Kellogg School of Management
Northwestern University

Intuit QuickBooks Upgrade: Moving to the Cloud

Quickbooks quickly became the mainstay of small business accounting software when it was introduced in 1992 by Intuit. It had grown quickly—to about 800,000 users after only 3 years. Since then Quickbooks had been continuously updated, most recently to Quickbooks 2015, and was being used by 4.8 million customers globally. Together with Quicken and Turbotax, QuickBooks embodied the Intuit success story.

While the success of Intuit was originally based on desktop software, Salesforce.com and other companies had shown the potential of software as a service (SaaS). This led Intuit to roll out "Quickbooks Online" in beta in July, 2012 in a move hailed as both important and strategic:

> "We view Intuit as a company still on a journey transforming itself from a good business to a great business," wrote Cowen analysts Peter Goldmacher and Joe del Callar, in a research note dated Sept. 19, 2012. "The decision to transition the model from desktop software to connected services is as profound a transition as we've seen in the software space. The data the company is now able to capture about who its customers are and how they behave is its core asset and should be the driving force behind its growth and margins for the next decade."[1]

Inuit charged a monthly subscription fee to use Quickbooks Online. To serve a wide variety of customers, Intuit offered QuickBooks Online in multiple versions with varying functionality, from "Simple Start" for $15 per month to "Plus" for $40 per month.

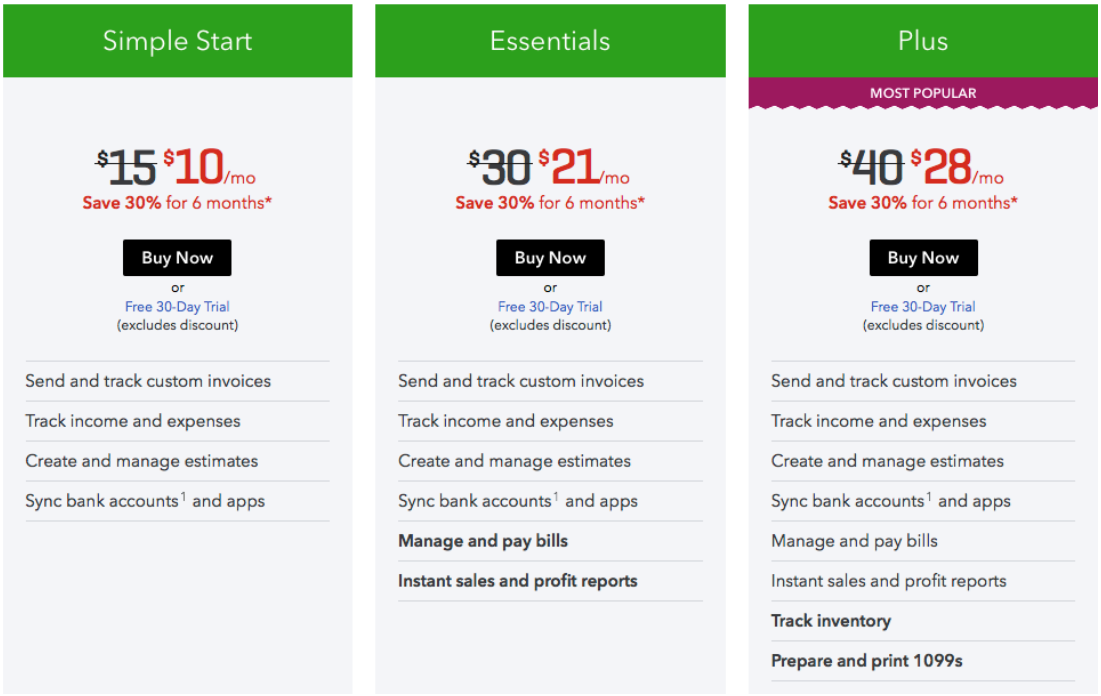
The Campaign

While Quickbooks Online had exceeded the company's expectations, there was still a sizeable segment of desktop users who had not moved online. Lucy Chen, marketing lead for Quickbooks, understood that there were some long-term Quickbooks customers who would never be ready to move to the cloud. However, she was puzzled by the number of recent customers of QuickBooks who still chose the desktop version. Could they have been unaware of QuickBooks Online when they purchased?

[1] http://www.zdnet.com/article/intuits-secret-to-going-global-lies-in-thinking-local/

Professor Florian Zettelmeyer prepared this case to provide material for class discussion rather than to illustrate either effective or ineffective handling of a business situation. The data used in the case were either publicly available or were simulated. Events and names have been fictionalized for pedagogical purposes.
This version: February 10, 2019

Lucy decided to fund a multi-channel campaign to encourage recent desktop customers to switch. Her idea was to target customers who had purchased the desktop version of Quickbooks during the last 36 months. This included customers who had purchased QuickBooks 2013 and QuickBooks 2015, the two most recent desktop versions. The core of the offer was a 6-month 33% discount on the subscription fee.



Intuit had determined that the profit of switching a customer from the desktop to the online version was on average $234 over the lifetime of the customer. However, this profit would have to be adjusted by the $54 average discount in the offer, resulting in a net benefit of $180 if a customer moved to the cloud as a result of the offer. The cost of making the offer was $1.60 because the offer included a direct mail piece. To stay within her budget, Lucy picked --- among customers who had purchased QuickBooks desktop during the last 36 months --- 50,000 at random to receive the offer.

The campaign was a big financial success. Out of the 50,000 targeted customers, 1,950 converted to Quickbooks Online, yielding an estimated profit of $ 271,000, or $5.42 per targeted customer.

Wave 2

Nine month later, Lucy received an e-mail from her boss Linnea Hennig.

> “Lucy – just closing out the budget year and noticed that we have extra funds available I need to spend asap. I can give you budget to target another set of customers with your Quickbooks online offer. Here is the catch: I have another team promising me $5.60 for every customer I allow them to target. I don’t want to spam customers with two offers back to back – so here is my suggestion: You pick which customers in the attached list (of 25,000

> customers) you want to target but you have to promise me that you will, on average, beat the $5.60 number. I will assign the remaining customers to the other team. Want to pick customers and try?
> --- Linnea."

Lucy decided to talk to Paul Heron, analytics lead for Quickbooks, about Linnea's offer. Paul shared his view:

> "Looking at the economics, I see that you made profits of $5.42 per customer during your prior campaign, let's call it "wave 1." Linnea is now asking you to beat $5.60 profit for every customer that you choose to target (since the other team is willing to pay that much for any customer they get from Linnea). This is bad news because it means that targeting all 25,000 consumers will not work, since the wave 1 results suggest you won't get to $5.60.
>
> However, there is also good news: Because you ran wave 1, we can use analytics on the responses to wave 1 to determine which type of customer is the most likely to accept an offer in wave 2. This way you can make wave 2 offers only to those customers for whom you expect to expect to make at least $5.60 in profit after your $1.60 mailing cost.
>
> If you like, I can analyze the data from wave 1, build some predictive models, and tell you which of the 25,000 customers in the wave 2 list you should pick. Also, I can give you an estimate of the profit that you can expect to make."

The Task

Put yourself in Paul's shoes. You have been provided with the customer data from Lucy's original mailing, including their response to the wave 1 offer.

Your job is to:

1. use this data to **build a model** that predicts how likely consumers respond to the Quickbooks Online offer,
2. **score** the 25,000 consumers on the wave 2 list and pick the subset that should be targeted with the Quickbooks Online offer, and
3. **forecast** how much profit (total and per targeted customer) Lucy can expect to generate by doing so.

The Data

The "intuit_online.RData" file contains two datasets.

1. `intuit` → The data from 50,000 customers in wave 1
2. `intuit.wave2` → List of 25,000 customers for possible inclusion in wave 2

The variable "res" in `intuit` denotes which of these customers responded to the offer. The remaining variables are available to help you predict who will respond to a wave 2 offer.

id	Customer ID

state	US State of customer
zip	5-Digit ZIP Code
speeddown	Average download broadband internet speed in ZIP code
speedup	Average upload broadband internet speed in ZIP code
last	Time (in months) since last order from intuitmarket.intuit.com[2] in last 36 months
numords	Number of orders from intuitmarket.intuit.com in last 36 months
dollars	Total $ ordered from intuitmarket.intuit.com in last 36 months
sincepurch	Time (in months) since first purchase of Quickbooks
version2013	Is 1 if customer's most recent version is Quickbooks 2013 (0 if the most recent version is Quickbooks 2015)
upgraded	Is 1 if customer upgraded from Quickbooks 2013 to 2015
payroll	Is 1 if customer has payroll add-on, 0 otherwise
bizflag	Is 1 if address contains a business name, 0 otherwise
sex	1=Male, 2=Female, 3=Unknown
income	Median household income (100k dollars) in ZIP code
medhvalue	Median value (100k dollars) for all owner-occupied housing units in ZIP code
res	Response to wave 1 offer (1=responded, 0=did not respond)
training	A dummy variable to create a intuit.train (70%) and intuit.test (30%) datasets

Assignment Instructions

1. Decide which of the 25,000 individuals in the `intuit.wave2` dataset you want to pick for wave 2. 1/3 of your grade (10 points) on this assignment will be based on the profit (not ROI) achieved on your offer by comparing your prediction with the actual findings from wave 2 (which are not included in your dataset but I have).

2. Two thirds of your grade (20 points) for this assignment will be based on write-up (R Notebook in both "pdf" and "nb.html") of your modelling approach.

   Answer the following questions in your R Notebook:

   - What is your final predictive model, how did you choose it, and what is its expected predictive performance (in terms of AUC, gains, etc.)? If you created new variables to include into the predictive model, please describe these as well.
   - How did you decide who should receive the wave 2 offer?
   - How much profit (both total and per targeted customer) should Lucy expect with her wave 2 offer?
   - What did you learn about what predicts whether consumers are likely to upgrade?

3. By 4 p.m. on the day before class please send me by e-mail a dataset containing the IDs of all 25,000 customers in `intuit.wave2`, the predicted scores of your best model, and dummy variable that tags all customer you want to target (**there is no need to send the assignment write-up early**).
   The reason for this is I want to compile all the results before class and ask some groups during class to discuss what they did. Thus, in the evening before class I will need to compile about 45 different data files, line up all the IDs, and evaluate the results. This takes

[2] intuitmarket.intuit.com is an online store in which Quickbooks customers can order checks, tax forms, envelopes, deposit slips, and other supplies.

quite a bit of time and you can help me a lot if you stick to the following instructions **exactly**!

Create the dataset:

- Please create a dataframe that contain 5 variables (columns).
  Make sure that all the variable names are lowercase.

  column 1: **`id`** → the original "id" variable from the `intuit.wave2` dataframe.

  column 2: **`best.score`** → the score (prediction) from your final (best) model

  column 3: **`target.wave2`** → a variable that contains a "1" if you want to target the customer in wave 2 and "0" if you don't

  column 4: **`group`** → a variable that contains in every row the **first** names of **all** your group members separated by underscores "_" (e.g. Nancy_Sam_Manuel). You create this variable by using `mutate(group = "Nancy_Sam_Manuel")`

  column 5: **`section`** → a variable that contains your section number. You create this variable by typing, for example: `mutate(section = 31)`. Please make sure this variable is numeric, not a string (i.e. **don't** use `section = "31"`, notice the quotes). The 8:30 section is section 31, the 10:30 section is section 32.

Save the dataset:

- For illustration, suppose your dataframe is called `intuit.wave2.submit`
- Please use the `saveRDS()` command to save your dataframe under the same name as you used to fill in the group variable in the dataset (e.g. Nancy_Sam_Manuel).
  For example:

```
saveRDS(intuit.wave2.submit, "Nancy_Sam_Manuel.RDS")
```

- Please notice the ".RDS" extension, it is important!
- The ".RDS" file will be in your project directory.

Send me an e-mail:

Please include in the e-mail:

- The first and last names of all of your group members
- Your section number, e.g. 31, 32, etc.
- Put in the subject line of the e-mail "Intuit:" followed by the same name as group variable (e.g. "Intuit: Nancy_Sam_Manuel ")
- Don't forget to attach the ".RDS" file with the data.
- Sorry to be so particular about this, but you help me out a lot if you follow these instructions!

Hints

1. You are welcome to use any model we have discussed in class, i.e. **RFM**, **logistic regression**, **neural networks**, or **random forests**.

2. Neural networks and random forests will require tuning to find the hyperparameters that allow the model to predict well.

    a. For **neural networks** you need to set the "size" parameter (the number of hidden nodes). Reasonable values for the number of variables we have is from 1 to 10. You also need to set the "decay" parameter (it controls how much large weights get penalized). Reasonable values are 0.001, 0.01, and 0.1. In theory, the higher the decay parameter, the less prone the neural network is to overfitting at the expense of not fitting as well in training.

    b. For **random forests** you need to set the "mtry" parameter (the number of randomly selected variables that are considered at each split). The suggestion by the inventors of this machine learning algorithm is to pick the square root of the number of variables (rounded up). So, for 9 variables, mtry = 3. You might want to explore decreasing or increasing this number. In theory, the lower the mtry parameter, the less prone the random forest is to overfitting at the expense of not fitting as well in training.

3. Following Data Science best practice, you should train all models in a training sample and evaluate their performance (by looking at the gains-curve and/or AUC) in a test sample. The full "intuit" dataframe contains a "training" variable you should use to create "intuit.train" and "intuit.test" dataframes. Please see the handout "Class 11: How to Evaluate Models using Gainsplot on Test samples" for how to evaluate performance for different types of predictive models.

4. If you want to generate new variables, I suggest you do this on the full "intuit" dataset before the split into train/test.

5. If one model has a higher gains-curve it will generally have a higher AUC, unless the gains curves cross. If you encounter crossing gains curves, compare expected profits.

6. Once you have picked your favorite model based on the performance in the test sample, best practice is to <u>retrain</u> that model using the <u>full</u> "intuit" dataframe to utilize as much data as possible. Use this final trained model to score your wave 2 data.

7. If you have created new variables in the "intuit" dataframe, make sure to also create them in the "intuit.wave2" dataframe before you score the wave 2 data. Otherwise the prediction will fail.

8. Contact me if you have problems or questions!