# Contents

## 1. Executive Summary

This project analyzed Wells Fargo customer outreach data by following the following analytical process:

1. **Feature Engineering** - Understanding the business requirements and creating new features which were relevant from business standpoint
2. **Top Down Approach** - Working through the problem by narrowing the scope of the problem throughout the analysis
3. **Business Understanding** - Identifying the drivers which help enhance the portfolio and balance of its customers. Business interpretation was as a key driver while creating the models.
4. **Choosing the right model** - The final model provides the odds of a customer changing their balance or portfolio. This helps Wells Fargo rank its consumers according to higher likelihood of increasing/reducing portfolio or balance

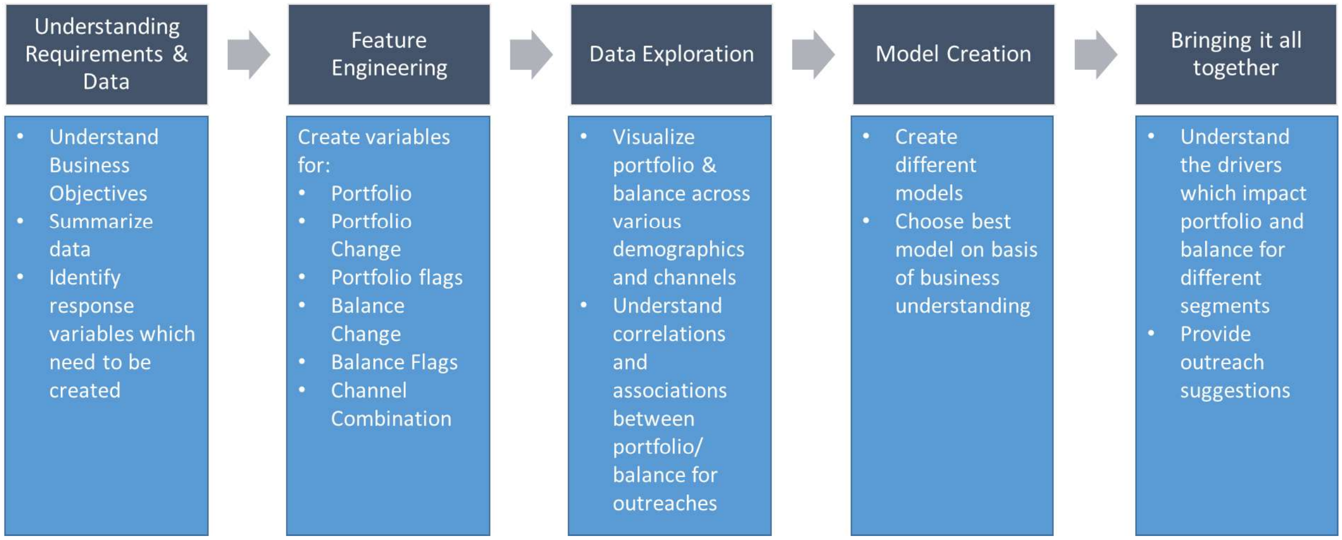
The analysis demonstrated the following results:

- customer demographics A 2, 3, and 4, combined with demographics B 3, 4, and 5 had a higher likelihood of closing accounts or decreasing balance category
- Customer outreach method 1 and market outreach through channel 1 were determined to consistently increase the likelihood of opening accounts C, D, and F, while decreasing the likelihood of closing an account for these demographics of interest

# 2. Response Section

## 2.1. Methodology & Approach

### 2.1.1. Approach




### 2.1.2. Methodologies Considered

| | Considerations | Caveats | Selected? |
|---|---|---|---|
| Decision Tree | Tested in initial exploration, but provided poor insights | Larger outreach numbers skewed the results | -- |
| Association Rules | Useful for gauging unrealized relationships | No temporal inference (customer may have account before outreach) | Exploratory Analysis |
| Regression | ○ Allows for quantitative ranking and magnitude of factors for each account<br>○ Ordinal Logistic Regression<br>▪ Determines how much more likely a customer is to open account or go up a balance level<br>○ Binary Logistic Regression<br>▪ Determines how much more likely a customer is to open account | ○ Can only be interpreted in terms of odds<br>○ Adding complexity increases accuracy, but decreases interpretability<br>○ Most basic model used for this reason | Final Model |

## 2.2. Business Objective

To find the optimal "sequence, frequency and combination of channel outreach" (as specified in the Challenge Business Objective), I focused on finding the "at-risk" demographics with a higher probability of decreasing their portfolio/balance. Armed with this information, I examined the outreach methods and channels that increase the likelihood of adding an account and decrease the likelihood of closing an account for these "at-risk" demographics.

### 2.3. Data Discussion

**Associations:** The following combinations of outreaches and channels have associations with account/balance changes:

Confidence – The primary mode of sorting associations. Determines which combinations of customer outreaches and channels occur most frequently with account/balance increases.

Support – How often the combinations occur together for account/balance changes.

Lift (x) – The combinations of outreaches and channels are (x00%) more likely to have portfolio/balance increases.

| | Portfolio | | | | Balance | | | |
|---|---|---|---|---|---|---|---|---|
| No. | Combinations | Support | Confidence | Lift | Combinations | Support | Confidence | Lift |
| 1 | Outreach: 1,3,4,6 | 0.0012166 | 0.100000 | 9.13 | Outreach: 2,3 | 0.078508 | 0.54007 | 1.051114 |
| 2 | Outreach: 1,4,6 | 0.0012666 | 0.091842 | 8.39 | Outreach: 2 | 0.095900 | 0.53722 | 1.045584 |
| 3 | Outreach: 1,3,5,6 Channel: 2 | 0.0012250 | 0.087292 | 7.97 | Outreach: 3 | 0.111333 | 0.53337 | 1.038083 |

### 2.4. Data modification and feature engineering

| No. | Feature | Combination |
|---|---|---|
| 1. | Portfolio | Sum of Accounts A, B and Flags C, D, E, F, G |
| 2. | Portfolio Change | Portfolio for ‘n’$^{th}$ month - Portfolio for ‘n-1’$^{th}$ month<br>Portfolio Change for 0$^{th}$ month is 0. |
| 3. | Portfolio Change Flag | 1: Portfolio Change > 0<br>0: Portfolio Change = 0<br>-1: Portfolio Change < 0 |
| 4. | Balance Change | Total balance (normalized) for ‘n’$^{th}$ month - Total balance for ‘n-1’$^{th}$ month |
| 5. | Balance Change Flag | 1: Balance Change > 0<br>0: Balance Change = 0<br>-1: Balance Change < 0 |
| 6. | Channel Combinations | Combine Flags for channels i,ii and iv to create a binary column.<br>Ignore channel iii as all flags are ‘0’ for all observations.<br>0 – 000 – No channels are selected<br>1 – 001 – Only channel iv selected<br>2 – 010 – Only channel ii selected<br>3 – 011 – channel ii + iv selected<br>4 – 100 – Only channel i selected<br>5 – 101 – channel i + iv selected<br>6 – 110 – channel i + ii selected<br>7 – 111 – channel i + ii + iv selected |

## 2.5. Final Model R Code

```
## demographic regressions
bank_data$balance_change_flag_new <- NA
bank_data$balance_change_flag_new[which(bank_data$balance_change < 0)] <- -1
bank_data$balance_change_flag_new[which(bank_data$balance_change > 0)] <- 1
bank_data$balance_change_flag_new[which(bank_data$balance_change == 0)] <- 0

bank_data$portfolio_change_flag_new <- NA
bank_data$portfolio_change_flag_new[which(bank_data$portfolio_change < 0)]<-1
bank_data$portfolio_change_flag_new[which(bank_data$portfolio_change > 0)]<-1
bank_data$portfolio_change_flag_new[which(bank_data$portfolio_change== 0)]<-0

bank_data$balance_change_flag_new<-as.factor(bank_data$balance_change_flag_ne
w)
bank_data$portfolio_change_flag_new<-as.factor(bank_data$balance_change_flag_
new)

bank_data$month<-as.integer(bank_data$month)

portfolio_order_model<-polr(portfolio_change_flag_new~cust_demographics_ai+cu
st_demographics_aii+month, data=bank_data, Hess=TRUE)
```

```
#Ordinal Logistic Regression
#Account A
A<-polr(typeA_bal_cat~month+cust_demographics_ai+cust_demographics_aii+
cust_outreach_ai+cust_outreach_aii+cust_outreach_aiii+cust_outreach_aiv+
cust_outreach_av+cust_outreach_avi+cust_outreach_avii+channel_combination,
data=bank_data, Hess=TRUE)
#Check p-values (all predictors are well below 0.05 significance level
ctableA<-coef(summary(A))
pc<-pnorm(abs(ctableA[, "t value"]), lower.tail=FALSE *2)
(ctableA <- cbind(ctableA, "p value" = pc))
#check odds (every factor above 1 is of interest)
sort(exp(coef(A)), decreasing = TRUE)
## cust_demographics_aii5 cust_demographics_aii4 cust_demographics_aii3
##              2.9751122              2.2610350              1.8374821
```

```
## cust_demographics_aii2        cust_outreach_ai      cust_outreach_aiii
##               1.3537018               1.0646607               1.0581115
##     channel_combination       cust_outreach_avi      cust_outreach_avii
##               1.0303387               1.0097238               0.9944910
##        cust_outreach_av                   month       cust_outreach_aii
##               0.9942498               0.9877769               0.9582489
##       cust_outreach_aiv  cust_demographics_ai3  cust_demographics_ai5
##               0.9563817               0.8994020               0.8907172
##   cust_demographics_ai4  cust_demographics_ai2  cust_demographics_ai1
##               0.8141342               0.7710761               0.1667269
```

*Code for similar models to Account A (Accounts B-E) and Account G (Account F) can be found in the attached code repository. Note odds ratio for Account B are for increasing value, as all customers have account B open.

```
#Account G

G_account<-glm(typeG_flag~month+cust_demographics_aii+cust_outreach_ai+cust_o
utreach_aiv+cust_outreach_avi+cust_outreach_avii+channel_combination, data=ba
nk_data,family="binomial")

#Example interpretation: Demographic B5 5.3 times as likely as Demographic B1
to open account G.

sort(exp(coef(G_account)), decreasing=TRUE)
```

```
## cust_demographics_aii5 cust_demographics_aii3 cust_demographics_aii4
##            5.317196950            3.716206031            3.601648872
## cust_demographics_aii2      cust_outreach_aiv       cust_outreach_ai
##            3.285788350            1.155693813            1.076386375
##      cust_outreach_avii    channel_combination      cust_outreach_avi
##            1.038977044            1.022213512            1.007306130
##                   month            (Intercept)
##            0.932536974            0.007611338
```

### 2.6. Insights, Analytics, and Quantitative Results

#### 2.6.1. What drives growth in accounts and/or balance between month 0 and month 12.

4 models were built to explain the factors that affect the change in total balance and change in accounts in a portfolio. The first two models are concerned with the net change in balance and portfolio, whereas the other two models are concerned with only increases in balance and portfolio.

Balance Change:

- Customer outreach 1 and 3 were found to have a significant impact on total change in balance, with every outreach of type 1 having a slight negative impact and every and outreach of type 3 having a slight positive impact.
- The month of September (assuming month 9 is September) also had a significantly positive impact in balance change.

Portfolio Change:

- Portfolio changes had similar results for the month of September, albeit a negative impact.
- Most demographics types affected likelihood of increasing portfolio level
- For every outreach of type 3 the odds of a portfolio increase increased by 41%.

Balance Growth:

- Only outreach 8 did the trick, increasing the odds of a balance increase by 6%.

Portfolio Growth:

- A channel combination of 1, 2, and 4 had 80% higher odds of a portfolio increase compared with using no channels.
- For every outreach of type 3, the odds of a portfolio increase increased by 41%.

More specific information for portfolio and balance change across demographics and accounts can be found in the Appendix Section 3.3.

#### 2.6.2. What demographic types, if any, are more likely to increase (or reduce) their number of accounts and/or balance between month 0 and month 12?

Two models were constructed for both portfolio change and balance change. However, due to the relationship between the two (when an account is added, balance increases), both models were nearly identical. For brevity, only the portfolio change model is discussed.

- Demographic A1 has 33% higher odds of increasing their portfolio.
- Across all demographics, for every month from January, the odds of increasing their portfolio rise by 2%.

Below is a table showcasing the probability of a portfolio increase for every combination of demographics for the month of January. A full demographics matrix for all months can be found in the Appendix.

| | Cust Demographics Ai | | | | | |
|---|---|---|---|---|---|---|
| Cust De.. | 0 | 1 | 2 | 3 | 4 | 5 |
| 1 | 0.5574 | 0.6261 | 0.4807 | 0.4810 | 0.4802 | 0.5169 |
| 2 | 0.5466 | 0.6159 | 0.4699 | 0.4702 | 0.4694 | 0.5061 |
| 3 | 0.5244 | 0.5945 | 0.4477 | 0.4480 | 0.4472 | 0.4837 |
| 4 | 0.5151 | 0.5855 | 0.4385 | 0.4388 | 0.4380 | 0.4745 |
| 5 | 0.5208 | 0.5910 | 0.4441 | 0.4444 | 0.4436 | 0.4801 |

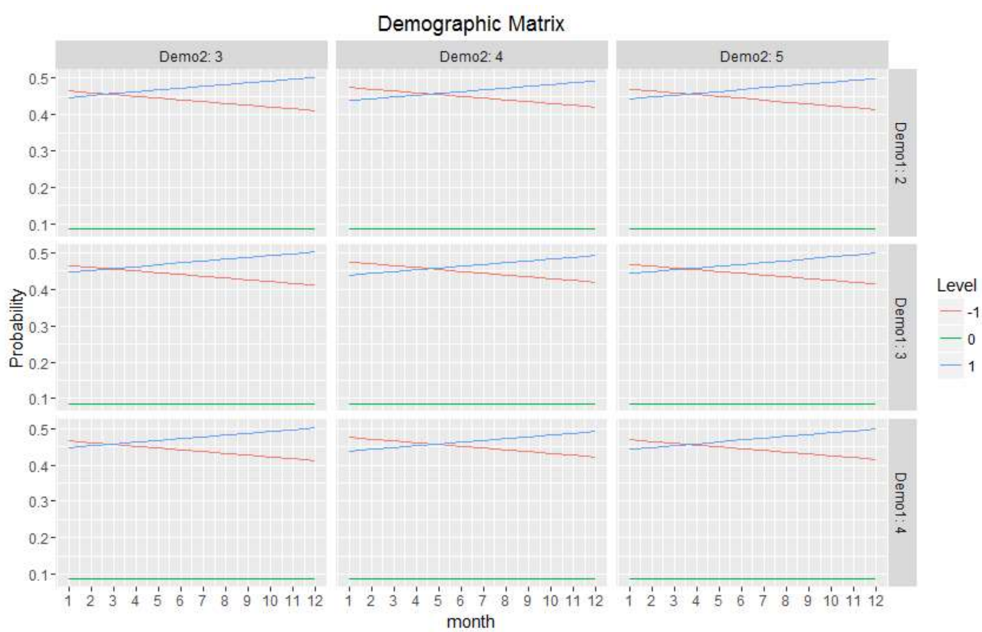
1

0.4380 0.6261

Sum of 1 broken down by Cust Demographics Ai vs. Cust Demographics Aii. Color shows sum of 1.

For most demographic combinations:

- the probability of a portfolio increase is greater than the probability of a decrease
- the probability of a portfolio increase rises as the year progresses
- the probability of a portfolio decrease falls as the year progresses
- after June, all demographic combinations have a higher likelihood of adding accounts




Of note are the above demographics:

- For the first half of the year, these demographic combinations are more likely to *decrease* their portfolio rather than increase. I call these demographics "at-risk."

This is the primary target of the analysis going forward: determining what outreach methods increase the likelihood of these nine demographics adding to their portfolio.

Target demographics are:

- Demographic A: 2,3, and 4
- Demographic B: 3,4, and 5

### 2.6.3. What types of accounts, customer interactions, customer events, or Wells Fargo outreach, are more correlated with account and/or balance change?

A model was created for every account, giving seven models in total.

- Five of the models utilized ordinal logistic regression for determining the odds of increasing in balance category (or creating an account if the account is not open) from an outreach or customer characteristic.

- The other two models used binary logistic regression for determining the odds of opening an account from an outreach or customer characteristic.

A comprehensive list of factor influence on odds for each account, in addition to odds ratio graphs, can be found in the Appendix.

The following accounts and corresponding channel combinations and outreach methods were determined optimal for the nine "at-risk" demographics.

| Account | Channel Combination (ranked) | Outreach (ranked) |
|---|---|---|
| **C** | 4, 5, 6 | 1, 8, 6 |
| **D** | 4, 5 6 | 1, 3 |
| **F** | 2, 4, 1 | 4, 1, 7 |

### 2.7. Summary of Analysis

From the analytics results, the business objective of finding the optimal "sequence, frequency and combination of channel outreach" was met by:

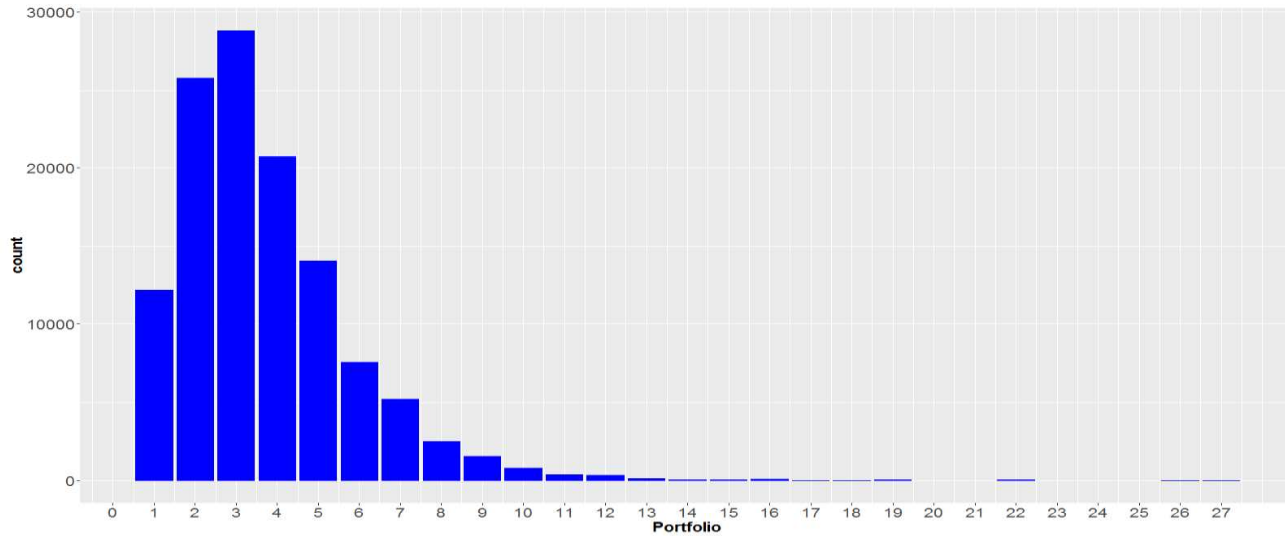
- Identifying demographics most likely to decrease their portfolio or balance
- Identifying which outreach methods and channels increase the likelihood of increasing portfolio/balance for the demographics of interest

The analytical methodology I followed of understanding the data, feature engineering, data exploration, and model building was successfully used to narrow the overall scope and increase the focus of the project.

The provided models enable Wells Fargo to differentiate between "at-risk" and high-likelihood customers. Additionally, the focused outreach and channel approach helps Wells Fargo reduce the costs of approaching customers.

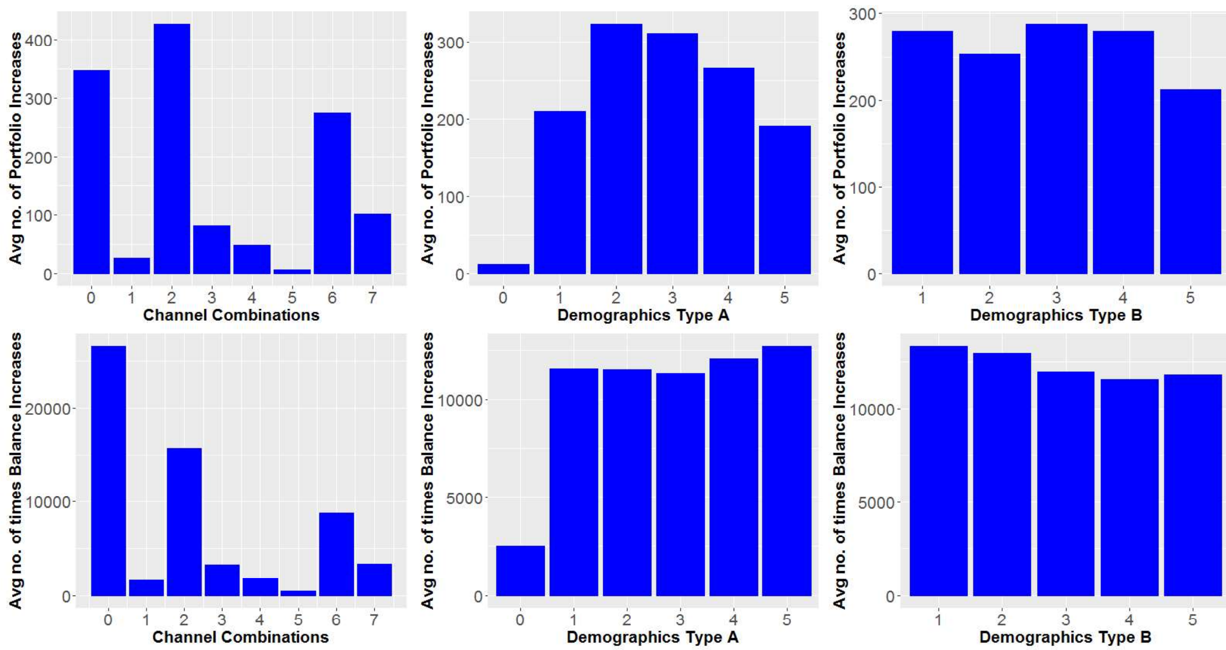
# 3. Appendix

## 3.1. Histogram of Portfolio across all demographics




Insights: Majority customers have a portfolio size of 3

## 3.2. Average number of Portfolio or Balance increases




Insights:

Portfolio changes are higher for channel combinations 0 (No channel used), 2 and 6.

Balance changes are higher for channel combinations 0, 2 and 6.

### 3.3. Correlations between portfolio or balance change with customer outreaches

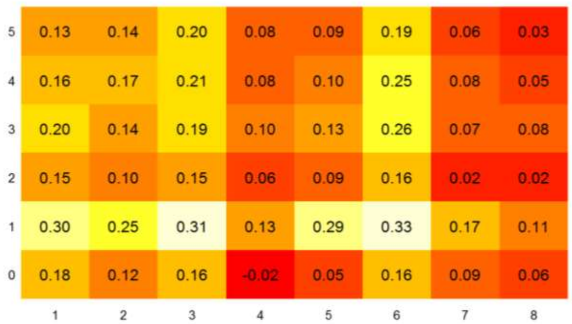
The below plots show the correlation between portfolio / balance change with customer outreaches:

X-axis – Customer Outreaches

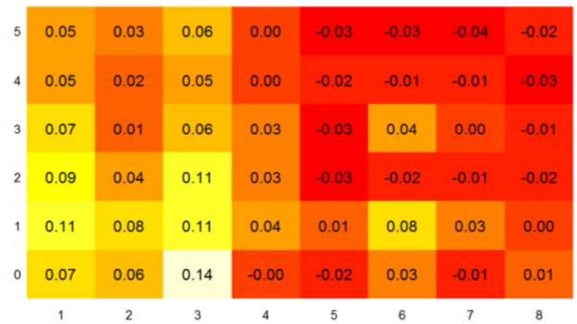
Y-axis – Types of Demographics

Cell values – Correlation co-efficient between portfolio / balance and customer outreaches

Demographics – A




Portfolio

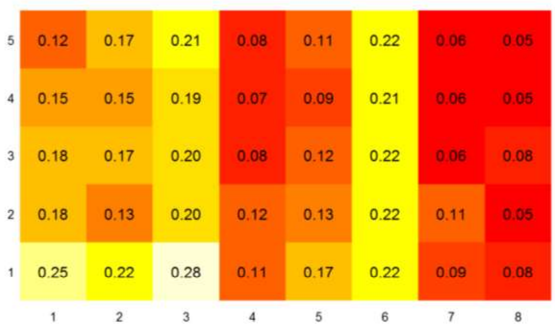



Balance

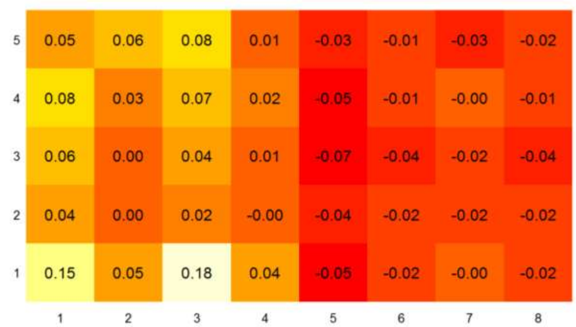
Insights –

- The correlations between customer outreach 1,3 and 6 are higher for portfolio
- The correlations between customer outreach 1 and 3 are higher for portfolio

Demographics – B




Portfolio

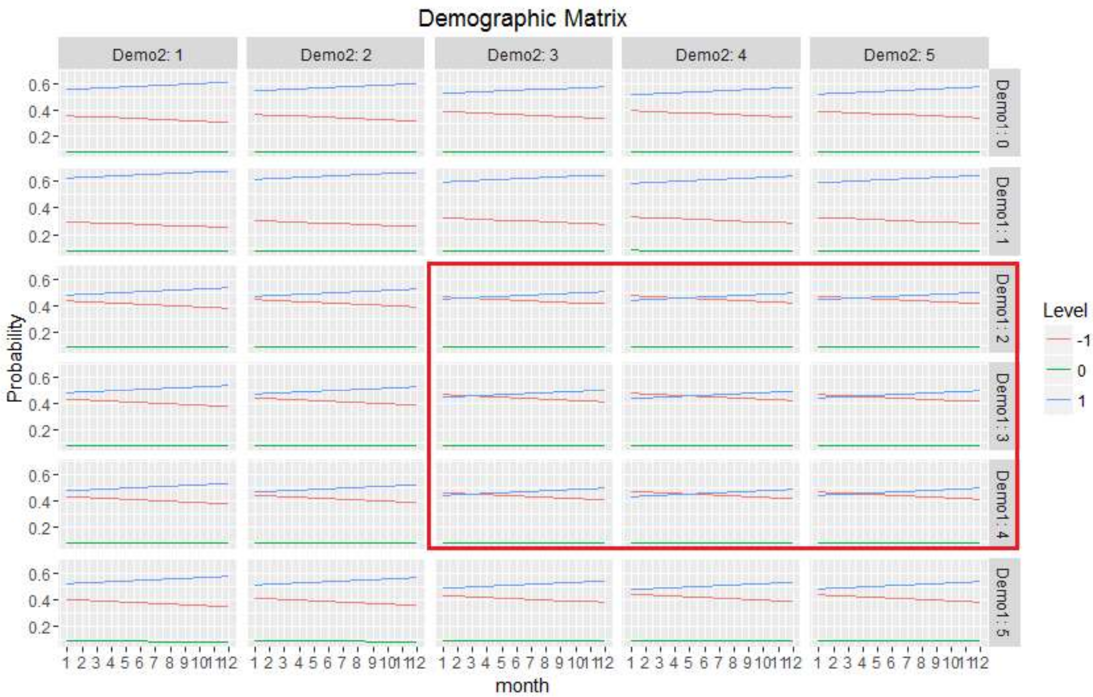



Balance

Insights –

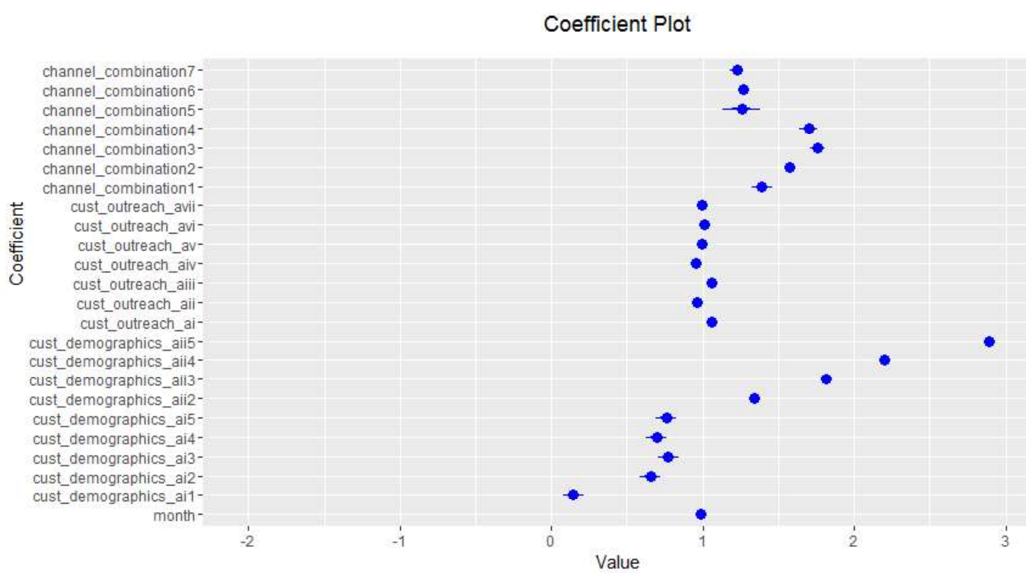
- The correlations between customer outreach 1,3 and 6 are higher for portfolio
- The correlations between customer outreach 1 and 3 are higher for balance

## 3.4. Demographics Matrix



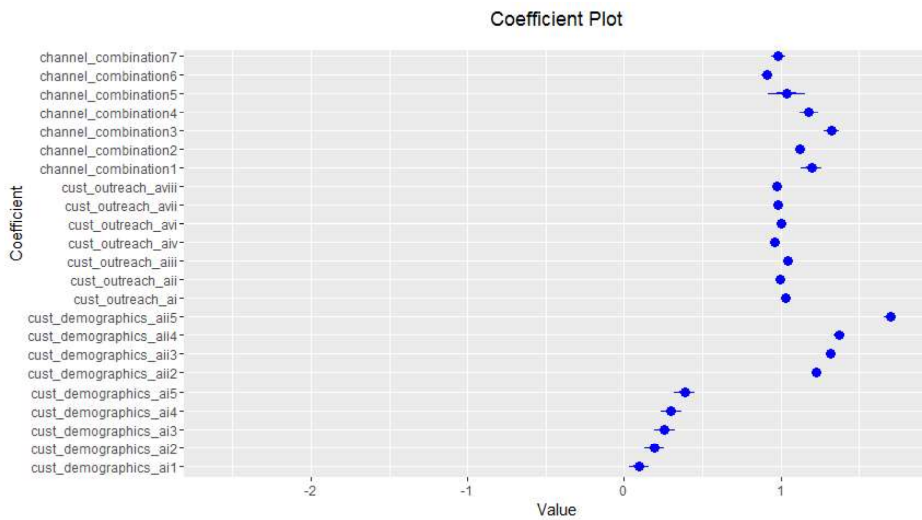

Zoomed-in red box can be found in section 2.5.2. of report body (Demo1 = Demographic A, Demo2 = Demographic B)

## 3.5. Coefficient Odds Ratio Plots

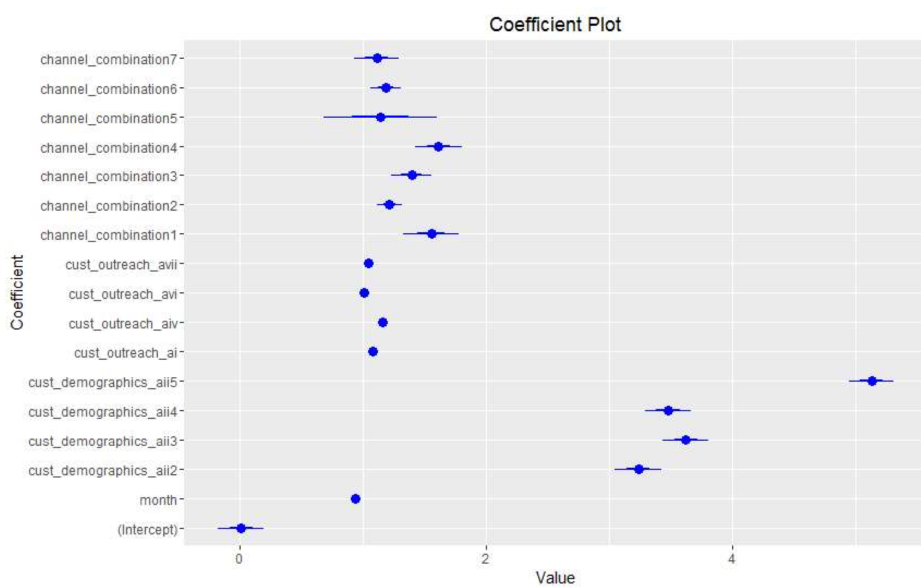



Odds Ratio of every variable for Account A

For factors, this is interpreted as the odds above the baseline (0 for Demographic A, 1 for Demographic 2). For example, customers in demographic B5 are 2.98 times as likely as customers in demographic B0 to increase their portfolio level. Continuous attributes can be interpreted in a similar fashion. For each customer outreach 1, customers are 1.06 times as likely to increase their portfolio level.




Odds Ratio for Account B (Note that since every customer in the data has an open account B, this provides the Odds Ratio for increasing in balance category)




Odds Ratio for Account B (Note customer demographics A had no effect on the odds of opening the account)

Complete coefficient information may be found in section 2.4, or in the complete R code attached to this report.