

Data and analytics case study

Background

- The Data and Analytics team has been asked to evaluate the recent revenue performance for a health insurance product. We are particularly interested in how long customers are keeping these insurance policies

Data description

- The data consists of transactions from 7/1/2018 to 6/7/2019
- It includes:
 - Customer ID
 - Membership Start Date – start of policy
 - Transaction Date – date of payment
 - Payment Type (Annual Fee or Monthly Fee)
 - Revenue – amount of payment received

Questions to be answered

1. Which month had the highest revenue (by "Transaction Date")? What was the percentage of revenue for that month by payment type?
2. How many members started their memberships in December 2018 (by "Membership Start Date")?
3. On average, for how many months did members who started their memberships in July 2018 continue to pay the monthly fee?
4. Assume that transactions for June 2019 are incomplete and that transactions beyond June 2019 are unobserved. What type of statistical or machine learning modeling technique(s) would you use to build a predictive model of customer retention?
5. How many customers paid the Monthly Fee in May 2019? Predict the Membership End Date for these customers.

Submission instructions

- Please place all of the files used in your analysis in a zip file and return to your recruiting contact

Note: The data file (dataset.csv) and this instruction page are property of GoHealth, LLC. Please do not distribute.