PENNSTATE

1855

08/04/2024

From: Suzanne Papik, *Graduate Student Consultant*

To: Management, *Alpine Pools and Spas*

Make a Splash: Analyzing Small Business Transaction Count Patterns to Improve Business Operations

EXECUTIVE SUMMARY

In response to a request from the management team at Alpine Pools and Spas, statistical consultants conducted a comprehensive analysis of three years of transaction and sales data to derive actionable insights for their business operations. The dataset, originating directly from the business's 'point of sale' software, provided rich information for investigating transaction patterns regarding various factors such as the day of the week, time of day, season, and weather conditions.

There were two primary objectives of this analysis. The first was to develop a model to obtain the expected number of transactions based on variables identified by the management team. The second objective was to explore suspicions that the team wanted to formally test, which were to explore seasonal variations in sales categories and investigate the impact of rainfall on transactions involving pool cleaning equipment. These questions were addressed through statistical methodologies including Poisson regression, chi-square tests of independence, and exploratory data analysis. The transaction dataset provided by the client was used in union with a weather data sourced from The National Centers for Environmental Information (NCEI).

The day of the week, time of day, summer season, and precipitation amounts were all found to be significant predictors of the expected transaction count, and the final model can be found on page 5 of this report, or in the footnote below. Additionally, a decision tree that can be used to model the expected transaction count can be found on page 6. The chi-square tests revealed significant differences in sales of product categories across different months, noting a general increase in sales of all product categories in the summer months. Pool chemicals were the most purchased product in the summer months, and winter & spa products were the most purchased products in the winter months. Additionally, Poisson regression analysis yielded the finding of a negative impact of rainfall on the count of transactions of pool cleaning equipment.

These insights will be useful for the management team at Alpine Pools and Spas to enhance inventory management strategies, optimize operational efficiency, and align staffing resources with fluctuating transaction volumes influenced by seasonal and weather patterns.

$$\textit{Expected Transaction Count} = 1.416 + (0.189 * \textit{Mon}) + (0.06825 * \textit{Tues}) + (0.21922 * \textit{Wed}) + (0.16865 * \textit{Thurs}) + (-0.051 * \textit{Sat}) + (-0.179 * \textit{Sun}) + (0.28350 * \textit{Morning}) + (0.44665 * \textit{SummerMonth}) + (-0.26346 * \textit{PRCP}) + (-0.36191 * \textit{Snowfall})$$

1.0 - PROJECT DESCRIPTION

The Management of Alpine Pools and Spas engaged the author of this report to assist with drawing insights from three years' worth (1/1/2020- 12/30/2023) of transaction and sales data from their small business. The data that they are interested in analyzing comes directly from their 'point of sale' software and includes information about sales transactions and service calls. Since this dataset comes directly from the software without intervention or treatment, this is considered an observational analysis.

The first question this report aims to address is a request for a model that the client can used to predict transaction counts based on factors such as the day of the week, time of day, seasonality, and weather conditions. The second and third questions that this report aims to address are based on suspicions that the client wants to formally test, which explore seasonal variations in sales categories as well as assess the impact of rainfall on transactions of pool cleaning equipment.

The findings from this report will inform strategies for inventory management, operational efficiency, staffing considerations, and the ability to better align with expected transaction volume based on weather influences.

1.1 - RESEARCH QUESTIONS

The following research questions will be addressed in this report:

1. Given the factors that the client suspects affect the number of sales in a given day, are we able to determine the expected number of transactions that will occur? The specific factors the clients are interested in are: the day of the week, the time of day, if it is during the summer season or not, the amount of rain for the day, if it snowed that day, and the average temperature of the day.
2. Are transaction numbers of different products consistent throughout the year, or is there variation of the category of the sale based on the month?
3. Is there an increase in the number of transactions of pool cleaning equipment (pool maintenance, cleaning equipment, and balancing chemicals) on days that it rains?

1.2 – STATISTICAL QUESTIONS

1. Can we model the count of expected transactions using the day of the week, the time of day, if it is the summer season or not, the amount of rain, if it snowed that day, and what the average temperature is? Can this information be conveyed in both an equation and tree format?
    a. This question will be assessed using Poisson regression and decision tree modeling.

2. Are there statistically significant differences in the number of transactions across different categories of products and months of the year?
    a. This question will be assessed using chi-square test of independence and exploratory data analysis.
3. Is there an increase in the number of transactions of pool cleaning equipment (pool maintenance, cleaning equipment, and balancing chemicals) on days that it rains?
    a. This question will be assessed using Poisson regression.

1.3 – VARIABLES

The main dataset utilized for this report was provided from the client and included all transaction information for the business from January 1st, 2020 through December 30th, 2023. This data was exported directly from the software that is used to facilitate sales and scheduling of service. This dataset included eighteen different variables detailing information about the product, the customer, the salesperson, and the price of the item. For purposes of this study, only four variables pertaining to the transaction and product type were initially retained from this dataset, which were *ID*, *Category*, *Date*, and *Quantity*. Six variables were created from these retained variables either as intermediate variables or to assist with addressing the research questions, which were *SalesDate*, *Time*, *Morning*, *SummerMonth*, *DayOfWeek*, *Month*, and *CategoryGroup*. The *CategoryGroup* variable was created to combine similar groups from the *Category* variable to reduce the number of unique categories for the analysis (see Table A.1.2 in Appendix A). The categories were grouped based on feedback from the client. The variables from the transaction dataset that are in the final dataset and utilized in the analysis can be seen in more detail in Table A.1.1 in Appendix A.

Weather and precipitation information for the area that the business is located was obtained in a separate dataset from The National Centers for Environmental Information (NCEI)[1]. This dataset included twenty-three different variables detailing information on the precipitation, the snowfall, and the temperature of the local area. For purposes of this study, only four variables were retained from this dataset, which were the *DATE, PRCP*, *SNOW*, and *TAVG*. These retained variables were used to create two additional variables used to assist with addressing the research questions, which were *Rain* and *Snowfall*. The variables from the weather dataset that are in the final dataset utilized in the analysis can be seen in more detail in Table A.1.1 in Appendix A.

The transaction dataset was transformed so that every observation represents the count of transactions of a specific category of product for a given day, split by morning and afternoon sales. This count is

[1] Dataset Documentation: https://www.ncei.noaa.gov/data/daily-summaries/doc/GHCND_documentation.pdf

represented with the *TransactionCount* variable. The weather information was appended to each observation based on a matching date so that the weather information for that day was available for analysis. To address the research questions, the *TransactionCount* variable was treated as the response variable and all other variables were treated as explanatory variables.

2.0 - EXPLORATORY DATA ANALYSIS (EDA)

In total, there are 47,187 unique transactions in the dataset. The final dataset consists of 8,609 observations, where each observation contains information on the count of transactions for products within a given category in the morning or afternoon of a given day. There are no missing values in the dataset, and all 8,609 observations have a value for each of the variables.

The *TransactionCount* variable represents the count of transactions for a given product category per time period (Morning/Afternoon), therefore it is not surprising that this variable exhibits a right-skewed distribution which can be seen in Figure 1. In the span of 1/1/2020 through 12/30/2023, there was an average of 5.48 transactions per product category and time of day (Morning/Afternoon). With a standard deviation of 11.339, this transaction count varies quite considerably across days and times of the day.

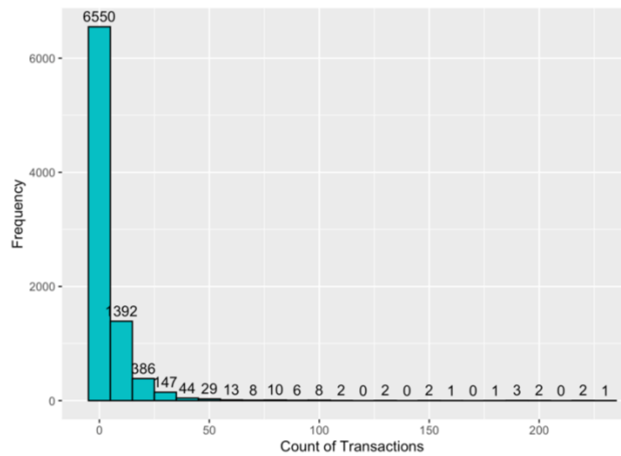

Figure 1: Distribution of TransactionCount

The *PRCP* and *SNOW* variables also exhibited right-skewness to their distributions. Both variables peak at a frequency of 0, and as the amount of rain or snow increases, the frequency sharply declines. This aligns with the climate expected in western PA. On average, the Pittsburgh area experiences rain or snow 140 days a year[2], therefore, most of the days do not experience precipitation which leads to a peak at 0 for both variables. The *TAVG* (average temperature) variable follows a slightly left-skewed distribution, which is also expected for this area and data. Temperature in Pittsburgh often levels around 55-75 degrees F, however it can get colder in the winter months leading to the left skew in the distribution. The distributions and summary statistics of the four continuous variables, *TransactionCount*, *PRCP*, *SNOW*, *TAVG*, can be seen in Table A.2.1 and Figures A.2.8-A.2.11 in Appendix A.

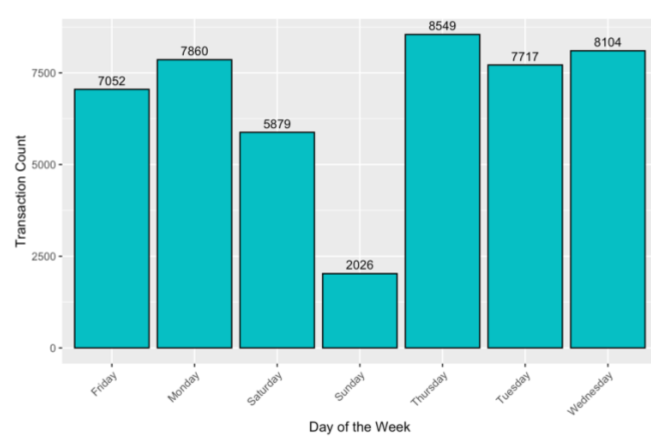

Figure 2: Total Transaction Count by DayOfWeek

[2] https://www.bestplaces.net/climate/city/pennsylvania/pittsburgh

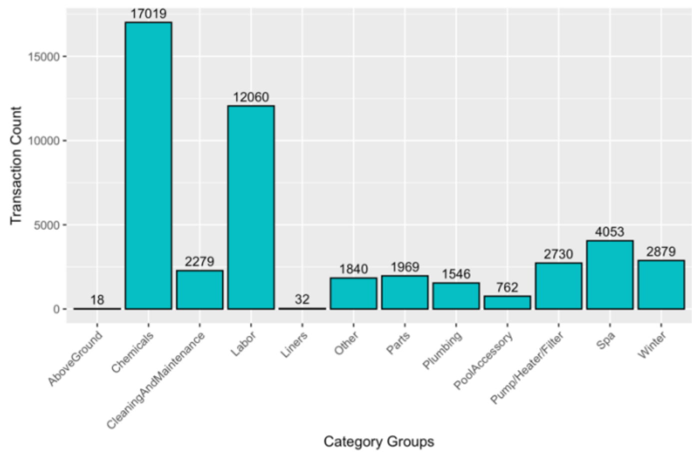

Figure 3: Total Transaction Count by CategoryGroup

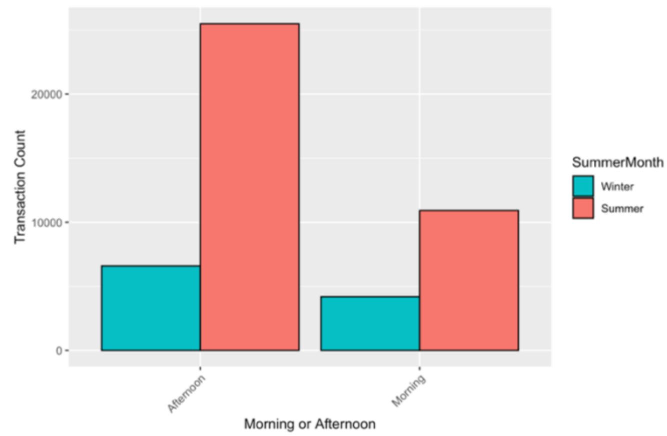

Figure 4: Frequency of Transactions by Morning/Afternoon and Summer/Winter Month

As seen in Figure 2, the total count of transactions appears to be similar for each day of the week, with a slight decrease in transactions on Saturday, and a larger decrease on Sunday. Thursdays had the largest number of unique transactions with 8,549 transactions, and Sundays had the fewest with 2,026. As seen in Figure 3, the categories of products that see the most transactions are the chemicals and labor categories. Both aboveground pool products and liner products have a significantly smaller total transaction count when compared to the other categories. One last observation can be seen in Figure 4, where it can be observed that total count of transactions was largest in the afternoon during the summer months. Relatedly, the total count of transactions was smallest in the morning during the winter. Note that Figures 2 through 4 utilize the total transaction number for the given category, not the frequency of the *TransactionCount* variable in the dataset (which is different compared to Figure 1).

More information about the breakdown, distribution, and summary statistics of these variables can be found in Appendix A.2. The distribution of the categorical variables based on the observation frequency in the dataset can be seen in Tables A.2.2-A.2.6, while the distribution of the categorical variables based on the count of unique transactions can be seen in Tables A.2.12-A.2.16 in Appendix A.

3.0 –STATISTICAL ANALYSIS

3.1- Predicting the Number of Expected Transactions

Poisson Regression was used to produce a model that can be used to predict the expected transactions for a given day based on the following factors: day of the week, time of day, if it is during the summer season or not, the amount of rain, if it snowed that day, and the average temperature of the day. Fitting a Poisson model using *TransactionCount* as the dependent variable led to the conclusion that all these factors were significant in predicting the transaction count. A summary of this model can be found in Table A.3.1 in Appendix A.3.

However, overdispersion was found to be significant (Chi-Square: 8.6, p-value: < 0.001). To adjust for overdispersion, the model was re-fit to utilize a quasi-poisson distribution (See Table A.3.2 for a summary of this model). In this model, the average temperature variable (*TAVG*) was not a significant predictor of the transaction count and was therefore removed. Additionally, the values Saturday, Sunday, and Tuesday for the *DayOfWeek* variable were found to not be significant. The *DayOfWeek* variable was retained since the other levels were significant. Model assumptions were verified in Table A.3.4 in Appendix A.3, and a summary of the final model output can be found in Table A.3.3 in Appendix A.3.

The final model that can be used to predict the expected number of transactions is:

Expected Transaction Count=1.416+(0.189*_Mon_)+(0.06825*_Tues_)+(0.21922*_Wed_)+ (0.16865*_Thurs_)+(-0.051 *_Sat_)+(-0.179*_Sun_)+(-0.28350*_Morning_)+(0.44665*_SummerMonth_)+ (-0.26346*_PRCP_)+(-0.36191*_Snowfall_)

The following table can be used in tandem with the equation above to predict the estimated transaction count based on the factors in the model:

Variable	Value in the equation	Variable	Value in the equation
Mon	1 if Monday, 0 otherwise	**Morning**	1 if morning (before noon), 0 if afternoon
Tues	1 if Tuesday, 0 otherwise	**SummerMonth**	1 if summer months, 0 if winter months
Wed	1 if Wednesday, 0 otherwise	**PRCP**	The amount of rainfall for the day (in mm)
Thur	1 if Thursday, 0 otherwise	**Snowfall**	1 if it snows that day, 0 if there is no snow
Fri	Mon=Tues=Wed=Thurs=Sat=Sun=0		
Sat	1 if Saturday, 0 otherwise		
Sun	1 if Sunday, 0 otherwise		

Table 5: Variable Values to Input in the Predictive Model Equation

From the final model we can conclude that, on average, the expected transaction counts are smaller in the morning time period compared to the afternoon, the expected transaction counts are smaller in the winter months compared to the summer, and the expected transaction counts decrease if it snows compared to if it does not. Additionally, with an increase in the amount of rain received in a day, the expected transaction count decreases. Finally, estimated transaction counts generally are higher on weekdays as opposed to weekends. Based on the coefficients in the model above, we could expect the highest number of expected transactions on Wednesday in the afternoon during the summer when there is not any precipitation or snowfall.

Another way to visualize and determine the predicted expected transaction count is with a decision tree, which can be found on the following page in Figure 6. The expected number of transactions can be found in the final blue circle at the bottom of the tree. Each branch taken is determined by the values of the variable factors that the user is interested in. For example, if the client were interested in the expected transaction count for afternoon hours during the summer on a Wednesday, and we didn't expect any precipitation, we would follow the branches until we arrived at 8. Therefore, for this situation we would expect, on average, 8 transactions.

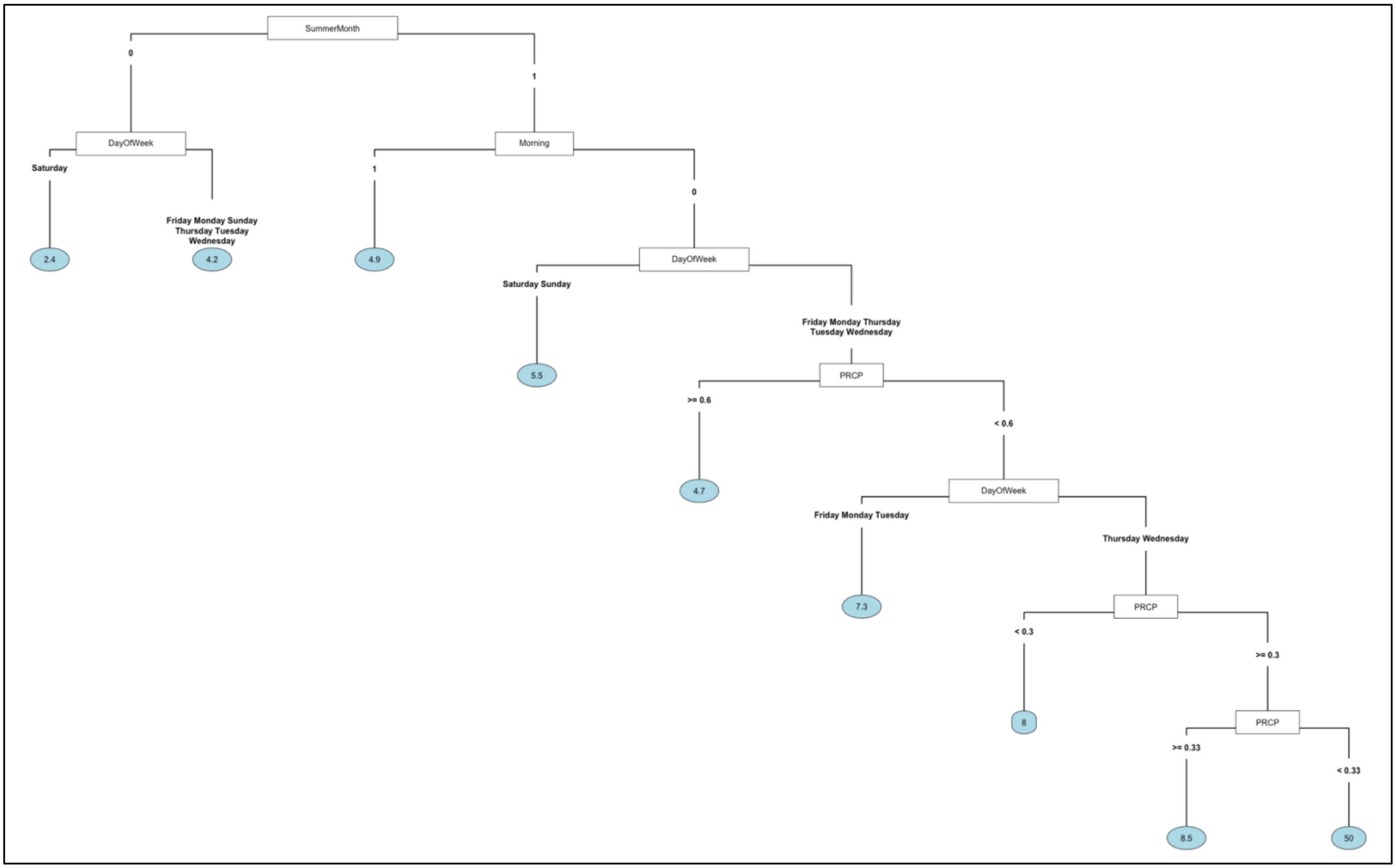

Figure 6: Decision Tree to Determine Expected Transaction Count

3.2- Analyzing Transactions or Product Categories Through the Year

Chi-square tests of independence were conducted to assess if transaction numbers are consistent throughout the year or if there is variation of the category of the sale based on the month. To ensure the expected count of each cell was greater than 5, the 'liner' and 'labor' categories were combined into the 'labor' category, and the 'aboveground' and 'other' categories were combined into the 'other' category. This was acceptable for the client since liner work can be considered outdoor labor sales, and aboveground items are no longer sold so they can be treated in the other category. A summary of the *CategoryGroup* variable after this change can be seen in A.3.5 in Appendix A.3

The Chi-square test yielded the result that there is a statistical difference in the products sold across the twelve months of the year (Chi-square: 1696.1, df: 99, p-value: < 0.001). This indicates that the distribution of the category of products transaction volume changes depending on what month it is (See A.3.6 in Appendix A to review model assumptions).

The next logical question is what trends do we see in the data when it comes to transaction counts within product categories across the twelve months? The distribution of sales for a given month can be seen in Figure 7 below. From this figure we can see that pool chemicals are consistently the most bought product in April through September. Winter products sales start to increase in August and significantly increase in September and stay elevated until November. Generally, the number of transactions across all categories, outside of winter products, peaks in the summer months, decreases through the fall into December, and is at the lowest through the first three months of the year.

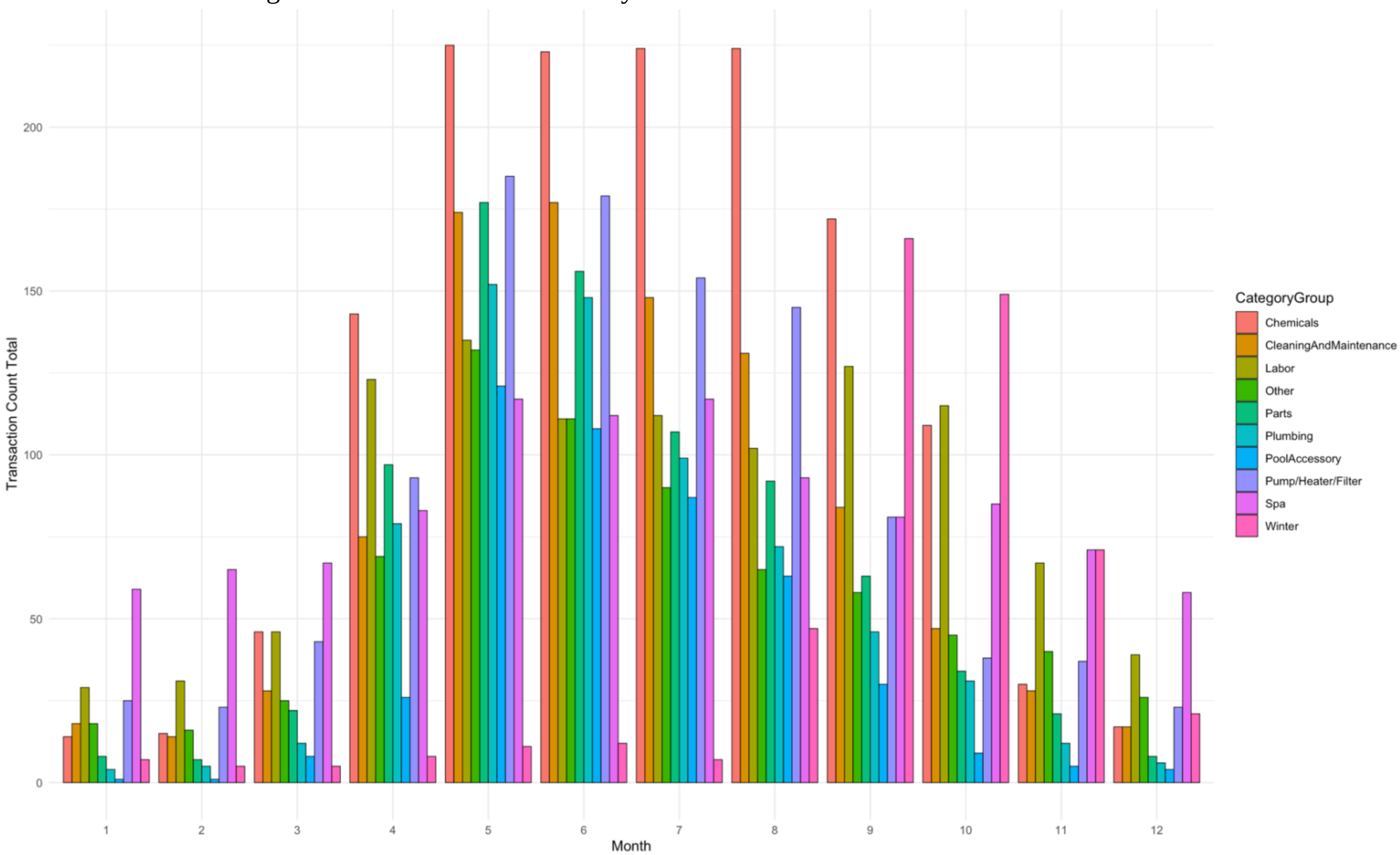

Figure 7: Transaction Count by Month and Category Group

3.3- Analyzing Precipitation and the Number of Expected Transactions

Poisson regression can be used to address the last research question which asks if there an increase in the number of transactions of pool cleaning equipment (pool maintenance, cleaning equipment, and balancing chemicals) on days that it rains. The dataset was transformed where each observation was the count of transactions on a given day for only pool cleaning equipment products. The response of the model was again *TransactionCount*, and the explanatory variable was the *Rain* variable, which was found to be significant at a 0.1 level of significance. A summary of the model can be found in Table A.3.7 in Appendix A.

However, overdispersion was found to be significant (Chi-Square:12.33, p-value: < 0.001). To adjust for overdispersion, the model was re-fit to utilize a quasi-poisson distribution. In this final model, which can be seen in Table A.3.8 in Appendix A, the *Rain* variable was found to be a significant predictor of the count of transactions, with an estimated coefficient value of -0.16443. This indicates that if it were to rain on a given day, the expected number of transactions of cleaning equipment is smaller when compared to days where it doesn't rain. Model assumptions were verified in Table A.3.9 in Appendix A.3

4.0 – RECOMMENDATIONS

1. The first research question asked if we able to predict the number of transactions that will occur based on the day of the week, the time of day, if it is during the summer season or not, the amount of rain for the day, if it snowed that day, and the average daily temperature. The final model that can be used for prediction is:

   Expected Transaction Count=1.416+ (0.189**Mon*)+ (0.06825**Tues*)+ (0.21922**Wed*)+ (0.16865**Thurs*)+ (-0.051 **Sat*)+ (-0.179**Sun*)+ (0.28350**Morning*)+ (0.44665**SummerMonth*)+ (-0.26346**PRCP*)+ (-0.36191**Snowfall*)

   A decision tree that can be used in a similar fashion to this equation can be seen in Figure 6 on Page 6 of this report.
2. The second research question addressed if transaction numbers are consistent throughout the year or is there variation of the category of the sale based on the month. Based on the results of the Chi-Square test of independence, there is significant evidence that transactions are not consistent throughout the year and the distribution of sales for a given category depends on the month.
3. The final research question looked at if there is an increase in the number of transactions of pool cleaning equipment (pool maintenance, cleaning equipment, and balancing chemicals) on days that it rains. Based on the results of the Poisson regression analysis, there is significant evidence that the expected number of transactions of cleaning equipment decreases on days that it rains (as compared to days where it does not rain).

5.0 – RESOURCES

The software used for this analysis was R (http://www.r-project.org/) and Minitab (www.minitab.com/en-us/)

6.0 – CONSIDERATIONS

There are several items with this analysis that should be considered. First, the primary dataset used in this analysis originated from a 'point of sale' software tailored for swimming pool businesses. This software primarily captures sales transactions and service calls, limiting the number of variables available for use in this analysis. Consequently, there may be confounding factors that are not accounted for in the analysis. Additionally, this software and output relies on the data being correctly recorded in the system. Data quality may be compromised due to human error, therefore extensive validation of the dataset was needed.

Another consideration is that this dataset includes transaction information from the peak of the coronavirus pandemic in 2020. Transaction patterns were different from March 2020 through roughly April 2021 compared to what they have been in prior years or in the period since the height of the pandemic. The unique circumstances during the pandemic may have influenced sales distributions and patterns, therefore extending the findings in this analysis to the current economic period must be done with care.

One final consideration is that correlation does not equal causation. Results outlined in this analysis may be due to confounding variables not accounted for in the analysis. For example, the analysis for the third research question found that the expected number of transactions of cleaning equipment decreases on days that it rains. However, there may be the confounding factor that all sales decrease on days that it rains, not just necessarily sales of cleaning equipment. Further analysis is needed.

Finally, R, and Minitab were used to conduct the analysis, produce the figures, and deliver the recommendations. A level of significance of 0.1 was used unless specified otherwise.

It has been a pleasure to work on this analysis and report. Please feel free to reach out with any further questions and I would be happy to address them!

Appendix A.1- Additional Figures and Tables: Variables

Variable	Source	Description	Valid Values
SaleDate	Transaction dataset	Date of the Transactions	01/01/2020 – 12/30/2023
DayOfWeek	Created from Date (transaction dataset)	The day of the week that the transactions occurred on	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday
Morning	Created from Time (transaction dataset)	A binary variable to specify if the transactions occurred in the morning or in the afternoon	1: Morning Sale (before 12pm) 0: Afternoon Sale (12pm and after)
SummerMonth	Created from Date (transaction dataset)	A binary variable to specify if the transactions occurred during peak season (the summer) or during the off season	1: Transactions were during summer months (May through September) 0: Transactions were during winter months (October through April)
CategoryGroup	Created from Category (transaction dataset)	The category of product that was sold	AboveGround, Chemicals, CleaningAndMaintenance, Labor, Liners, Other, Parts, Plumbing, PoolAccessory, Pump/Heater/Filter, Spa, Winter
TransactionCount	Created from ID (transaction dataset)	The count of transactions for the given combination of the other variables.	This can take any countable positive number, including 0 The range of values seen in this dataset is 1 to 232
PRCP	Weather Dataset	Daily precipitation (mm)	This can take any positive number, including 0 The range of values in this dataset is 0 to 2.85
SNOW	Weather Dataset	Daily Snowfall (mm)	This can take any positive number, including 0. The range of values in this dataset is 0 to 8
TAVG	Weather Dataset	The average daily temperature (degrees Fahrenheit)	This can take any value from $(-\infty, \infty)$, however the range of values in this dataset is [14, 83]
Rain	Created from PRCP (Weather Dataset)	A binary variable to specify if it rained during the day	1: There was rain on that day (>0 mm) 0: There was no rain on that day (0 mm)
Snowfall	Created from SNOW (Weather Dataset)	A binary variable to specify if it snowed during the day	1: There was snowfall on that day (>0 mm) 0: There was no snowfall on that day (0 mm)

Table A.1.1 Final Analysis Variables and Their Descriptions

CategoryGroup Variable	Category Variable	CategoryGroup Variable	Category Variable
AboveGround	Abg Pool Parts 13	**Plumbing**	Coping/Joints/Forms
	Abg Pools 14		Hoses/Pipes/Connectors
	Radiant Pools (31)		Plumbing/Glues/Primers
	X Abg Pool Parts		Seal Sets/Bearings
Chemicals	Chlorinators/Brominators	**PoolAccessory**	Diving Boards/Parts
	Package Chemicals		Floating Lounge Chairs
	Pool Chemicals (35)		Furniture/Accessories
CleaningAndMaintenance	Aquabot Parts (91)		Grills & Accessories
	Auto Pool Cleaners		Handrails
	Brushes (06)		Ladders(Inpool/A-Frame) Drop In Steps (23)
	Kreepy Krauly Parts (71)		Lights/Color Lens
	Lab Reagents/Supplies		Masks/Goggles/Snorkels
	Leaf Nets/ Leaf Baggers (24)		Polaris Mini Jets And Water Falls
	Maintenance Kits		Rope/Floats/Attachments
	Patch Kits/Dye Testers		Skimmers/Main Drains
	Test Kits/Reagents/Strips		Slides/Parts
	Vacuum Heads/Parts		Solar Controls
	Vacuum Poles/Parts		Thermometers
Labor	Labor		Toys
Liner	Liners	**Pump/Heater/Filter**	Cartridges Replacement (07)
Other	Close Outs		Filter Media (Sand De)/Pool Base
	Misc		Filters/Hose Kits/Bases (12)
	X Not Used		Heaters/Heat Pumps (18)
Parts	Comfortzone Parts		Pumps
	Fiber Works Parts	**Spa**	Artesian Spa Parts (82)
	Goldline Parts (99)		Artic Spa Parts
	Hayward Parts (65)		Four Winds/Mira Spa Parts (78)
	Ig Pool Parts		Misc Spa Part & Spa Plus
	King Technology Parts (80)		Spa Accessories
	Minimax Heater Parts		Spa Chemicals/Fragrances
	Nature 2 Parts		Spa Covers
	Olympic Parts		Spa Parts Plus
	Ozone/Salt Sys/Ionozer & Parts		Spas-Artesian
	Pentair Parts (67)		Spas-Artic
	Polaris Parts (69)		Spas-Mira Spas
	Premier Parts		Vita Spa Parts
	Raypac Parts	**Winter**	Winter
	Safety Cover Parts (81)		
	Solar Covers/Parts		
	Teledyne Laars Parts		
	Waterway Parts (89)		
	Watkins Parts		

Table A.1.2 The Values of CategoryGroup based on the Category Variable

Appendix A.2- Additional Figures and Tables: EDA

Variable	N	Mean	Std Dev	Min	Max
TransactionCount	8,609	5.48112	11.3390	1	232
PRCP	8,609	0.0955117	0.241038	0	2.85
SNOW	8,609	0.0234870	0.209507	0	8
TAVG	8,609	61.5581	13.7654	14	83

Table A.2.1 Summary Statistics Quantitative Variables

DayOfWeek	N	Percentage of Total
Monday	1,334	15.5%
Tuesday	1,454	16.9%
Wednesday	1,302	15.1%
Thursday	1,469	17.1%
Friday	1,418	16.5%
Saturday	1,229	14.3%
Sunday	403	4.7%

Table A.2.2 Summary of the DayOfWeek Variable

CategoryGroup	N	Percentage of Total
AboveGround	14	0.16%
Chemicals	1,442	16.75%
CleaningAndMaintenance	941	10.93%
Labor	1,035	12.02%
Liners	22	0.26%
Other	691	8.03%
Parts	792	9.2%
Plumbing	666	7.74%
PoolAccessory	463	5.38%
Pump/Heater/Filter	1,026	11.92%
Spa	1,008	11.71%
Winter	509	5.91%

Table A.2.2 Summary of the CategoryGroup Variable

Morning	N	Percentage of Total
1	3,294	38.3%
0	5,315	61.7%

Table A.2.3 Summary of the Morning Variable

Rain	N	Percentage of Total
1	3,330	38.7%
0	5,279	61.3%

Table A.2.4 Summary of the Rain Variable

SummerMonth	N	Percentage of Total
1	5,876	68.3%
0	2,733	31.7%

Table A.2.5 Summary of the SummerMonth Variable

Snowfall	N	Percentage of Total
1	274	3.2%
0	8,335	96.8%

Table A.2.6 Summary of the Snowfall Variable

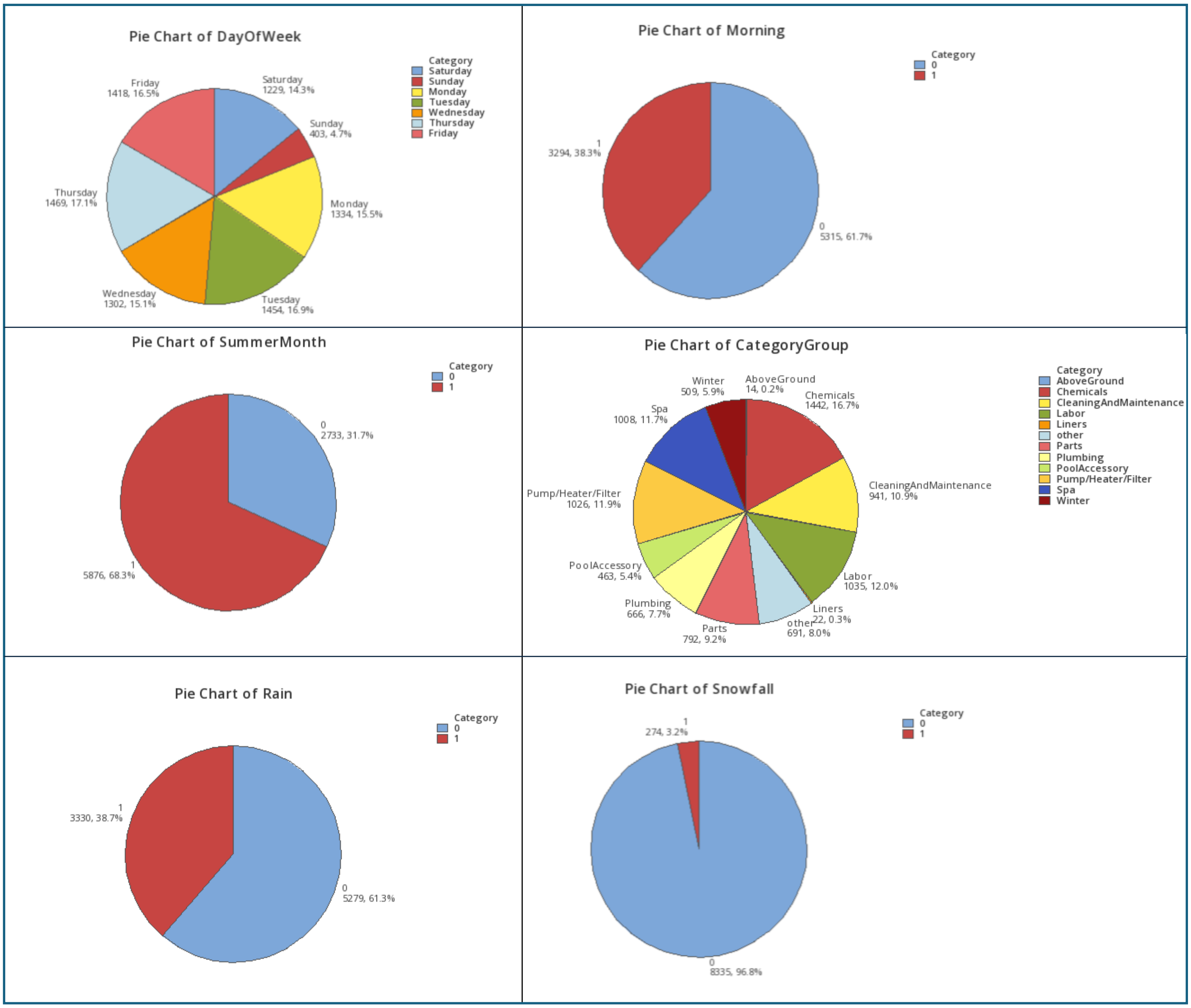

Figure A.2.7 Pie Charts of Categorical Variables- Distribution in Final Dataset

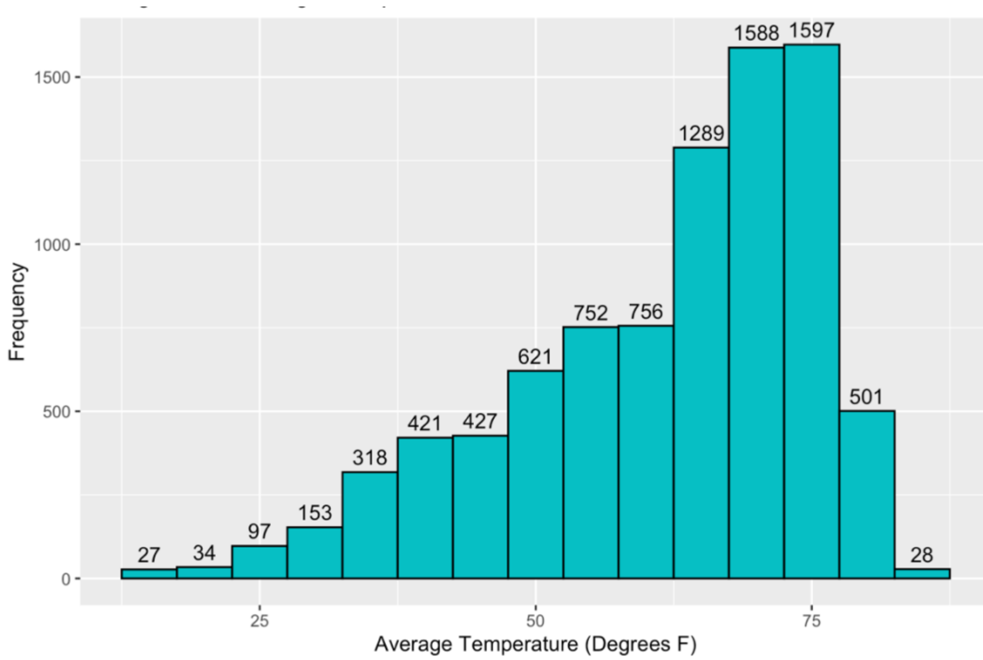

Figure A.2.8 Distribution of the TAVG Variable

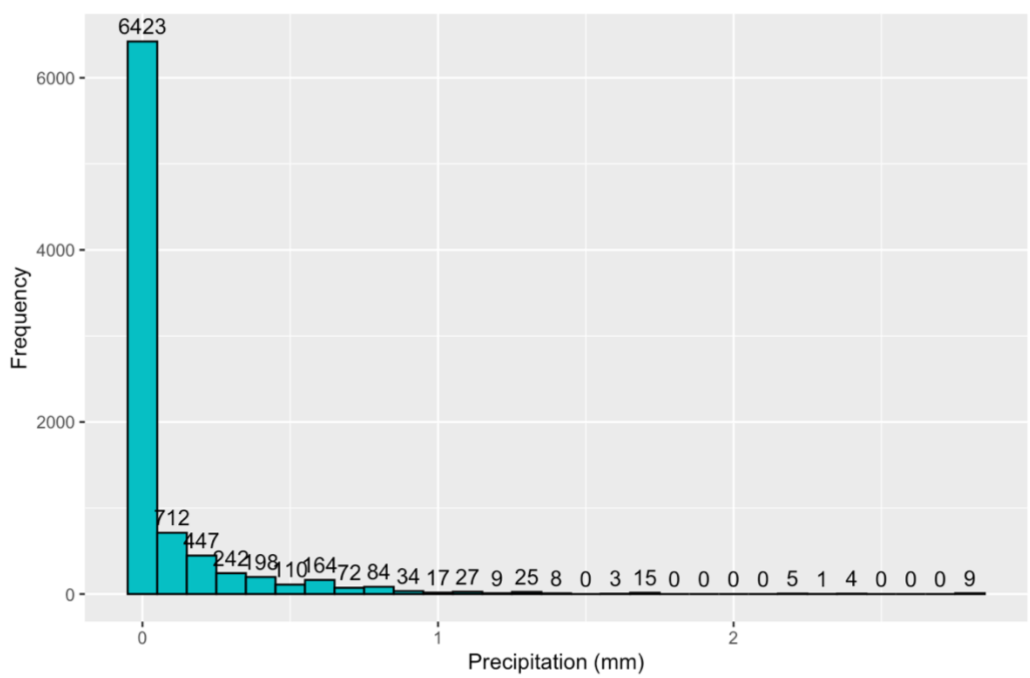

Figure A.2.9 Distribution of the PRCP Variable

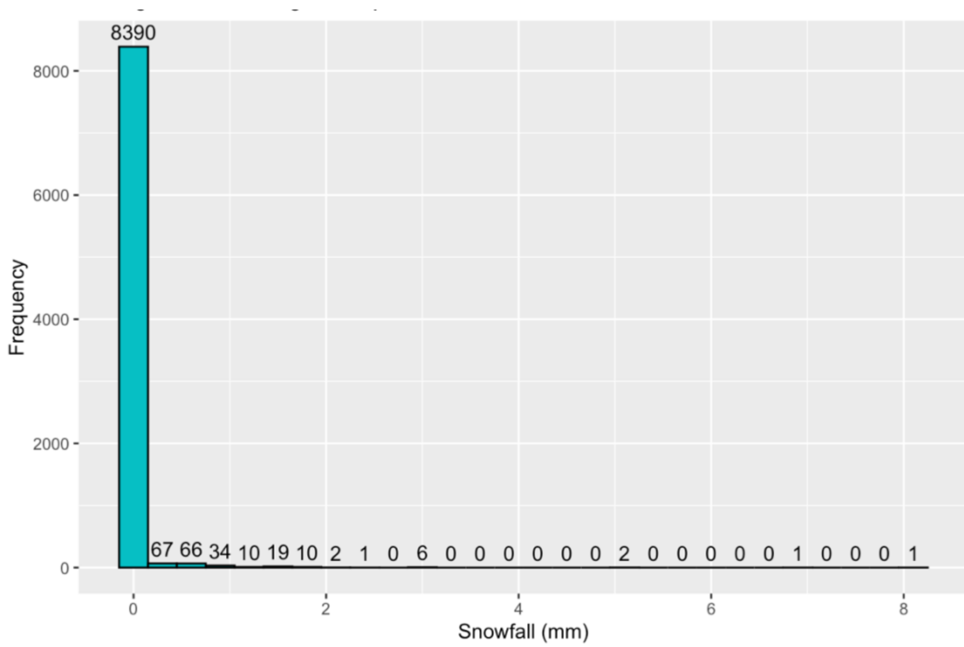

Figure A.2.10 Distribution of the SNOW Variable

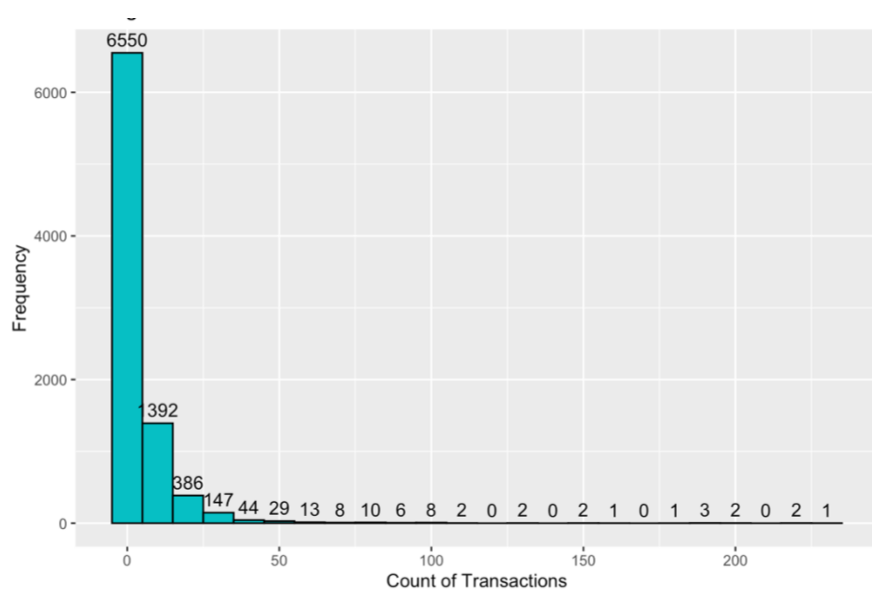

Figure A.2.11 Distribution of the TransactionCount Variable

DayOfWeek	Total Transaction Count	Percentage of Total Transaction Count
Monday	7860	16.66%
Tuesday	7717	16.35%
Wednesday	8104	17.17%
Thursday	8549	18.11%
Friday	7052	14.94%
Saturday	5879	12.46%
Sunday	2026	4.29%

Table A.2.12 Total Number of Transactions by DayOfWeek Variable

CategoryGroup	Total Transaction Count	Percentage of Total Transaction Count
AboveGround	18	0.038%
Chemicals	17019	36.07%
CleaningAndMaintenance	2279	4.83%
Labor	12060	25.56%
Liners	32	0.068%
Other	1840	3.9%
Parts	1969	4.17%
Plumbing	1546	3.28%
PoolAccessory	762	1.61%
Pump/Heater/Filter	2730	5.79%
Spa	4053	8.59%
Winter	2879	6.1%

Table A.2.14 Total Number of Transactions by CategoryGroup Variable

Morning	Total Transaction Count	Percentage of Total Transaction Count
1	15111	32.02%
0	32076	67.98%

Table A.2.15 Total Number of Transactions by Morning Variable

SummerMonth	Total Transaction Count	Percentage of Total Transaction Count
1	36404	77.15%
0	10783	22.85%

Table A.2.16 Total Number of Transactions by SummerMonth Variable

Appendix A.3- Additional Figures and Tables: Analysis

	Estimate	Std. Error	Zvalue	Pr(>|z|)
(Intercept)	1.2502926	0.0298865	41.835	< 0.001
DayOfWeekMonday	0.1912208	0.0164158	11.649	< 0.001
DayOfWeekSaturday	-0.0465528	0.0177068	-2.629	0.00856
DayOfWeekSunday	-0.1815242	0.0254185	-7.141	< 0.001
DayOfWeekThursday	0.1691730	0.0161011	10.507	< 0.001
DayOfWeekTuesday	0.0696562	0.0164809	4.226	< 0.001
DayOfWeekWednesday	0.2199487	0.0162924	13.500	< 0.001
Morning	-0.2847146	0.0099414	-28.639	< 0.001
SummerMonth	0.3818731	0.0151151	25.264	< 0.001
PRCP	-0.2731587	0.0217930	-12.534	< 0.001
Snowfall	-0.3021967	0.0383162	-7.887	< 0.001
TAVG	0.0033657	0.0005201	6.471	< 0.001

Table A.3.1: Summary of the Initial Poisson Model for Research Question 1

	Estimate	Std. Error	tvalue	Pr(>|z|)
(Intercept)	1.250293	0.134896	9.269	< 0.001
DayOfWeekMonday	0.191221	0.074094	2.581	0.00987
DayOfWeekSaturday	-0.046553	0.079922	-0.582	0.56026
DayOfWeekSunday	-0.181524	0.114729	-1.582	0.11364
DayOfWeekThursday	0.169173	0.072674	2.328	0.01994
DayOfWeekTuesday	0.069656	0.074388	0.936	0.34910
DayOfWeekWednesday	0.219949	0.073538	2.991	0.00279
Morning	-0.284715	0.044872	-6.345	< 0.001
SummerMonth	0.381873	0.068223	5.597	< 0.001
PRCP	-0.273159	0.098365	-2.777	0.00550
Snowfall	-0.302197	0.172944	-1.747	0.08061
TAVG	0.003366	0.002347	1.434	0.15168

Table A.3.2 Summary of the Intermediate Quasi-Poisson Model for Research Question 1

	Estimate	Std. Error	tvalue	Pr(>|z|)
(Intercept)	1.41565	0.06934	20.417	< 0.001
DayOfWeekMonday	0.18856	0.0742	2.541	0.01107
DayOfWeekSaturday	-0.05142	0.07999	-0.643	0.52035
DayOfWeekSunday	-0.17881	0.11491	-1.556	0.11974
DayOfWeekThursday	0.16865	0.07279	2.317	0.02054
DayOfWeekTuesday	0.06825	0.07451	0.916	0.35973
DayOfWeekWednesday	0.21922	0.07366	2.976	0.00293
Morning	-0.2835	0.04494	-6.308	< 0.001
SummerMonth	0.44665	0.05141	8.689	< 0.001
PRCP	-0.26346	0.09823	-2.682	0.00733
Snowfall	-0.36191	0.16807	-2.153	0.03131

Table A.3.3 Summary of the Final Quasi-Poisson Model for Research Question 1

Assumption	Notes
The dependent variable consists of count data	The dependent variable is the count of sales transactions.
Response variable Assumptions	$E(Y) = \mu$, $Var(Y) = \phi\mu$, Where μ is the expected mean transaction count, and ϕ is the dispersion factor determined from the sample data.

Table A.3.4: Model Assumptions: Quasi-Poisson Research Question 1

Chi-Sq Test Statistic	Df	p-value
1696.1	99	< 0.001

Table A.3.4 Summary Chi-Square Test for Research Question 2

CategoryGroup	N	Percentage of Total
Chemicals	1,442	16.75%
CleaningAndMaintenance	941	10.93%
Labor	1057	12.28%
Other	705	8.19%
Parts	792	9.20%
Plumbing	666	7.74%
PoolAccessory	463	5.38%
Pump/Heater/Filter	1,026	11.92%
Spa	1,008	11.71%
Winter	509	5.91%

Table A.3.5 Summary of CategoryGrouping after Combining AboveGround and Liners Categories

Assumption	Notes
Sample is randomly drawn from the population	The sample is three continuous years of sales data from a population of all years the business was open. However, the sample size is large.
Expected values >5	Groups were combined to ensure the minimum expected count is greater than 5. The minimum expected count is 9.788
Mutually Exclusive Groups	All groups with observed counts are mutually exclusive. A given transaction can only occur in one product category in one month.

Table A.3.6: Model Assumptions: Chi-Square Test Research Question 2

	Estimate	Std. Error	zvalue	Pr(>|z|)
(Intercept)	2.151084	0.008879	242.27	< 0.001
Rain	-0.164434	0.015167	-10.84	< 0.001

Table A.3.7: Summary of the Initial Poisson Model for Research Question 3

	Estimate	Std. Error	tvalue	Pr(>|z|)
(Intercept)	2.15108	0.04812	44.706	< 0.001
Rain	0.16443	0.08220	-2.001	0.0456

Table A.3.8: Summary of the Final Quasi Poisson Model for Research Question 3

Assumption	Notes
The dependent variable consists of count data	The dependent variable is the count of sales transactions.
Response variable Assumptions	$E(Y) = \mu$, $Var(Y) = \phi\mu$, Where μ is the expected mean transaction count, and ϕ is the dispersion factor determined from the sample data.

Table A.3.9: Model Assumptions: Quasi-Poisson Research Question 3