Acme Insurance Co. Marketing Analytics

Making Predictions Using Linear and Logistic Regression

The problem

Company

Acme Insurance a small west coast insurance provider is looking to expand into a new market.

Context

One proposed marketing strategy for this expansion is to target customers that are most likely to respond to marketing messages and further incentivize those that are most likely remain loyal.

Problem statement

We need to be able to predict weather a prospect will respond to our marketing campaign and we need to be able to predict customer lifetime value for our new customers.

Challenges

Challenge 1

Categorical Variables

We have many categorical variables in the data that we have. It may be tough to find linear relationships with this data.

Challenge 2

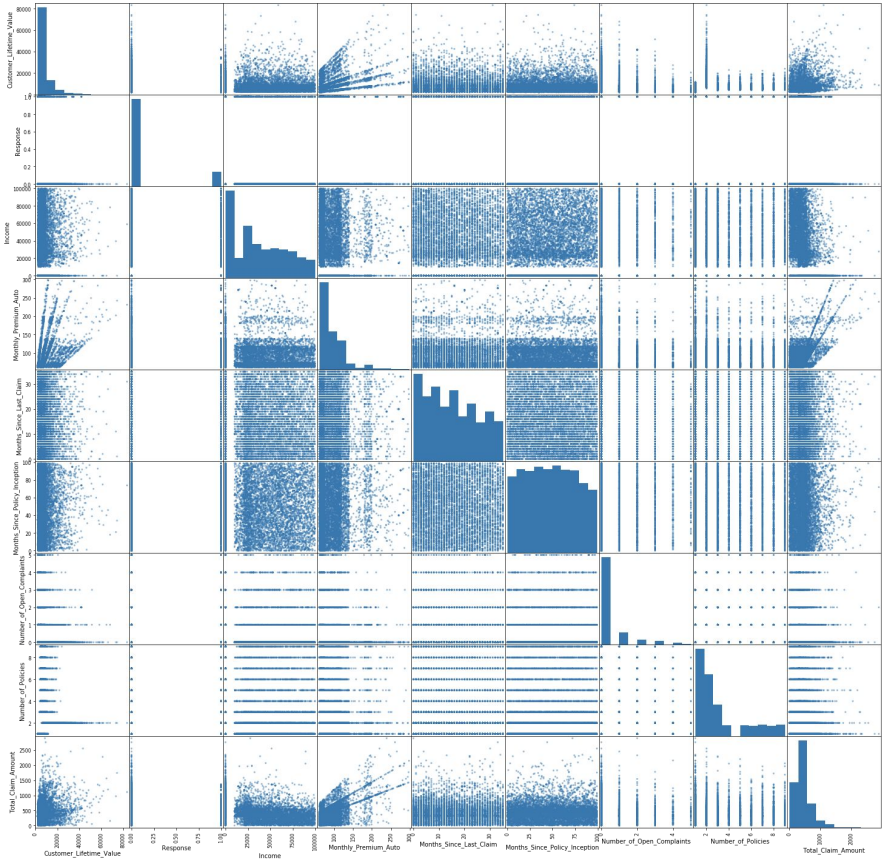
No Normal Distribution

Our data is not normally distributed and we have a significant class imbalance between yeses and noes in our marketing response data.

Challenge 3

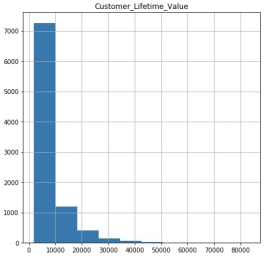
Limited Models

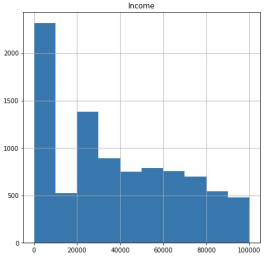
Given challenges 1 and 2 along with the models we currently have at our disposal, making reliable predictions will be a challenge.

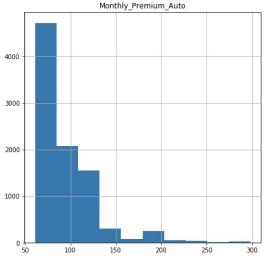


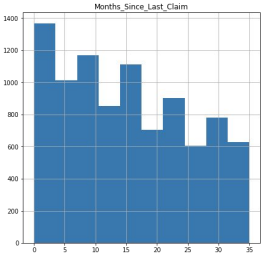
Categorical Variables

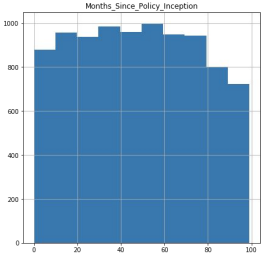
At first glance there are clearly many categorical variables and not many linear relationships.

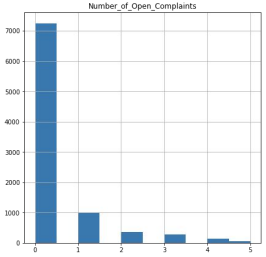


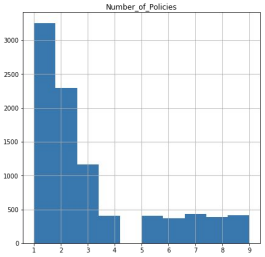


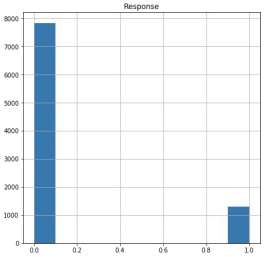


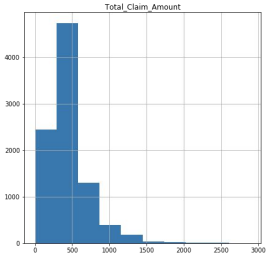






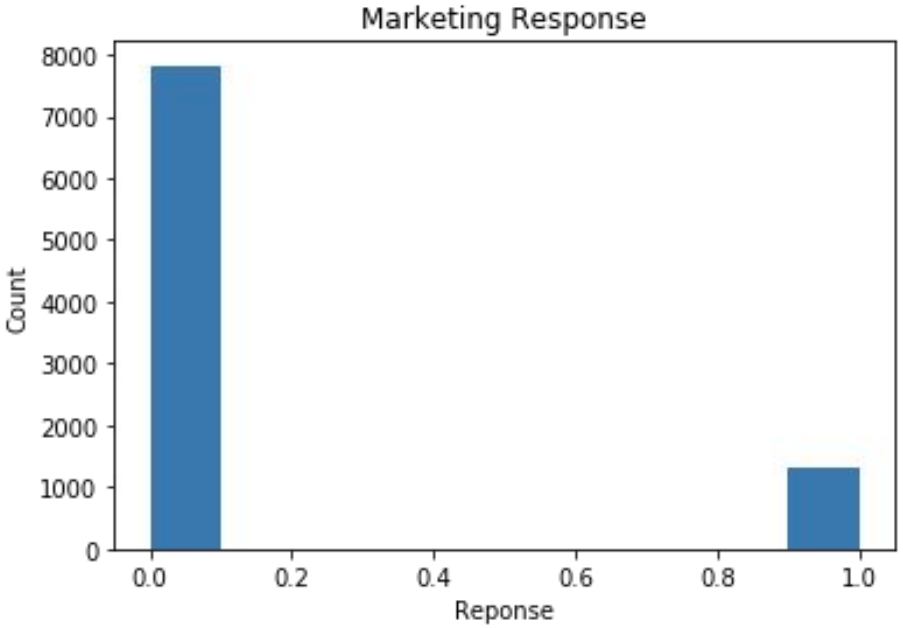






Lack of Normal Distributions

Along the diagonal we see that none of our data is normally distributed.



Class Imbalance

There are significantly more noes than yeses

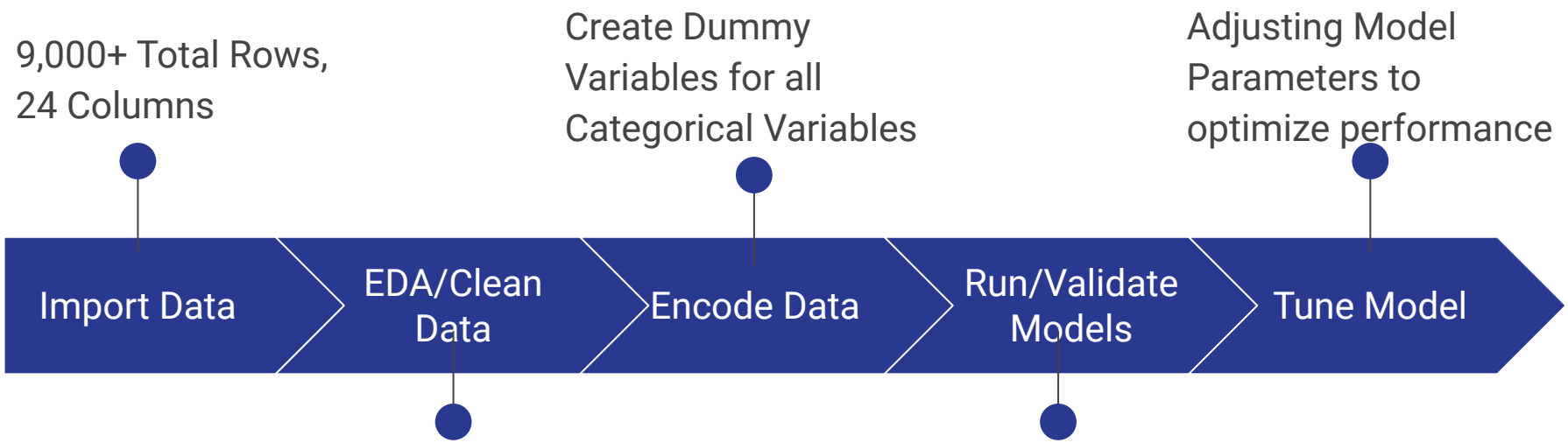
Available Models

The models that we have available given our resources are linear and logistic regression.

Our models have assumptions of normal distributions within given variables and linear relationships from one variable to another.

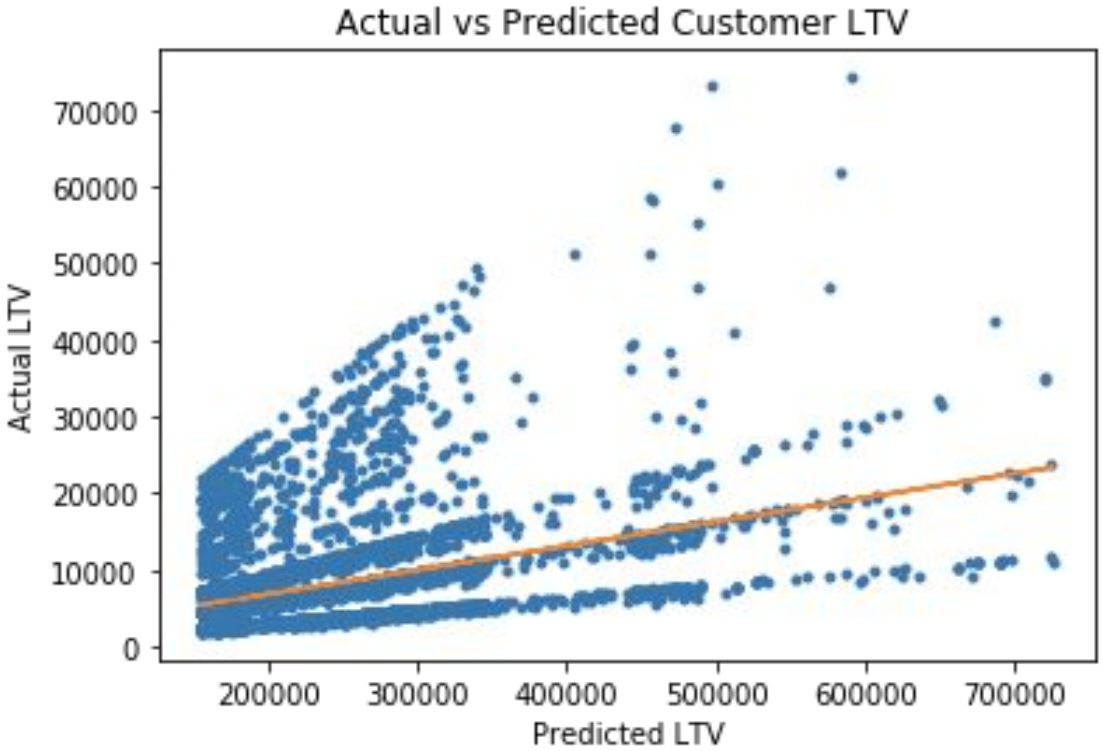
The assumptions for these models have not been met and this will affect the model's performance.

Project Roadmap



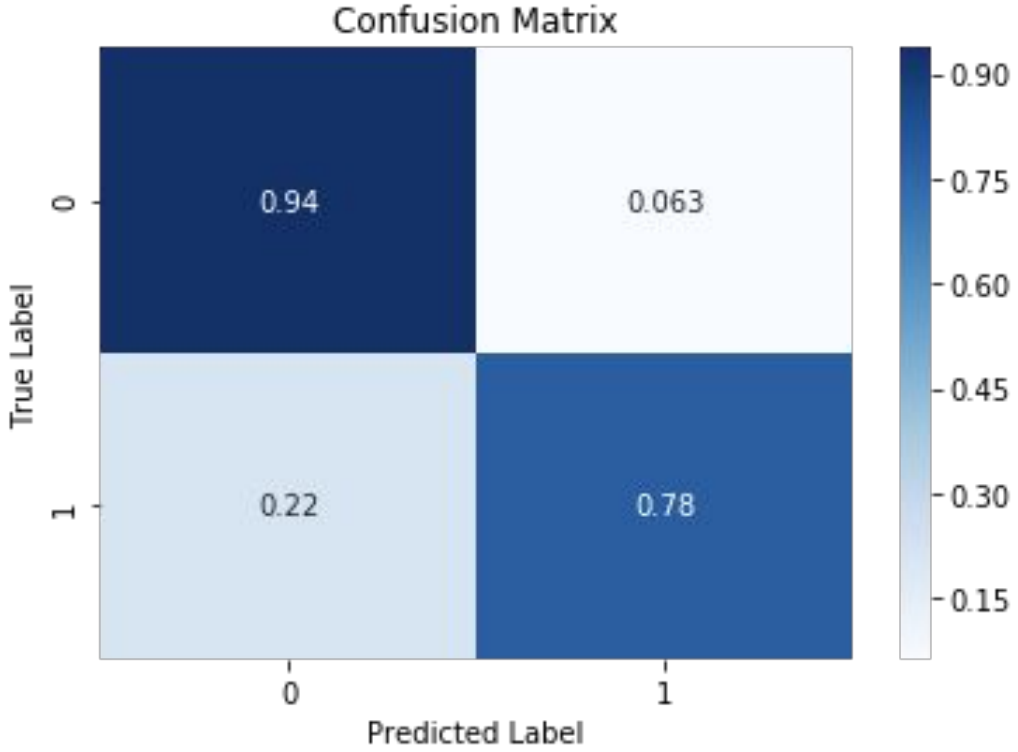
Results

Customer Lifetime Value Predictions

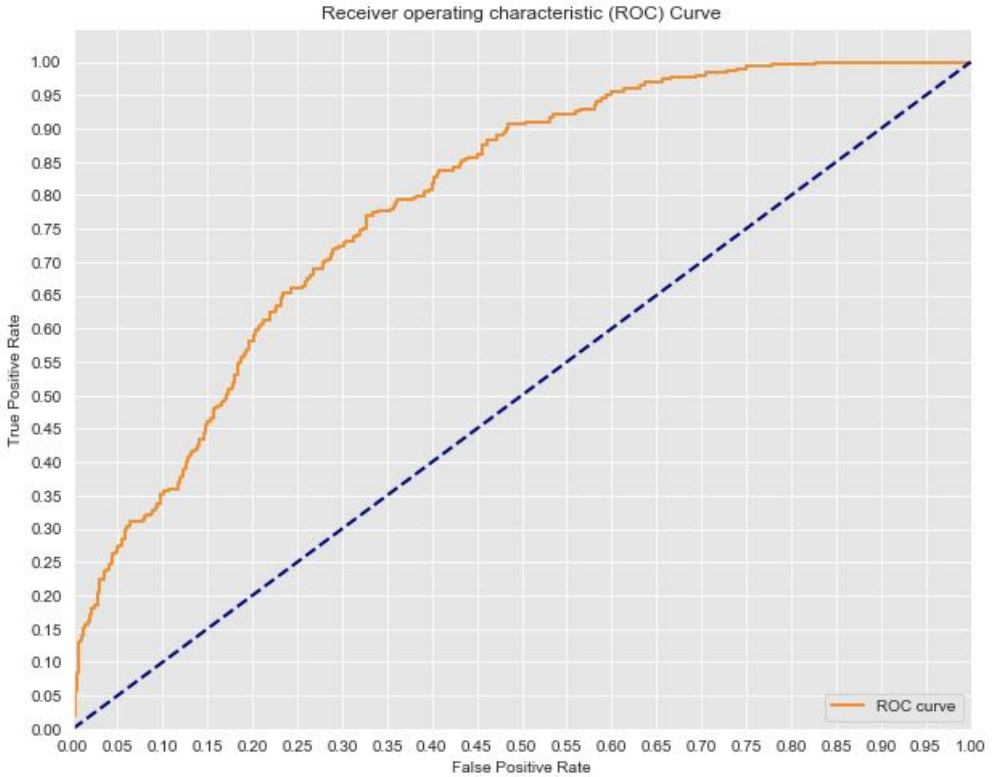


R^2 Score:
~16% on
Testing Data

Marketing Responsiveness Prediction



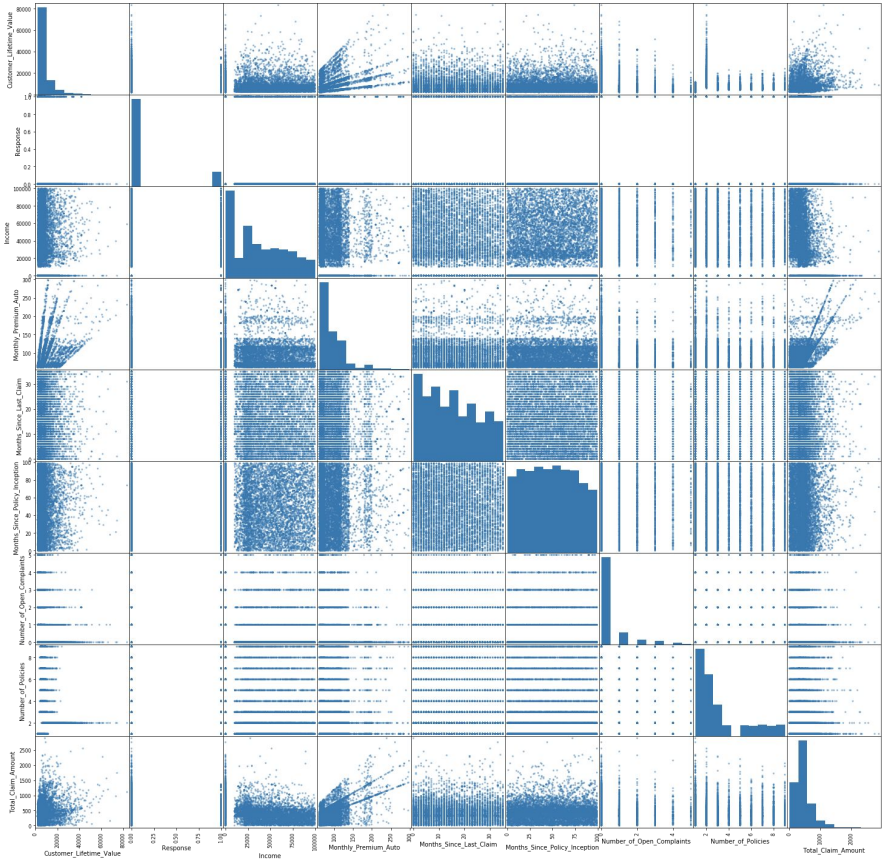
Marketing Responsiveness Prediction Cont.



AUC: 79%

Given two customers, one that said yes, to marketing and one that said no, there is a 79% probability that our model will correctly classify both

Takeaways & Next Steps



Our model assumptions were not met

There are too many categorical variables, no strong linear relationships, and our variables were not normally distributed.

Next steps:

Use more advanced modeling techniques that don’t hold our previous assumptions

- ★ Decision Trees
- ★ SVC
- ★ XG Boost
- ★ Neural Networks

Fin