USING MULTI-TOUCH ATTRIBUTION & ML TO OPTIMIZE AND PREDICT HCP JOURNEY

Purdue Data Mine Team, Abbvie

Business Problem

The challenges of efficiently utilizing marketing channels to maximize conversion rate from each touchpoint

Project Goal

To construct a predictive model utilizing XGBoost to identifies the optimal sequence of marketing channels and delve into the customer journey to comprehensively evaluate and refine marketing strategies.

Analytical Problem

The depth of the business problem highlights the need for a data- driven approach to evaluate customer journey and optimize marketing efforts.

- Conversion Rate
- Multi touch Attribution Model
- Key Variables
- Strategy Optimization.

Literature Review

Multi Touch Attribution Model	XGBoost Model
Additional Multi-Touch Attribution for Online Advertising | Proceedings of the AAAI Conference on Artificial Intelligence, Ji, W., & Wang, X. (2017).	**Simple Exploration+Baseline - GA Customer Revenue, Google Analytics Customer Revenue Prediction, Kaggle Competiton, SRK (2018)**
Data-driven Multi-touch Attribution Models, Xuhui Shao & Lexin Li (2011)	

We utilized the MTA model, which integrates machine learning and probabilistic methods for comprehensive analysis and predictions. Additionally, we explore data processing and feature engineering techniques to gain further insights.

Exploratory Data Analysis

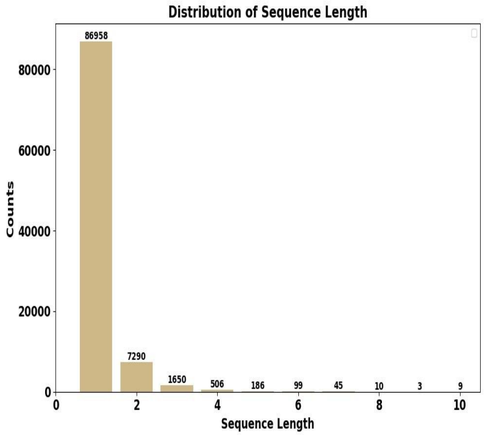


Figure 1 shows the lengths of these sequences and their quantities, including both successful and unsuccessful outcomes. Sequences longer than 6 are uncommon. Therefore, we set 6 as the maximum length to make processing smoother.

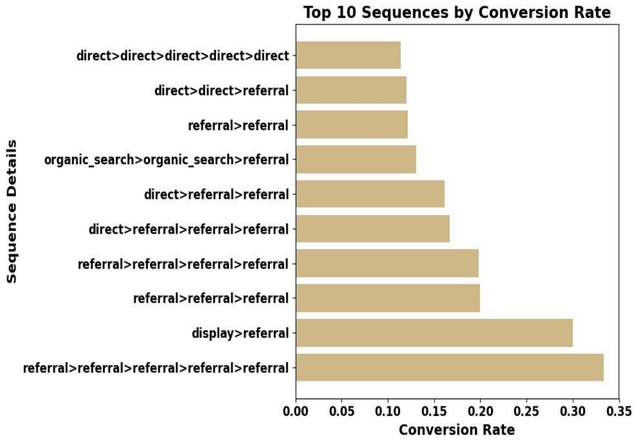


Figure 2 reveals that when we remove unique touchpoint combinations and sequences under 6, referrals stand out as a key element for conversion. Specifically, sequences ending with a referral touchpoint often have a higher rate of conversion.

Methodology / Model Building Process

EDA

- Gain deep insights into the dataset, identifying significant trends, patterns, and potential relationships between variables.
- Understand existing conversion rates of each channel - customer journey to understand marketing tactics effectiveness.

Data Preparation

- Build a *data dictionary* to identify and understand the key variables of interest (channels and additional control variables).
- Create a *dataframe* with information that determines the distance of each element in the sequence from the reference point
  - The number in position means the distance from the conversion (i.e., the larger the number, the further away from conversion.)
  - The dataframe has total 54 columns, including 8 channels × 5 positions and 14 additional variables.

Take the last 5 touchpoints

Ex: ***direct > referral > referral > direct > referral*** with additional information

Given a sequence, convert it into the data frame by its distance from the reference (conversion)

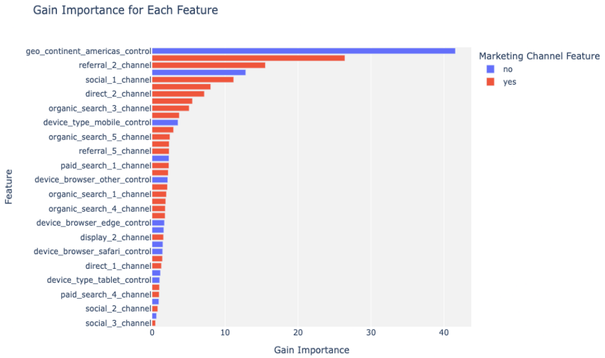
	channels (40)																other variables (14)		
Variables	Referral					Direct					Social	Organic Search	Paid search	Display	Affiliates	Other	geo	device	etc.
Position	1	2	3	4	5	1	2	3	4	5	1, …, 5	1, …, 5	1, …, 5	1, …, 5	1, …, 5	1, …, 5			
Data	1	0	1	1	0	0	1	0	0	1	…	…	…	…	…	…	1	0	…

Modeling

Use XGBoost algorithm to predict conversion probabilities or attribution weights for each tactics in a customer journey.

Data Insights

Gain importance measures the relative contribution of each feature to the model's predictions by calculating the average gain of splits which use the feature. It signifies how much a feature improves the model's ability to make accurate predictions, with higher values indicating more important features in predicting the target variable.



Outcome

Applying certain variables like social_1, referal_2, direct_2 can increase conversion rate. And whichever has the highest impact, more resources could be deploy to that variable to help enhance customer's journey.