Intro

Sales data for a manufacturer is used as data set for this project to Analyzing Sales Data to Identify Trends and Develop Effective Strategies.

Variables

1. Product: This refers to the name of the product being analyzed.
2. UPC 13 digit: This refers to the Universal Product Code, which is a unique 13-digit number that is used to identify a specific product.
3. Dollar Sales No Merch: This refers to the dollar sales generated by the product without any merchandising support.
4. Dollar Sales Any Merch: This refers to the dollar sales generated by the product with any type of merchandising support.
5. Unit Sales No Merch: This refers to the number of units of the product sold without any merchandising support.
6. Unit Sales Any Merch: This refers to the number of units of the product sold with any type of merchandising support.
7. Volume Sales No Merch: This refers to the volume of the product sold (in fluid ounces, for example) without any merchandising support.
8. Volume Sales Any Merch: This refers to the volume of the product sold with any type of merchandising support.
9. Price per Unit: This refers to the price of the product per unit (such as per ounce or per piece).
10. Price per Unit No Merch: This refers to the price of the product per unit without any merchandising support.
11. Price per Unit Any Merch: This refers to the price of the product per unit with any type of merchandising support.
12. Price per Volume: This refers to the price of the product per unit of volume (such as per ounce or per liter).
13. Price per Volume No Merch: This refers to the price of the product per unit of volume without any merchandising support.
14. Price per Volume Any Merch: This refers to the price of the product per unit of volume with any type of merchandising support.
15. ACV Weighted Distribution No Merch: This refers to the distribution of the product in stores without any merchandising support, weighted by the size of each store (in terms of total annual sales).
16. ACV Weighted Distribution Any Merch: This refers to the distribution of the product in stores with any type of merchandising support, weighted by the size of each store.
17. Base Unit Sales: This refers to the number of units of the product sold in the absence of any merchandising support.

18. Base Volume Sales: This refers to the volume of the product sold in the absence of any merchandising support.
19. Base Dollar Sales: This refers to the dollar sales generated by the product in the absence of any merchandising support.
20. Incremental Units: This refers to the additional number of units of the product sold as a result of the merchandising support.
21. Incremental Volume: This refers to the additional volume of the product sold as a result of the merchandising support.
22. Incremental Dollars: This refers to the additional dollar sales generated by the product as a result of the merchandising support.

Methodology

A linear regression model is used to predicts "Dollar Sales Any Merch" based on several predictors such as "ACV Weighted Distribution Any Merch", "Price per Unit", "Unit Sales Any Merch", "Time", "Geography", and "Brand"

Result of regression -

Residuals:

Min	1Q	Median	3Q	Max
-1863669	-797	89	857	1560167

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-4.619e+02	1.174e+04	-0.039	0.968622	
data$`ACV Weighted Distribution Any Merch`	2.855e+02	9.148e+00	31.206	< 2e-16	***
data$`Price per Unit`	5.462e+02	1.840e+01	29.688	< 2e-16	***
data$`Unit Sales Any Merch`	2.883e+00	1.334e-03	2161.892	< 2e-16	***
data$TimeWeek Ending 01-14-18	-1.869e+01	3.271e+02	-0.057	0.954433	
data$TimeWeek Ending 01-21-18	-1.050e+02	3.269e+02	-0.321	0.747973	
data$TimeWeek Ending 01-28-18	7.852e+01	3.271e+02	0.240	0.810324	
data$TimeWeek Ending 02-04-18	6.007e+01	3.267e+02	0.184	0.854110	
data$TimeWeek Ending 02-11-18	-5.170e+02	3.268e+02	-1.582	0.113645	
data$TimeWeek Ending 02-18-18	1.130e+01	3.264e+02	0.035	0.972381	
data$TimeWeek Ending 02-25-18	1.085e+02	3.270e+02	0.332	0.740054	
data$TimeWeek Ending 03-04-18	4.754e+01	3.268e+02	0.145	0.884343	
data$TimeWeek Ending 03-11-18	5.789e+01	3.263e+02	0.177	0.859190	
data$TimeWeek Ending 03-18-18	-1.275e+02	3.266e+02	-0.390	0.696297	
data$TimeWeek Ending 03-25-18	-6.757e+02	3.264e+02	-2.070	0.038473	*
data$TimeWeek Ending 04-01-18	-1.530e+03	3.264e+02	-4.687	2.77e-06	***
data$TimeWeek Ending 04-08-18	-2.476e+02	3.265e+02	-0.758	0.448248	
data$TimeWeek Ending 04-15-18	-1.089e+02	3.260e+02	-0.334	0.738393	
data$TimeWeek Ending 04-22-18	-1.418e+02	3.264e+02	-0.434	0.663963	
data$TimeWeek Ending 04-29-18	1.657e+02	3.265e+02	0.507	0.611831	
data$TimeWeek Ending 05-06-18	-1.425e+02	3.269e+02	-0.436	0.662780	
data$TimeWeek Ending 05-13-18	-1.835e+02	3.262e+02	-0.563	0.573721	
data$TimeWeek Ending 05-20-18	-5.180e+02	3.264e+02	-1.587	0.112571	
data$TimeWeek Ending 05-27-18	1.450e+02	3.256e+02	0.445	0.656149	
data$TimeWeek Ending 06-03-18	-1.032e+02	3.264e+02	-0.316	0.751835	
data$TimeWeek Ending 06-10-18	6.798e+01	3.258e+02	0.209	0.834696	
data$TimeWeek Ending 06-17-18	1.957e+01	3.260e+02	0.060	0.952130	
data$TimeWeek Ending 06-24-18	8.717e+01	3.266e+02	0.267	0.789558	
data$TimeWeek Ending 07-01-18	-1.839e+02	3.262e+02	-0.564	0.572899	
data$TimeWeek Ending 07-08-18	-5.213e+02	3.265e+02	-1.597	0.110374	
data$TimeWeek Ending 07-15-18	-2.466e+02	3.262e+02	-0.756	0.449726	
data$TimeWeek Ending 07-22-18	-2.526e+02	3.263e+02	-0.774	0.438811	
data$TimeWeek Ending 07-29-18	-2.678e+02	3.264e+02	-0.821	0.411837	
data$TimeWeek Ending 08-05-18	2.528e+01	3.257e+02	0.078	0.938133	
data$TimeWeek Ending 08-12-18	-2.432e+02	3.252e+02	-0.748	0.454449	
data$TimeWeek Ending 08-19-18	1.040e+02	3.252e+02	0.320	0.749066	
data$TimeWeek Ending 08-26-18	-1.455e+02	3.257e+02	-0.447	0.654927	
data$TimeWeek Ending 09-02-18	9.406e+00	3.258e+02	0.029	0.976969	
data$TimeWeek Ending 09-09-18	-5.503e+01	3.259e+02	-0.169	0.865890	
data$TimeWeek Ending 09-16-18	1.045e+01	3.262e+02	0.032	0.974436	
data$TimeWeek Ending 09-23-18	-2.253e+01	3.255e+02	-0.069	0.944823	
data$TimeWeek Ending 09-30-18	-1.319e+01	3.257e+02	-0.040	0.967704	
data$TimeWeek Ending 10-07-18	-4.919e+01	3.255e+02	-0.151	0.879876	
data$TimeWeek Ending 10-14-18	-1.914e+02	3.258e+02	-0.587	0.556915	
data$TimeWeek Ending 10-21-18	-1.538e+02	3.256e+02	-0.472	0.636693	
data$TimeWeek Ending 10-28-18	1.997e+02	3.257e+02	0.613	0.539883	
data$TimeWeek Ending 11-04-18	-1.424e+02	3.251e+02	-0.438	0.661325	
data$TimeWeek Ending 11-11-18	-4.630e+02	3.256e+02	-1.422	0.155048	
data$TimeWeek Ending 11-18-18	-1.205e+02	3.250e+02	-0.371	0.710809	
data$TimeWeek Ending 11-25-18	-3.407e+02	3.239e+02	-1.052	0.292831	
data$TimeWeek Ending 12-02-18	-2.721e+02	3.249e+02	-0.837	0.402431	

```
                                                                  -3.383e+02  3.248e+02   -1.042 0.297499
                                                                   1.770e+02  3.251e+02    0.545 0.586072
                                                                  -2.705e+02  3.248e+02   -0.833 0.404971
                                                                   1.103e+02  3.255e+02    0.339 0.734755
data$GeographyGreat Lakes - IRI Standard - Multi Outlet + Conv    -9.495e+02  1.605e+02   -5.915 3.32e-09 ***
data$GeographyMid-South - IRI Standard - Multi Outlet + Conv      -6.234e+02  1.604e+02   -3.887 0.000101 ***
data$GeographyNortheast - IRI Standard - Multi Outlet + Conv      -1.577e+02  1.558e+02   -1.012 0.311508
data$GeographyPlains - IRI Standard - Multi Outlet + Conv         -8.448e+02  1.691e+02   -4.995 5.90e-07 ***
data$GeographySouth Central - IRI Standard - Multi Outlet + Conv -7.156e+02  1.729e+02   -4.140 3.47e-05 ***
data$GeographySoutheast - IRI Standard - Multi Outlet + Conv      -7.281e+02  1.669e+02   -4.362 1.29e-05 ***
data$GeographyTotal US - Multi Outlet + Conv                      -2.184e+02  1.431e+02   -1.526 0.126925
data$GeographyWest - IRI Standard - Multi Outlet + Conv           -5.565e+02  1.677e+02   -3.319 0.000904 ***
data$brandADMIRATION                                               8.708e+01  1.180e+04    0.007 0.994112
data$brandAL HALOUB COW                                           -8.527e+03  1.181e+04   -0.722 0.470291
data$brandALCAM                                                   -1.243e+03  1.180e+04   -0.105 0.916119
.....and more
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16610 on 270262 degrees of freedom
Multiple R-squared: 0.9509, Adjusted R-squared: 0.9508
F-statistic: 1.346e+04 on 389 and 270262 DF, p-value: < 2.2e-16

Findings

- The intercept is -461.9, but it is not statistically significant (p-value > 0.05).
- ACV Weighted Distribution Any Merch has a positive coefficient of 285.5, which means that a one-unit increase in ACV Weighted Distribution Any Merch is associated with an increase of $285.5 in Dollar Sales Any Merch. This predictor is statistically significant (p-value < 0.001).
- Price per Unit has a positive coefficient of 546.2, which means that a one-unit increase in Price per Unit is associated with an increase of $546.2 in Dollar Sales Any Merch. This predictor is also statistically significant (p-value < 0.001).
- Unit Sales Any Merch has a positive coefficient of 2.88, which means that a one-unit increase in Unit Sales Any Merch is associated with an increase of $2.88 in Dollar Sales Any Merch. This predictor is highly significant (p-value < 0.001).
- The Time predictor includes several weeks ending at different dates.

Week Ending	Significance
3/25/18	0.038473
4/1/18	2.77E-06

- The Geography predictor includes several regions. The coefficients for Great Lakes, Mid-South, Plains, and South Central are all negative and statistically significant, indicating that these regions have lower Dollar Sales Any Merch than the reference region (West - IRI Standard - Multi Outlet + Conv). The Northeast region is not statistically significant.

Significant regions	Significance
GeographyGreat Lakes - IRI Standard - Multi Outlet + Conv	3.32E-09

GeographyMid-South - IRI Standard - Multi Outlet + Conv	0.000101
GeographyPlains - IRI Standard - Multi Outlet + Conv	5.90E-07
GeographySouth Central - IRI Standard - Multi Outlet + Conv	3.47E-05
GeographySoutheast - IRI Standard - Multi Outlet + Conv	1.29E-05
GeographyWest - IRI Standard - Multi Outlet + Conv	0.000904

- Overall, this model suggests that ACV Weighted Distribution Any Merch, Price per Unit, and Unit Sales Any Merch are all positively associated with Dollar Sales Any Merch. The impact of Time and Geography on Dollar Sales Any Merch varies depending on the specific time period and region.

Effect of geography on sales-

```
> anova_sales_geography <- aov(data$`Unit Sales Any Merch` ~ Geography, data = data)
> summary(anova_sales_geography)
               Df    Sum Sq   Mean Sq F value Pr(>F)
Geography       8 8.570e+11 1.071e+11   169.3 <2e-16 ***
Residuals   270643 1.713e+14 6.329e+08
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

The results of the ANOVA table indicate that the variable "Geography" is significant with a p-value < 0.05. This suggests that there is a significant difference in sales between the different geographical regions. The F-value of 169.3 with 8 degrees of freedom for Geography indicates that the variation in sales between the geographical regions is much higher than the variation within each region.
Therefore, we can conclude that Geography is an important factor that influences sales. The results of the ANOVA suggest that the company should consider developing different marketing and merchandising strategies for each geographical region in order to optimize sales.

Effect of brand on sales-

```
> anova_sales_brand <- aov(data$`Unit Sales Any Merch` ~ brand, data = data)
> summary(anova_sales_brand)
               Df    Sum Sq   Mean Sq F value Pr(>F)
brand         327 2.171e+12 6.638e+09   10.56 <2e-16 ***
Residuals   270324 1.700e+14 6.288e+08
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

The results of the ANOVA suggest that the brand variable is statistically significant in explaining the variation in sales. The F-value of 10.56 is significant at the 0.01 level, which means that there is a very low probability that the observed relationship between brand and sales is due to chance.
The null hypothesis in the ANOVA is that there is no significant difference in sales across different brands, while the alternative hypothesis is that there is a significant difference in sales across brands. Since the p-value is less than 0.05, we can reject the null hypothesis and conclude that there is a significant difference in sales across brands.
In other words, the brand variable is an important predictor of sales, and there is evidence to suggest that companies should consider varying their merchandising strategies based on the brand.

Result

The F-value for the geography is higher than the F-value of the brand model , this indicating that merchandising strategies should be based on geography when compared to the brand.