Full Code

Xandre Clementsmith and Jacob Mattox

10/15/2017

Using Customer Data to Individualize Credit Card Rewards

The customer data are displayed as spending in different categories over a six month period in both categorical and histogram form (uses masked_ID 1 initially, but can be changed to show any individual or group as needed).

```
library(readxl) # loads data sheets for use with 'read_excel' and assigns them to variables
library(ggplot2) #loads the plot library used for bar graph
library(randomForest)#loads library for regression
```

```
## randomForest 4.6-10
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(dplyr)#loads library for displaying user interaction with Wells Fargo
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:randomForest':
##
##     combine
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#initializing variables for use in code
category = "AUTO / GAS" #used for default category (can be changed to alter data outputs)
userID = 1 #change this for desired masked Id
keywordForSearch = "Credit" #keyword used to find web traffic

month_end_bal <- read_excel("~/Fake+Data+and+Metadata+-+Final+no+pass.xlsx", #change the directo
ry location as needed
    sheet = "Month end balances ")
daily_use_cc <- read_excel("~/Fake+Data+and+Metadata+-+Final+no+pass.xlsx",
    sheet = "Daily use of a WF credit card ")
web_traffic <- read_excel("~/Fake+Data+and+Metadata+-+Final+no+pass.xlsx",
    sheet = "Daily WellsFargo.com traffic")
```

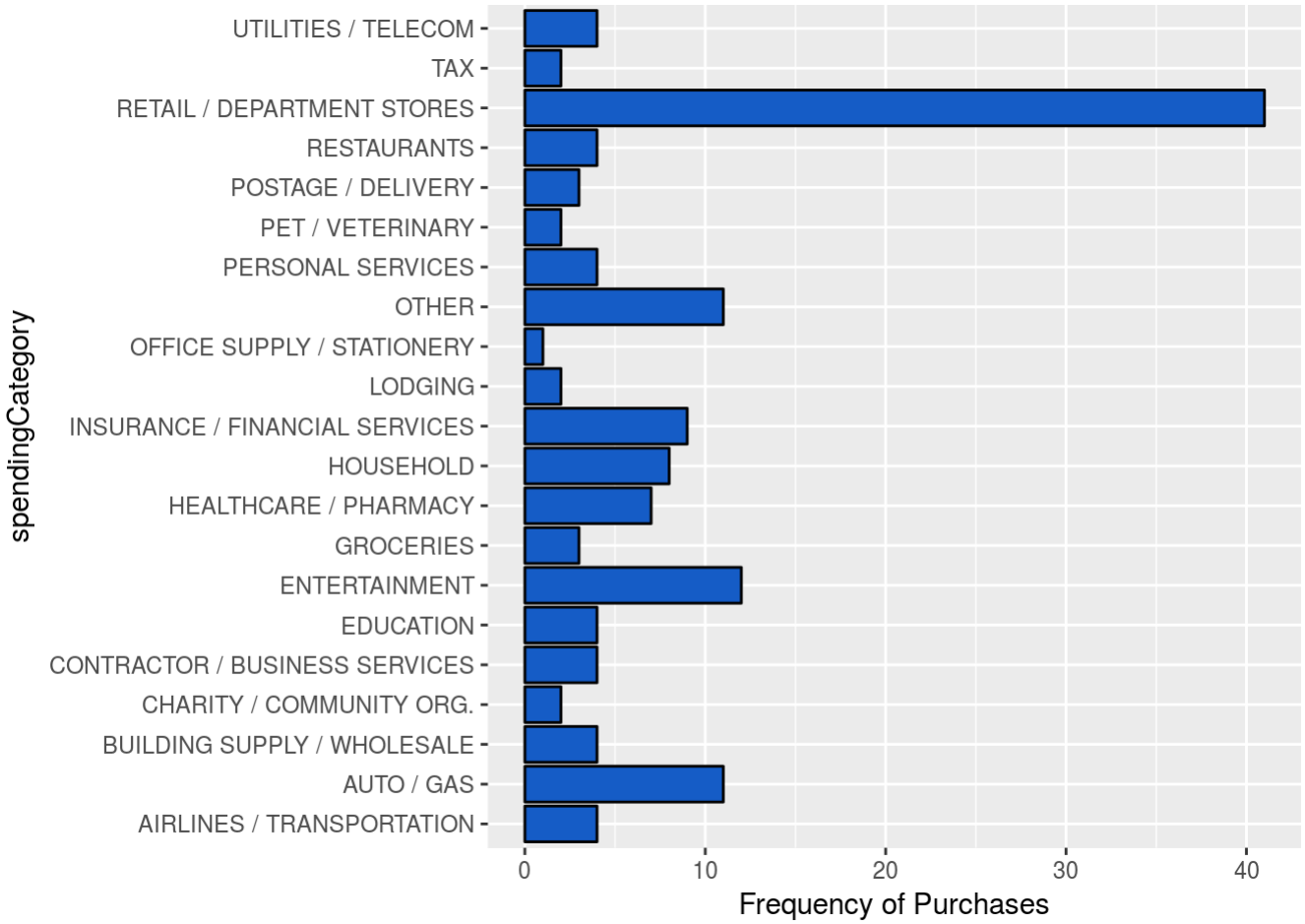
```
for(i in userID) { #if a range is desired use 1:5 notation in place of userID
  #id <- paste("id", i, sep = "") # if you have multiple users' data being shown
  if (i %in% daily_use_cc$masked_id) { # checks to see if user has credit card data
    spendingCategory <- subset(daily_use_cc, # creates a subset with Des2 based on a given maske
d_id
                           masked_id==paste(i),
                           select = Des2)
    dataframe_ft <- as.data.frame(table(spendingCategory)) # altering the form of the subset int
o a dataframe

    print(dataframe_ft) # displays the user's spending in a numerical format

    plt <- ggplot(data = dataframe_ft, aes(x = spendingCategory, y = Freq)) +
      geom_bar(stat = "identity", colour = "black", fill="#155cc6") + coord_flip() + # code for
 creating the graphic
      xlab("spendingCategory") + ylab("Frequency of Purchases")

    print(plt)
  }
}
```

```
##                  spendingCategory Freq
## 1       AIRLINES / TRANSPORTATION    4
## 2                       AUTO / GAS   11
## 3     BUILDING SUPPLY / WHOLESALE    4
## 4        CHARITY / COMMUNITY ORG.    2
## 5  CONTRACTOR / BUSINESS SERVICES    4
## 6                       EDUCATION    4
## 7                   ENTERTAINMENT   12
## 8                       GROCERIES    3
## 9           HEALTHCARE / PHARMACY    7
## 10                      HOUSEHOLD    8
## 11 INSURANCE / FINANCIAL SERVICES    9
## 12                        LODGING    2
## 13     OFFICE SUPPLY / STATIONERY    1
## 14                          OTHER   11
## 15              PERSONAL SERVICES    4
## 16               PET / VETERINARY    2
## 17             POSTAGE / DELIVERY    3
## 18                     RESTAURANTS    4
## 19     RETAIL / DEPARTMENT STORES   41
## 20                            TAX    2
## 21            UTILITIES / TELECOM    4
```



```
# Creates an aggregate of month_end_balances by taking the mean of every value
month_end_bal_agg <- aggregate(. ~ masked_id, data = as.data.frame(month_end_bal), FUN = mean)

# Merges the two data sheets together based on masked_id
merged_data <- merge(as.data.frame(daily_use_cc), month_end_bal_agg, by = "masked_id")
```

Determining the Quantity Spent on a Given Category

This segment allows another way for the user to see the spending in an individual category (Auto/Gas used as default) for an inputted masked_ID.

```
# the following is a function to determine the quantity spend on a given category for a given ma
sked_id
# the default is currently listed as AUTO / GAS

quan_spent <- function(id, cat = category) { #uses category variable set in beginning of code
  if (id %in% daily_use_cc$masked_id) { # checks to see if the masked_id has credit card usage d
ata
    id_merged_data <- merged_data[merged_data$masked_id == id,] # creates a subset based on the
 given id

    id_fact_des2 <- factor(id_merged_data$Des2) # factors the categories based on the subset

    fact_des2_df <- as.data.frame(table(id_fact_des2))# changes the data from a subset into a da
taframe
    fact_des2_df[fact_des2_df$id_fact_des2 == cat, 2]

  }
  else{
    print("No credit card data for user")#catches users without credit cards
  }
}
print(paste("Number of times masked Id ", userID, " spent money on", category,": ", quan_spent(u
serID)))
```

```
## [1] "Number of times masked Id  1  spent money on AUTO / GAS :  11"
```

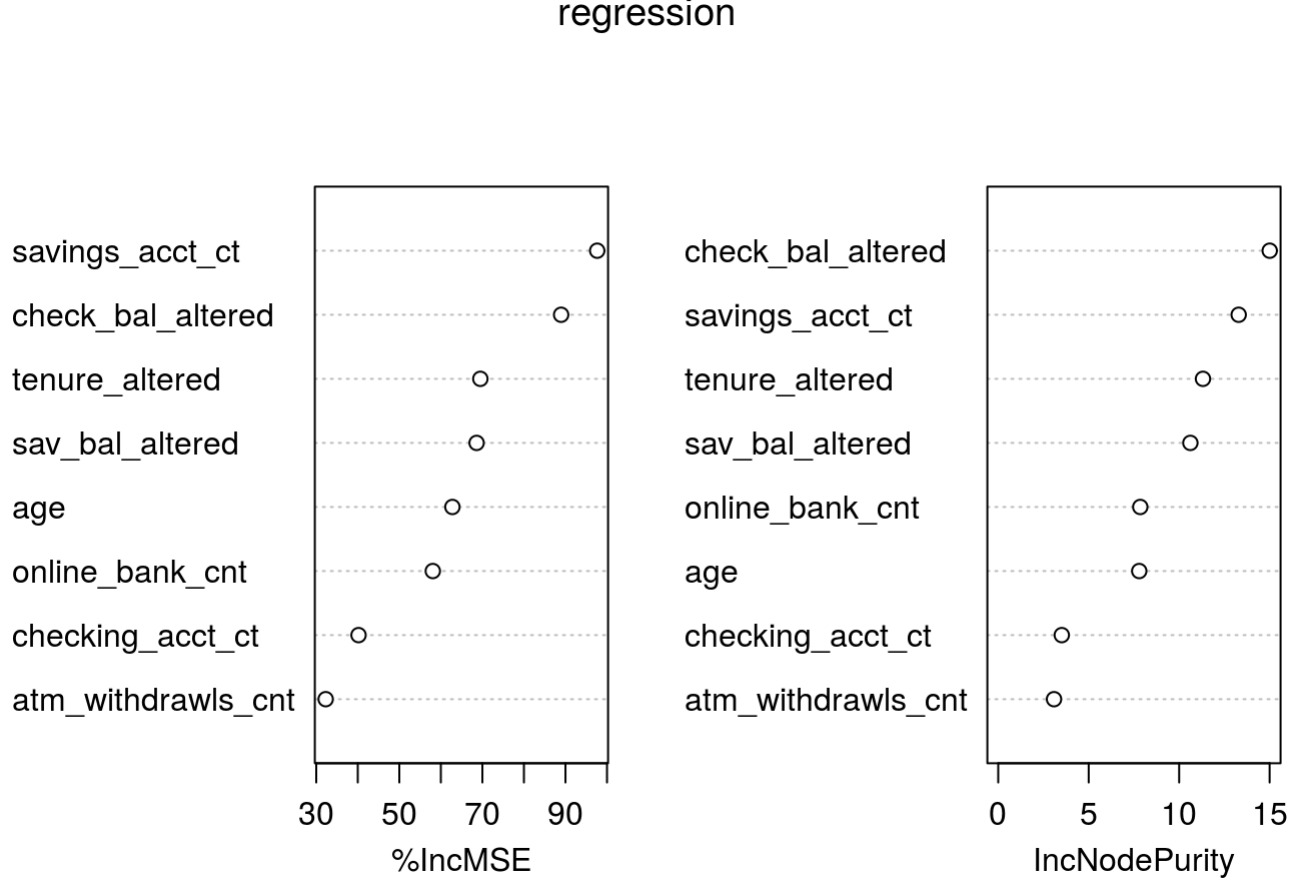
```
 # change this number based on what masked_id you're using

#This gets a vector of Auto/Gas for statistical analysis
# most of this is the code from above looped to get a vector
vec <- c()
for (id in 1:50) { # goes through every id up to masked_id 50
  id_merged_data <- merged_data[merged_data$masked_id == id,] # creates a subset for each singul
ar masked_id
  id_fact_des2 <- factor(id_merged_data$Des2) # factors the subset
  fact_des2_df <- as.data.frame(table(id_fact_des2)) # changes the factor into a dataframe
  value <- fact_des2_df[fact_des2_df$id_fact_des2 == category, 2] # change AUTO / GAS to your de
sired category
  vec <- c(vec, value) # adds each quantity spent to the final vector (list) for all users (as s
hown below)
}
```

Using Regression to Find Identifying Factors of Credit Card Usage

Shows an output of important identifying factors for determining with 89.75% accuracy whether someone would be a good fit for a credit card.

```
regression <- suppressWarnings(randomForest(cc_flag ~ savings_acct_ct + checking_acct_ct + age +
 atm_withdrawls_cnt + online_bank_cnt + tenure_altered + check_bal_altered + sav_bal_altered,
                      data=month_end_bal,
                      importance=TRUE,
                      ntree=2000))
varImpPlot(regression)
```




```
print(regression)
```

```
##
## Call:
##  randomForest(formula = cc_flag ~ savings_acct_ct + checking_acct_ct +      age + atm_withdra
wls_cnt + online_bank_cnt + tenure_altered +      check_bal_altered + sav_bal_altered, data = mo
nth_end_bal,      importance = TRUE, ntree = 2000)
##                Type of random forest: regression
##                      Number of trees: 2000
## No. of variables tried at each split: 2
##
##           Mean of squared residuals: 0.02576156
##                     % Var explained: 89.69
```

Show Which Users Spend More Than Average (and what do they spend it on)

Using a previously created vector for all spending in Auto/Gas we determine who spends more than average in the chosen category. This makes for an excellent target for credit card rewards.

```
mean <- mean(vec) # determines the mean of the vector
sd <- sd(vec) # determines the standard deviation of the vector

# this creates a vector with all user_ids who spend more than one standard deviation than the me
an
sig <- c()
for (i in vec) {
  if (i >= mean + sd) {
    sig = c(sig, i)
  }
}

# this for loop goes through several if statements to determine if the user has credit card info
rmation and spends
# more often than one standard deviation above the mean
for (i in userID) {
  if (i %in% daily_use_cc$masked_id) {
    if (quan_spent(i) %in% sig) {
      # Change the string here from AUTO / GAS to whatever category you're looking for data abou
t
      print(paste("Id", i, " spends more often in AUTO / GAS than the majority of users.", sep =
 ""))
    }
  else print("The user does not spend significantly more in AUTO / GAS.")
  }
}
```

```
## [1] "Id1 spends more often in AUTO / GAS than the majority of users."
```

Searching Web Data for a Keyword

In this last code, we're searching their web data for the keyword: "Credit". Glancing at this output can put perspective on interests and future spending habits.

```
for (i in userID) {#change user to a range using 1:5 notation
  wtm <- web_traffic$masked_id == i
  if (exists("wtm")) { # determines if there's web traffic data for the user
    credit_subset <- subset(web_traffic, masked_id == i, select = wf_page) # creates a subset fo
r masked_id
    credit_df <- credit_subset[1,1] # further narrows the data to only the web_traffic of the gi
ven id

    # this code determines if the word "Credit" appears in any of the user's web traffic data
    # it creates a dataframe to hold all of these results
    for (i in userID:nrow(credit_subset)) {#change userID to a range using 1:5 notation
      if (grepl(keywordForSearch, credit_subset[i,])) {
        credit_df <- rbind(credit_df, credit_subset[i,])
      }
    }

    # finally, we print the dataframe to see all of their search history regarding "Credit"
    credit_df <- credit_df[2:nrow(credit_df),]
    print(credit_df)
  }
  else {
    print("There is no web data for this user.")
  }
}
```

```
## Source: local data frame [35 x 1]
##
##                                                                  wf_page
##                                                                    (chr)
## 1                Financial Education/You and Your Family/Credit Management
## 2                Financial Education/You and Your Family/Credit Management
## 3                Financial Education/You and Your Family/Credit Management
## 4  Accounts and Services/Credit Cards/Account preference set or maintained
## 5         Accounts and Services/Credit Cards/Account or service maintained
## 6                Mortgage Loans/Your Financial Goals /Borrowing and Credit
## 7                     Mortgage Loans/Personal Lines and Loans/Credit Cards
## 8                Mortgage Loans/Your Financial Goals /Borrowing and Credit
## 9         Accounts and Services/Credit Cards/Account or service maintained
## 10                    Mortgage Loans/Personal Lines and Loans/Credit Cards
## ..                                                                     ...
```

Determining What Methods By Which the User Communicates

This code can be used to determine which means of communication the user uses frequently, and which they never use. As a result, you can reach the user by their preferred mean. This will help in preventing customer dissatisfaction and ensure that both the credit card advertisement and customer support can easily reach the user.

```
# uses dplyr to select a subset of the following categories
selection <- select(month_end_bal, masked_id, phone_banker_cnt, mobile_bank_cnt,
online_bank_cnt, direct_mail_cnt, direct_email_cnt, direct_phone_cnt)

groups <- group_by(selection, masked_id) # groups the data into smaller subsets based on masked_
id
groups <- arrange(groups, masked_id) # arranges the subsets by masked_id
group_sums <- summarize(groups, "Call Access " = sum(phone_banker_cnt), "Online Access " = sum(o
nline_bank_cnt),
                  "Contact by Mail " = sum(direct_mail_cnt), "Contact by Email " = sum(direct_em
ail_cnt),
                  "Contact by Phone " = sum(direct_phone_cnt)) # finds the summation of each cat
egory
group_sums[userID,] #change variable at beginning of doc to show different userID
```

```
## Source: local data frame [1 x 6]
##
##   masked_id Call Access  Online Access  Contact by Mail  Contact by Email
##       (dbl)        (dbl)          (dbl)            (dbl)             (dbl)
## 1         1            0            234                5                72
## Variables not shown: Contact by Phone (dbl)
```