General Assembly Final Project

Influential Factors of Medicare Social Media Performance¶

Natalie Kampf

Background Information

- The data sampled was collected from a social media campaign for a private insurance provider located in New Jersey.
- Campaign data was run for Medicare Advantage plans offered only through private insurance companies rather than original Medicare offered by the government.
- The data was pulled from 7 months of a year long campaign that is targeted to people ageing into Medicare (turning 65).

Problem Statement

Medicare Advantage enrollment is very concentrated in a small number of private companies, which include UnitedHealthcare and Humana together accounting for 44 percent of all Medicare Advantage enrollees nationwide, and the BCBS affiliates (including Anthem BCBS plans) account for another 15 percent of enrollment in 2019.

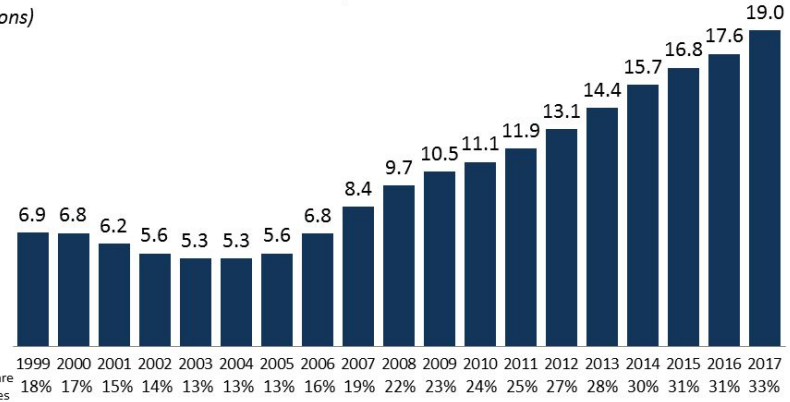
With a competitive landscape for enrolling members in Medicare Advantage plans, it's important to evaluate how you can capture your audience in the most effective way possible. Specifically looking at social media channels.

Increase In Medicare Advantage Enrollment

Figure 1

Enrollment in Medicare Advantage plans has steadily increased since 2004

Total Medicare Private Health Plan Enrollment, 1999-2017
(in millions)



NOTE: Includes MSAs, cost plans, demonstration plans, and Special Needs Plans as well as other Medicare Advantage plans. Excludes beneficiaries with unknown county addresses and beneficiaries in territories other than Puerto Rico.
SOURCE: Authors' analysis of CMS Medicare Advantage enrollment files, 2008-2017, and MPR, "Tracking Medicare Health and Prescription Drug Plans Monthly Report," 1999-2007; enrollment numbers from March of the respective year, with the exception of 2006, which is from April.

THE HENRY J. KAISER FAMILY FOUNDATION

With Medicare advantage enrollment nearly doubling over the last decade, in 2019 there were approximately 22 Million people enrolled in Medicare Advantage plans.

Questions

- What factors determine the amount of leads generated by a social media ad?
- Which models would perform best on the data to predict the amount of leads a campaign will generate?
- How well does it predict and how might we better predict leads in the future?

Goal

- By using social media data from a private insurance company we can train an algorithm on the dataset that can predict lead generation based on multiple variables.

- With the ability to predict lead generation the model can hopefully inform the company about how to make better business decisions and win in the competitive landscape

Taking A Look At The Data

	Day	Ai_trk	GP Job Name	Impressions	Media Cost	visits	Call Leads	Web Leads	Total Leads
0	18-Jan-19	19SOC-BDA1-C1	Age-In NTM Social Facebook (64+6 mo) Free Guide	333	5.85	3.0	0.0	0.0	0.0
1	18-Jan-19	19SOC-BDA1-C2	Age-In NTM Social Facebook (64+6 mo) Three Mon...	340	6.87	3.0	0.0	0.0	0.0
2	18-Jan-19	19SOC-BDA1-C3	Age-In NTM Social Facebook (64+6 mo) Carousel	156	2.15	2.0	0.0	0.0	0.0
3	18-Jan-19	19SOC-BDA2-C3	Age-In NTM Social Facebook (64+9 mo) Image Car...	303	6.00	0.0	0.0	0.0	0.0
4	18-Jan-19	19SOC-BDA3-C1	Age-In NTM Social Facebook (64+11mo) See Any D...	177	1.56	0.0	0.0	0.0	0.0

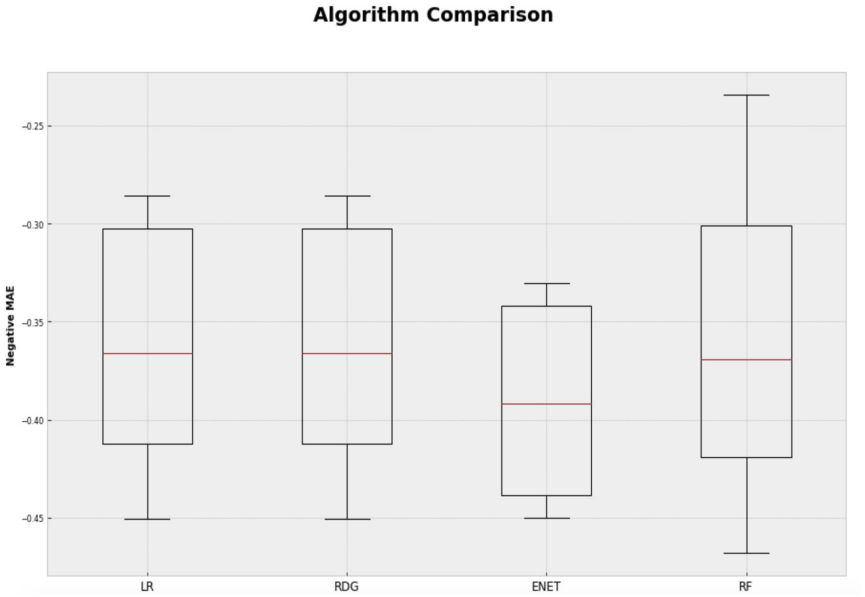
	Impressions	Media Cost	visits	Call Leads	Web Leads	Total Leads
0	333	5.85	3.0	0.0	0.0	0.0
1	340	6.87	3.0	0.0	0.0	0.0
2	156	2.15	2.0	0.0	0.0	0.0
3	303	6.00	0.0	0.0	0.0	0.0
4	177	1.56	0.0	0.0	0.0	0.0

Initial Data Exploration

- Starting out with a limited amount of variables it was important to take a an in-depth look at the correlations that appear among the variables.
- The web and call leads had the highest sense of correlation to the target variable Total Leads
- Media Costs and visits had the second highest correlation in regards to correlation with Total Leads
- Impressions shared the lowest correlation with Total Leads.

Evaluation of Models

- Performed cross validation of models to understand which approach would best support the data I was looking at
- Evaluated each model on a clean dataset
- Chose the Random Forest (RF) model based on comparison and effectiveness



The Models

- I evaluated the models on both R-squared and RMSE.
- By utilizing RMSE I wanted to ensure that any larger errors would be penalized, which was important in the model process due to making decisions that could affect business outcomes for the insurance provider.

Results

Initial Run Scores (without Web Leads)

- Linear Regression
  - R-Squared: 15%
  - RMSE: 0.824
- Random Forest
  - Score: 84%

Best Scores

- Linear Regression
  - R-Squared: 96%
  - RMSE: 0.15
- Random Forest
  - Score: 98%