**GROUP 5 - FINAL SUBMISSION**

****

Tony Bejos

Matthew Flynn

Shivani Gupta

Robert Reed

Tanya Mendes

**TABLE OF CONTENTS**

[Introduction 3](#_Toc39084446)

[Overview of Analysis 3](#_Toc39084447)

[Data Collection 4](#_Toc39084448)

[Data Cleaning 5](#_Toc39084449)

[Dataset Characteristics 5](#_Toc39084450)

[Modeling 6](#_Toc39084451)

[Model Findings/Discoveries 6](#_Toc39084452)

[Industry and Market Analysis 7](#_Toc39084453)

[Sentiment Analysis 8](#_Toc39084454)

[Sales Analysis 9](#_Toc39084455)

[Recommendations 9](#_Toc39084456)

[Conclusion 10](#_Toc39084457)

[Appendix 11](#_Toc39084458)

[Exhibit 1 12](#_Toc39084459)

[Exhibit 2 13](#_Toc39084460)

[Exhibit 3 14](#_Toc39084461)

[Exhibit 4 15](#_Toc39084462)

[Exhibit 5 16](#_Toc39084463)

[Exhibit 6 17](#_Toc39084464)

[Exhibit 7 18](#_Toc39084465)

[Exhibit 8 19](#_Toc39084466)

[Exhibit 9 20](#_Toc39084467)

[Exhibit 10 21](#_Toc39084468)

[Works Cited 22](#_Toc39084469)

# Introduction

Emerald Green Inc. is a lawn care and landscaping service that has been operating in the Piedmont Triad in North Carolina since 1999. It is a family-run business, with a large proportion of their customers being long-time loyal customers. The company provides a variety of services such as lawn care, landscaping, and tree and shrub control; and customers have the option to purchase a single service, or sign up for a subscription, with the latter being more lucrative. Emerald Green Inc. has been growing in sales and customer numbers year-over-year, but after getting to know more about the business and operations, our group decided that we could help the company by providing data-driven recommendations for their upcoming marketing campaign and outlining realistic key performance indicators to measure success. We were able to analyze Emerald Green Inc.’s sales and customer data and supplement the analysis with external sources of data, such as census data and results of a customer survey to arrive at our recommendations.

# Overview of Analysis

We ran a broad range of analyses on the data extracted from the CRM tool to attempt to understand consumer demand for the various products and services offered by Emerald Green Inc. A sales analysis was performed to review areas of sales volume and growth across time and across different product and service offerings. Growth was measured on a year-over-year and quarter-over-quarter basis to analyze seasonal and annual trends. A churn analysis was conducted on the subscription services to understand the extent of customer turnover. These services were analyzed because this segment has recurring customers. A marketing tool analysis was conducted to examine the most effective way of promoting sales. Identifying the most demanded services explains what the customer wants, but it does not explain the most effective way for Emerald Green Inc. to reach them. This analysis attempts to resolve this issue and sharpen the focus of the marketing campaign by identifying the most effective form of promotion. A predictive model was created, refined, and shocked to reveal historical trends and extrapolate them. Finally, the data attributes were organized and presented in interactive Tableau visualizations to reveal insights that may not have been immediately obvious when in the original chart and table format.

# Data Collection

Data collection for this project only took up a small portion of the total time spent due to the data being readily available. We had three main datasets that we analyzed. The first and main source of data was Emerald Green Inc.’s proprietary data. Secondly, we pulled relevant zip code demographics from the United States census. Finally, we conducted a customer survey in order to run a sentiment analysis.

Emerald Green’s proprietary data is housed in a CRM tool called Service Assistant (Emerald Green's Service Assistant, 2020). This tool provided us with structured data exports in CSV format. We utilized the Census data portal (Explore Census Data, 2020) to create CSV files with relevant financial and demographic data grouped by zip code. This tool is extremely user friendly and allowed us to enter the specific zip codes serviced by Emerald Green Inc. in order to group the data. The survey was conducted on the popular website Survey Monkey. We created an eight-question survey and sent it to all 400 customer emails that were provided to us by Emerald Green Inc. 43 customers responded, and their answers were used to conduct our sentiment analysis.

# Data Cleaning

Emerald Green Inc. has manually entered all the data in Service Assistant over a 10 to 20-year period, so there were a few typographical errors in their data that needed cleaning. We were aiming to help Emerald Green Inc. boost their residential sales only, so in order to eliminate commercial entities, we removed all customer names that did not begin with the prefix *Mr.* or *Mrs.* per direction from Emerald Green Inc. We also deleted all rows that did not have a zip code in the address field since these would not be joinable to the Census data.

# Dataset Characteristics

We housed more than 110,000 rows of data in ten tables and created seven relevant views using those tables. An entire database diagram can be seen in [*Exhibit 1*](#_Data_Cleaning). Some of the most relevant tables were *cancel\_report*, *sources* (new customers and their source of acquisition), *allsalesdata2004* (every sale since 2004), *census\_finance* (income and rent)*,* and *census\_demographics* (age, race, and education). Indexes were created on all tables to order by their primary key; customer number. Those five main tables allowed us to create views that summarized quarterly and yearly results in cancels, new customers, and total sales. All were grouped by zip code and quarterly vintage (i.e. *Q32010*) in order to join Emerald Green’s data to the zip code demographics we pulled from the Census. To transform this clean data, we aggregated the job count, customer lawn size, and total sales data by zip code and vintage in order to create a linear regression model that aimed to forecast sales quarter over quarter.

# Modeling

We decided that a multiple linear regression model would be the best way to forecast sales over a four-quarter period given the linear relationship between sales and time. After creating the previously described views, we loaded our quarterly sales view into SAS Enterprise Guide to create regression analytics. The total sales field was used as the dependent variable and the independent variables were 25 fields containing zip code demographics from the Census and customer demographics from Emerald Green’s tables. After obtaining a model with a desired significance, we plotted (in Excel) the past eight years to judge model accuracy. We then extrapolated the following four quarters in 2020 to forecast sales over the next year. Lastly, we shocked our model’s job count variable during our prescriptive analysis in order to measure the potential impact of our marketing recommendation. [*Exhibit 3*](#_Exhibit_3) outlines the shocks applied and results.

# Model Findings/Discoveries

[*Exhibit 2*](#_Exhibit_2) outlines the linear regression model obtained in SAS. Our R-squared came to 0.8771 and the p-stat was highly significant at less than 0.0001. The parameter estimates generated a linear regression formula of:

y-hat = 50.663 + 50.867(job count) + 38.096(avg lawn size) + 0.0175(median income) – 1.253(median rent)

All variables were found to be significant with p-stats of less than 0.05. This was crucial because we discovered that median income and rent per zip code could be used to forecast sales and would serve as valuable data points for our prescriptive analytics.

[*Exhibit 4*](#_Exhibit_4) shows our forward looking forecast, while [*Exhibit 5*](#_Exhibit_5) is our accuracy test. We found that our model was 99.986% accurate over the periods dating 2011–2018 (the only years available for Census data) and never less than 80% accurate over any given year in that same range. This accuracy level allowed us to be confident that our model would accurately forecast Emerald Green’s sales over a four-quarter period.

# Industry and Market Analysis

As diverse and dynamic as the lawn care service industry is, there were some interesting and common traits this market shares with other customer service businesses. For instance, word of mouth (e.g. neighbor referral) proved to be the highest source of customer acquisition across our analysis. This marketing tip was closely followed by discount coupons and seasonal promotions, which always work for a medium-scale enterprise in any customer service industry. Emerald Green’s most popular services include the basic lawn care and partial lawn care programs, which according to Emerald Green Inc. are popular offerings among competitors as well. We were also informed that contrary to popular belief, the lawn care industry is trending positively in the current COVID-19 pandemic with more people spending time at home in their backyard.

Target area demography plays a big role in lawn care business as its customers typically share the traits of higher age group, income, and education. Analyzing areas with the highest Emerald Green Inc. consumer base, the following social pattern clearly stood out. More than 80% of residents belong to the $75,000 to $150,000 income group and are associated with a college degree. Additionally, these target consumers are aged either under 35, or 65+ ([*Exhibit 8*](#_Overview_of_Analysis)). The previously mentioned social statistics of age, education and income have grown positively over the years in the zip codes found in [*Exhibit 8*](#_Exhibit_8). By industry standards, Emerald Green Inc. is in the right areas to aggressively grow its business, and its loyal client base provides a competitive advantage.

Key competitors identified in Emerald Green’s service area included franchises like Lawn Doctor, Weed Doctor of Greensboro, Lush Lawn, Growing Green, and True green of Greensboro. Unfortunately, specific pricing data for competitors was unavailable for analysis due to being obtainable only through estimates and consultations.

# Sentiment Analysis

Customer sentiment at Emerald Green Inc. is generally positive, as showcased by the company’s first customer from over 20 years ago *still existing as an active customer* through the current year*.* However, to gain further and broader insight into customer sentiment, our team employed two primary tools: a churn analysis and a customer survey. The results and details of both are described below.

The churn analysis was conducted on two service offerings: the basic lawn care program and the partial lawn care program. The results of the churn analysis revealed that the churn rate was generally between 0% and 6% annually for the majority of the past ten years, with the exception being the most recent two years (2018 and 2019), which had *negative* churn, indicating strong customer growth ([*Exhibit 10*](#_Overview_of_Analysis)).

The survey results revealed generally positive customer sentiment. One key survey takeaway was that 81% of responders indicated that they were either “Likely” or “Very Likely” to recommend the business to others. The survey response indicates an Emerald Green Inc. Net Promoter Score of 72 ([*Exhibit 7*](#_Exhibit_7)).

# Sales Analysis

We analyzed the company’s customer sales data in the file allsalesdata2004. This data set went back for over sixteen years and contained information on over 107,000 transactions. While looking into this we were able to gain some insights on the productivity of the company. The first thing we were able to determine was the top performing products for the company. These products included the aeration and seeding programs, the tree and shrub program, and most notably the basic lawn care program, which brings in a significant majority of sales most quarters.

We were also able to determine the amount of business Emerald Green Inc. was generating in each zip code where the company operated. From this, we were able to determine the zip codes that were performing the best, which was later used to fuel our predictive models as well as our prescriptive analytics. The health of the business in each territory can be seen in our geographic model, which provides visual representation of sales in each region ([*Exhibit 9*](#_Data_Cleaning)).

# Recommendations

Our group used a combination of descriptive and predictive methodologies to arrive at focus ideas (such as target areas, products, and referral sources) around which we based our recommendations. A descriptive analysis of customer referral sources ([*Exhibit 8*)](#_Data_Cleaning) and lawn sizes yielded the recommendation that the coupon packet should be further enhanced to include a neighbor referral bonus to increase neighbor referral numbers. The service emphasis should be on lucrative subscription services—basic and partial lawn care services. In addition, the coupon packet should also include discount coupons and spring season incentives in the immediate seasonal campaign. The coupon packet strategy would be most effective when disseminated in appropriate locations. According to our predictive model ([*Exhibit 4*](#_Exhibit_4)), we recommended an existing-customer focus in high-income zip codes that have large customer numbers (27358, 27310), whereas new customers should be the focus in high-income zip codes where there is scope to increase market share (27235, 27012, 27278, 27023). Further, our recommendations were accompanied by assumptions and a risk analysis to highlight factors that would affect successful implementation.

# Conclusion

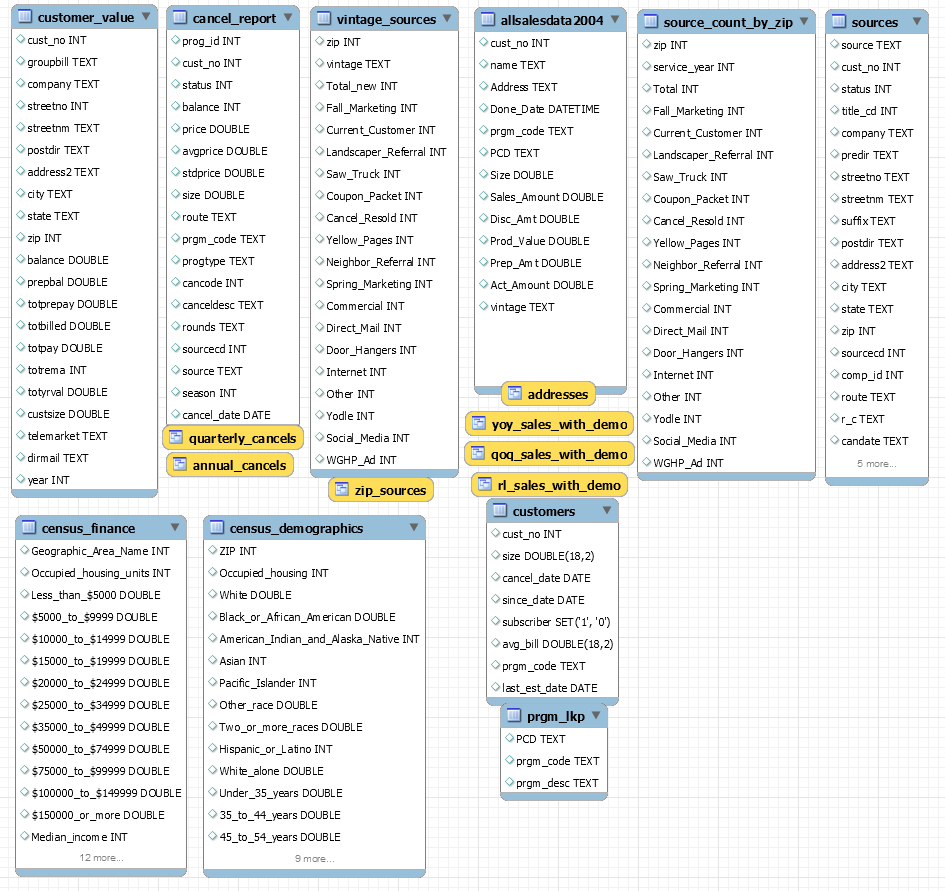
Throughout the project, our focus was to work with the problem statement in mind—how to grow the business through the marketing campaign. One of the strong points in the project was the extent and quality of Emerald Green Inc.’s data, which allowed us to flexibly apply descriptive, predictive, and prescriptive analytics to arrive at recommendations that we are confident of and able to justify to the business owner. We were able to analyze the target markets, product focus and how to market these products in the target markets. Further, we believe that the key performance indicators identified—lead and customer conversion rates per campaign, Q2 2020 sales growth, and year-over-year growth in neighbor referrals--would help Emerald Green Inc. measure the success of our recommendations.

# 

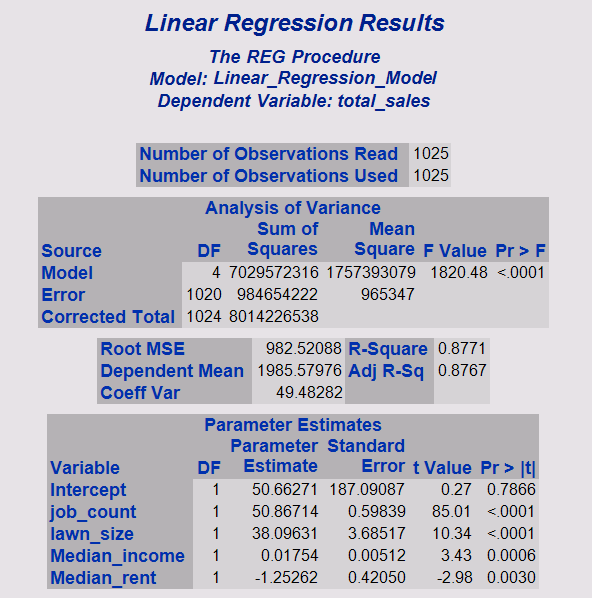
# 

# Appendix

# Exhibit 1



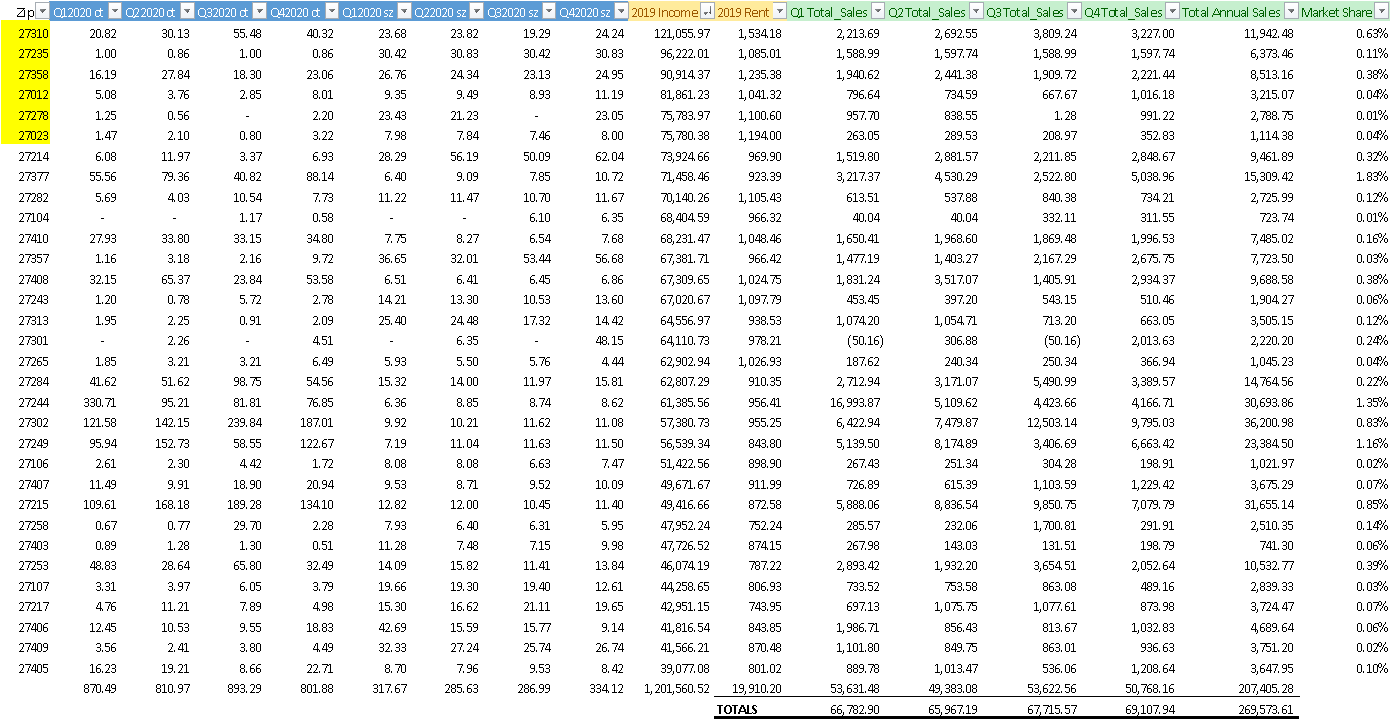
# Exhibit 2



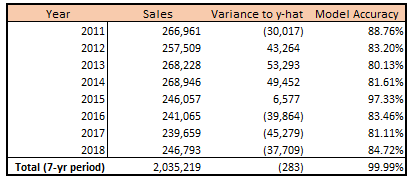
# Exhibit 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Performance** |  | **Shock** | **Q2 Sales Growth** | **Annual Sales Growth** |
| **Poor** |  | 0.5% | 3.76% | 0.92% |
| **Low** |  | 1% | 7.52% | 1.84% |
| **Below Average** |  | 1.5% | 11.28% | 2.76% |
| **Average** |  | 2% | 15.04% | 3.68% |
| **Above Average** |  | 2.5% | 18.80% | 4.60% |
| **High** |  | 3% | 22.55% | 5.52% |
| **Exceptional** |  | 3.5% | 26.31% | 6.44% |

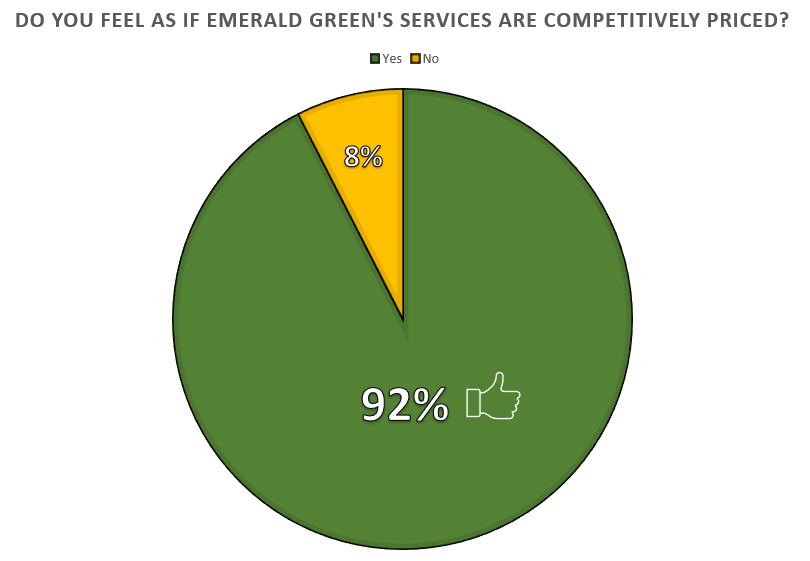
# Exhibit 4



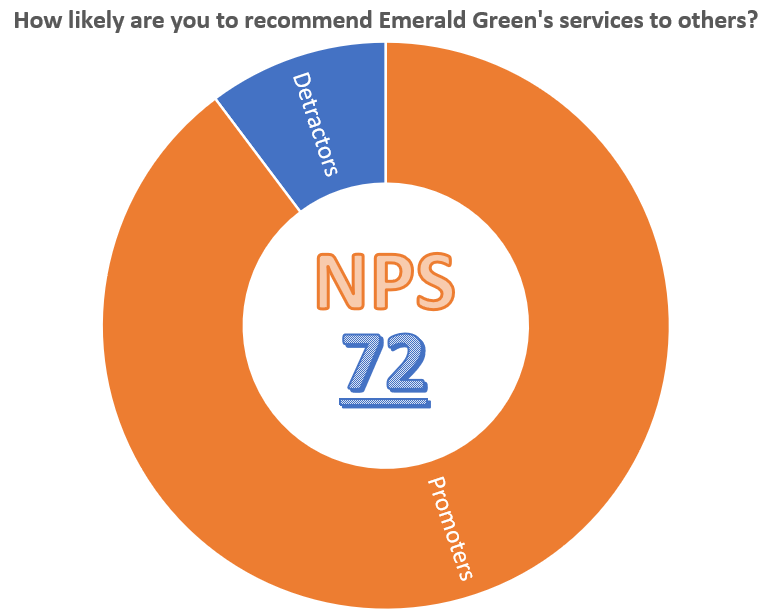
# Exhibit 5



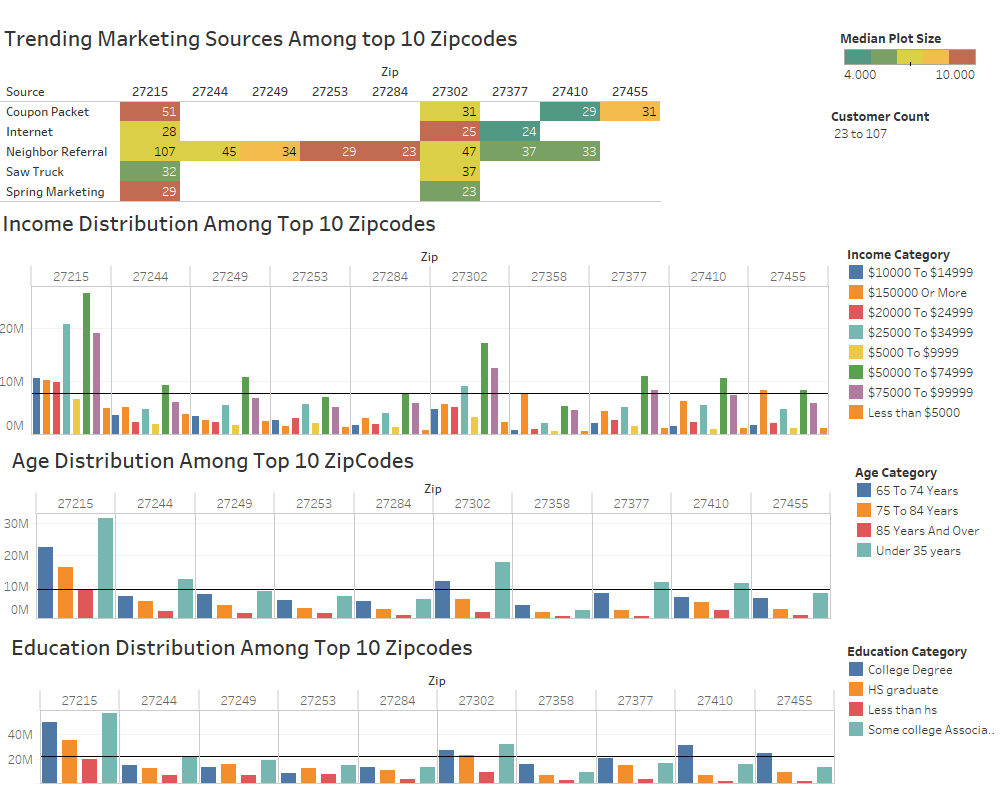
# Exhibit 6



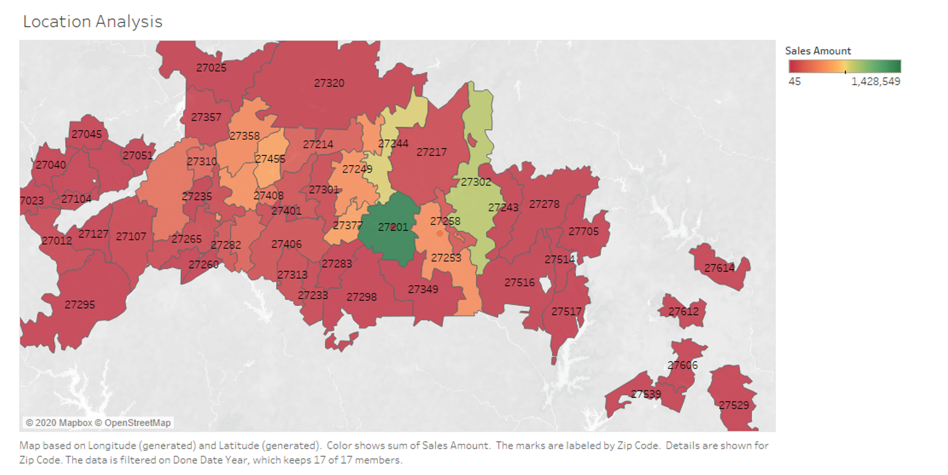
# Exhibit 7



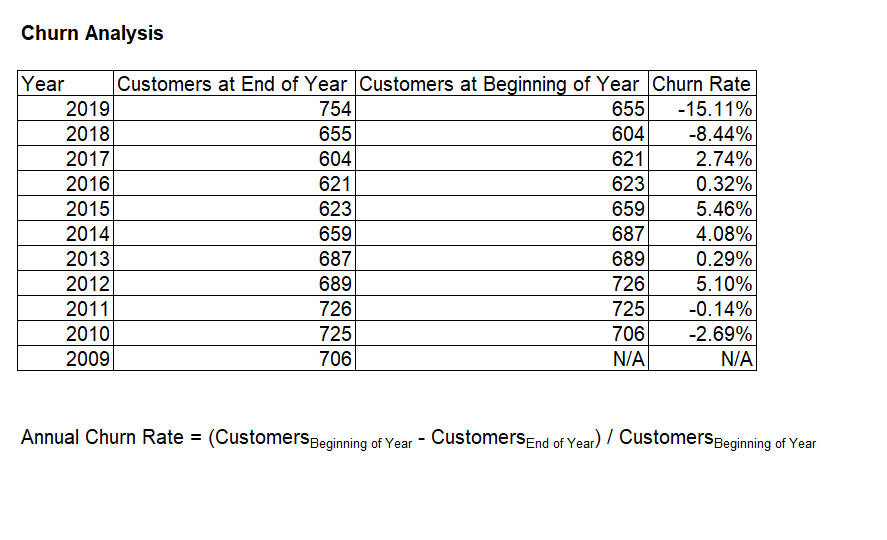
# Exhibit 8



# Exhibit 9



# Exhibit 10



# Works Cited

*Emerald Green's Service Assistant*. (2020, 03). Retrieved from Service Assistant: https://leia.serviceassistant.com/145911

*Explore Census Data*. (2020, 03). Retrieved from United States Census Bureau: https://data.census.gov/cedsci/