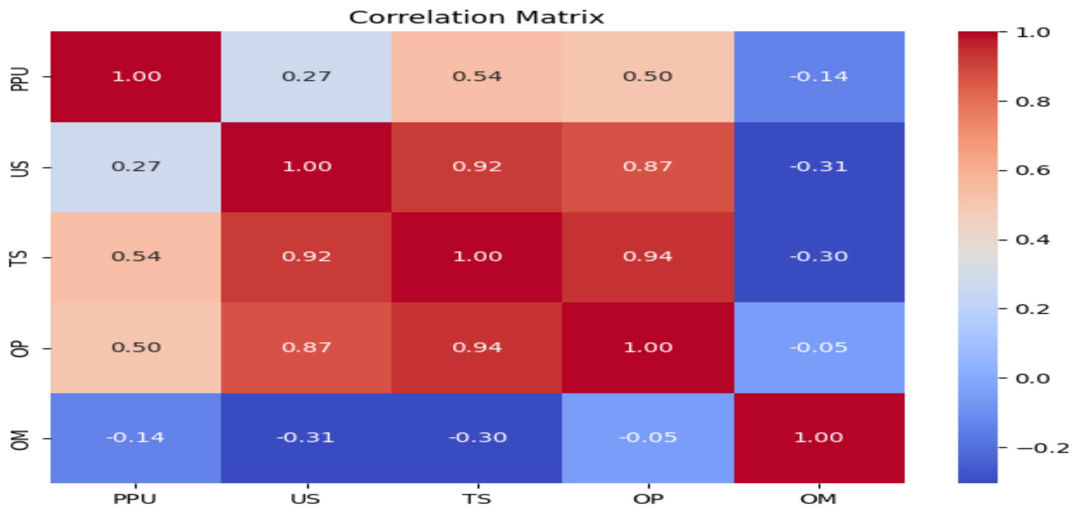

Fig 5.1

10. As PPU (Price Per Unit), TS (Total Sales), and US (Units Sold) increase Operating Margin (OM) will decrease which is saying that PPU, TS, and US are inversely proportional to OM.[from fig 5.1]

11. As PPU (Price Per Unit), US (Units Sold), and OP (Operating Profit) increase, Total Sales (TS) also increase, indicating that PPU, US, and OP are directly proportional to TS.[from fig 5.1]

```
=== Mean OM by Retailer ===
                mean       std  count
Retailer
FizzyCo     0.445399  0.105449   2017
Target      0.419291  0.101299   1030
BevCo       0.418484  0.087807   2618
West Soda   0.417856  0.100466   2374
Amazon      0.413891  0.085159   1005
Walmart     0.409338  0.094756    604

=== Mean OM by Region ===
                mean       std  count
Region
South       0.466898  0.090724   1728
Midwest     0.435272  0.092230   1872
Southeast   0.419167  0.092130   1224
Northeast   0.410450  0.084233   2376
West        0.396691  0.107119   2448

=== Mean OM by Beverage_Brand ===
                     mean       std  count
Beverage_Brand
Coca-Cola        0.446130  0.081071   1610
Dasani Water     0.441318  0.129305   1608
Fanta            0.424359  0.087936   1606
Powerade         0.413225  0.104695   1606
Sprite           0.410199  0.080085   1608
Diet Coke        0.402702  0.082307   1610
```

Fig-5.2

12.FizzyCo has the highest average Operating Margin (0.445), Walmart has the lowest (0.409). [from fig 5.2]

13.South has the highest average Operating Margin (0.467),West has the lowest (0.397) .[from fig 5.2]

14.Coca-Cola has the highest average Operating Margin (0.446),Diet Coke has the lowest (0.402).[from fig 5.2]

15.Walmart has the highest average Total Sales(1523.43) ,BevCo has the lowest(1067.45).[from fig 5.3]

16.Southeast has the highest average Total Sales(1746.35),Midwest has the lowest(890.80). [from fig 5.3]

17.Coca-Cola has the highest average Total Sales(1719.38),Fanta has the lowest(891.46).[from fig 5.3]

=== Mean TS by Retailer ===

Retailer	mean	std	count
Walmart	1523.430464	1417.923133	604
West Soda	1365.263269	1410.325773	2374
Target	1311.972816	1112.294716	1030
Amazon	1279.598010	1375.174431	1005
FizzyCo	1201.882499	1167.565299	2017
BevCo	1067.451872	1168.773356	2618

=== Mean TS by Region ===

Region	mean	std	count
Southeast	1746.354575	1616.340842	1224
West	1488.474265	1339.961385	2448
South	1192.390625	1216.789658	1728
Northeast	1055.574916	1163.737613	2376
Midwest	890.801816	883.553025	1872

=== Mean TS by Beverage_Brand ===

Beverage_Brand	mean	std	count
Coca-Cola	1719.381366	1497.920340	1610
Dasani Water	1484.573383	1419.970107	1608
Diet Coke	1278.162112	1270.819750	1610
Sprite	1069.828980	1119.696860	1608
Powerade	1028.759651	1078.406883	1606
Fanta	891.462640	963.071355	1606

fig 5.3

=== Mean OP by Retailer ===

Retailer	mean	std	count
Walmart	574.339404	501.206007	604
FizzyCo	520.608329	517.506574	2017
West Soda	513.807077	504.391564	2374
Target	503.152427	404.650230	1030
Amazon	491.102488	513.834567	1005
BevCo	417.989305	452.096992	2618

=== Mean OP by Region ===

Region	mean	std	count
Southeast	685.740196	628.603434	1224
South	533.684606	529.894163	1728
West	531.805556	447.628796	2448
Northeast	409.659933	447.717052	2376
Midwest	366.496795	366.952001	1872

=== Mean OP by Beverage_Brand ===

Beverage_Brand	mean	std	count
Coca-Cola	722.340373	607.107413	1610
Dasani Water	602.351368	584.286172	1608
Diet Coke	461.996273	408.113622	1610
Sprite	403.894279	395.758377	1608
Powerade	397.383562	405.500663	1606
Fanta	348.582814	347.052923	1606

Fig 5.4

18. Coca-Cola has the highest average Operating Profit(722.34), Fanta has the lowest(348.58).[From fig 5.4]

19. Southeast has the highest average Operating Profit(685.74),Mideast has the lowest(366.49).[from fig 5.4]

20. Walmart has the highest average Opearting Profit(574.33),BevCo has the lowest(417.98).[from fig 5.4]

21. Washington has the highest average Total Sales (932.72), Florida has the lowest average

Total Sales (~–705.33).

22. South Carolina has the highest average Operating Profit (338.86), Florida again has the lowest average Operating Profit (–246.42).

23. Alaska has the highest average Operating Margin (0.067), Alabama has the lowest average Operating Margin (–0.133).

24.West Soda in the West regions records the highest record of the units sold and Target in the Southeast regions records the lowest.

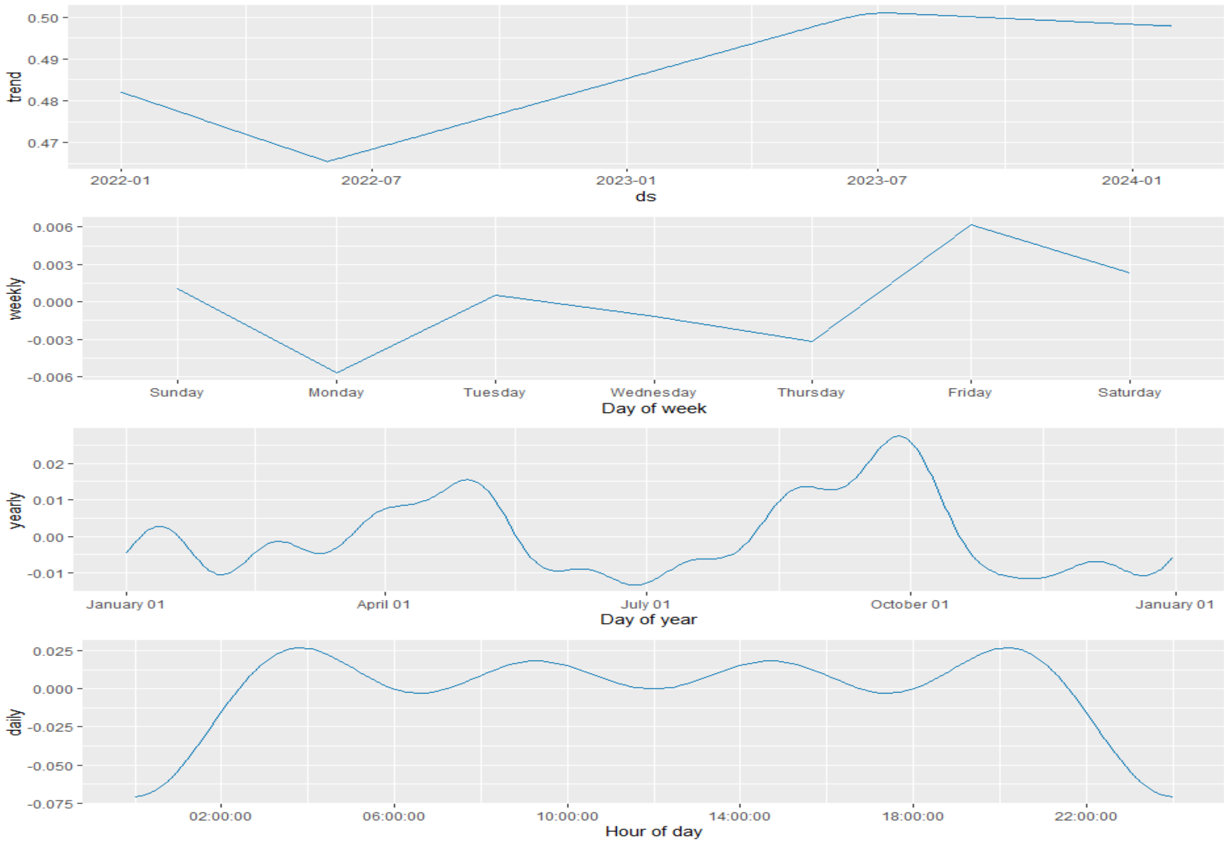

Fig -5.5

25. It is observed that Monday has lowest sales and the highest on Friday. Indicates a clear annual pattern due to climatic, politics and seasonal activities. Hourly pattern is exhibited. [From fig 5.5]

6.Interpretation of Results

6 .1 Overview of Dataset Integrity and Suitability

The dataset which is used in this project is in exceptional quality and structure, which will tell the ideal foundation which is used in data analysis. Which has a total of **9,647 records** and **15 well-defined variables** (7 discrete and 8 continuous), the data **is free from missing values.** They are **zero empty fields** in the dataset, which consists of a memory footprint of **706.2KB** .The dataset is well-structured and lightweight, which makes easy to load the dataset ,explore, and model on any system without slowing things down.

The meaning of data cleanliness is not only preprocessing but also **improving the reliability and interpretability** of all further analysis. Because no error correction was needed, the patterns and relationships identified during exploratory and statistical **analysis** are useful to reflect on the **true business realities** not like the artifacts of data issues. Therefore, all insights and recommendations derived from this dataset are built on a **solid foundation**.

6.2Descriptive and Exploratory Insights

The descriptive analysis will give us of the following things :

Brand Performance: From the data, **Coca-Cola appears to have the highest unit sales across all regions** based on the region wise bar charts we plotted during EDA. It's followed closely by **Dasani Water** and **Diet Coke**, which also show strong performance. Some drinks like **Sprite** and **Fanta** do better in specific regions say, the Northeast and West while **Powerade stands out mostly in the South**, suggesting a regional taste preference. These patterns clearly show that customer demand varies by location and that future inventory or marketing strategies must account for **regional preferences**, not just national trends.

Retailer Contribution: The dataset consists of six retailers and BevCo has the largest number of recorded transactions, with a Retailer ID (1185732). This points to a strong link between BevCo's sales lead and its presence on the ground—likely driven by broader distribution or deeper market reach.

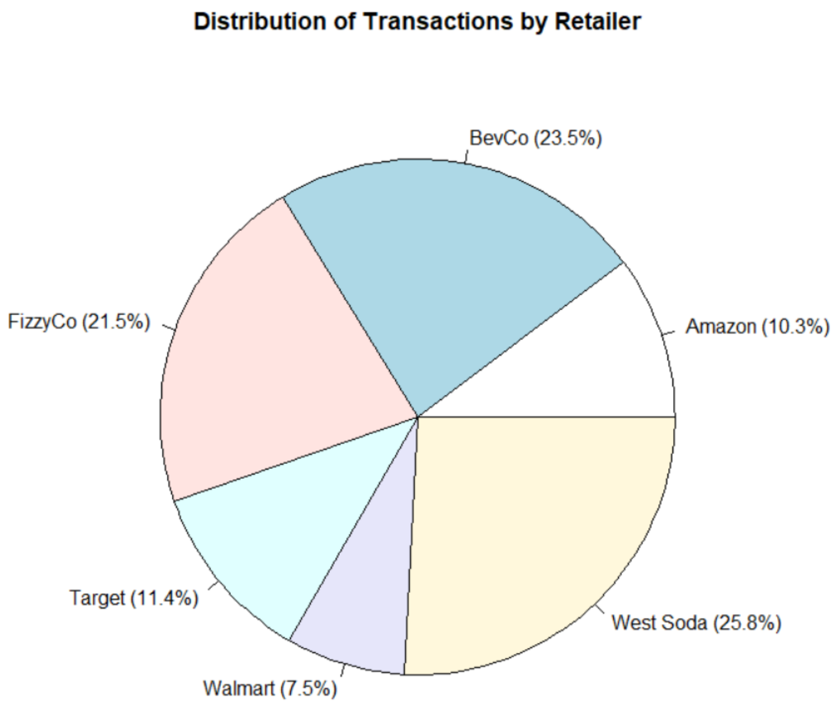

Fig-6.1

Regional Differences: The **West and Northeast regions** have the greatest number of transactions in the given dataset where the **Southeast region** has the lowest. The imbalance in the data may reflect on the actual sales performance. Still it tells us the importance of difference in pricing, marketing, and product based on different regions.

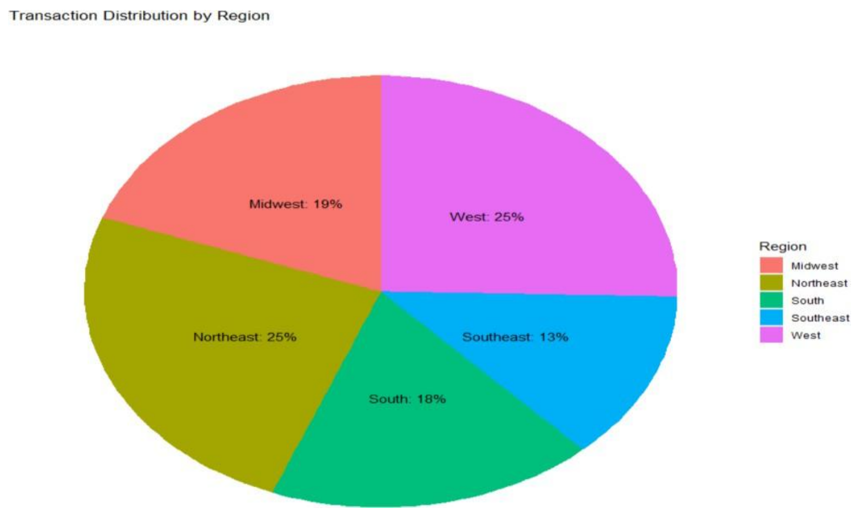

Fig-6.2

Time Trends: The dataset tells us about the two **full years (2022–2023)** and which are distributed across all four quarters. The dataset makes it easy to analyze seasonal trends and patterns over time. The huge number of transactions in 2023 tells us the growth in business operations or data collection and supports more confident modelling and forecasting based on recent trends.

6.3Problem Interpretation and Real-World Implications

From this analysis we can define three problem statements and there insights are as follows:

Problem 1: Uneven Sales Across Regions

Some beverages (like Coca-Cola and Dasani Water) have highest sales in some regions like the West and Northeast, but others (e.g., Powerade) have the least number of sales. This imbalance leads to some beverages are selling like crazy and requesting for more stock, while the other beverages are not selling properly and getting wasted. This uneven sales trend is causing waste and missing out on profits.

Insight: The visualizations (regional bar charts and heatmaps) clearly illustrated this pattern, supporting the need for **data-driven inventory planning** based on **region-brand sales alignment**.

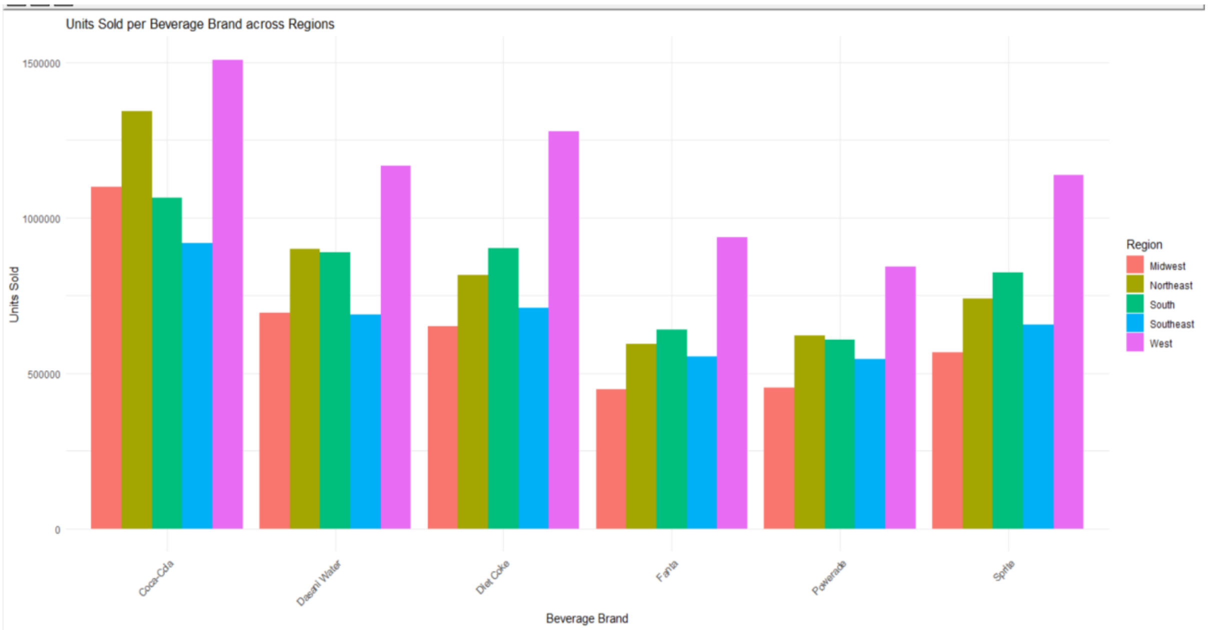

Fig-6.3

Problem 2: Uniform Pricing and Promotions

The dataset uses the same marketing and pricing strategy across all regions failing to consider **local economic conditions, preferences and price sensitivity. Which** may also lead to lower ROI, can hurt customer satisfaction and loss in revenue. But the people are different have different tastes, culture, lifestyle and so on. This may lead to loss in market.

Insight: The correlation and regression tell us that margins and sales behave differently depending upon the price and region. It makes us one thing clear that we should not treat every region the same. Pricing and promotional strategies need to fit how people shop in each region.

Problem 3: Mismatch of Product with Market Needs

Regions with high demand for specific beverages are sometimes underserved or misaligned in product offerings, leading to **poor performance of otherwise high potential products**.
This means that the beverages that are available do not always match the needs of people in different regions. Because of this, some beverages have low sales even if people love them in other areas. Not all regions get the beverages they actually need So, even if a drink is loved in one region, it does not mean it will sell properly in all the regions.

Insight: The data makes it clear just by taking bestsellers from one region and pushing them into another doesn't work. Each region needs its own mix of products, based on what people actually want and how they shop there.

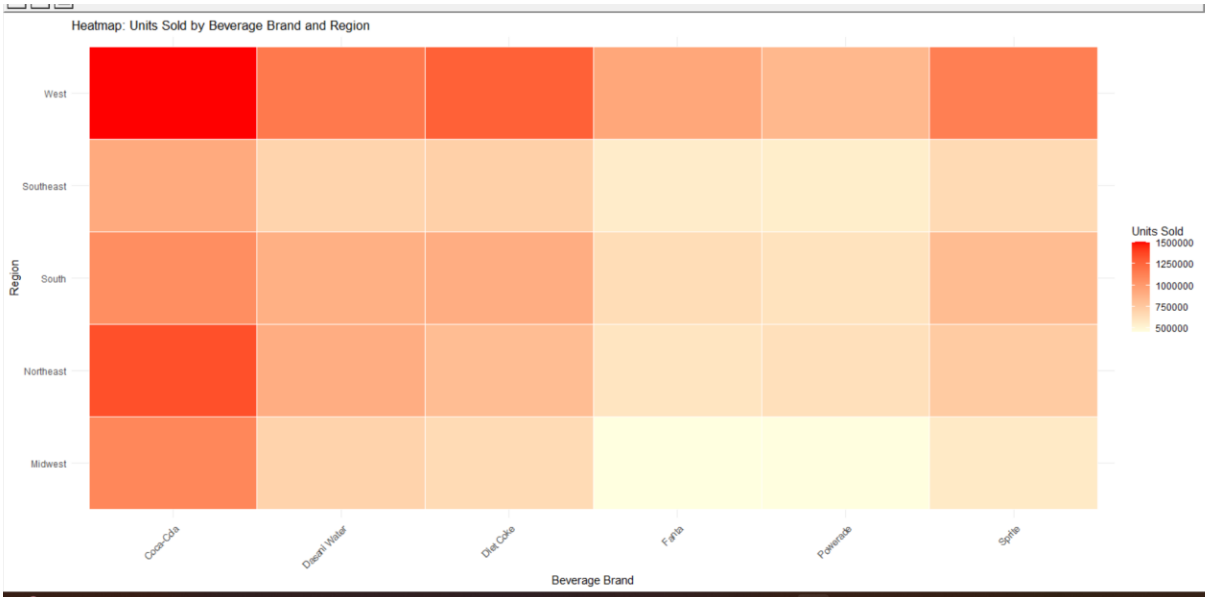

Fig-6.4

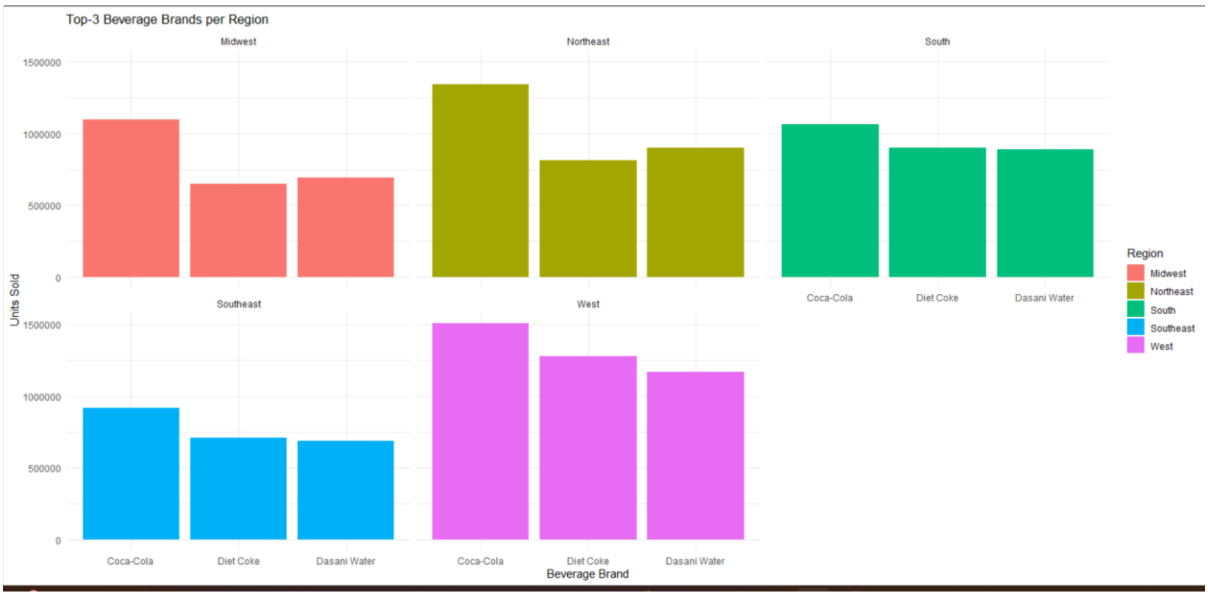

Fig-6.5

6.4 Recommendations

Based on the analysis of beverage sales performance, statistical modeling, and correlation insights, the following recommendations are proposed to improve operational efficiency, reduce waste, and enhance profitability:

1. We should decrease the Price per unit, so that the Operating Margin and the Operating Profit will increase but the Total Sales will also decrease.

2. Customer needs changes based on the regions. From the analysis we can say that Powerade is selling well in South, while Sprite and Fanta are popular in the Northeast and West. If you tailor the product selection and stock levels to match what people actually want in different regions we will make them happy and reduce the wastage of stock. We can sell more of what works and waste less on what is not need.

3. Uniform pricing in all the regions is ignoring the regional income levels, lifestyle , culture and many other. Pricing and promotional strategies should be made for each region separately to increase the sales and profit.

4. The Southeast regions are showing lower sales, so it might help to try more local friendly approaches like targeted marketing, special combo offers, or adjusting prices to boost customer interest and increase demand in those areas.

5. We should focus on the Walmart more to increase the Operating margin as it has the lowest.

6. To strengthen our presence in the West and Northeast regions, we recommend focusing on targeted promotions, leveraging regional market insights, and improving distribution strategies.

7. Coca-Cola is having more Operating Margin, Operating Profit, and Total Sales which says that the strong overall performance. Dasani Water also having a better performance across all metrics. Fanta is having less values in OP and TS , we should target more advisement, offers on the Fanta Beverage brand.

8. Different retailers like BevCo, Target, and Walmart are having their own customer buying patterns. Customizing the products and pricing strategies to increase the each retailer's profile can boost performance and cut down on logistics issues.

9. Beverage Brands like Sprite and Fanta are have moderate sales and perform well in specific regions. With proper branding and promotion, they can become strong contributors to overall sales.