Customer Churn in the airline industry

Introduction

BigNorth Airlines is one of the top four airlines in the United States. Like the other large airlines in the U.S., customers buying a BigNorth plane ticket fly on BigNorth Airlines primary routes as well as on BigNorth's regional partner airlines. Regional airlines act as feeder airlines to major airlines by connecting smaller airports to the airline's main hubs. Hub airports are always located in major cities, whereas the regionals serve smaller cities and rural areas.

Organizational Structure

BigNorth's organizational structure is fairly simple with a sales & marketing group, a group in charge of partner airlines, an IT group as well as a fairly new data science team. The head of the data science team currently reports to the CEO, in part because the CEO wanted to make sure the data science grew to critical mass. As a result of this structure, the data science team mainly works on projects that the CEO thinks might have value and typically doesn't get input from the sales and marketing team. When needed, the IT team is used to deploy the data science models. Occasionally, VPs from other groups ask the data science team to do various, fairly self-contained analyses or projects.

BigNorth and its Regional Airline Partners

Like other airlines, BigNorth contracts out to regional carriers because it allows them to lower their risks related to capacity and pricing. Specifically, regional airline contracts last for a number of years, after which, BigNorth can renegotiate to adjust (up or down) the number of flights provided by that partner. This enables BigNorth to more easily reflect their current market conditions. It is possible, for example, that if demand falls, BigNorth would not renew some of their regional contracts. On the other hand, if demand rises, BigNorth can expand their contract, and bring more planes into service more quickly than they could on their own.

The Data Science Team

The team consists of 12 people. In addition to the head of the data science team, there are three team leads and rest are team members, with skills ranging from data munging to model building. The team uses a shared project management tool where a project is broken into tasks and assigned to individuals. These tasks are tracked via "task cards".   The team meets once a week to discuss status, where for example, it is noted which task cards are done, and which new cards should be assigned. The new task cards are created by one of the team leads (TLs). Cards are assigned to people, by a TL, based on their skill set of the team members. So, a data munging task would go to a data engineer, etc. If someone finishes a task card early they would ask the TL what they should work on next.

The team does not have its own QA or deployment personnel, so instead when a project is ready for QA they send it to the IT department for QA review, approval, and deployment. Once every month the TL discusses the status of each project with the main stakeholder of that project, and where they share any insights that the team has generated that are actionable.

Current organizational situation

A consultant recently conducted interviews across the organization, and several items were noted. First, the CEO thinks that the DS team is not productive. The head of the DS group pushed back on this claim, noting that DS team members were working 80+ hour weeks to meet some of the key project milestones. A different challenge was noted by the Head of Sales and Marketing, who didn't trust some of the model predictions, and thought that some of other models were simplistic and didn't add any incremental value. Another challenge noted by the Head of Sales and Marketing was that the insights sometimes were not useful. For example, the DS team recently built a model to predict if customers would be unhappy if they were charged for soft drinks. But the model wasn't useful, since there was no way to

know if their unhappiness would translate into a loss of overall customer satisfaction and ultimately, to some customers choosing other airlines.

The head of the data science team was also frustrated, in that, it was difficult to get resources from other groups to help with their projects (e.g., domain experts from sales, or folks from IT to help think through how to put a model into production). Another challenge was that everyone wanted the DS team to provide an ROI (project cost, time to implement and expected benefits). This was requested so any effort (e.g., adding additional data fields, such as loyalty) could be prioritized with other (non data science) efforts. However, the challenge was that many of the data science efforts were exploratory in nature, and hence, it was often difficult to commit to specific incremental value for the data science efforts.

New Project – Reducing Churn

The CEO reached out to the data science team to start a project to lower customer churn (sometimes referred to as customer attrition). In fact, the CEO suggested that this new effort should be started right away, as customer churn was a key issue he wanted the airline to better understand and manage.

Like many airlines, BigNorth has, up until now, believed that the best way to minimize customer churn was to have a robust loyalty program for frequent flyers. The basics of BigNorth's loyalty program is similar to other airlines. In short, the airline rewards repeat business and as a customer flies more, they rise through levels of service and "bank" miles that can be redeemed for free or discounted travel. There is no data to back up this line of thinking -- it is just the "accepted industry best practice". However, customers are starting to value the loyalty program less, which is one reason why just relying on their loyalty program might not be sufficient in lowering customer churn. In fact, frequent flier points are slowly devaluing, while the overall balance of the miles banked by customers is increasing.

The CEO pointed out that customer churn is actually a lagging indicator, meaning the loss has already occurred. As such, it is a measurement of the damage inflicted. The real goal is to reduce churn by getting ahead of the loss (of the customer) by identifying some leading indicators, or metrics, that might help keep a customer. In other words, these leading indicators, or metrics, could help identify when a customer was about to stop flying BigNorth. These insights could provide actionable suggestions as to how to avoid having the customers leave and go to another airline.

Net Promoter Score (NPS)

NPS surveys ask customers to respond, on a scale of 1–10, to one simple question: "How likely is it that you will recommend our airline to a friend or colleague?". If respondents score less than 7, they're detractors. If they scored above an 8, they're promoters. In the middle range (a score of 7 or 8), then they're "passive". In a given group, subtracting the percent of respondents who are detractors from the percent of respondents who were promoters provides the overall NPS score. The concept of NPS is that customers who are promoters are good customers to keep. Such customers may sometimes even provide free "word of mouth" advertising. Customers who are detractors are problematic in that they may actively tell their social connections not to use the product or service (i.e., they would tell people not to fly BigNorth). It has often been suggested that NPS provides a good proxy for understanding how likely a customer is to churn. For example, according to one source, NPS is nearly three times more sensitive at predicting customer churn than customer satisfaction. In addition, detractors are 1.5 times more likely to stop using a service as compared to promoters. In short, analytics could be a key to the success of BigNorth.

Current NPS Practices

BigNorth has been using the surveys to calculate NPS. They would increase their focus on providing good customer service when their NPS score went down. This is currently typically done via a memo to customer facing staff, where they are encouraged to "smile more". Note that NPS is not currently used as part of BigNorth's partner airline strategy (i.e., it is not part of the data BigNorth used to help decide which partners to keep, which partners to drop and which regional airlines should become new partners).

The Data Available

BigNorth often surveyed their customers, and in fact, possesses thousands of recently completed customer surveys. The survey dataset contains thousands of observations of flight segment data collected by BigNorth Airlines. Each survey contains a simple rating of each customer's likelihood to recommend

the airline that they just flew, the flight number, the date, the customer frequent flyer number, as well as a field for open-ended text comments. Below are the field that have been collected:

Attributes Name:
1. **Likelihood to Recommend** – rated on a scale of 1 to 10, which shows how likely the customer is to recommend the airline to their friends (10 is very likely, and 1 is not very likely).
2. **Airline Flyer Status** – each customer has a different type of airline status or package, which are platinum, gold, silver, and blue.
3. **Age** – the specific customer's age. That is starting from 15 to 85 years old.
4. **Gender** – male or female.
5. **Price Sensitivity** – the grade to which the price affects to customers purchasing. The price sensitivity has a range from 0 to 5.
6. **Year of First Flight** – this attribute shows the first flight of each single customer.
7. **Flights Per Year** – The number of flights that each customer has taken in the most recent 12 months. The range starting from 0 to 100.
8. **Loyalty** – An index of loyalty ranging from -1 to 1 that reflects the proportion of flights taken on other airlines versus flights taken on this airline. A higher index means more loyalty.
9. **Type of Travel** – is provide three traveling purpose for each consumer, which are business travel, mileage tickets that based on loyalty card, and personal travel like to see the family or in vacation
10. **Total Frequent Flyer Accounts** – How many frequent flyer accounts the customer has.
11. **Shopping Amount at Airport** – The spending in dollars on non-food/drink goods and services at the airport(s) where the customer was before, between, or after flights.
12. **Eating and Drinking at Airport** – The spending in dollars on food/drink goods and services at the airport(s) where the customer was before, between, or after flights.
13. **Class** – it consisted of three different kinds of service level such as, business, and economy plus, economy. Moreover, customers have optional to choose their seat.
14. **Day of Month** – it means the traveling day of each costumer. In this attribute, shows total of 31 days of the month.
15. **Flight date** – the passenger's flight date travel.
16. **Partner Code** – This airline works with wholly- and partially-owned subsidiary companies to deliver regional flights. For example, AA, AS, B6, and DL.
17. **Partner Name** – These are the full names of the partner airline companies.
18. **Origin City** – refers to actual city that customers have departed from. For example, Yuma AZ, Waco TX, and Toledo HO.
19. **Origin State** – same thing as origin city such as, what state that customers have departed from? A good example, Texas, Ohio, Alaska, and Utah.
20. **Destination City** – the place to which passenger travels to. For example, Akron HO, Alpena MI, Austin TX, and Boston MA.
21. **Destination State** – also, it is the same thing as origin city, such as, to what state passenger travel to? Some example of destination states, Alaska, Kentucky, Iowa, and Florida.
22. **Scheduled Departure Hour** – the specific time at which passengers are scheduled to depart. In this data in scheduled departure hour is starting at 1 am until 23 pm.
23. **Departure Delay in Minutes** – which are minutes of departure delayed for each passenger, when compared to schedule. In this data the rage are starting from 0 until 1128 minutes.
24. **Arrival Delay in Minutes** – how many minutes of arrival delayed of each passenger. Rang of delayed minutes in this data are starting from 0 until 1115 minutes.
25. **Flight Cancelled** – occurs when the airline dose not operates the flight at all, and that is for a certain reason.
26. **Flight time in minutes** – indicate to period time to the destination.
27. **Flight Distance** – the extent of space between two places. Also, that means how many minutes are passenger traveling between two different places. Rang in this data starting from 31 until 4983 minutes.

Other Projects

Beyond this newly suggested project, the team has five other projects currently in progress, at various stages of completion. Three of these projects were requested from the CEO and two were requested by the sales and marketing team. The CEO is only vaguely aware of the sales and marketing projects and doesn't know the details of any of these efforts.