Auto/Mate®

Auto/Mate - RPI Project Summary Report

Prepared by:

Preetham Jagadeeshan, Mengqi Fan, Wanruo Li, Seyma Yurduseven

Contents

1. Project Summary

We are working with the Salesforce data of Auto/mate, to uncover the behavioral patterns of the leads through the sales funnel. As per the data, major changes are at the stage of the proposal submitted to closed won. Our project will limit our analysis on two broad areas as mentioned below:

1. Conversion of leads to opportunities (Image 1)
2. Conversion of opportunities into closed won and closed lost (Image 2)

We will identify the patterns and findings using appropriate algorithms in the above 2 categories of the sales process. More details can be found in our solution approach.

Image 1: Converting leads to accounts

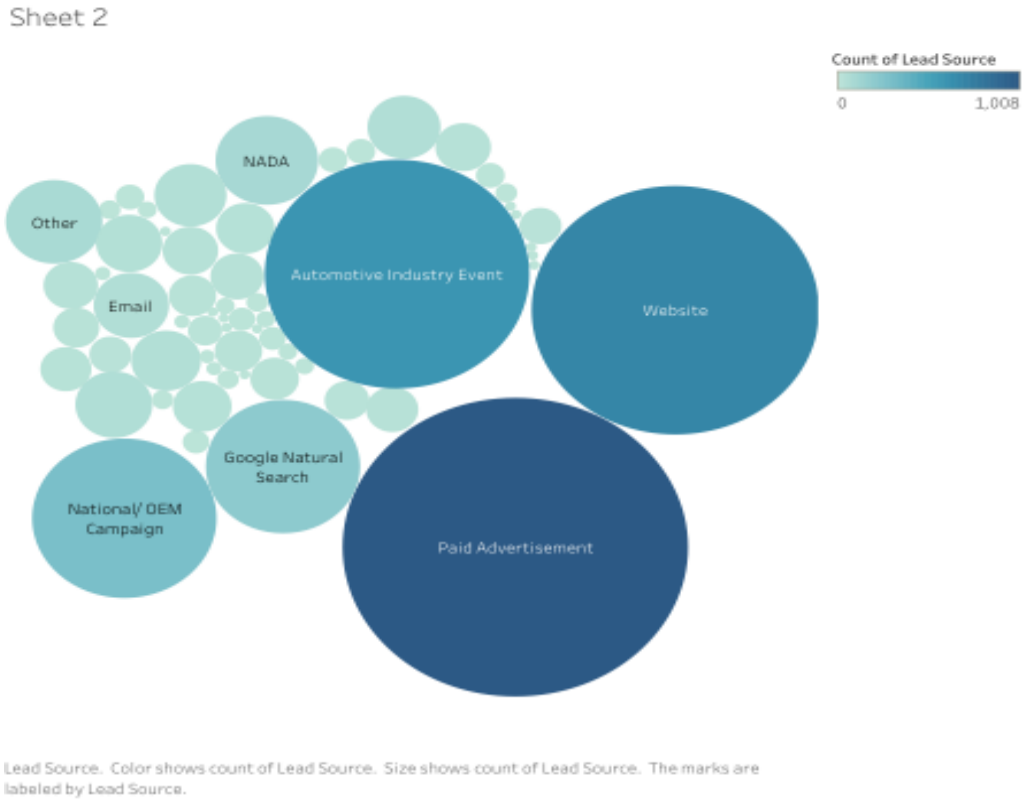
Image 2: Qualifying leads

2. Data Understanding and Business Insights

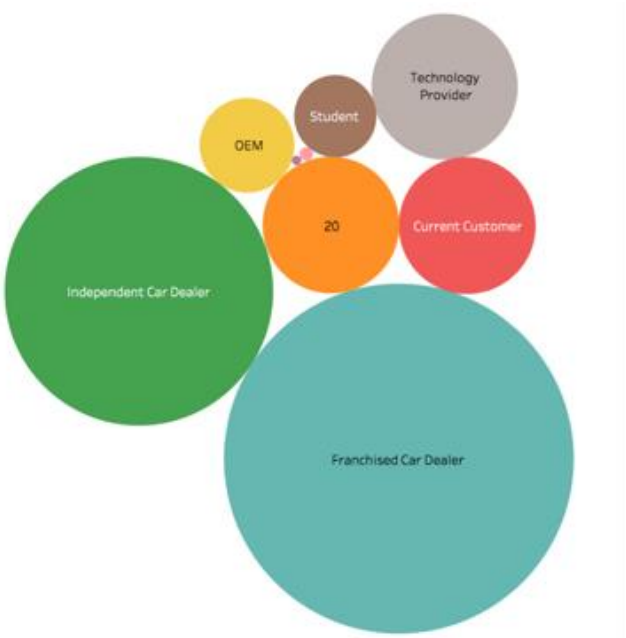
a) Leads Data (Click on icon for supporting numbers)

Leads _ Summary reports.xlsx

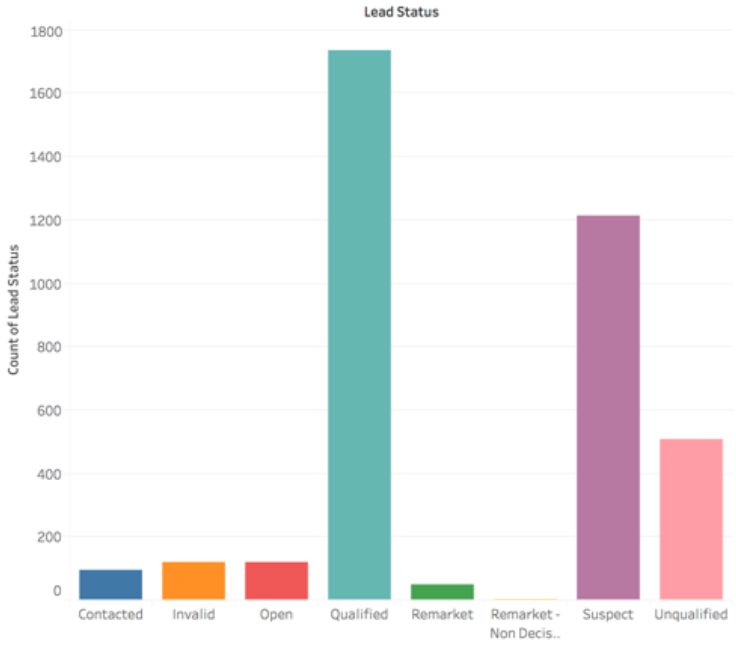
1. The raw data has 3844 leads record but due to the updated information, we deleted the duplicates (only kept the most recent one according to the column: (Last.Modified.Date) and there are 3305 unique leads left. The analysis will be based on records of these 3305 unique leads.
2. We found that the lead source is a valuable variable to explore. In the below graph we can see that “paid advertisement”, “website” and “automotive industry event” are really effective at getting new leads.



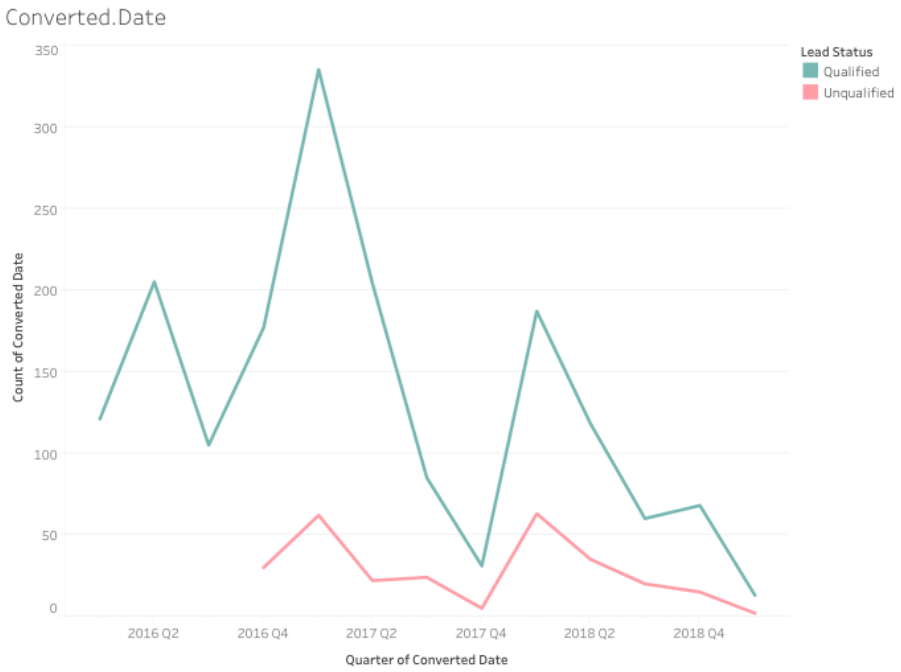
3. Inquiry type shows the type of leads we are getting. The major type of leads who reach out to us are Franchises car dealer, Independent car dealers, Technology providers, and current customers. Exploring the percentage of our current customers reaching back to us during the end of contract periods will help us sense the customer's pulse and align our efforts better.



4. The lead status will be our target variable while evaluating the successful conversion of leads to opportunities. There are no obvious relationships with converted variable and lead status.

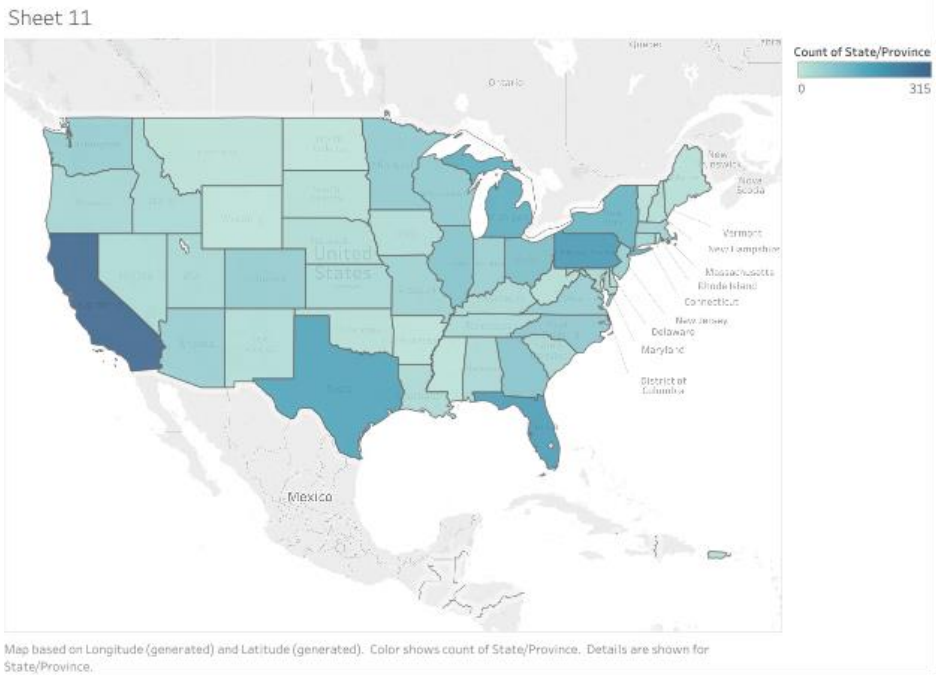


5. There is an overall decrease in the trend of the number of leads qualified and unqualified. 2016 Q4 sees the most number of qualified leads and those spikes are seen every alternative quarters



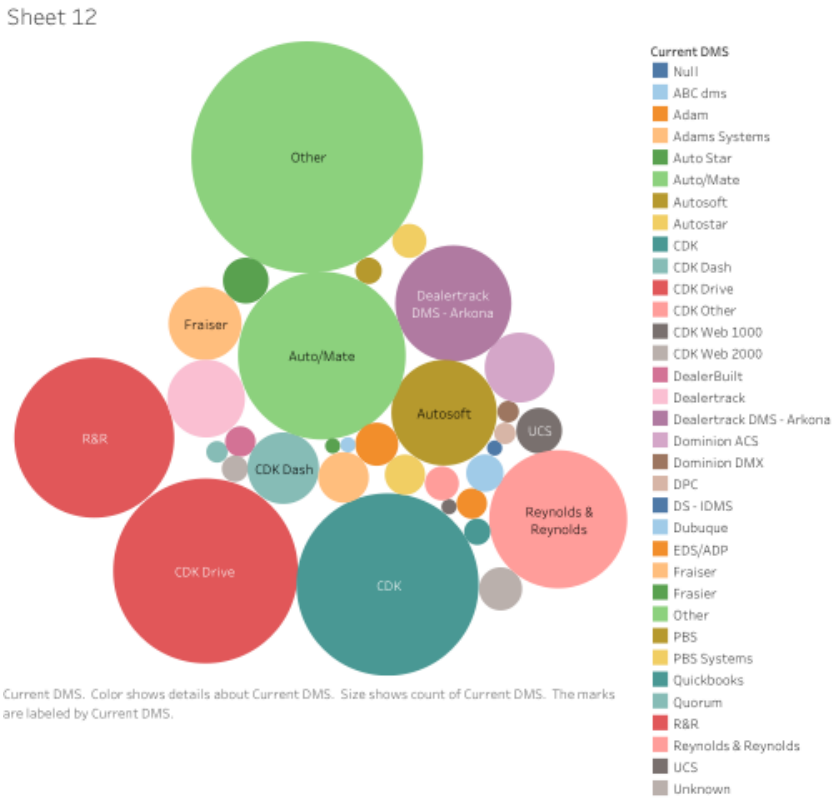
The trend of count of Converted Date for Converted Date Quarter. Color shows details about Lead Status. The data is filtered on Converted Date Year, which keeps 2016, 2017, 2018 and 2019. The view is filtered on Lead Status, which keeps Qualified and Unqualified.

6. Below is the geographical distribution of our customers across the United States.



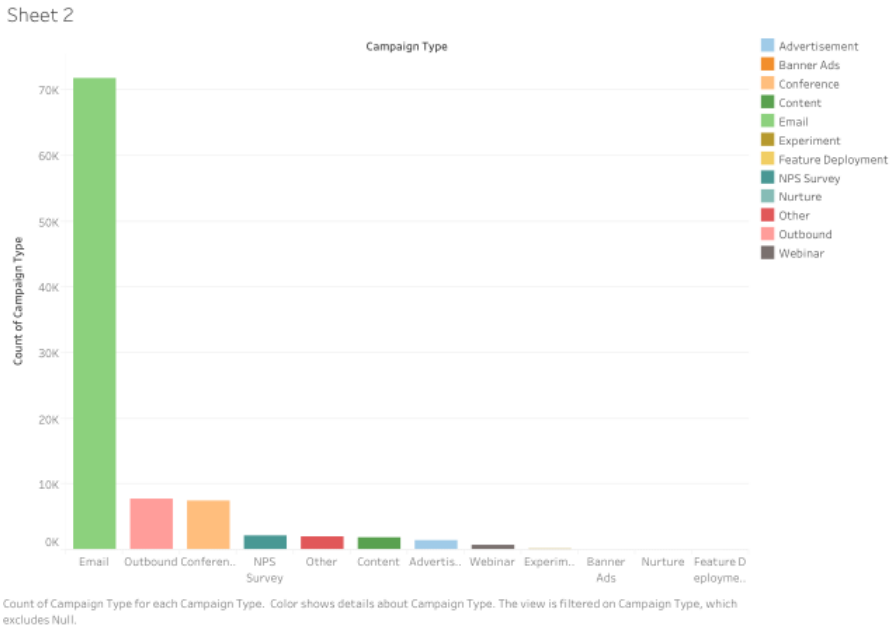
Map based on Longitude (generated) and Latitude (generated). Color shows count of State/Province. Details are shown for State/Province.

7. We have an equal volume of leads from CDK and R&R. Exploring further into how many of our competitor's leads are we able to converting will show us our capability to grab the customers.



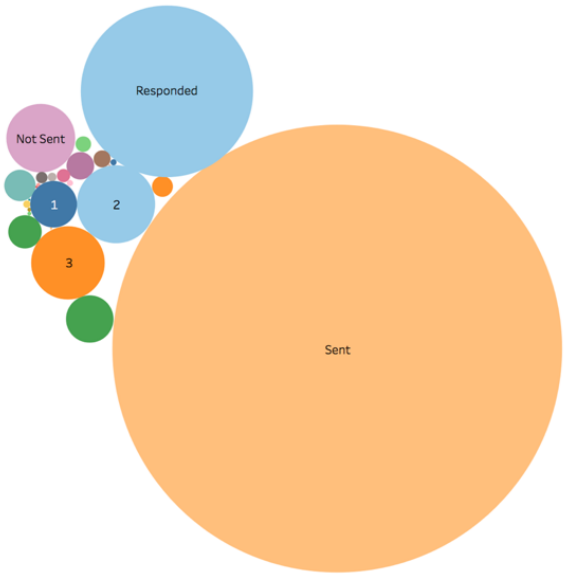
b) Campaigns Data:

1. We have 12 Campaign Types > 30 Parent Campaign Name > 229 Campaign Names (ID) > 95799 Records. We will use this dataset by combining the data from here with either leads data or opportunities data. It is important to see the success of different campaigns.
2. Apart from the IDs, the main variable we need to focus on is the campaign type. We can gather from our initial analysis that “email” is by far the most common campaign type followed by “outbound” and “conference”.



Campaign Type Distribution

3. It is also important to see how the leads are responding to different campaigns. From this analysis we can gather insights into the responses we get to our campaign actions.



Member Status Distribution

c) Contacts and Accounts Data:

1. There are 14695 unique Contact ID and 6940 unique Account ID. It means multiple Contact ID corresponds to one Account ID. An account may be contacted by multiple people to gain success or failure, or it is a way to keep old clients. We can use this to see whether contacting time will influence the loyalty of clients.
2. There are more than 74% of Account ID is suspect type. Nearly 11% of Account ID is previous prospect type. 7.19% of Account ID is prospect type. 2.13% of Account ID is third party type. 1.69% of Account iD is previous customer type. 0.04% of Account ID

is opportunity type. 0.24% of Account ID is lead type. 0.07% of Account ID is mate partner type. We may focus on opportunity type, lead type and mate partner type. Although the data is very small, it may provide us with useful information about how to gain clients successfully.

3. There are 23 details of sources. The dataset only has 392 values in this column. Among them, 165 clients came from a digital advertisement; 137 clients came from a national event. These two sources are very important to attract clients. We may compare the campaigns that being used in these two resources with campaigns being used in other resources to discover interesting things. Besides campaigns, there are still other things we need to think about, such as the money spend on these two resources, the range of affected clients in these two resources and so on. So that we can know the power of campaigns in these two resources accurately.

Row Labels	Count of Lead Source Detail 2
Blog Signup Form	
Buy/Sell	
Cold-Call	4
Content Download Form	2
Cross-sell/Upsell	1
Dealer Association	
Digital Advertisement	165
Existing Customer	3
Inbound Call	2
Industry Influencer	6
Linkedin	2
Manufacturer Event	54
National Event	137
One to One Email	
Other Event	
Outbound Call	
Print Advertisement	2
Regional Event	5
RSM Research	
Support	
Walk-In	1
Web Lead Form	3
(blank)	5
Grand Total	**392**

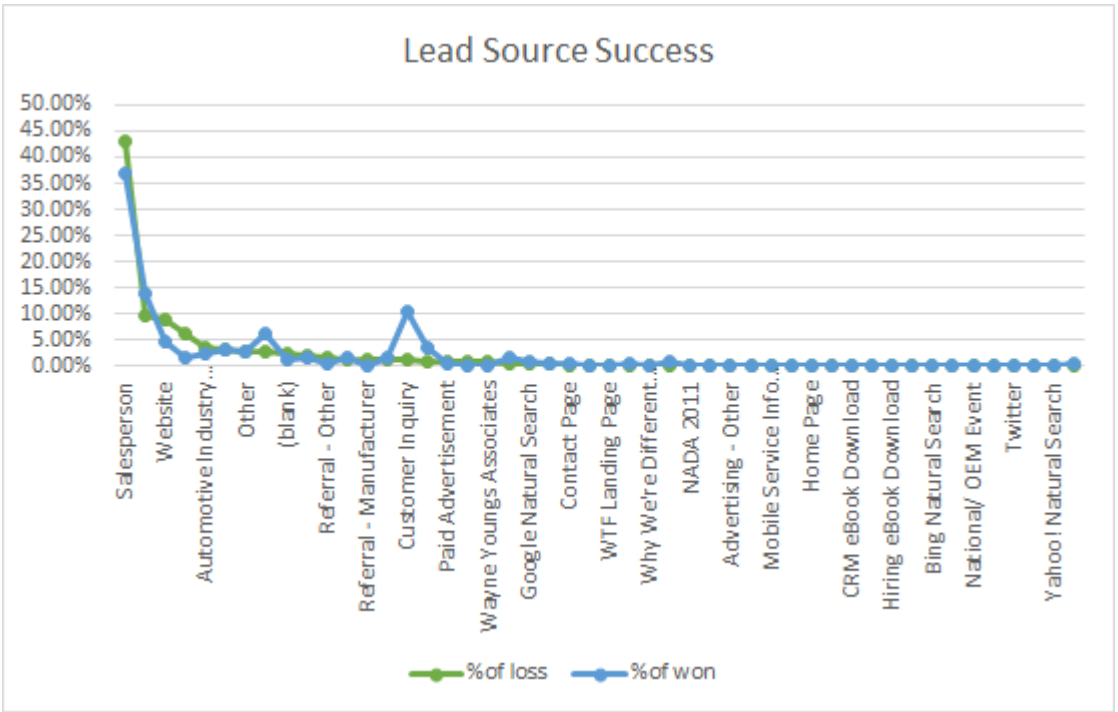
4. There is no much information about the lead source of current Auto Mate clients. Most of the data are null values. Except for that, top sources of current clients are a salesperson, referral. Since the work of salesperson is difficult for us to analysis in this project. We may try to find out what kind of clients are more likely to refer DMS to other dealers.

d) Opportunities Data:

1. This dataset captures information about the leads that has been converted into opportunities. We can use this dataset to get insights about which opportunities were closed as won and which were closed as loss. This is the last step of the funnel.
2. This dataset helps us identify the factors that contributes to a successful sale. From our initial analysis, we have reached the numbers in below graph about the lead source. From this data, we can gather that as a lead source, salesperson action is the biggest contributor to the total number of sales since it constitutes 37% of all our Closed Won variable. However, the success rate of salesperson variable is 28% while Customer Inquiry has 81% success rate. We can also see that existing customer referrals has a high success rate, but

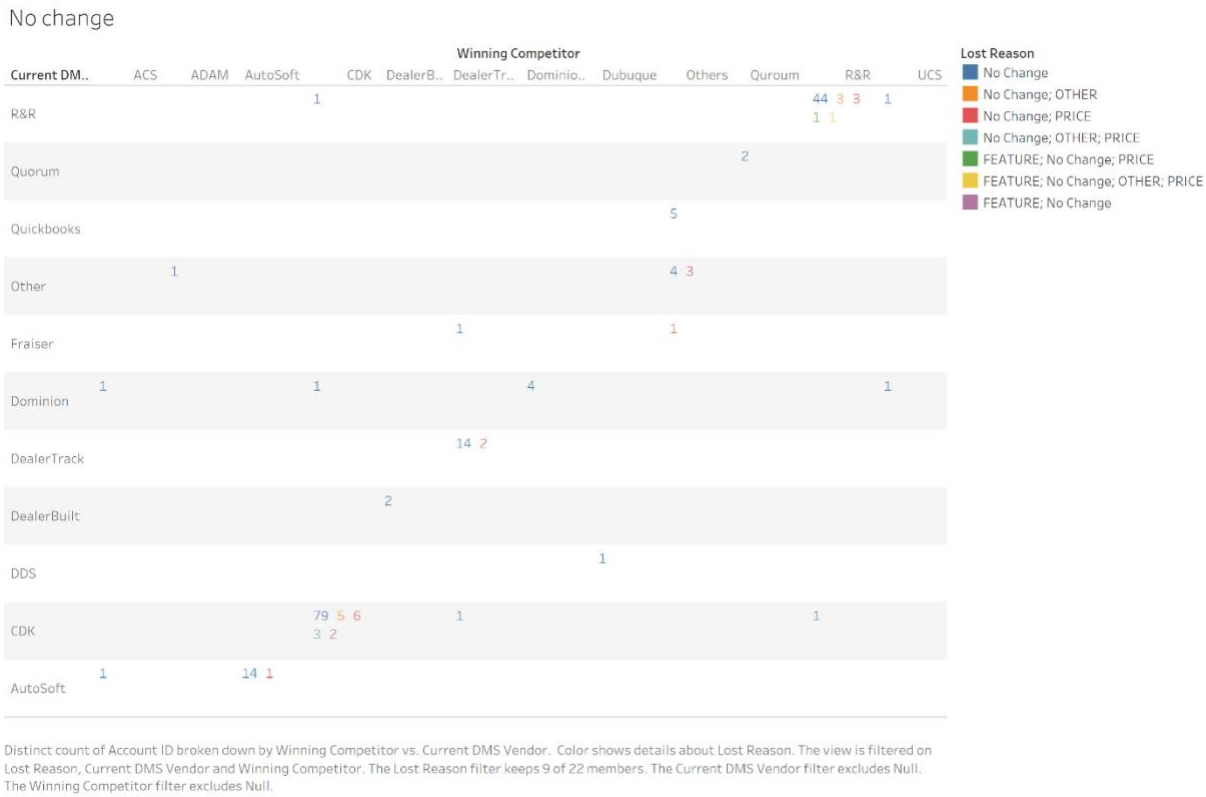
in the end, salesperson contributing the most to the total sales, and being under direct control of the company, makes it the most important asset we have, and if company works on improving salesperson performance, it might have direct influence on the sales. However, it is important to see other factors that affect the sales and try to find ways to improve them or make them more frequent, too.

Lead Source	Loss	Won	Grand Tota	%of los	%of wo	Success rat
Salesperson	1066	410	1476	42.93%	37.07%	27.78%
Referral	240	153	393	9.67%	13.83%	38.93%
Website	226	51	277	9.10%	4.61%	18.41%
NADA	158	20	178	6.36%	1.81%	11.24%
Automotive Industry Event	90	27	117	3.62%	2.44%	23.08%
Paul Gillrie	78	35	113	3.14%	3.16%	30.97%
Other	72	32	104	2.90%	2.89%	30.77%
Referral - Existing Customer	68	70	138	2.74%	6.33%	50.72%
(blank)	61	16	77	2.46%	1.45%	20.78%
Phone Call	48	20	68	1.93%	1.81%	29.41%
Referral - Other	44	8	52	1.77%	0.72%	15.38%
DMS Section	34	17	51	1.37%	1.54%	33.33%
Referral - Manufacturer	31	3	34	1.25%	0.27%	8.82%
State ADA	29	17	46	1.17%	1.54%	36.96%
Customer Inquiry	28	117	145	1.13%	10.58%	80.69%
Buy/Sell - Add-on Location	21	40	61	0.85%	3.62%	65.57%
Paid Advertisement	20	5	25	0.81%	0.45%	20.00%
Marketing Campaign	20		20	0.81%	0.00%	0.00%
Wayne Youngs Associates	19		19	0.77%	0.00%	0.00%
Email	18	17	35	0.72%	1.54%	48.57%
Google Natural Search	17	12	29	0.68%	1.08%	41.38%



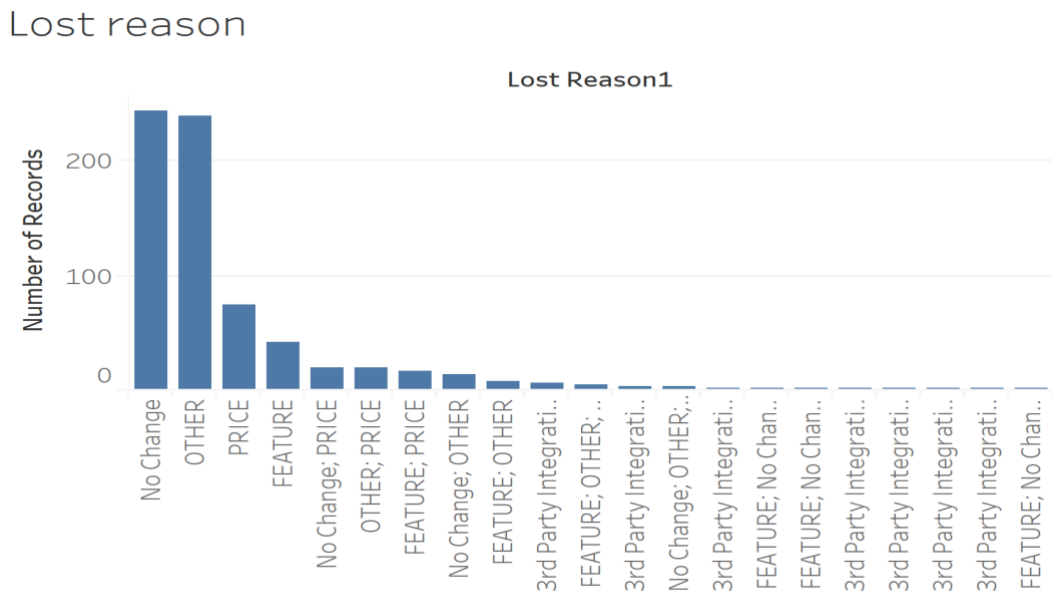
3. Lost Reason

Why the company loses an opportunity and has to close the account as a “loss” is a very important variable. We have two columns to capture this data. One is the lost reason and the other one is “Reason Comments”. Unfortunately, we have a lot of missing values for these variables since they were not entered into the system. However, we are hoping to have enough to get some insights. For the initial analysis, we have looked to see if the reason written matches the reality. Logically, if the lost reason is that the customer does not want to change their system, the winning competitor should be the same as the current vendor they are using. From the graph below, we can see that except for singular cases that can be considered as outliers, these lost accounts mainly stayed with their current vendors. So we can mostly trust the integrity of this variable.



Distinct count of Account ID broken down by Winning Competitor vs. Current DMS Vendor. Color shows details about Lost Reason. The view is filtered on Lost Reason, Current DMS Vendor and Winning Competitor. The Lost Reason filter keeps 9 of 22 members. The Current DMS Vendor filter excludes Null. The Winning Competitor filter excludes Null.

After the initial analysis of the lost reason, we have found that the biggest issue while trying to close a sale is that the dealers not wanting to change their current DMS. This reason is followed closely by “Other” value that we cannot reach any insights to and then by price and feature. We can suppose that any improvements on “no change” factor would have a high impact on the total sales. However, this factor is not under the control of the company and it is hard to improve. Price and feature factors would also be effective if improved.

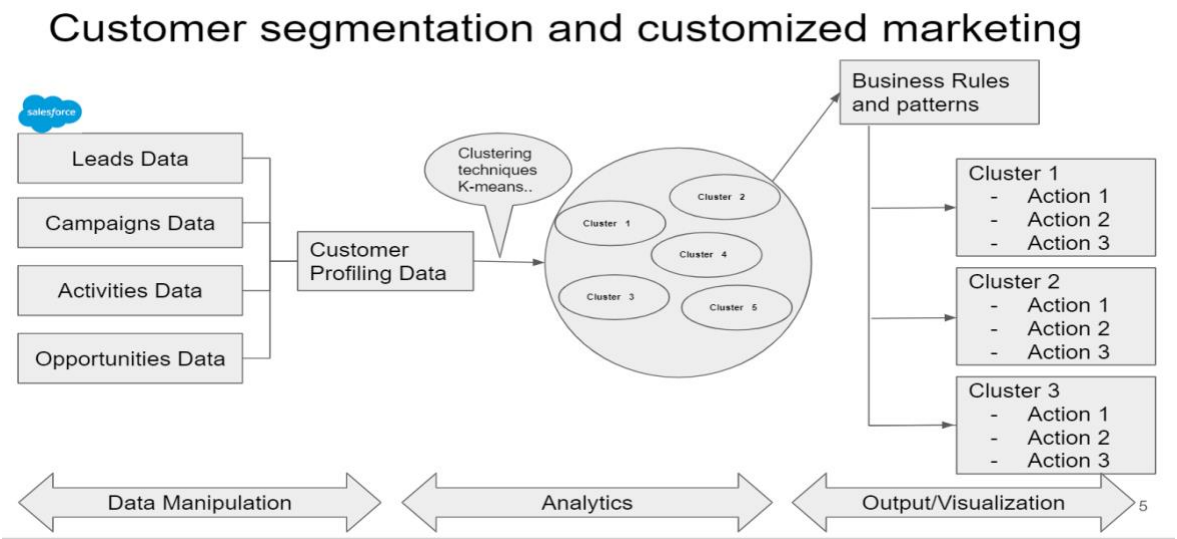


Sum of Number of Records for each Lost Reason1. The view is filtered on Lost Reason1, which excludes Null.

3)Proposed Solution

Our proposed solution has two approaches. At the first stage, we will try to diagnose which factors affect the successful conversion of a lead into an opportunity and try to suggest improvements on marketing actions towards this goal. At the second stage we will try to diagnose the factors that affect the successful sale of the DMS to those opportunities and come up with improvements.

We will also implement customer segmentation and customized marketing to see if we can further optimize the marketing actions towards a successful sale.



4)Key Milestones

Week 1: Team created, understand the Auto/Mate company and the project: the company background, product research, project goal/ initial outline.

Week 2: Initial kickoff meeting at Auto/Mate Dealership Systems: deeply understand the selling process/stages, report the proposed approach to our client and get some ideas

Week 3-4: Get access to data: a collection of data from different Salesforce objects. Creating a comprehensive data dictionary for the 5 datasets. Doing the basis summary analysis of data, delete the duplicate IDs and generate 4 data summary report with visualizations.

Week 5-6: Data preparation: creating two datasets by combining the 5 initial dataset: leads to opportunities, opportunities to closed with basic data cleaning and feature selection.

Week 7-8: Based on the prepared dataset, using R/EXCEL to continue feature selection, choose the variables prepared for modeling. Try multiple basic algorithms to train and test the model and compare the fit result.

Week 9-11: Try other different more complex algorithm with continuing feature selection & feature creation. Model selection by comparing fit result and try stacking model if we have time.

Week 12-15: final hyper-parameter adjustment and writing final report

5)Project Timelines:

From week 1 to week 4, we plan to understand the whole project better, to meet the clients, to gain data and to be familiar with the data set. That is, we need to get insights from data, prepare the data for analysis and deal with the questions we have in the datasets.
From week 5 to week 9, we plan to prepare the data for modeling. So that we need to clean the data; we need to select appropriate models and evaluate them. Besides, in order to evaluate models from business sight, we need to learn business theory which is matched with our project. From week 10 to week 13, we plan to test our model and evaluate our models. And then create documents to explain models and our final findings on Auto/Mate business process. From week 14 to week 15, we will convey our findings to the client.

Project plan draft