Option #1: Business Analytics - Retail Clothing Store

Scott Miner

Colorado State University – Global Campus

Option #1: Business Analytics - Retail Clothing Store

For his final *Portfolio Project* in *MIS543 – Enterprise Performance Management*, the student uses SAS to analyze the **clothing_store.csv** dataset, a dataset containing customer information for a retail clothing organization that wants to increase sales and maximize profits by utilizing more effective direct marketing techniques. Specifically, the organization wants to identify customers who are responding to direct marketing promotions. Also, the organization wants to gain insight into how they can better forecast future business growth. In this paper, the student conducts the appropriate descriptive analytics tests to summarize the organization's dataset. Additionally, the student provides business questions the dataset can answer, null and alternative hypotheses for each business question, and conducts the necessary predictive analytics tests to help decision-makers achieve their business goals. Lastly, the student analyzes his findings concerning the business questions and hypotheses and gives his recommendations for what other data sources the organization can utilize to achieve its business goals.

The organization's dataset contains observations from 28,799 customers and 29 variables. Figure 1 provides descriptive statistics for all variables at the end of this paper. To summarize, the average time that a customer has been on file with the organization is 436.91 days, with values ranging between 1 and 717 days. Each customer has an average of 11.53 promotions on file. Some customers have as few as 0 promotions on file, while others have as many as 38. Customers make an average of 5.04 purchase visits to the store, and the average number of days between purchase visits is 4.79. Each customer spends an average of $113.59 at the store per visit. In total, customers spend an average of $473.21 at the store across all visits. The average gross margin percentage (GMP) per customer is 52%, which is "the amount of money the company pockets after accounting for costs, expressed as a percentage of total sales" (Sable, 2019). Customers spend the most on sweaters (21.39%) and the least on jewelry (0.9%). There are 17,768 customers in the dataset who do not use credit cards (61.70%) and 11,031 who do (38.30%). Also, there are 51 MicroVision lifestyle cluster types in the dataset, including type 0. The six most common lifestyle types account for 45.62% of all lifestyle cluster types. Figures 19 – 20 show the count of credit card users and lifestyle cluster types, respectively. Finally, three normality tests returned by SAS, including the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling, indicate that all dataset variables are distributed nonnormally (Elliott & Woodward, 2016).

One question the dataset can answer to help the organization achieve its business goals is, "Do any variables correlate strongly with the total net sales or the gross margin percentage (GMP)?" The null hypothesis, H_0, states there is no linear relationship between any variables in the dataset and the total net sales or GMP. The alternate hypothesis, H_a, states there is a linear relationship between any dataset variables and the total net sales or GMP. The student conducts a correlation analysis to determine whether such an association exists, examining the Spearman correlation coefficient, R_s, which is the most popular nonparametric correlation measure. Across various nonnormal bivariate distributions, R_s is more potent than Pearson's coefficient, R_p, and, through rank-ordering, causes outliers in a variable's distribution to converge toward the center (Bishara & Hittner, 2012; Croux & Dehon, 2010). The higher the absolute value of R_s, which ranges between -1 and 1, the stronger the association between two variables (Puth *et al.*, 2015).

Figures 2 – 3 show the top ten variables in the dataset most strongly correlated with total net sales and GMP. All variables return p-values < .0001. The most strongly correlated variable with total net sales is the frequency of purchase visits (R_s = 0.74040). Both the average time between visits in days (R_s = -0.64230) and the number of days between purchases (R_s = -0.57981) correlate negatively with total net sales, while the number of promos on file for a customer (R_s = 0.54563) and whether the customer uses a credit card (R_s = 0.45697) correlate positively with total net sales. In terms of the GMP, the markdown percentage on customer purchases (R_s = -0.80566) is the most strongly correlated variable, followed by the average amount spent per visit (R_s = 0.38599), followed by the percent spent on collectibles (R_s = 0.28847). Figures 4 – 5 show scatterplot matrices for all these correlations. As a result, we can reject the null hypothesis and conclude that there are significant linear relationships between several variables in the dataset and the total net sales and GMP. What this means for the organization is that to increase total net sales per customer, the organization should send customers with credit cards more promotions to get them in the store more often. To increase the GMP, the organization should scale back markdowns, attempt to increase the average amount each customer spends, and market collectibles more heavily.

A related business question this dataset can answer is, "Are there any differences among the distributions of the GMP, the average amount that customers spend per visit, or the percentage of total sales spent by the customer on the respective product categories for any of the six most common lifestyle cluster types in the dataset?" The null hypothesis, H_0, says there are no differences among the distributions of these variables for any of the six most common lifestyle cluster types. In contrast, the alternate hypothesis, H_a, says there is a difference among these variables' distributions for any of the six most common lifestyle cluster types. The appropriate statistical analysis test in this circumstance is the Kruskal-Wallis (KW) test, a nonparametric test that compares differences between the mean ranks of *k* independent groups. The results of the KW test indicate differences in the distribution of the GMP (p = 0.0057), the average amount spent per visit (p = 0.0031), and the percent spent on sweaters (p = 0.0013).

When rejecting a KW test's null hypothesis, Dunn's test is the appropriate nonparametric pairwise multiple comparison procedure to determine which groups are different from which (Dinno, 2015). Dunn's test results indicate that the "Upper Crust" lifestyle cluster type produces a significantly higher GMP for the organization than the "Great Beginnings" lifestyle cluster type. Also, the "Mid-life Success" lifestyle cluster type spends significantly more per visit on average than the "Home Sweet Home" and "Great Beginnings" cluster types. Further, the "Upper Crust" and "Mid-life Success" cluster types spend significantly more on sweaters than the "Movers and Shakers." Figures 6 – 11 convey these results. What does this mean for the organization? To increase the GMP, the organization should target the "Upper Crust" and "Mid-life Success" cluster types more heavily. The "Mid-life Success" cluster type spends significantly more on average per visit than the other cluster types mentioned above. The average amount spent per visit correlates heavily with the GMP, as discussed previously. One way to get the "Mid-life Success" cluster type into the store more often would be to promote sweaters to them more frequently.

Another related business question this dataset can answer is, "Are there any differences among the distribution of the gross margin percentage (GMP) according to customer credit card usage?" The null hypothesis, H_0, says there is no difference between the distribution of the GMP according to customer credit card usage. The alternate hypothesis, H_a, says there is a difference between the GMP's distribution according to whether a customer uses a credit card. The appropriate statistical test to use in this instance is the Wilcoxon test, a nonparametric test that detects differences in central tendency between two independent, nonnormally distributed samples (Feltovich, 2003). The KW test is also applicable in this scenario since it contains two or more independent samples (Elliott & Woodward, 2016). The KW test reports $p < 0.0001$, indicating a significant difference in the groups' GMP distribution. We, therefore, reject the null hypothesis. Notably, the results indicate that customers who do *not* use credit cards yield a higher GMP for the organization than those who do. Figures 12 – 13 illustrate these results. On which product categories are customers who do not use credit cards most likely to spend their money? Figure 14 shows these results, indicating that customers without credit cards are most likely to spend their money on sweaters (22.64%), followed by jackets (12.78%), followed by blouses (9.68%). Therefore, the organization should advertise these products more often to customers without credit cards to increase its GMP.

Lastly, the organization wants to gain insight into how they can better predict future business growth. Therefore, we can ask the business question, "Do any variables serve as significant predictor variables for total net sales?" The null hypothesis, H_0, states there will be no significant prediction of total net sales from any variables in the dataset. The alternate hypothesis, H_a, states there will be a significant predictor of total net sales from any dataset variables. Multiple linear regression (MLR) allows analysts to predict a single dependent variable (Y) from multiple independent variables (X_1, X_2, X_3, …, X_k) (Berger, 2003). The coefficient of determination, R^2, ranges between 0 and 1 and measures how well the regression model fits the data (Ostertagová, 2012). More specifically, R^2 portrays the proportion of the variation of the outcome variables explained by the explanatory variables (Kasuya, 2019).

Using MLR, the student finds a parsimonious model capable of explaining 64.14% of the variance in total sales per customer (R^2 = 0.6414; p < .0001) using just two predictor variables, namely, the number of purchase visits per customer (F = 47132.3; p < .0001) and the average amount each customer spends per visit (F = 10835.2; p < .0001). The following equation illustrates this model: $Total\ Sales\ = -255.78780\ +\ 80.87016\ \times \mathrm{fre} +\ 2.83034\ \times avrg.$

An R^2 value above .6, as returned here, indicates a satisfactory predictive model, though there may be a relatively high number of errors in predictions (Ostertagová, 2012). Performing MLR according to the six most common lifestyle cluster types, we can increase the R^2 value by 0.0987 (9.87%) and most accurately predict total sales per customer for the "Upper Crust" lifestyle cluster type (R^2 = 0.7401; $p < .0001$) using the following equation: $Total\ Sales\ = -218.07848 +\ 78.43632\ \times \text{fre} +\ 2.53933\ \times avrg.$ Therefore, we can reject the null hypothesis and conclude that the organization can significantly predict total net sales based on the frequency of purchase visits and the average amount each customer spends per visit. Figures 15 – 18 display these results and the fit diagnostics for each model.

These analyses provide a good starting point for the organization to achieve its business goals and predict future product sales. However, the organization could benefit by including additional data sources, including big data sources. Examples of big data sources include social media data, clickstream data, sensor data, and video data. The organization could run sentiment analysis on social media data to determine how customers are talking about their products. Clickstream data will help the organization determine whether customers abandon webpages after viewing specific products. Sensors can be installed in outlets to determine how much foot traffic comes into shops. Likewise, CCTV video data can be analyzed to study which products customers view in the shop to boost sales further (Marr, 2015). These big data sources will better help the organization achieve its business goals. In conclusion, this paper analyzed the **clothing_sales.csv** dataset. The student provided four business questions the dataset can answer, their null and alternate hypotheses, and the appropriate predictive tests to aid decision-makers. Lastly, the student gave his recommendation for what other data sources the organization could use to achieve its business goals.

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
customer_id	a unique customer identification number	28799	24360.00	8313.70	9961.00	38759.00
days	Number of days the customer has been on file	28799	436.9161776	192.9708984	1.0000000	717.0000000
promos	Number of Marketing Promos on File	28799	11.5391159	7.1393560	0	38.0000000
ln_lifetime_ave_time_betw_visits	Average time between visits in days	28799	3.9237425	1.0204171	-2.4100000	5.9000000
ln_days_between_purchases	Number of days between purchases	28799	4.7932341	0.8727006	0	6.5800000
fre	Total Number of purchase visits	28799	5.0390291	6.3491216	1.0000000	115.0000000
avrg	Avg. Amount Spent Per Visit	28799	113.5883176	86.9808026	0.4900000	1919.88
MON	Total Net Sales	28799	473.2124633	659.3274137	0.9900000	24140.33
GMP	Gross Margin Percentage	28799	0.5179412	0.1722468	-6.4600000	0.9900000
MARKDOWN	Markdown percentage on customer purchases	28799	0.1871020	0.1292032	0	0.9500000
PERCRET	Percent of Returns	28799	0.1291021	0.5431292	0	40.9200000
clustype	Microvision Lifestyle Cluster Type	28799	15.1638599	12.2464390	0	50.0000000
cc_card	Credit Card Usage	28799	0.3830341	0.4861350	0	1.0000000
PSWEATERS	% Spent on Sweaters	28799	0.2139460	0.2311677	-0.9700000	1.0000000
PKNIT_TOPS	% Spent on Knit tops	28799	0.0272138	0.0680677	-0.3100000	1.0000000
PKNIT_DRES	% Spent on Knit dresses	28799	0.0411240	0.1109860	-0.7100000	1.0000000
PBLOUSES	% Spent on Blouses	28799	0.0930296	0.1355609	-0.6600000	1.0000000
PJACKETS	% Spent on Jackets	28799	0.1356939	0.1841386	-0.3600000	1.0000000
PCAR_PNTS	% Spent on Career Pants	28799	0.0851193	0.1411686	-0.7700000	1.0000000
PCAS_PNTS	% Spent on Casual Pants	28799	0.0686125	0.1327096	-0.5000000	1.0000000
PSHIRTS	% Spent on Shirts	28799	0.0657492	0.1167469	-0.7500000	1.0000000
PDRESSES	% Spent on Dresses	28799	0.0683635	0.1579639	-0.4200000	1.0000000
PSUITS	% Spent on Suits	28799	0.0333671	0.1300946	-0.5900000	1.0000000
POUTERWEAR	% Spent on Outerwear	28799	0.0181944	0.1000981	-0.7300000	1.0000000
PJEWELRY	% Spent on Jewelry	28799	0.0097621	0.0364999	-0.1100000	1.0000000
PFASHION	% Spent on Fashionable Wear	28799	0.0300010	0.0796572	-0.6700000	1.0000000
PLEGWEAR	% Spent on Leg Wear	28799	0.0127216	0.0500886	-0.1000000	1.0000000
PCOLLSPND	% Spent on Collectibles	28799	0.0735088	0.1765617	-0.4400000	1.0000000
zip_code	Customers's Zip Code	28799	49023.47	24084.64	0	99687.00

Figure 1. Descriptive statistics for the customer_sales.csv dataset

Spearman Correlation Coefficients, N = 28799
Prob > |r| under H0: Rho=0

	fre	ln_lifetime_ave_time_betw_visits	ln_days_between_purchases	promos	cc_card	PJACKETS	avrg	days	PCAR_PNTS	PSHIRTS
MON **Total Net Sales**	0.74040	-0.64230	-0.57981	0.54563	0.45697	0.38059	0.36673	0.36496	0.34437	0.33642
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

Figure 2. Spearman Correlation Coefficients for the top 10 variables in the dataset most correlated with total net sales

Spearman Correlation Coefficients, N = 28799
Prob > |r| under H0: Rho=0

	MARKDOWN	avrg	PCOLLSPND	fre	ln_lifetime_ave_time_betw_visits	ln_days_between_purchases	PKNIT_TOPS	PSWEATERS	PCAS_PNTS	PERCRET
GMP **Gross Margin Percentage**	-0.80566	0.38599	0.28847	-0.24796	0.23269	0.20806	-0.16907	-0.16109	-0.13692	-0.12530
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

Figure 3. Spearman Correlation Coefficients for the top 10 variables in the dataset most correlated with GMP

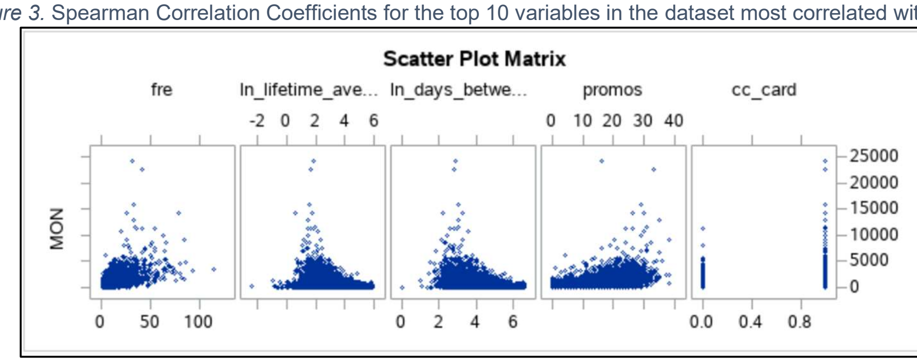

Figure 4. Scatterplot matrix of the top five most correlated variables with total net sales

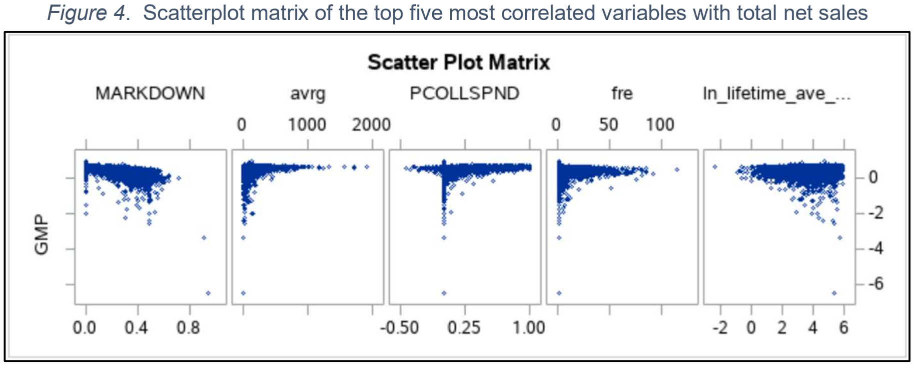

Figure 5. Scatterplot matrix of the top five most correlated variables with GMP

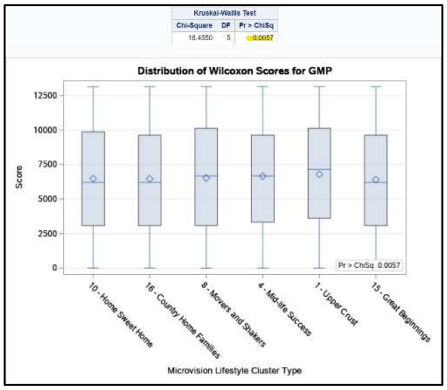

Figure 6. KW test results and boxplots showing GMP distribution by lifestyle cluster type

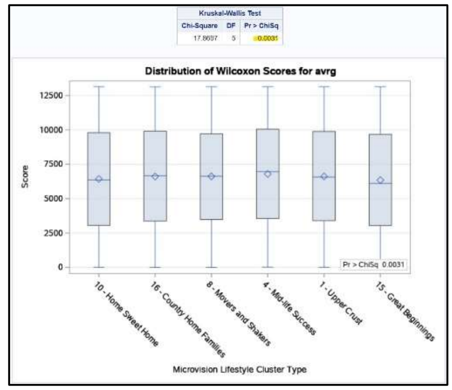

Figure 7. KW test results and boxplots showing AVRG distribution by lifestyle cluster type

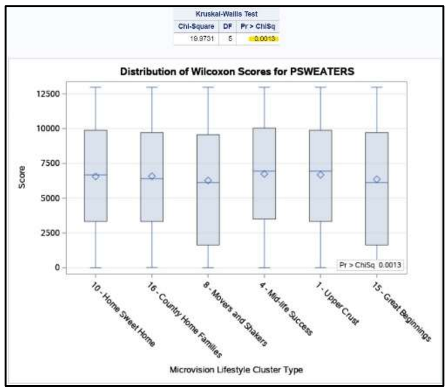

Figure 8. KW test results and boxplots for percent spent on sweaters by lifestyle cluster type

```
Group sample sizes not equal, or some ranks tied. Performed Dunn's test, alpha=0.05
Comparison group = clustype
Compare   Diff      SE        q          q(0.05)    Conclude
------------------------------------------------------------
1 vs  15  377.47    126.98    2.97       2.935      Reject
1 vs  16  324.51    113.51    2.86       2.935      Do not reject
1 vs  10  Do not reject (within non-sig. comparison)
1 vs  8   Do not reject (within non-sig. comparison)
1 vs  4   Do not reject (within non-sig. comparison)
4 vs  15  248.19    130.86    1.9        2.935      Do not reject
4 vs  16  Do not reject (within non-sig. comparison)
4 vs  10  Do not reject (within non-sig. comparison)
4 vs  8   Do not reject (within non-sig. comparison)
8 vs  15  Do not reject (within non-sig. comparison)
8 vs  16  Do not reject (within non-sig. comparison)
8 vs  10  Do not reject (within non-sig. comparison)
10 vs 15  Do not reject (within non-sig. comparison)
10 vs 16  Do not reject (within non-sig. comparison)
16 vs 15  Do not reject (within non-sig. comparison)
Note: "Do not reject (within non-sig. comparison)" indicates that any comparison
within the range of a non-significant comparison must also be non-significant.
Reference: Biostatistical Analysis, 4th Edition, J. Zar, 2010.
```

Figure 9. Dunn's test pairwise comparison for GMP distribution across lifestyle cluster types

```
Group sample sizes not equal, or some ranks tied. Performed Dunn's test, alpha=0.05
Comparison group = clustype
Compare   Diff      SE        q          q(0.05)    Conclude
------------------------------------------------------------
4 vs  15  444.73    130.91    3.4        2.935      Reject
4 vs  10  359.07    102.09    3.52       2.935      Reject
4 vs  16  181.93    117.89    1.54       2.935      Do not reject
4 vs  1   Do not reject (within non-sig. comparison)
4 vs  8   Do not reject (within non-sig. comparison)
8 vs  15  273.73    144.57    1.89       2.935      Do not reject
8 vs  10  Do not reject (within non-sig. comparison)
8 vs  16  Do not reject (within non-sig. comparison)
8 vs  1   Do not reject (within non-sig. comparison)
1 vs  15  Do not reject (within non-sig. comparison)
1 vs  10  Do not reject (within non-sig. comparison)
1 vs  16  Do not reject (within non-sig. comparison)
16 vs 15  Do not reject (within non-sig. comparison)
16 vs 10  Do not reject (within non-sig. comparison)
10 vs 15  Do not reject (within non-sig. comparison)
Note: "Do not reject (within non-sig. comparison)" indicates that any comparison
within the range of a non-significant comparison must also be non-significant.
Reference: Biostatistical Analysis, 4th Edition, J. Zar, 2010.
```

Figure 10. Dunn's test pairwise comparison for AVRG distribution across lifestyle cluster types

```
Group sample sizes not equal, or some ranks tied. Performed Dunn's test, alpha=0.05
Comparison group = clustype
Compare   Diff      SE        q          q(0.05)    Conclude
------------------------------------------------------------
4 vs  8   462       126.97    3.64       2.935      Reject
4 vs  15  379.98    129.97    2.92       2.935      Do not reject
4 vs  10  Do not reject (within non-sig. comparison)
4 vs  16  Do not reject (within non-sig. comparison)
4 vs  1   Do not reject (within non-sig. comparison)
1 vs  8   407.65    123.02    3.31       2.935      Reject
1 vs  15  Do not reject (within non-sig. comparison)
1 vs  10  Do not reject (within non-sig. comparison)
1 vs  16  Do not reject (within non-sig. comparison)
16 vs 8   303.69    131.93    2.3        2.935      Do not reject
16 vs 15  Do not reject (within non-sig. comparison)
16 vs 10  Do not reject (within non-sig. comparison)
10 vs 8   Do not reject (within non-sig. comparison)
10 vs 15  Do not reject (within non-sig. comparison)
15 vs 8   Do not reject (within non-sig. comparison)
Note: "Do not reject (within non-sig. comparison)" indicates that any comparison
within the range of a non-significant comparison must also be non-significant.
Reference: Biostatistical Analysis, 4th Edition, J. Zar, 2010.
```

Figure 11. Dunn's test pairwise comparison for percent spent on sweaters across lifestyle cluster types

Wilcoxon Two-Sample Test					
				t Approximation	
Statistic	Z	Pr < Z	Pr > |Z|	Pr < Z	Pr > |Z|
1.4545E8	-19.5415	<.0001	<.0001	<.0001	<.0001

Z includes a continuity correction of 0.5.

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
381.8720	1	<.0001

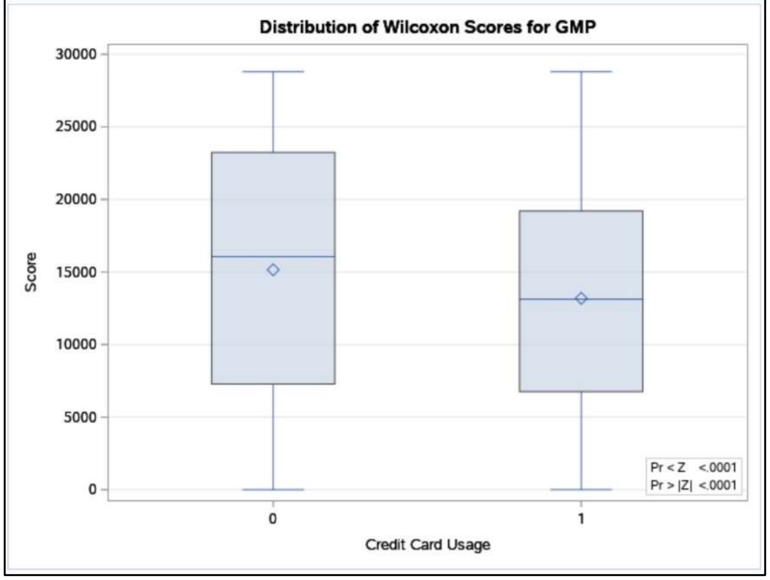

Figure 12. KW test and boxplots comparing GMP across credit card usage categories

```
Group sample sizes not equal, or some ranks tied. Performed Dunn's test, alpha=0.05
Comparison group = cc_card
Compare   Diff      SE        q         q(0.05)     Conclude
------------------------------------------------------------
0 vs  1   1968.58   100.74   19.54      1.960      Reject
Reference: Biostatistical Analysis, 4th Edition, J. Zar, 2010.
```

Figure 13. Dunn's Pairwise comparison test for GMP distribution by customer credit card usage

Product Categories That Customers Without Credit Cards Spend the Most On

Obs	Variable	Label	N	Mean
1	PSWEATERS	% Spent on Sweaters	17768	0.2264
2	PJACKETS	% Spent on Jackets	17768	0.1278
3	PBLOUSES	% Spent on Blouses	17768	0.0968
4	PCAR_PNTS	% Spent on Career Pants	17768	0.0865
5	PCOLLSPND	% Spent on Collectibles	17768	0.0776
6	PDRESSES	% Spent on Dresses	17768	0.0743
7	PCAS_PNTS	% Spent on Casual Pants	17768	0.0727
8	PSHIRTS	% Spent on Shirts	17768	0.0656
9	PKNIT_DRES	% Spent on Knit dresses	17768	0.0420
10	PSUITS	% Spent on Suits	17768	0.0345
11	PFASHION	% Spent on Fashionable Wear	17768	0.0297
12	PKNIT_TOPS	% Spent on Knit tops	17768	0.0277
13	POUTERWEAR	% Spent on Outerwear	17768	0.0193
14	PLEGWEAR	% Spent on Leg Wear	17768	0.0128
15	PJEWELRY	% Spent on Jewelry	17768	0.0101

Figure 14. The product categories that customers without credit cards spend the most on

Maximum R-Square Improvement: Step 2

Variable avrg Entered: R-Square = 0.6414 and C(p) = 3154.953

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	8029955065	4014977533	25755.8	<.0001
Error	28796	4488899495	155886		
Corrected Total	28798	12518854561			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-255.78780	4.53406	496126341	3182.62	<.0001
fre	80.87016	0.37250	7347273568	47132.3	<.0001
avrg	2.83034	0.02719	1689060779	10835.2	<.0001

Figure 15. MLR results for predicting total net sales from the frequency of purchase visits and the average amount spent per visit

Maximum R-Square Improvement: Step 2

Variable avrg Entered: R-Square = 0.7401 and C(p) = 499.4130

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	739591508	369795754	3863.78	<.0001
Error	2713	259656506	95708		
Corrected Total	2715	999248013			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-218.07848	11.72185	33127056	346.13	<.0001
fre	78.43632	0.91542	702656969	7341.65	<.0001
avrg	2.53933	0.07003	125824538	1314.67	<.0001

Figure 16. MLR results for predicting total net sales from the frequency of purchase visits and the average amount spent per visit for the "Upper Crust" lifestyle cluster type

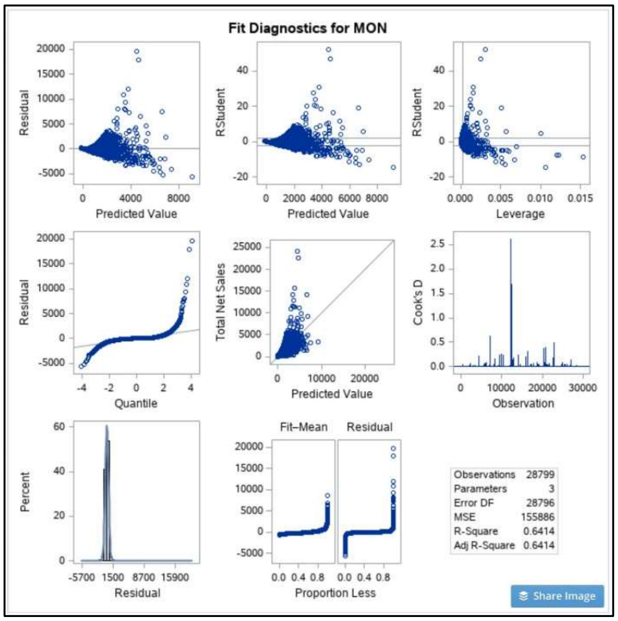

Figure 17. Fit Diagnostics for MLR model predicting total net sales from the frequency of purchase visits and the average amount spent per visit

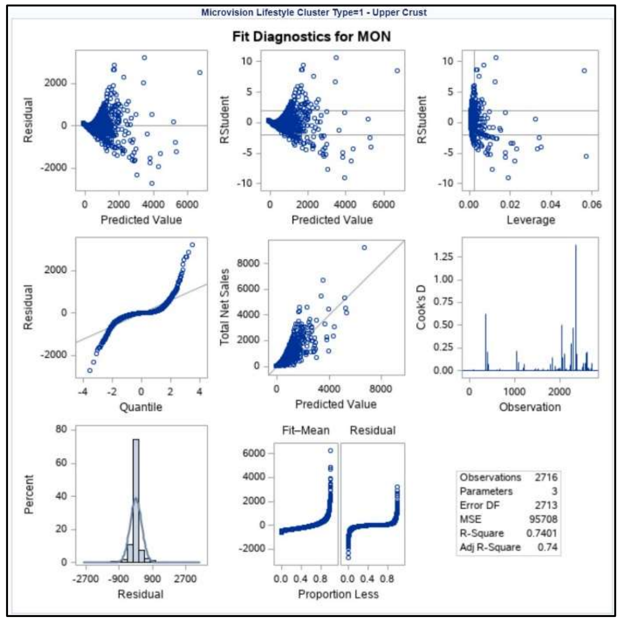

Figure 18. Fit Diagnostics for MLR model predicting total net sales from the frequency of purchase visits and the average amount spent per visit for the "Upper Crust" lifestyle cluster type

Credit Card Usage				
cc_card	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	17768	61.70	17768	61.70
1	11031	38.30	28799	100.00

Figure 19. Count of credit card users in the dataset

Microvision Lifestyle Cluster Type				
clustype	Frequency	Percent	Cumulative Frequency	Cumulative Percent
10 - Home Sweet Home	3488	12.11	3488	12.11
1 - Upper Crust	2716	9.43	6204	21.54
4 - Mid-life Success	2284	7.93	8488	29.47
16 - Country Home Families	1893	6.57	10381	36.05
8 - Movers and Shakers	1430	4.97	11811	41.01
15 - Great Beginnings	1327	4.61	13138	45.62

Figure 20. Six most common lifestyle cluster types in the dataset

References

Berger, D. E. (2003). Introduction to multiple regression. *USA: Claremont Graduate University*.

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods*, *17*(3), 399–417. https://doi.org/10.1037/a0028087

Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Applications*, *19*(4), 497–515.

Dinno, A. (2015). Nonparametric Pairwise Multiple Comparisons in Independent Groups using Dunn's Test. *The Stata Journal*, *15*(1), 292–300. https://doi.org/10.1177/1536867X1501500117

Elliott, A. C., & Woodward, W. A. (2016). *SAS Essentials: Mastering SAS for data analytics* (Second edition). John Wiley and Sons, Inc.

Feltovich, N. (2003). Nonparametric tests of differences in medians: Comparison of the Wilcoxon–Mann–Whitney and robust rank-order tests. *Experimental Economics*, *6*(3), 273–297.

Kasuya, E. (2019). On the use of r and r squared in correlation and regression. *Ecological Research*, *34*(1), 235–236. https://doi.org/10.1111/1440-1703.1011

Marr, B. (2015). *Big data: Using smart big data, analytics, and metrics to make better decisions and improve performance*. Wiley.

Ostertagová, E. (2012). Modeling using polynomial regression. *Procedia Engineering*, *48*, 500–506.

Puth, M.-T., Neuhäuser, M., & Ruxton, G. D. (2015). Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, *102*, 77–84.

Sable, J. (2019, March 12). *What Is the Profit Margin for Retail Clothes?* Small Business - Chron.Com. https://smallbusiness.chron.com/profit-margin-retail-clothes-25123.html