Statistical Computing and Data Visualization

Final Project Report

—

You Luen (Alex) Chang

Weikang (Iris) Cui

Wei-Wei Lee

Lily Zhang

Executive Summary

Atri Industries, a mechanical parts manufacturer, is looking to determine their marketing budget allocation for an upcoming sales period. Specifically, their marketing managers are looking to evaluate the company's current customers to determine whether more marketing spend should be alloted for acquiring new customers or maintaining relationships with existing customers. Our team analyzed a select list of order information and determined existing customers from the Defense sector appear to be the most valuable clients for the company, and recommend for the marketing team to allocate proper funding for maintaining those client relationships.

Introduction

For our project we will be using R to analyze the historical sales data of a fictional company called Atri Industries, a mechanical parts manufacturer. We will be using data visualization techniques learned throughout the DSO-545 course to manipulate and evaluate the dataset.

We intend to segment Atri Industries' current customers by characteristics such as: customer type (new vs retained customer), business size, geographical location, business sector, and acquisition channel. We will then evaluate the performance of each customer segment and use this information to form recommendations for the marketing department in how to allocate marketing budget with the goal of maximizing ROI and total revenue.

Data Description

The dataset ("data.csv") we will be working with contains many different variables, not all of which will be used for our purposes. The following is a description of the variables:

Variable	Description
Customer.ID	Unique identifying number for each customer (both active and inactive)
Period.Active	How many periods the customer stays with the company
Current.Customer	Whether the customer is an active customer
Period.Acquired	The period during which a customer first made a purchase from Atri Industries
Period.Terminated	The period during which a customer first made their final purchase

Acquisition.Channel	Method through which a customer was acquired
Sector	Industry the customer's company belongs to
Premium	Whether the customer took the special premium offers in period 8
State	State the customer's company is located in
Region	Region of the USA the customer's company is located in
Co.Bio	Age and size of the customer's company
No..Maker.Contracts	Number of contracts signed for "Maker Products" in period 8
No..Advanced.Contracts	Number of contracts signed for "Advanced Products" in period 8
Maker.Revenue	Revenue from maker models from customer in period 8
Advanced.Revenue	Revenue from advanced models from customer in period 8
Expected.CLV	Calculated CLV from historical data

Data Cleaning

The dataset we are working with contains improperly formatted data, so before performing any analysis we cleaned up the data through the following ways:

1. Separate

    a. Separated company year and size data from "Co.Bio" into two seperate new variables called "Age" and "Size".

    b. "No..Maker.Contracts" and "No..Advanced.Contracts" both contain numerical information we want mixed with a letter. The letter have no significance to us so we removed it by separating the letters and numbers within those two variables into different new variables: "Maker_No." and "Adv_No." and then saved this as "data_clean".

2. Add

    a. "Maker.Revenue" and "Advanced.Revenue" both contain information that counts towards the total revenue generated by the company. So we saved the two into a new variable called "Total.Revenue".

    b. We are currently consider to be in period 9, so any customer acquired during period 8 need to be classified as new customers. To do this, we mutated "Period.Acquired" and save those with 8 to "New.customer".

3. Change

"State" is changed to all lowercase so it will be compatible with a mapping function we will be using as a part of the analysis.

4. Filter

Some of the companies listed in the data are no longer active customers of Atri Industries for various reasons. For the purpose of this analysis, we want to only evaluate current customer data. To do so, we filtered "data_clean" for active current customers and saved it as "final_data".

Data Analysis

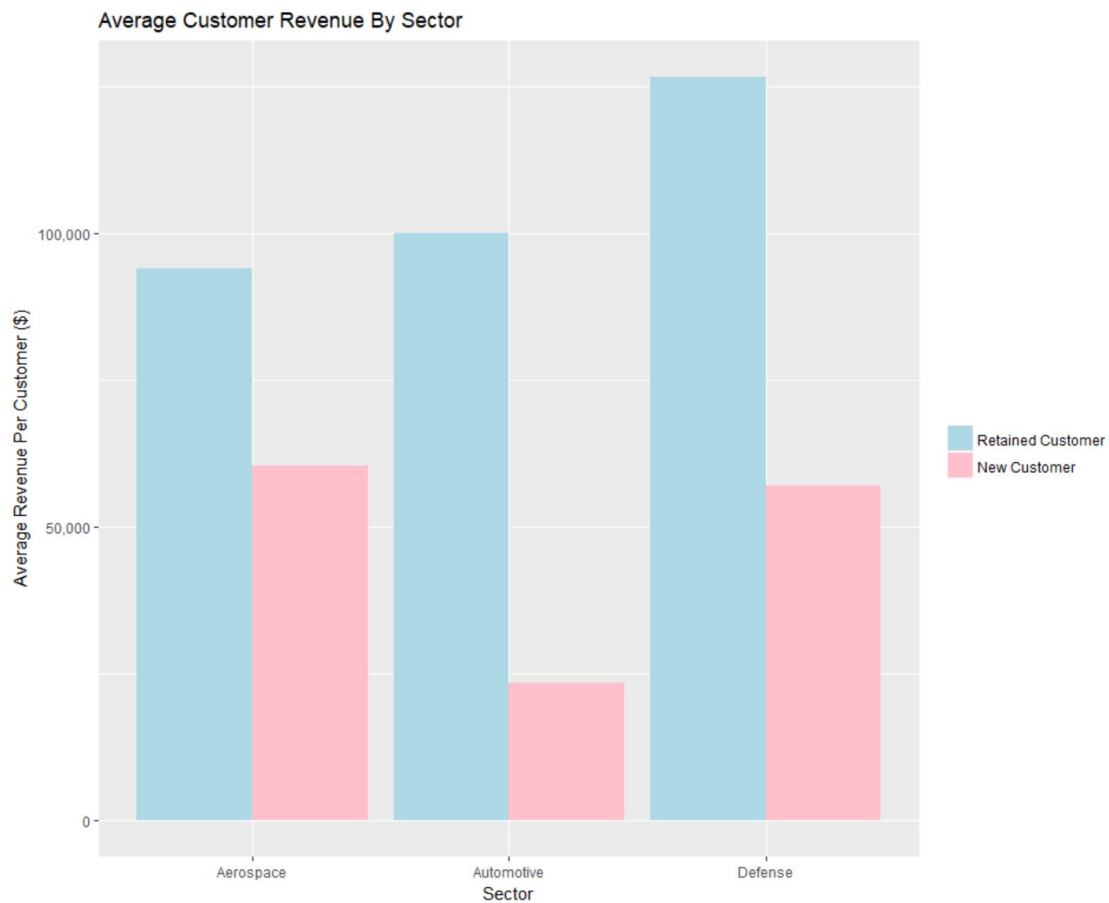### 1. New Customers vs. Retained Customers

The key question we hope to answer is whether Atri Industries should spend more money trying to acquire new customers or more money on maintaining their current customer base. We first compared the average revenue from new customers (acquired during period 8) and retained customers (acquired before period 8) by generating the following graph:

Average Revenue for New Customers vs. Retained Customers
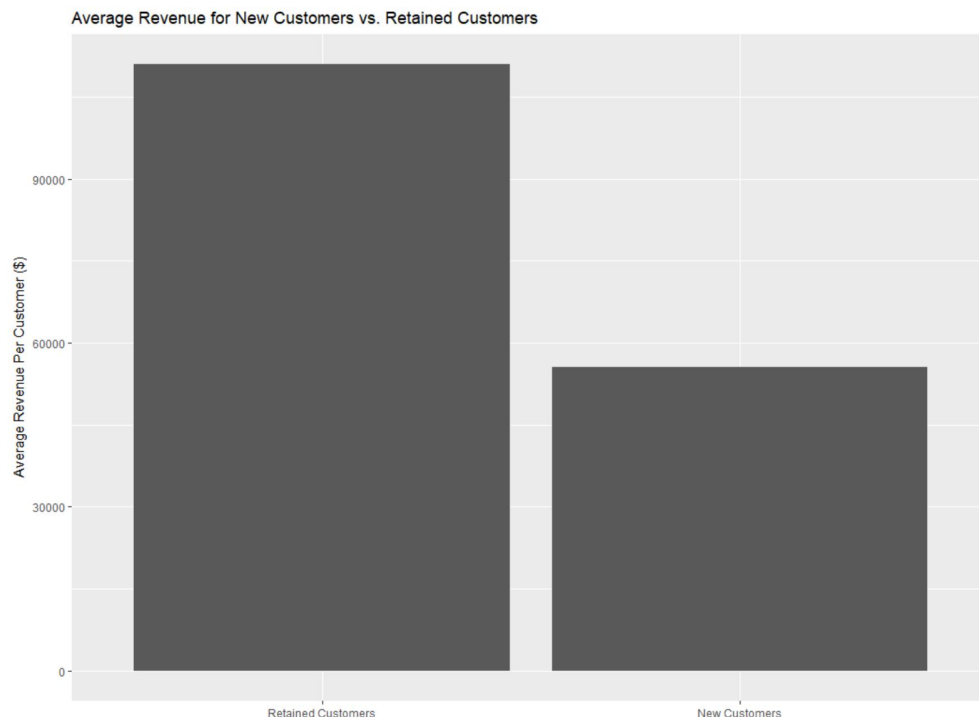
As indicated by the graph, retained customers contribute twice as much revenue as new customers.

2. Business Sector

Next we segment the customers by their business sector to see if we can gain more information about our customer base.

The bar graph we generated again shows retained customers to be the bigger spenders. However, between the new customers, we noticed new automotive customers are purchasing significantly less than new customers from other sectors.

3. Regional Performance

We then looked into segmenting current customers by region to see if it will offer us any additional insights about the customers.

Regional Performance: New Customers vs. Retained Customers

Surprisingly the spending of new customers and retained customers from the same region don't correlate – Northeast has the second lowest revenue for retained customers but the highest revenue for new customers.

4. Acquisition Channel

Lastly we looked at the sources through which the company acquired new customers to see if there are any significant patterns.

Average Customer Revenue By Acquisition Channel
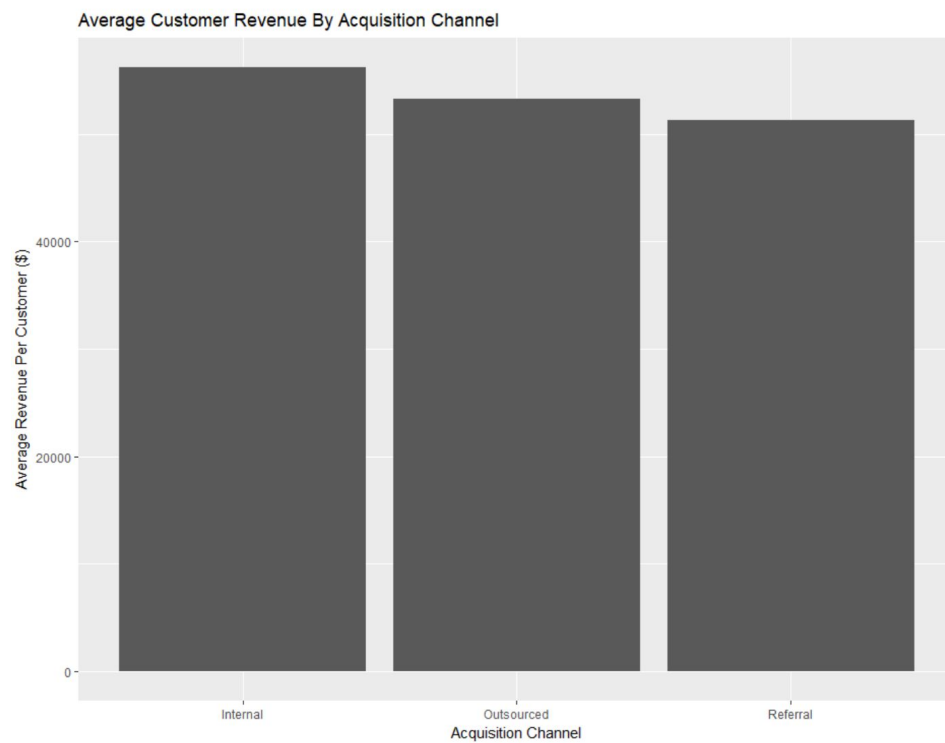
From the graph we can see Internal acquisition produces the most profitable customers, but the difference between the three are not drastic.

5. Revenue Overview

Lastly, we generated a pie chart to show exactly how the different customer categories are contributing to the total revenue of Atri Industries. Defense sector customers contribute to a large portion of the total.

Revenue Contribution % by Customer Category
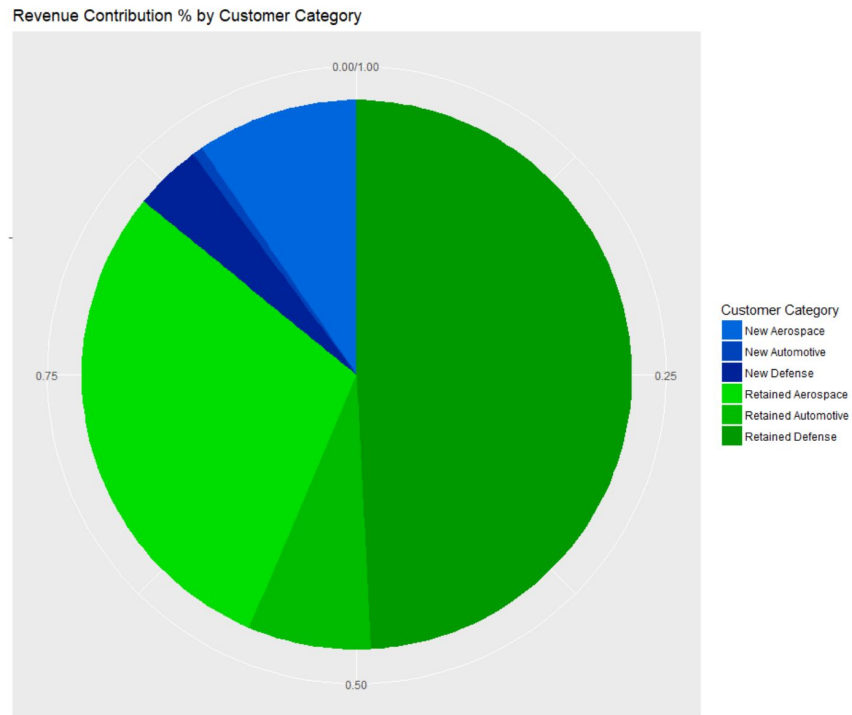
Dashboard

We also created a simple interactive dashboard using Shiny R. This dashboard will allow the user to easily select and see the performance of different customer segments. In our previous data analysis, we only looked at currently active customers. For the dashboard, we decided to use the whole dataset (containing both active and inactive customers) to see if we can uncover additional insights from the data. Below is a screenshot of the dashboard:
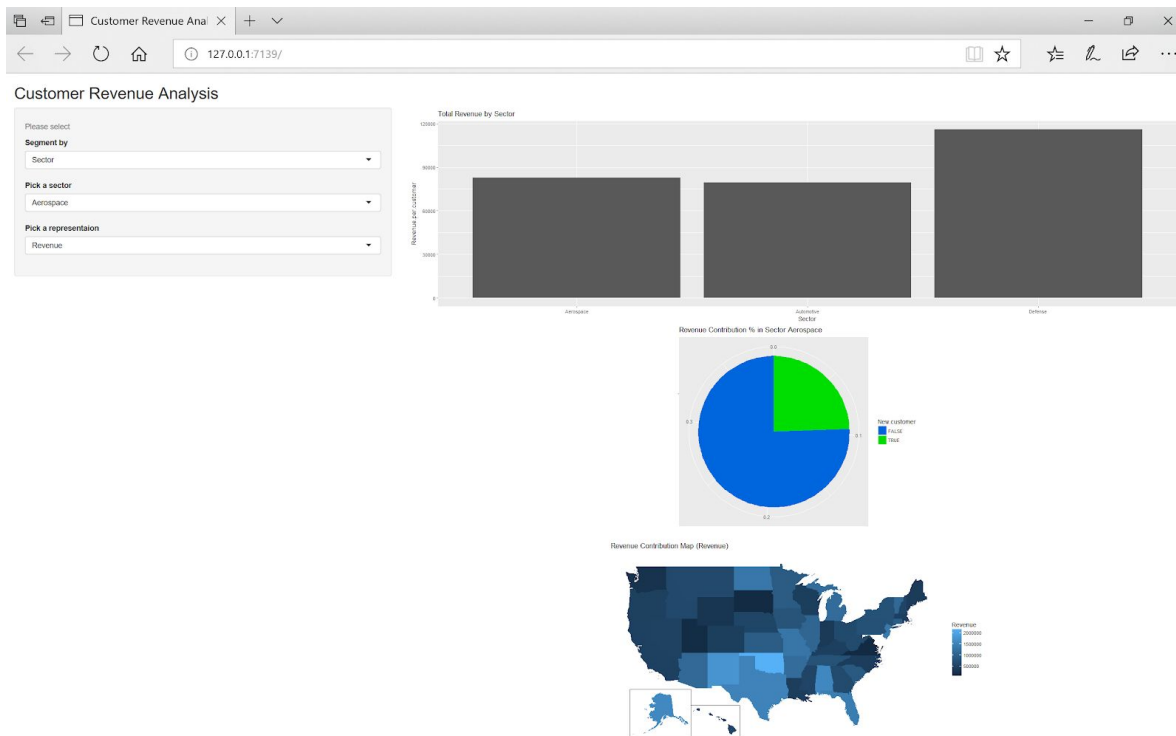
Users are able to access three different drop down lists, each linked to visual representation that will change based on the user's input choice.

The first drop down list allows users to segment customers based on 4 characteristics and see how much revenue is associated with each segment.

Segment by	
New Customer	Bar graph of average revenue per customer for new customers and not-new customers
Sector	Bar graph of average revenue per customer for Aerospace, Automotive, and Defense sectors
Acquisition Channel	Bar graph of average revenue per customer for customers acquired through Internal, Outsourced, and Referral channels
Region	Bar graph of average revenue per customer for five regions of the USA

The second drop down list allows users to select between three sectors and view a pie chart of new and retained customers' spendings from that sector. The whole pie represents the total percentage that sector contributes to the total revenue of Atri Industries. Sales to customers in the Defense industry makeup over 50% of the company's revenue.

Pick a sector	
Aerospace	Pie chart of revenue contributions from new and retained customers in the aerospace sector
Automobile	Pie chart of revenue contributions from new and retained customers in the automobile sector
Defense	Pie chart of revenue contributions from new and retained customers in the defense sector

The third drop down list allows users to select between two USA maps which show two ways of plotting our customer revenue on maps.

Pick a representation	
Revenue	Revenue map shows level of revenue in gradient colors
Revenue (Tier)	Tiered map segments revenue into 4 tiers: Tier 1: 0% ~ 25%: $87000 ~ $452000 Tier 2: 25% ~ 50%: $452000 ~ $622500 Tier 3: 50% ~ 75%: $622500 ~ $1103250 Tier 4: 75% ~ 100%: $1103250 ~ $2164000

Conclusion

Recommendations for Atri Industries

Through our analysis, we have determined that the most valuable customers are retained customers in the Defense sector. As such, Atri Industries need to allocate sufficient marketing budget to maintain relationships with said customers.

Comparing the acquisition channel revenue (active customers only vs entire dataset), we noticed many now inactive (but previously valuable) customers were acquired through referrals. Amongst active customers, those acquired through referrals are the worst performing. We recommend Atri Industries examine their referral program to determine if there are any deterrents that are causing this low performance.

Future Work

We also noted several areas of improvement to keep in mind for future analysis work. First, we wasted time on unnecessary data clean up as several columns of data we cleaning up were never used in our final analysis. Second, by focusing on the average customer revenue, we treated all customers in a segment as the same. Looking at the following summary statistics, it is very possible that one or two outliers in each segment are skewing the data. Moving forward, this is something we should address.

	New.customer	Avg.Revenue	Stdev.Rev	Min.Rev	Max.Rev
1	FALSE	110948.10	94092.67	0	552000
2	TRUE	55547.37	45836.61	0	192000

	Acquisition.Channel	Avg.Revenue	Stdev.Rev	Min.Rev	Max.Rev
1	Internal	92436.09	86972.29	0	552000
2	Outsourced	101064.52	87229.93	0	392000
3	Referral	134160.00	95874.78	32000	480000

	Sector	Avg.Revenue	Stdev.Rev	Min.Rev	Max.Rev
1	Aerospace	82738.64	79753.98	0	480000
2	Automotive	79351.35	70771.63	0	216000
3	Defense	116040.94	95916.53	0	552000

Additionally, the data we worked with was only for one period's sales of Atri Industries. We would like to look at more time series data to compare the revenue performance as time changes. For example, it is possible that new customers tend to place smaller initial orders and then ramp up purchasing amounts in the following periods. We would also want to access the cost data in order to analyze the profit gains.

There are also multiple variables in our dataset that we did not utilize in our current analysis; it would be worthwhile to further evaluate the information contained in those variables to see if additional insights can be derived.