

[Introduction](#)  
[Total Opportunity](#)  
[Engagement](#)  
[Events](#)  
[Membership](#)  
[Emails](#)

If any issues occur, please contact [musch.sam@gmail.com](mailto:musch.sam@gmail.com).

## Introduction

---

We set the working directory

```
## set working directory  
setwd('D:/Group Folder/SEQUENTIAL FILES')
```

```
library(tidyverse)  
# Contains lots of very useful libraries  
# dplyr, ggplot, etc  
  
library(data.table)    # for large data files  
library(lubridate)     # for dates
```

Our first file: clusters

```
clusters <- read_csv('3 Data Generated  
Files/per_person_clusters.csv') %>%  
  select(ID_DEMO, cluster = cluster)
```

Our second file: individual info.

```
indy <- read_csv('2 Data Cleaned  
Files/individual_info_cleaned.csv') %>%  
  filter(MEMBERSHIP_TYPE_CODE != 'X')
```

There are many people who are "annual" but they are past members. For our purpose, **we only care about what someone is right now.**

```
# We only care about current status
# Annual, lifetime, non-member
indy <- indy %>%
  mutate(MEMBERSHIP_TYPE_CODE =
    ifelse(MEMBERSHIP_TYPE_CODE == 'L' &
      MEMBERSHIP_STATUS_CODE == 'C', 'L',
    ifelse(MEMBERSHIP_TYPE_CODE == 'A' &
      MEMBERSHIP_STATUS_CODE == 'C', 'A',
    NA)))
```

To add in a bit more detail, we split people by average age.

```
avg_age <-
  indy %>%
    summarize(ages = mean(age))
# 43 years old

indy <- indy %>%
  select(ID_DEMO, age) %>%
  mutate(age_cuts =
    ifelse(age > avg_age$ages, 'older', 'Younger'))
```

## Total Opportunity

---

This is our `membership` table. We are just selecting the 3 relevant columns and keeping from 2015-present.

```
# Adding in clusters to our membership file
membership_clean <- read_csv('4
Tableau/membership_EDA_new_v4.csv')

membership_clean <-
  membership_clean %>%
  inner_join(indy, by = 'ID_DEMO') %>%
  inner_join(clusters, by = 'ID_DEMO') %>%
```

How many people had "x" characteristics at the time? We will be combining this with our other dataframes later on to compare opportunity vs results.

As an example, maybe we have 7 annual members who are older people and are in the sports cluster attend some event in 2019. Is this good? How many people from that group could have attended?

```
mem_ind_clu <-
  membership_clean %>%
  count(year_tableau = YEAR_FISCAL,
        member,
        age_cuts,
        cluster)
```

## Engagement

---

This whole section is creating the "engagement" piece of the dashboard.

This is just reading in the file and making sure we have the correct information.

```
engage <- read_csv('2 Data Cleaned
Files/engagement_cleaned.csv') %>%
  filter(YEAR_FISCAL > 2014) %>%
  inner_join(clusters, by = 'ID_DEMO')
```

Engagement originally had each engagement type in its own separate column. For our purposes, we need to bring each engage type into 1 single column to compute summaries.

```
engage_gather <-  
  engage %>%  
  gather(key = engage_type,  
         value = given_value,  
         DONOR_ANNUAL : UMAA_MEMBER_ANNUAL)
```

### Before

	DONOR_ANNUAL	UMAA_MEMBER_ANNUAL
Sam	3	2

### After

	ENGAGE_TYPE	GIVEN_VALUE
Sam	DONOR_ANNUAL	3
Sam	UMAA_MEMBER_ANNUAL	2

This section is just adding in person-level details to our engagement file.

```
engage_join <-  
  engage_gather %>%  
  inner_join(clusters, by = 'ID_DEMO') %>%  
  inner_join(membership_clean,  
            by = c('ID_DEMO', 'YEAR_FISCAL')) %>%  
  inner_join(indy, by = 'ID_DEMO')
```

This is where we compute our actual values. We are saying: per year, per cluster, per age, per engagement type, and per member status, **what was the total engagement?**

```
# Summary - how did the cluster perform? (The results)
engage_for_tableau <-
  engage_join %>%
  group_by(year_tableau = YEAR_FISCAL,
            cluster,
            age_cuts,
            engage_type,
            member) %>%
  dplyr::summarize(engage_values = sum(given_value))
```

This directly connects the previous section (the results) with the opportunity that we had talked about before.

```
engage_for_tableau <-
  engage_for_tableau %>%
  left_join(mem_ind_clu,
            by = c('year_tableau', 'member',
                  'age_cuts', 'cluster'))
```

## Events

---

This is just adding in clusters and verifying we have the correct data

```
#setwd("D:/Group Folder/Data/Cleaned data sets")
events <- read_csv('2 Data Cleaned
Files/events_cleaned.csv') %>%
  inner_join(clusters, by = 'ID_DEMO') %>%
  mutate(DATE_EVENT = ymd(DATE_EVENT)) %>%
  filter(YEAR_FISCAL > 2014) %>%
  select(ID_DEMO, YEAR_FISCAL, broad_cat)

# Adding person's characteristics
events <-
```

```
events %>%
  inner_join(membership_clean,
             by = c('ID_DEMO', 'YEAR_FISCAL')) %>%
  inner_join(indy, by = 'ID_DEMO')
```

(Same as for engagement)

This is where we compute our actual values. We are saying: per year, per cluster, per age, per engagement type, and per member status, **what was the total events attended?**

```
# Number of events per group
events_for_tableau <- events %>%
  count(year_tableau = YEAR_FISCAL,
        cluster,
        age_cuts,
        Event_Category = broad_cat,
        member) %>%
  rename(event_values = n)
```

This directly connects the previous section (the results) with the opportunity that we had talked about before.

```
# How many people in the group? (Opportunity)
events_for_tableau <-
  events_for_tableau %>%
  inner_join(mem_ind_clu,
             by = c('year_tableau', 'member',
                   'age_cuts', 'cluster'))
```

## Membership

---

```
# Add persons characteristics
membership_join <-
  membership_clean %>%
  inner_join(indy, by = 'ID_DEMO') %>%
  inner_join(clusters, by = 'ID_DEMO')
```

```
# Summary per group - what is the membership makeup of the
cluster?
membership_tab <-
  membership_join %>%
  count(year_tableau = YEAR_FISCAL,
        cluster,
        age_cuts,
        member) %>%
  rename(member_values = n)
```

```
# Summary per group (Results)
ind_clu <-
  membership_join %>%
  count(year_tableau = YEAR_FISCAL,
        cluster,
        age_cuts) %>%
  rename(total_possible_people = n)
```

```
# How many people in the group? (Results and opportunity)
membership_tab <-
  membership_tab %>%
  inner_join(ind_clu, by =
    c('cluster', 'year_tableau', 'age_cuts'))
```

## Emails

---

```
emails2 <- fread('./2 Data Cleaned
Files/emails_cleaned.csv')
```

```
# Adding in all the details
emails <- emails2 %>%
  merge.data.table(clusters, by = 'ID_DEMO') %>%
  merge.data.table(membership_clean,
                  by = c('ID_DEMO', 'YEAR_FISCAL')) %>%
  merge.data.table(indy, by = 'ID_DEMO')
```

```
# CTR per cluster and per category
emails_groups <-
  emails %>%
  group_by(year_tableau = YEAR_FISCAL,
            cluster,
            age_cuts,
            Email_Category = broad_cat,
            member) %>%
  summarize(Total_Emails_Sent = n(),
            Total_Emails_Clicked =
              sum(Status_Clicked))
```

```
#setwd("D:/Group Folder/Data/Tableau Data")
write.csv(emails_groups, '4
Tableau/year_emails_for_tableau.csv')
write.csv(membership_tab, '4
Tableau/membership_for_tableau.csv')
write.csv(engage_for_tableau, '4
Tableau/engage_for_tableau.csv')
write.csv(events_for_tableau, '4
Tableau/year_events_for_tableau.csv')
```