NOVA IMS
Information Management School

Paralyzed Veterans of America (PVA)

A Data Mining approach to Customer Segmentation

Luís Fernando Rodrigues Agottani (m20200621)

Vilmar Adriano Bussolaro (m20200268)

Xavier Golaio Gonçalves (m20201090)

Professor: Fernando Lucas Bação

Assistant Professors: David Silva, João Fonseca

Subject: Data Mining

Lisbon (Portugal)

January 2021

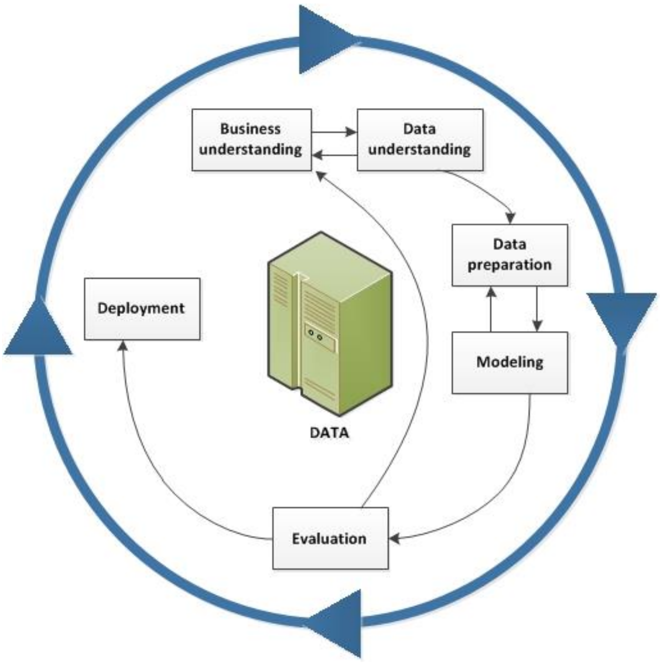
Introduction

In this practical group work for Data Mining class, we have been asked to work as Data Mining/Analytics Consultants to Paralyzed Veterans of America (PVA), a non-profit organization that provides programs and services for US veterans with spinal cord injuries or diseases.

Being one of the largest direct mail fundraisers in the United States, PVA has a very large and complex database. With the total of 13 million donors, the last round of mailing was sent to 3.5 million people. We are analyzing a small sample of that: 95.412 individuals, who were selected for being "Lapsed" donors according to the RFM analysis performed internally by the NGO.

Methodology

To perform the analysis, we will build upon a simplified version of the CRISP-DM framework recommended by Kelleher, Mac Namee, & D'arcy (2015, p. 53).



[1]

As a process, the life cycle model consists of six phases with arrows indicating the most important and frequent dependencies between phases. The sequence of the phases is not strict. In fact, most projects move back and forth between phases, as necessary. The figure emphasizes that data is at the heart of the process. Certain phases in CRISP-DM are more closely linked together than others. For example, Data Preparation and Modeling phases are closely linked, and analytics projects often spend some time iterating between these two phases. (Kelleher, Mac Namee & D´arcy, 2015, p. 54).

We will also refer to some steps and terminology presented in Berry and Linoff´s approach, from the reference book of this discipline, "Data Mining Techniques" (2004, p. 54-86) and Han, Kamber & Jian´s "Data Mining Concepts and Techniques" (2011, p. 83-123) that present a simple and elegant way to perform Data Preprocessing.

[1] The data mining life cycle. Image from IBM Knowledge Center, https://www.ibm.com/support/knowledgecenter/it/SS3RA7_sub/modeler_crispdm_ddita/clementine/crisp_help/crisp_overview.html

Since this is an iterative process where advances and insights acquired along the steps can influence the prior decisions, the steps are not entirely followed in a strictly linear/straightforward logic.

Step 1 – Translate the Business Problem into a Data Mining Problem

Before working with any Data Mining problem, we must understand that as Data Analysts, having Domain Knowledge in the subject we are evaluating is not only valuable for discussing the outcomes of the analysis with the target audience, but also to perform the analysis itself. According to Jain, Murty and Flynn (1999, p. 290), it was revealed through experimental studies that when using weak clustering algorithms, the use of domain knowledge would improve their performance. The authors defend the idea that every clustering algorithm uses some type of knowledge either implicitly or explicitly: whether in selecting a pattern representation scheme (e.g., using one's prior experience to select and encode features), choosing a similarity measure or selecting a grouping scheme.

To that extent, we started this group project by performing a thorough research on NOVA Discovery (EBSCOHost database), to find relevant articles and news pieces regarding "Fundraising using Direct Mail". We selected an initial set of 21 articles related to the matter, trimming down to 6 articles that provided information more relevant to our case with PVA. We summarized highlights from those articles in a "Domain Knowledge Summary" (please find attached to this report).

Our task in this project is to analyze a sample of the latest PVA´s fundraising campaign, organize the data we have in a proper way to bring information to the surface, and combine that information with our expertise to create knowledge. PVA requested a Customer Segmentation analysis, and that´s indeed the best approach to develop actionable efforts targeted at specific customers who might be churning or about to churn, for example. Based on the assumption made by Swearingen (2009) that "the best predictor of future customer behavior is past customer behavior", we will perform the segmentation of customers using RFM.

According to Wei, Lin & Wu (2010), the RFM model is one of the most frequently adopted segmentation techniques. It comprises three measures (recency, frequency and monetary), which are combined into a three-digit RFM cell code, covering five equal quintiles (20% group). Recency refers to the interval since the last consuming behavior (most-recent purchasers are more likely to purchase again). Frequency is the number of transactions that a customer has made (customers with more purchases are more likely to buy again). Monetary refers to the cumulative total of money spent by a particular customer. In the end of the process, all customers are represented by a concatenated value corresponding to their own RFM value.

In that way, the best customer segment is 555, whereas the worst customer segment is 111. Based on the assigned RFM behavior scores, customers can be grouped into segments and their profitability can be further analyzed. (Bult and Wansbeek, 1995)

To better understand our customers, we will work with undirected data mining (which in the Machine Learning literature is called *"unsupervised learning"*). We will perform clustering, a method that finds groups of similar records. (Berry & Linoff, 2004, p. 57). More specifically, let´s use K-Means++, a variation of the K-Means algorithm.

Arthur & Vassilvitskii (2007) defend that K-Means++ provides better clusters and fasters results by calculating the initial center points from the squared distance of the closest center already chosen (as opposed to randomly generated). Before we can proceed to segment customers, we must do a thorough analysis of the dataset and its features, followed by cleansing steps, which will be done on the next two steps.

Step 2 – Select appropriate data

The dataset received on the exercise has a massive number of variables. After being imported using Pandas, the shape of "Donors.csv" presented 476 columns. For a better understanding of the data, after carefully analyzing the metadata file that accompanied the .csv, our approach was to divide the dataset into 4 main categories, presented in the sequence data is found on the file.

PVA Primary Data *– Donors database probably built by PVA over the years*

From ODATEDW to HIT: Consists of variables like when the donor performed its first donation and entered the list (ODATEDW), the title of the donor (TCODE, includes Mr./Mrs./Doctor/Captain etc.), ZIP code, if they are homeowners or not (HOMEOWNR), if they have children or not (NUMCHLD).

PVA Secondary Data (MetroMail/Polk) *– Probably acquired from private mail list companies*

From MBCRAFT to PUBOPP: the number of times the donor has responded to other types of mail order offers.
DATASRCE regards which is the third-party company that provided the information.
From MALEMILI to GEOCODE: percentage of males active in military in the region of the donor, number of solicitations (mail requests) permitted by year.
From COLLECT1 to LIFESRC: hobbies and lifestyle. Code FISHER for people into fishing, BIBLE for people who reads the Bible, PETS for who is a pet owner.

PVA Secondary Data (US Census Bureau) *– Regarding demographic data from the region of donor*

From POP901 to AC2: data from the 2010 US Census, with variables reflecting characteristics of each of the donors´ neighborhoods.

So far, the three categories we saw above are going to be useful to make an analysis of profiles after we have done our cluster analysis, to identify similarities and patterns that can be used to provide market intelligence.
The next group of information is the basis on which we will build the cluster analysis.

PVA Primary Data *– RFM Analysis*

In this field, instead of using the commonly used Recency-Frequency-Monetary acronym (RFM), Paralyzed Veterans of America used Recency-Frequency-Amount (RFA). For the sake of simplicity, we will use the term RFA when referring to PVA´s analysis, and RFM when referring to the same analysis we did ourselves on the next steps, using primary data.
From ADATE_2 to ADATE_24: the date the promotion was mailed. The column indicates year and type of gift sent. The rows are the donors. When a gift was sent, we have the date that event happened.
From RFA_2 to RFA_24: Which was the RFA status of the donor at the date the mailing was sent.
From CARDPROM to NUMPRM12: summary variables from the promotion history file.
From RDATE_3 to RDATE_24: date the gift was received.
From RAMNT_3 to RAMNT_24: dollar amount of the gift.
From RAMNTALL to AVGGIFT: summary variables from the giving history file.

Feature Selection

To perform the clustering techniques, we have created a copy of the main *dataframe*, and set aside the information about the Neighborhoods, regarded before as ***PVA Secondary Data (US Census Bureau).*** To better characterize the donors, we think it makes sense to focus on their actual characteristics, instead of the characteristics of their neighborhoods (ZIP codes).

```
#Delete neighborhood characteristics
neighborhood_columns = donors.loc[:, 'POP901':'AC2'].columns.to_list()
donors_selected = donors.copy()
donors_selected.drop(columns = neighborhood_columns, inplace = True)
```

We also removed people who were in the list, but had no donation history, which corresponds to 1,4% of the database.

```
#Delete rows from non donors
delete_non_donors = donors_selected[donors_selected['Recency'].isnull()].index
# % of non_donors:
len(delete_non_donors)/len(donors_selected)
Output: 0.014075797593594097
```

To build clear and concise clusters, we have to remove noise and focus on what are the characteristics of the donors (***PVA Primary Data** – Donors database)* and the RFM (built on the next step – feature engineering). So non-related columns, mainly from third-party sources, that contain information about the neighborhoods/census bureau, were put aside from the clusterization. They will be used later only for characterization of the individual clusters.

```
non_related_columns = ['ODATEDW','OSOURCE', 'TCODE', 'STATE', 'ZIP','MAILCODE', 'DOB','NOEXCH', 'RECINHSE', 'RECP3', 'RECPGVG', 'RECSWEEP', 'MDMAUD', 'DOMAIN', 'HOMEOWNR', 'CHILD03', 'CHILD07', 'CHILD12', 'CHILD18', 'NUMCHLD', 'MBCOLECT','PUBDOITY', 'PUBGARDN', 'PUBHLTH', 'MAGMALE','PUBNEWFN', 'PUBPHOTO', 'PUBOPP', 'MAGFEM', 'MAGFAML','MBBOOKS','MBGARDEN','WWIIVETS','MBCRAFT','PUBCULIN','DATASRCE','MALEMILI','MALEVET','VIETVETS','LIFESRC', 'LOCALGOV', 'STATEGOV', 'FEDGOV', 'GEOCODE','COLLECT1', 'VETERANS', 'BIBLE', 'CATLG', 'HOMEE', 'PETS', 'CDPLAY','STEREO', 'PCOWNERS', 'PHOTO', 'CRAFTS', 'FISHER', 'GARDENIN', 'BOATS','WALKER', 'KIDSTUFF', 'CARDS', 'PLATES']
```

Feature Engineering

In this step a lot of transformation had to go through. We will write on the report the main changes, but the full set and steps can be found in the Jupyter Notebook using our GitHub link[2]. The "Age" variable was created using DOB (Date of Birth). RAMNT had all its values summed up by type of gift and years.

Going back to PVA´s RFA analysis, we saw that there were a lot of missing values because of the nature of how the features were built. But we also noticed some inconsistencies between what would be the Recency (taken from RFA variables) and the date that the last gift was received (RDATE). The "Lapsed" code didn´t match some donors, causing us to prefer to drop entirely the RFA analysis and build our own from the ground up, since we had all the required information.

[2] https://github.com/m20200268/DM-PROJECT-PVA-2020

Using RAMNT and RDATE we created RFM[3]:

```
#Recency: Last promotion date in months
lastpromotiondate = datetime.strptime('2017-06-01', '%Y-%m-%d')
Recency = []
for i in rdate['lastgiftdate']:
    if i == 0:
        Recency.append(None)
    else:
        Recency.append((lastpromotiondate.year - i.year)*12 + (lastpromotiondate.month - i.month))
#Monetary sum of all gifts from donors/rows
Monetary = ramnt.sum(axis=1)
#Frequency count of how many times the donor gifted.
Frequency = ramnt.count(axis=1)
```

In here, Recency is measured in months since the last donation. Frequency is how many times the donor has gifted. Monetary is the sum of all donations from RAMNT.

The next step was to create the "Star classification" for RFM, dividing in quintiles using pandas.qcut and assigning a number from 1 to 5 to each characteristic of the RFM, 1 being the worst and 5 the best. In that way, the concatenated customer who is 555 is the most valuable, and 111 the least valuable from our segmentation.

The process was repeated for RFM_Promotion (regarding all the campaigns/promotions, as opposed to just the RFM which is the whole historical data).

We will focus on RFM_Promotion, since we want to see the results regarding the campaigns (to better assess the type of material sent that resulted in more engagement and donations to PVA.

The output was a table[4] with the value of RFM, the corresponding "star" to each category and the concatenated value to see where that customer fits on the scale. In the image, showing just the first 5 results. The number on the index (left) is the donor code.

	Recency	Frequency	Monetary	Recency_star	Frequency_star	Monetary_star	Stars
0	18.0	7	73.0	4	5	5	455
1	18.0	2	37.0	4	2	3	423
2	21.0	7	58.0	2	5	5	255
3	18.0	7	65.0	4	5	5	455
4	17.0	2	25.0	4	2	2	422

Step 3 – Fix problems with data

The dataset received has a lot of missing values. We must understand what it means for our analysis, the possible cause of those missing values and how to tackle this problem.

Missing values

In this step we will work with 2 different groups: PVA´s RFA analysis, and all the others.

[3] Line #Recency: Last promotion date in months, in TheMasterpiece Notebook
[4] Line # Sum of RFM and the respective Score for every donor, in TheMasterpiece Notebook

It´s important to separate the RFA analysis because even though it might look like there are a lot of Null values, it´s in fact because when there was no response from the campaign, or the event didn´t happen (didn´t send the gift/didn´t e-mail the donor), there is no value on that specific cell.
So if we calculate the ratio between nan_count/len(donors), we will end up with very high percentages of null values, like in this example[5] below:

```
RAMNT_5     0.999906
RDATE_5     0.999906
RAMNT_3     0.997464
RDATE_3     0.997464
RDATE_4     0.997055
RAMNT_4     0.997055
RDATE_6     0.991867
RAMNT_6     0.991867
```

For missing values, we will continue calculating the ratio, excluding the RDATE_, RAMNT_ and ADATE_ variables.[6]
When dividing nan_count/len(donors), we get the following list of variables:

```
NUMCHLD     0.870184
MBCOLECT    0.554584
MBGARDEN    0.553955
MBCRAFT     0.553955
PUBOPP      0.553955
PUBPHOTO    0.553955
PUBNEWFN    0.553955
PUBDOITY    0.553955
PUBHLTH     0.553955
PUBCULIN    0.553955
PUBGARDN    0.553955
MAGMALE     0.553955
MAGFEM      0.553955
MAGFAML     0.553955
MBBOOKS     0.553955
WEALTH1     0.468830
WEALTH2     0.459303
DOB         0.250314
INCOME      0.223096
NEXTDATE    0.104526
TIMELAG     0.104526
ADI         0.001383
DMA         0.001383
GEOCODE2    0.001383
MSA         0.001383
FISTDATE    0.000021
   dtype: float64
```

We took into consideration that NUMCHLD with so many Null values could mean a lot of donors without children at all - but after realizing that only 15% of women remain childless at the age of 40 on the US[7], we thought it was very unlikely that 87% of donors had no progeny. Therefore, we could not assume the missing values were in fact 0s.
To define the threshold of where to cut the variables with too many missing variables, we were considering 40%, but later we crossed the line at 50%, because WEALTH1 and WEALTH2 are very important variables and should not be excluded from the analysis.

[5] Line #NaN Counting on Cleaning_ALL Notebook.
[6] Line #Drop RFM columns from nan_counts on Cleaning_ALL Notebook.
[7] According to Pew Research, 2015 - https://www.pewsocialtrends.org/2015/05/07/childlessness/

The selected features that had missing values and were important to our cluster analysis were imputed using K-Nearest Neighbors. KNN method replaces missing values using k-most similar non-missing subjects' values (Cover & Hart, 1967). It can impute both discrete attributes (using the most frequent value among the k-nearest neighbors) and continuous attributes (using the mean among the k-nearest neighbors). We will use the KNNImputer function from the impute module of the sklearn. This implementation fills missing values present in the observations by finding the nearest neighbors with the Euclidean distance matrix.

```
imputer = KNNImputer(n_neighbors=8, weights="uniform")
fill_na = imputer.fit_transform(donors_segmentation)
```

OneHotEncoding

The non-metric features Gender, Major (if the person has been flagged as a Major donor by PVA) and PEPSTRFL (PEP Star RFA Status) have all been encoded using OHE. It has been used parsimoniously, since OHE despite being easy to use and simple to implement, brings some aspects of the curse-of-dimensionality. (Ng, P. 2017)

Outliers Removal

In order to remove outliers, we have used the Local Outlier Factor (LOF) algorithm. It is an unsupervised anomaly detection method which computes the local density deviation of a given data point with respect to its neighbors. It considers as outliers the samples that have a substantially lower density than their neighbors. (Scikit Learn, 2020)
We saw the correlation between RFM and the variables, and removed the ones that had either a correlation too high, or too low:

```
#Delete low and high corr
corr_filter = ['AGE', 'INCOME', 'WEALTH1', 'WEALTH2', 'HIT', 'MAXADATE_days','Monetary','Recency','Frequency']
```

Scaling Data

We used the MinMaxScaler, an estimator that scales and translates each feature individually such that it is in the given range on the training set. In this case, between zero and one.

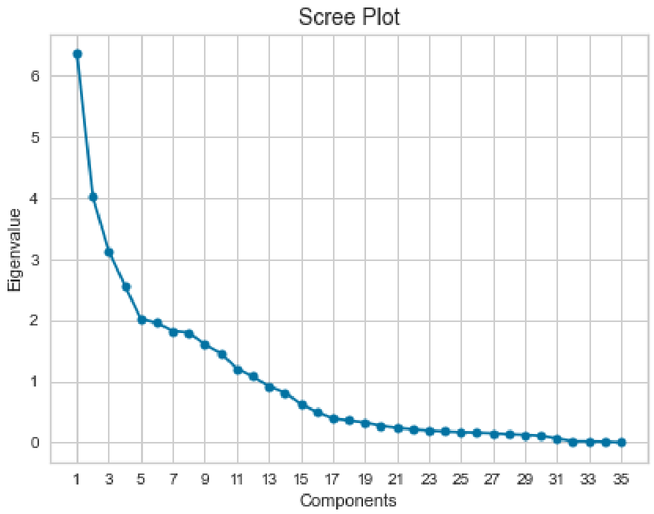
PCA

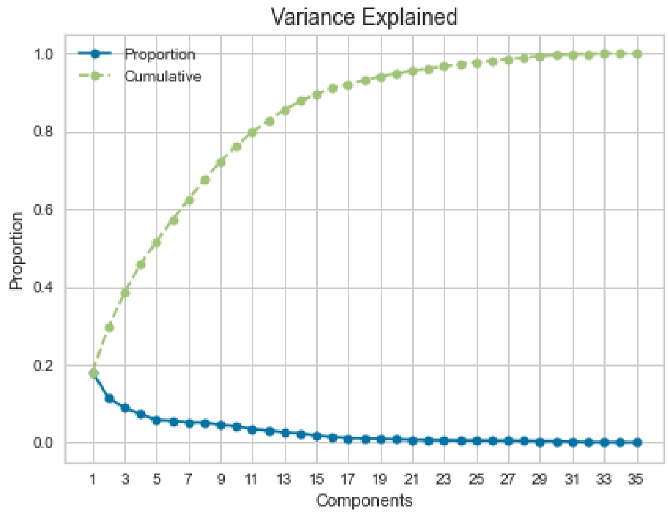
To reduce the dimensionality of data, we used Principal Component Analysis. Principal component analysis (PCA) is a widely used dimensionality reduction technique in data analysis. Its popularity comes from three important properties. First, it is the optimal (in terms of mean squared error) linear scheme for compressing a set of high dimensional vectors into a set of lower dimensional vectors and then reconstructing. Second, the model parameters can be computed directly from the data - for example by diagonalizing the sample covariance. Third, compression and decompression are easy operations to perform given the model parameters - they require only matrix multiplications. (Roweis, 1997)
A key factor to consider when performing PCA is the number of components to use. The table[8] below shows the Eigenvalue of the number of components, and the Cumulative value they have.

[8] Line # Output PCA table, in TheMasterpiece Notebook.

	Eigenvalue	Difference	Proportion	Cumulative
1	5.914333	0.000000	0.197142	0.197142
2	4.004780	-1.909553	0.133491	0.330633
3	2.907329	-1.097451	0.096910	0.427543
4	2.552182	-0.355147	0.085072	0.512615
5	2.002786	-0.549396	0.066759	0.579374
6	1.894548	-0.108238	0.063151	0.642525
7	1.772564	-0.121984	0.059085	0.701610
8	1.601062	-0.171502	0.053368	0.754978
9	1.455852	-0.145210	0.048528	0.803506
10	1.169595	-0.286257	0.038986	0.842492
11	0.836045	-0.333550	0.027868	0.870360
12	0.635252	-0.200793	0.021175	0.891534
13	0.449729	-0.185523	0.014991	0.906525
14	0.387475	-0.062254	0.012916	0.919441
15	0.333679	-0.053796	0.011123	0.930563
16	0.286552	-0.047127	0.009552	0.940115
17	0.238899	-0.047652	0.007963	0.948078
18	0.217456	-0.021443	0.007248	0.955327
19	0.194256	-0.023200	0.006475	0.961802
20	0.180755	-0.013502	0.006025	0.967827
21	0.168996	-0.011758	0.005633	0.973460
22	0.164277	-0.004719	0.005476	0.978936
23	0.151062	-0.013215	0.005035	0.983971
24	0.133339	-0.017722	0.004445	0.988416
25	0.124361	-0.008978	0.004145	0.99256

Similarly, we have a Scree Plot with the Eigenvalue, showing that the bigger the number of components, the smaller their delta on the impact in the variance.





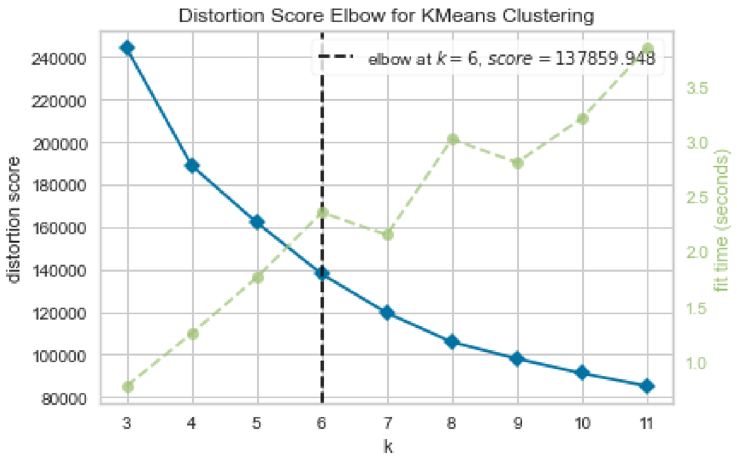
We decided on 25 components, since this would explain 99% of the variance.

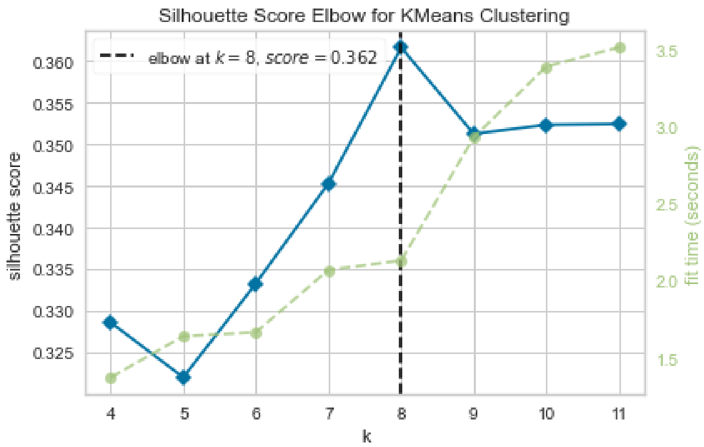
Clustering

Our next step is to make the clusterization of donors taking into account all steps we have done so far.

We performed the clusterization using Scikit Learn implementation of *K-Means++*, which is a partitioning cluster algorithm that works by grouping n vectors based on attributes into k partitions, where k < n, according to some measure. The name comes from the fact that k clusters are determined and the center of a cluster is the mean of all vectors within this cluster. The algorithm starts with determining k appropriate initial centroids, then assigns vectors to the nearest centroid using Euclidean distance and re-computes the new centroids as means of the assigned data vectors. This process is repeated over and over again until vectors no longer changed clusters between iterations. (Birant, D., 2011)

The Elbow method indicates an optimal number of K clusters as 6. But the Silhouette method indicates the optimal number of clusters of 8. We have run both situations and found that the results on K=8 had better defined customer types, so we sticked with the second method.





To check the performance of the clusters, we ran a DecisionTreeClassifier to check what set of input variables have the best predicting power.

The best results in defining the clusters was RFM Score itself, without the PCA variables. Therefore, we built the final cluster solution without the PCA variables.

```
It is estimated that in average, we are able to predict 85.59% of the customers correctly
Recency_score                        0.498527
Frequency_score                      0.340631
Monetary_score                       0.160842
PC24                                 0.000000
Blank folding cards with labels      0.000000
Christmas Cards with labels_count    0.000000
dtype: float64
```

Evaluation and Deployment: Cluster Analysis

So here´s the final set of variables used to make the clusters:

```
Index(['INCOME', 'WEALTH1', 'HIT', 'WEALTH2', 'CARDPROM', 'NUMPROM',
       'CARDPM12', 'NUMPRM12', 'MINRAMNT', 'MAXRAMNT', 'LASTGIFT', 'AGE',
       'MAXADATE_days', 'MINRDATE_days', 'MAXRDATE_days', 'FISTDATE_days',
       'NEXTDATE_days', 'Recency', 'Frequency', 'Monetary', 'Recency_score',
       'Frequency_score', 'Monetary_score', 'Just labels',
       'Calendars with stickers', 'Blank folding cards with labels',
       'Blank cards with labels', 'Thank you cards with labels',
       'Christmas Cards with labels', 'General Greeting Cards with Labels',
       'Notepad with labels', 'Just labels_count',
       'Calendars with stickers_count',
       'Blank folding cards with labels_count',
       'Blank cards with labels_count', 'Thank you cards with labels_count',
       'Christmas Cards with labels_count',
       'General Greeting Cards with Labels_count',
       'Notepad with labelss_count', 'x0_F', 'x0_M', 'x0_U', 'x1_MAJOR',
       'x2_PEPSTRFL', 'Cluster_RFM', 'Customer_RFM']
```

We ran the cluster again, with those 46 columns, and with optimal values:

```
#The best score was with de RFM variables, 85.59%.
kmeans = KMeans(init='k-means++', n_clusters=8, n_init=15, max_iter=300, random_state=1)
kmeans.fit(donors_seg_outlier.loc[:,['Recency_score','Frequency_score', 'Monetary_score']])
```

We can see clearly defined clusters, each one with their own characteristics.

The results are as follow:

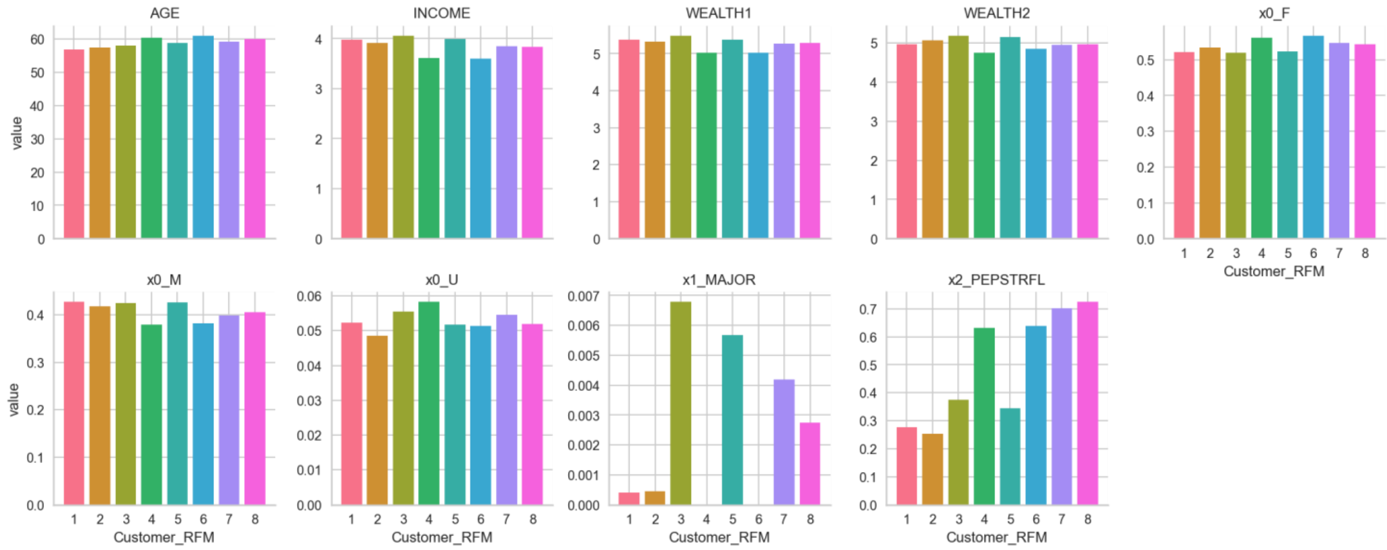
	Recency_score	Frequency_score	Monetary_score	
	mean	mean	mean	Total
Cluster_RFM				
0	4.56	4.54	4.59	13.70
5	2.14	4.43	4.59	11.17
2	4.58	4.03	2.39	11.00
3	3.75	2.45	3.73	9.92
7	2.43	3.98	2.28	8.69
1	1.35	2.55	3.50	7.40
4	4.01	1.41	1.40	6.82
6	1.44	1.36	1.36	4.16

Now we are renaming those 8 clusters so their name makes sense to us. We will give 1 to the worst customer (RFM close to 111), and 8 to the best customer (RFM close to 555).

```
bins = [0,5,2,3,7,1,4,6]
labels = ['8','7','6','5','4','3','2','1']
donors_seg_outlier['Customer_RFM'] = donors_seg_outlier['Cluster_RFM'].map(dict(zip(bins, labels)))
```

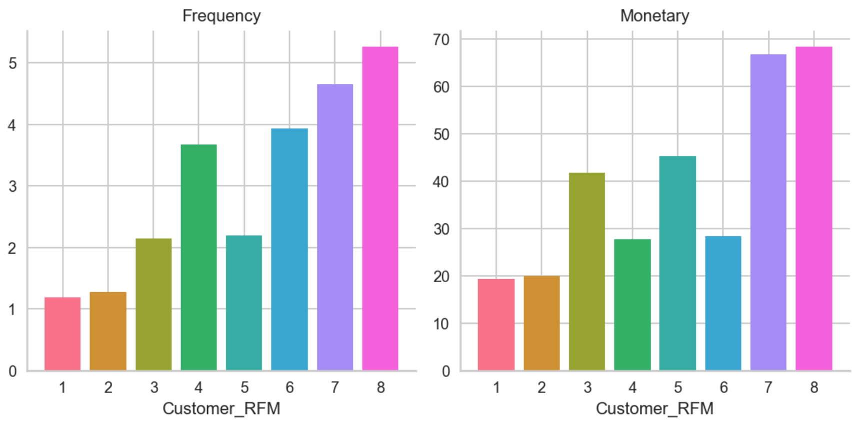
For us, what´s most important for characterizing who to pursue in marketing campaigns, is to identify the type of donors using RFM, and then analyze who they are and how they react to different gifts.

A thorough analysis can be done with the results, but we will perform a quick overview of some of the key differences spotted by the model.

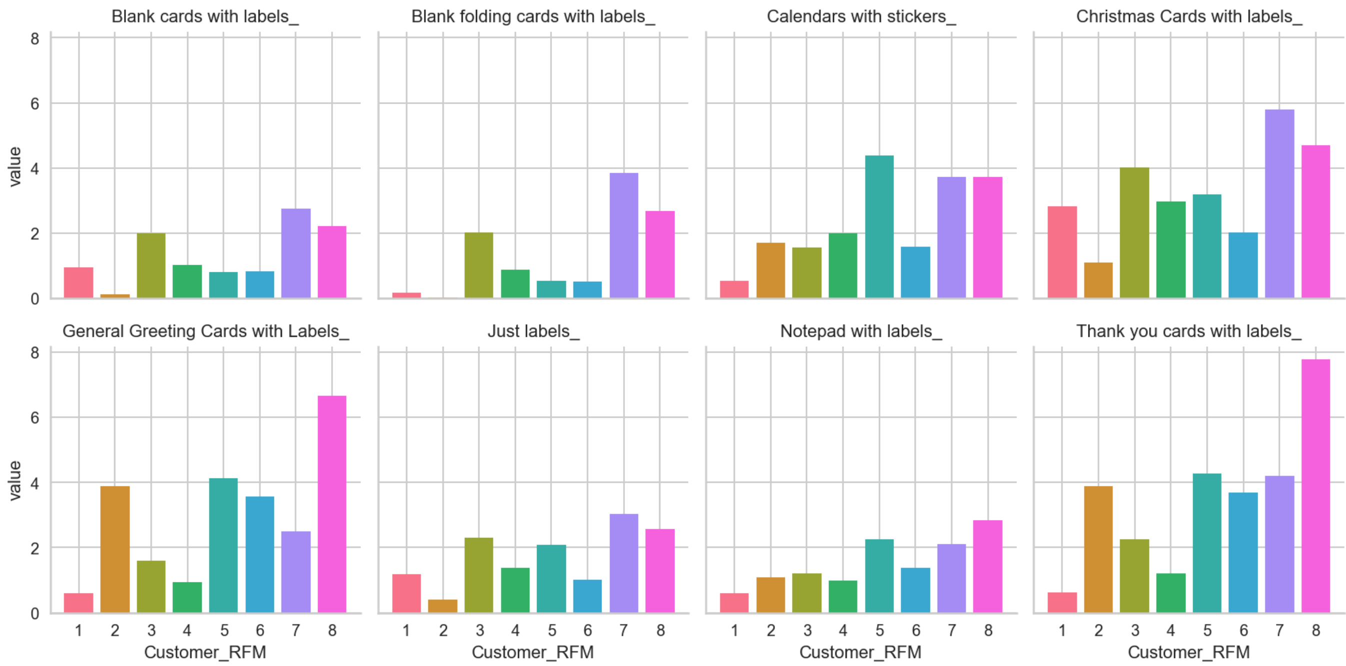


Above we have a Demographic Analysis, with the behavior of each cluster according to Gender (F for Female, M for Male, U for Unknown), Income, Wealth and Age. The flags Major and Pepstrfl had very distinctive behaviors, with Major in the groups 3,5,7 and 8 (indicating the Major donor flag was done a long time ago by PVA, since the best donors are now moving on the lower scales of the RFM analysis).

In the visualization below, we can see how much each cluster has donated (Monetary), and how often they have done so (Frequency), indicating the high value obtained from donors of clusters 7 and 8.

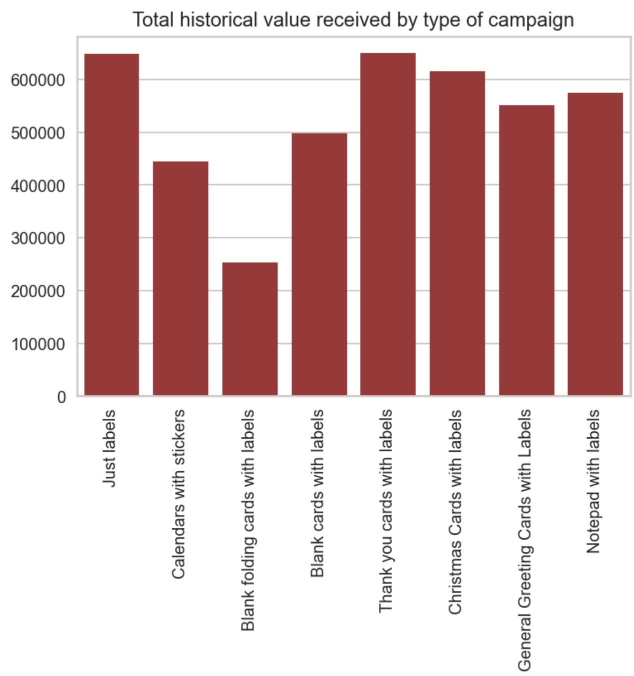


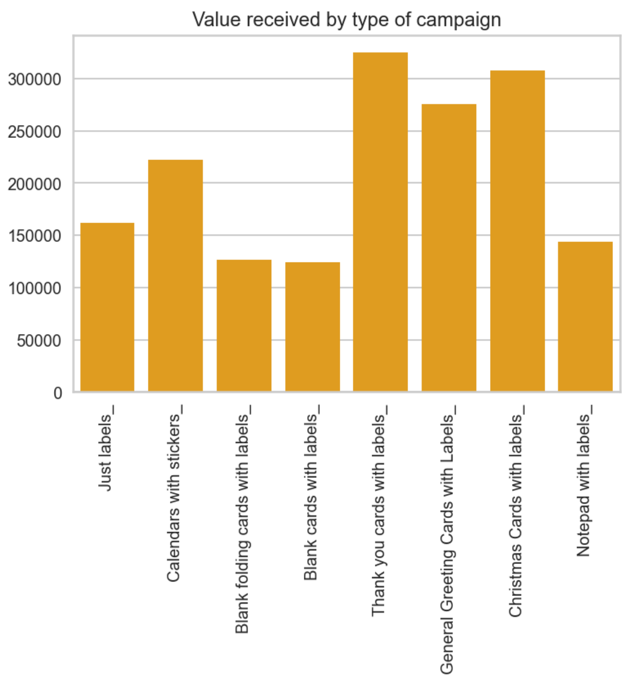
Regarding the type of campaign, we have the following information:



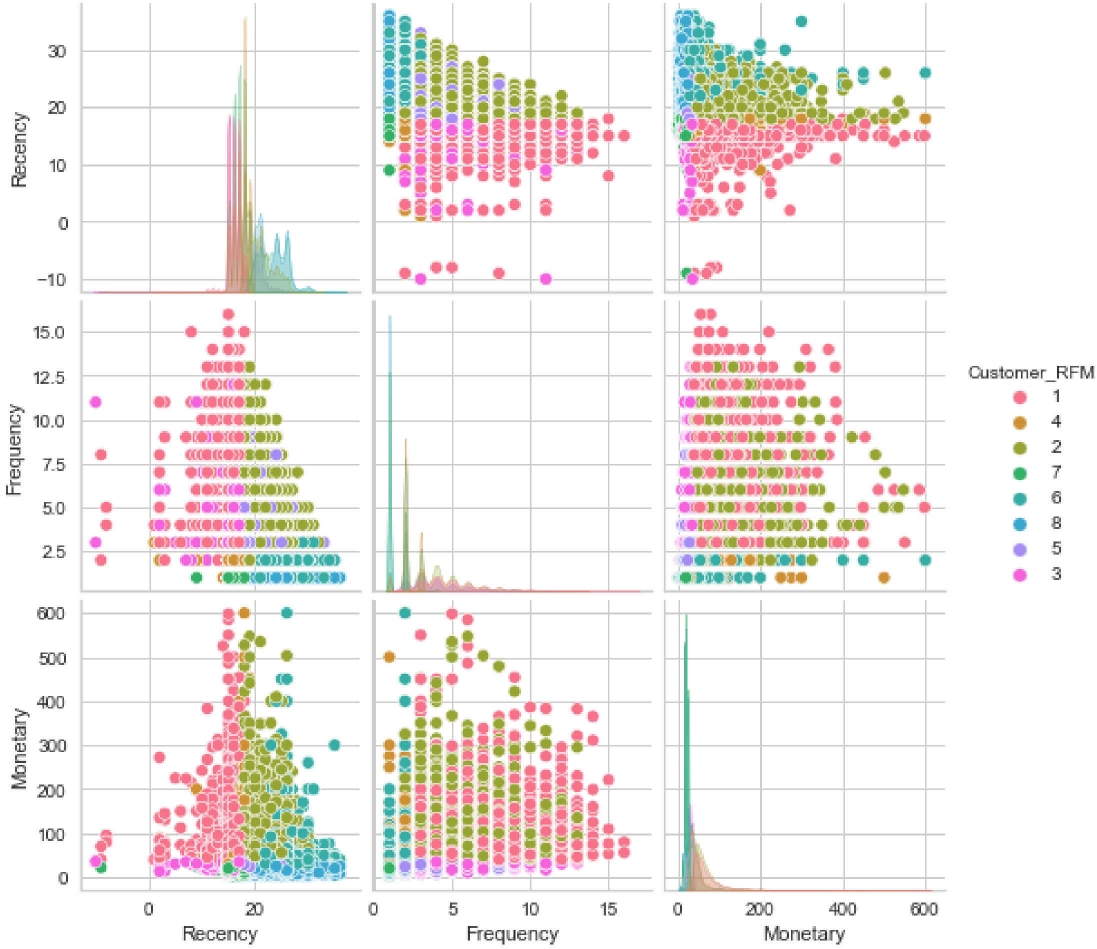
The graph above shows how each cluster reacts to different types of campaigns. And we can see clear different behaviors according to the received material. Thank you cards and Christmas cards had very good responses comparing to Blank cards, for example.

We also can analyze the total amount received according to the campaign that was done (in brown). And perhaps more interestingly, analyze the amount received dividing by type of campaigns (in yellow), since some of them ran more frequently than others.

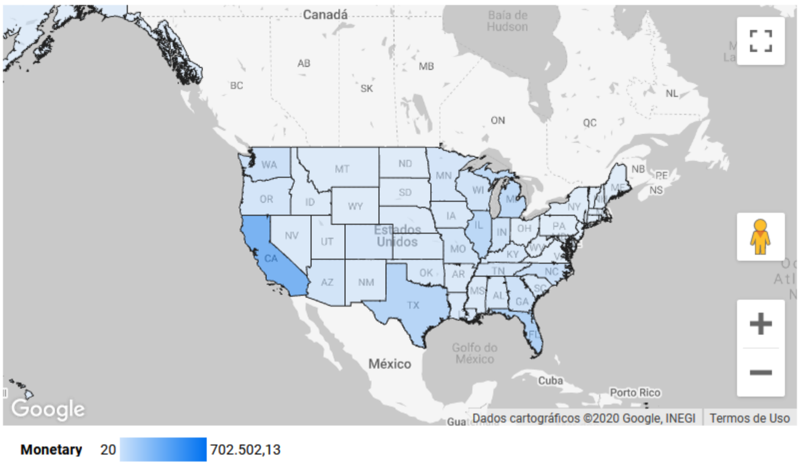


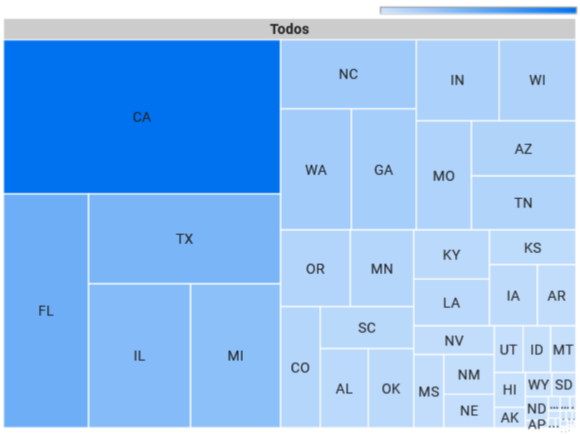


Here we can see the pairwise distribution of variables RFM. The clusters are characterized by colors.



Using Google Cloud Studio[9], we came up with a map showing the dispersion of the amounts of Monetary by states, and a *TreeMap* graph next to it showing the importance of states regarding total value donated. By it, we can see how California is the leader in donations, with Florida, Texas, Illinois and Missouri following.





After implementing the clusters in PVA´s database, a lot of information has come to the surface that can be used by the marketing and strategic departments of the NGO to gather more donors, work with the ones that have churned, and focus the campaign efforts on type of gifts that return more value (like Thank you cards, and Christmas cards).

Let´s perform a quick analysis of the clusters we came up with.

[9] To access the interactive map, go to https://datastudio.google.com/reporting/2f05aa41-144d-4eaa-8d21-8b1aa6ea7f11

	Recency_score	Frequency_score	Monetary_score	
	mean	mean	mean	Total
Cluster_RFM				
8	4.56	4.54	4.59	13.70
7	2.14	4.43	4.59	11.17
6	4.58	4.03	2.39	11.00
5	3.75	2.45	3.73	9.92
4	2.43	3.98	2.28	8.69
3	1.35	2.55	3.50	7.40
2	4.01	1.41	1.40	6.82
1	1.44	1.36	1.36	4.16

Cluster 8 – Champions

This cluster has the best donors in every possible way. They should be treated with special care and attention, with some of them receiving personalized "Thank you" calls after donating. We recommend sending a lot of information from how their donations have impacted real people´s lives, since they are correlated with gratitude: this segment has performed specifically well with Greeting cards, Thank you cards and Christmas cards. If treated well, this cluster can become "evangelists" and spread the word about PVA´s work and importance with their friends and family.

Cluster 7 – Great donors, slipping away

This cluster should be the center of attention of the marketing team. Why has this donor spent so much time without any contributions? It´s a high-value, frequent donor that should be treated in ways to understand why they are churning. Responds well to campaigns in general, but specially Christmas cards.

Maybe PVA should try a different approach: instead of sending gifts that remind people of money, they should send pictures of Veterans who they have helped, their history, some interviews, or even a Thank you note written by some veterans who have been helped by the donors contribution.

We will use some highlights collected from the Domain Knowledge Summary[10]:

- To illustrate how incentives might influence a relationship, imagine a situation in which the person you have a burgeoning romantic relationship with gives you a $20 bill before asking you to attend a family function with them. We predict that the mere introduction of money is likely to lead you to question the nature of the relationship. But what if, rather than cash, your partner gives you a gift like a bottle of wine or heartfelt greeting card?
- The plethora of diverse gifts that charities send to potential donors can be classified into two groups: monetary and nonmonetary.
- People following communal norms are motivated to care for others and are willing to incur costs to do so, regardless of whether they will receive anything in return. In comparison, exchange norms are those in which benefits are given with the expectation of receiving comparable benefits in return.
- Activating the concept of money leads people to infer that they are in a businesslike or exchange relationship and behave as though they are interacting with a business party. Nonmonetary gifts are preferred over monetary ones in communal

[10] Please read more about this approach on **Coins Are Cold and Cards Are Caring: The Effect of Pregiving Incentives on Charity Perceptions, Relationship Norms, and Donation Behavior** - Yin et al, 2020.[10], on the fourth page of our Domain Knowledge Summary appended to this Report.

relationships, as they do not evoke marketplace norms in a payment context (Heyman and Ariely 2004) and are consistent with communal norms.

Cluster 6 – Loyal donors, that donate a small amount

PVA has already created loyalty. They should try different types of General greeting cards, since this is the gift that returns more value by campaign. The “Major” flag indicates this donor was once one of the best donors PVA had, so maybe there´s space to win him back and increase the amount of each donation.

Cluster 5 – Faithful donors, right in the middle

Since they already make small repeated donations, there´s space to try a monthly-basis charge directly on their bank accounts, to reduce friction and build a solid income of value.

Cluster 4 – Have donated many times in the past, low value, but keep returning

This group has a good history of donations (Frequency) but are not the best customers because their Recency and Monetary are in the middle/lower ground, hovering around 2.5 in the RFM scale. They responded fairly on the same way no mater what gift was received by mail, indicating that maybe it´s not the gift itself that prompts the donation, but the value of helping PVAs mission.

Cluster 3 – Churned best donors

Those were good donors, who had frequent and large donations, but that stopped donating a while ago. It´s challenging to win them back, but the high value of this donors makes the effort worth pursuing. The marketing team should analyze their preferences based on earlier transaction data and try to approach them with a different technique/channel, since most direct mailing pieces get tossed in the trash bin unopened (Danner, 2020).

Cluster 2 – New donors

Continue to nourish the relationship with this group, since they are promising and can lead to anything between a Champion and a Faithful Donor. New donors responded very well to Christmas cards with labels.

Cluster 1 – Hibernating

Those donors purchased a long time ago, and the number of donations and amount are low. For this analysis they are to be considered the “worst donors”. Try winning them back with innovative approaches, different type of gifts or campaign. But don´t waste a lot of money with physical gifts being mailed to them. Digital requests which cost less are better suited, since most likely they won´t return easily.

That concludes our analysis of PVAs list of donors.

References:

Arthur, D. & Vassilvitskii, S. (2007). K-Means++ The advantages of careful seeding, Proceedings of ACM-SIAM Symposium on Discrete Algorithms, pp. 1027-1035, ISBN:978-0898716245, New Orleans, January 2007, Society for Industrial and Applied Mathematics, USA.

Berry, Michael JA, and Gordon S. Linoff. Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons, 2004.

Birant, D. (2011). Data mining using RFM analysis. In Knowledge-oriented applications in data mining. IntechOpen.

Bult JR, Wansbeek T (1995). Optimal seleciton for direct mail. Market Sci., 14(4): 378-394.

Cover T, Hart P (1967) Nearest neighbor pattern classification. Information Theory, IEEE Transactions on 13: 21–27.

Danner, G. (2020, December 24). Why Your Direct Mail is Really Being Thrown... NextPage. https://gonextpage.com/2012/11/29/why-your-direct-mail-is-really-being-thrown-away/

Jain, Anil K., Murty M. Narasimha, and Flynn, Patrick J.. (1999). "Data clustering: a review." ACM computing surveys (CSUR) 31.3

Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2015). Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT press.

Han, Jiawei; Kamber, Micheline and Jian Pei. "Data mining concepts and techniques: Third Edition." The Morgan Kaufmann Series in Data Management Systems, 2011.

Ng, P. (2017). dna2vec: Consistent vector representations of variable-length k-mers. arXiv preprint arXiv:1701.06279.
Outlier detection with Local Outlier Factor (LOF) — scikit-learn 0.24.0 documentation. (2020). Scikit Learn. https://scikit-learn.org/stable/auto_examples/neighbors/plot_lof_outlier_detection.html

Roweis, S. (1997). EM algorithms for PCA and SPCA. Advances in neural information processing systems, 10, 626-632.

Swearingen, C. (2009). 101 Powerful Marketing Strategies for Growing Your Business Now!, SmallBiz Marketing Services, pp. 24-27.

Wei, Jo-Ting, Shih-Yen Lin, and Hsin-Hung Wu. "A review of the application of RFM model." African Journal of Business Management 4.19 (2010): 4199-4206.

APPENDIX A

DOMAIN KNOWLEDGE SUMMARY

"FUNDRAISING USING DIRECT MAIL"

Highlights from research made on NovaDiscovery (EBSCO Host)

A Veteran Direct-Marketing Fund Raiser's Bright View of the Future - Blum, Debra E. 2004[1]

When Max Hart began work at the Disabled American Veterans in 1969, the organization was best known for the gifts it included in its mailed fund-raising appeals -- miniature license plates for motorists' key rings that were personalized to match the tags on each recipient's car.

The I-dento Tags, as the trinkets were called, had been helping the Cold Spring, Ky., organization bring in the money it needed for nearly 20 years. But by the end of the 1960s, donations were starting to slip. A growing number of states allowed motorists to keep the same plate numbers year after year, which made getting a new I-dento Tag each year a less appealing offer. Also, notes Mr. Hart -- who retired as head of the group's direct-mail fund-raising efforts in December -- Disabled American Veterans wasn't doing much to build loyalty among previous donors. The organization wasn't even keeping track of people who gave money, he says.

"They'd make the mailing, and then simply shred the list," says Mr. Hart, 70. "Year after year they just were blindly sending out millions of I-dento Tags."

Why does Disabled American Veterans rely on direct mail?

We've never been able to make anything else work. We've taken stabs at things over the years. Back in the early days, before my time, they tried door-to-door. Telemarketing, by choice, we've never done. We have a strong negative bias against it, don't like it, and I think that policy of not having telemarketing has served us well over the years.

What's the biggest challenge the direct-marketing industry will face in the near future?

I've always thought, going back years and years, that privacy is the greatest challenge we face, and will continue to face. It is an area that could literally put us out of business.

What is the privacy issue all about?

The crux of the issue is that if the direct-marketing industry were required to obtain positive consent from the individual in order to mail them -- what is known as opt-in -- it would just kill us. If we had to obtain

---

[1] A Veteran Direct-Marketing Fund Raiser's Bright Viewof the Future. By: Blum, Debra E. Chronicle of Philanthropy. 3/18/2004, Vol. 16 Issue 11, p58-58. 1p. 1 Color Photograph. Abstract: Interviews Max Hart, a fund raiser at the Disabled American Veterans in the U.S., on the nonprofit organization and the direct-mail industry. Reason Disabled American Veterans rely on direct mail; Challenges in the direct-marketing industry; Privacy issue of direct-marketing industry.
(AN: 12602500)

permission to solicit the individual or to provide their name to other organizations, most people would not opt-in. It would be like saying, Do you want more fundraising? And naturally the answer to that would be No. Donor files, which are our most-productive lists, would quickly dry up, and as those universes dried up, attrition would take our files down, and we would not have the new donor input to maintain the levels we have today, and that would very quickly put us out of business.

Is direct marketing all about statistics?

The data, and the statistics, and the models let you get very close to knowing what is going to happen with a mailing. There's no question you have to understand the numbers. But the statistics and the information reports need to be an extension of your mind and body. You have to have a feel for what is going to work.

To recruit donors, charities revamp mailings; try new tactics - Hall, Holly & Gardyn, Rebecca. 2004[2]

Even in good times, mailings to prospective donors typically don't prompt more than 1 percent of recipients to give, so charities don't recover their costs for the mailings until the new donors make repeat gifts over the years. While most fundraisers have tried other ways to attract new donors, they have yet to discover a working alternative to direct mail. E-mail solicitations aimed at new donors, for instance, are often considered spam. Other approaches, such as 30-minute television programs used by charities seeking donors to "sponsor" needy children, are usually too expensive to justify dropping direct mail entirely.
Still, many charities are finding ways to make their direct mail more efficient or supplementing it with other methods.

Among the tactics:

Using statistics to predict returns. A growing number of charities are hiring outside companies to analyze characteristics of their direct-mail donors--such as how recently they gave, type of appeals they respond to, average gift, number of gifts per year, and total amount given, as well as other demographic characteristics gleaned from public records, such as age, sex, and home ownership. Using a process known as "predictive modeling," the companies then use those characteristics to create a statistical model that enables charities to predict which people on a mailing list are most like their best donors and therefore most likely to give.
The Nature Conservancy, which sends appeals to 20 million potential donors a year, has increased the size of the average gift made in response to some mailings by as much as 70 percent over the last two years by using a predictive model.
"These donors tend to be more loyal over time, and they give more frequently and upgrade more quickly,"

---

[2] Seeking Better Returns. By: Hall, Holly; Gardyn, Rebecca. Chronicle of Philanthropy. 9/30/2004, Vol. 16 Issue 24, p25-27. 3p. 3 Color Photographs. Abstract: Presents information on the efforts of charities in the U.S. to attract donors through direct mail. Analysis of direct-mail returns conducted by Cambridge, Massachusetts-based direct-marketing consulting firm Target Analysis Group; Alternative approaches to direct-mail to attract donors including the use of statistics to predict returns; Advantages of paying for television commercials. INSET: Charities Urged to Stop 'Crass and irrelevant' Appeals. (AN: 14728696)

says Lisa Steen, the charity's senior manager of new-member acquisition.

That, she adds, enables the Nature Conservancy to recover its costs more rapidly.

Because most recruiting mailings are also sent to previous donors who have stopped giving, modeling can also help fund raisers figure out how likely or unlikely those people are to renew their support.

Catholic Relief Services, which sends 11 million appeals to former and potential donors each year, has used modeling over the past 15 months to reduce the number of people it mails to each year. "It's not the holy grail, but we see modeling as a way to screen out the bottom 20 percent," says Kevin Whorton, who oversees mailings for the charity. As a result, Catholic Relief Services has managed the economic downturn better than many. "We're holding steady," Mr. Whorton says.

While modeling can save money for charities with big mailings, it is less cost-effective for smaller organizations, experts say. A charity that sends 100,000 appeals to potential donors might reduce its costs by 10 percent to 15 percent or about $10,000, says Ray Grace, chairman of Creative Direct Response, in Crofton, Md. But because modeling generally costs $5,000 to $7,000, regardless of the size of the mailing list, the charity's savings would be negligible unless it sent millions of pieces.

"A lot of small- to medium-sized nonprofits think modeling is the answer, but it's not cheap enough to satisfy everyone's needs," says Mr. Grace.

Including gifts in mailings. Nearly 40 percent of all charity mail to prospective donors last year contained address labels, greeting cards, pins, and other small gifts, according to Paradysz-Matera, a New York company that assists nonprofit groups with direct mail, online marketing, and related services. The company, analyzing more than 6,000 different recruiting mailings by more than 1,000 charities from 2001 to 2003, found that the use of such gifts, often called premiums, was on the rise. Address labels were the most common, accounting for nearly 25 percent of the token gifts last year.

Mothers Against Drunk Driving, which has long enclosed premiums in letters it sends to 9 million to 12 million potential donors annually, has been doing better than most charities in the past 12 months by including multiple premiums in its appeals. Unlike most mailings to win new donors, which lose money, many of MADD's new appeals have more than covered their costs.

One version of the new approach, an elaborate box mailing about the size of a legal envelope but a quarter inch thick, includes a pen bearing the advocacy organization's name, a red ribbon reminding people not to drink and drive that recipients can attach to their cars, a membership card, and a calendar. The mailings have drawn gifts from 3.5 percent to 8 percent of recipients, says Bobby Heard, MADD's director of marketing and development. "We have a unique set of packages that get more attention in mailboxes, and our acquisition has been phenomenal."

But most charities aren't that fortunate. Using premiums can be tricky because they increase the cost of a mailing and typically bring in smaller gifts than mailings without them.

What's more, people who respond to premiums tend to be less loyal than other donors; even when they do make repeat gifts, they are apt to do so only in response to another premium.

Still, some groups like the American Heart Association have successfully used premiums to offset shrinking

returns from mailings without them.
About four years ago, when response to its long-used appeal, a simple letter with a return envelope, started to dip, the association began replacing some of the letters with solicitations that included address labels or other premiums. Now more than half its donor-recruiting appeals contain such enclosures.

Tailoring appeals to men and women. Last year, when fundraisers at the National Trust for Historic Preservation decided that they needed to win more new donors, they began testing alternatives to the organization's 20-year-old appeal: a four-page letter accompanied by a formal-looking invitation.
In testing a less expensive two-page letter, Dolores McDonagh, vice president for membership development, says the organization discovered that when both men and women were asked to give $15 to join the trust and then add on some money to support the organization, men gave $23.55 on average, while the women's average gift was $18.44.
Using those findings and results from additional tests, the organization has developed new "male" and "female" appeals. While both are two pages long and request a minimum of $15, the women's letter features a more colorful design and personal tone. In initial tests, although the gender gap remains, both men and women are giving more than they did in response to the 20-year-old appeal: The average gift from new female donors is now 25 percent higher, while men's is 37 percent higher, on average. Based on those returns, the National Trust recently mailed nearly 850,000 appeals with different approaches for men and women.

Recruiting donors in person. Not all nonprofit groups believe in-person recruitment is cost-effective. Greenpeace USA, for example, gave up direct-mail appeals a year ago in favor of sidewalk solicitors, but is now doing a study to determine if donors recruited that way give enough over time to justify the cost.
"Our preliminary figures are telling us there is a high attrition rate with these donors, so we're checking this out," says Rick Gentry, the charity's acquisition coordinator. He says that Greenpeace has hired a new fundraising director from the international organization Oxfam, in Britain, who is planning year-end tests of a new mailing to potential donors. "He's trying to put the 'wow' factor back into Greenpeace," Mr. Gentry says, "and it could be that our old package was stale."

Coins Are Cold and Cards Are Caring: The Effect of Pregiving Incentives on Charity Perceptions, Relationship Norms, and Donation Behavior - Yin et al, 2020.[3]
Charities often include low-value monetary (e.g., coins) and nonmonetary (e.g., greeting cards) pregiving incentives (PGIs) in their donation request letters. People are more likely to open and read a letter containing a monetary PGI (vs. a nonmonetary PGI or no PGI). In addition, monetary PGIs increase response rates in donor acquisition campaigns. However, the return on investment for direct mail campaigns drops

---

3 Yin, Bingqing (Miranda), et al. "Coins Are Cold and Cards Are Caring: The Effect of Pregiving Incentives on Charity Perceptions, Relationship Norms, and Donation Behavior." Journal of Marketing, vol. 84, no. 6, Nov. 2020, pp. 57–73, doi:10.1177/0022242920931451.

significantly when PGIs are included. Furthermore, average donations for appeals with a nonmonetary PGI or no PGI are similar, while those with a monetary PGI are actually lower than when a nonmonetary PGI or no PGI is included. This is because monetary PGIs increase exchange norms while decreasing communal norms.

- Over 90% of nonprofit organizations use direct mail as one of their primary fundraising methods (Non-Profit Research Collaborative 2016). However, most direct mail campaigns achieve modest success, with an average response rate of 1% to 3.7% (Data & Marketing Association 2015)
- To illustrate how incentives might influence a relationship, imagine a situation in which the person you have a burgeoning romantic relationship with gives you a $20 bill before asking you to attend a family function with them. We predict that the mere introduction of money is likely to lead you to question the nature of the relationship. But what if, rather than cash, your partner gives you a gift like a bottle of wine or heartfelt greeting card?
- The plethora of diverse gifts that charities send to potential donors can be classified into two groups: monetary and nonmonetary.
- People following communal norms are motivated to care for others and are willing to incur costs to do so, regardless of whether they will receive anything in return. In comparison, exchange norms are those in which benefits are given with the expectation of receiving comparable benefits in return.
- Activating the concept of money leads people to infer that they are in a businesslike or exchange relationship and behave as though they are interacting with a business party. Nonmonetary PGIs are preferred over monetary ones in communal relationships, as they do not evoke marketplace norms in a payment context (Heyman and Ariely 2004) and are consistent with communal norms.
- Thus, monetary PGIs could be effective in promoting initial engagement with a piece of mail, especially among individuals who are unfamiliar with the charity. *(Research on this article corroborates with this)*.

Direct Mail Response Rate Guidelines - Successful Fundraising © 2018[4]

What's a typical response rate to a direct mail appeal? Because of multiple factors that can impact results (e.g., timing, audience makeup, the message, etc.), that's tough to predict.

With that in mind, however, here's one guide to which you can refer:

- For donor acquisition mailings, the response rate may fall between 0.5 percent and 2.5 percent.
- For renewal solicitation mailings the response rate may fall between six percent and 12 percent.

Direct Mail Dos and Don'ts - Successful Fundraising © 2018

Are you satisfied with the response rate of your direct mail appeals?

If not, consider these direct mail dos and don'ts:

[4] 'Just Getting Started? Find Yourself One or More Mentors' 2018, Successful Fundraising, vol. 26, no. 2, p. 6, viewed 14 December 2020, <http://search.ebscohost.com/login.aspx?direct=true&db=edo&AN=ejs44407463&lang=pt-pt&site=eds-live&scope=site>.

- **Do:** Pay strict attention to your offer. Responses hinge on that. If your offer is too vague or uncompelling, your results will reflect that.
- **Do:** Constantly test and measure results. All sorts of variables can impact results — time of year, your target audience, the nature of your request. Try mailing to a smaller audience to determine what's working and what's not.
- **Don't:** Be too shy about multiple mailings. If your appeals are achieving the desired result, mail more often. You owe it to those who want to invest in your cause.
- **Don't:** Fail to segment your mailing list for particular types and frequencies of appeals. Too many charities send out one or two general appeals during the year to their entire mailing list. Wrong. Break your database into various types of groups and direct offers that most appeal to each group based on their interests, demographics, etc.

The effect of effectiveness: Donor response to aid effectiveness in a direct mail fundraising experiment
- Karlan, Dean & Wood, Daniel H., 2017.[5]

Freedom from Hunger implemented a test of its direct marketing solicitations, varying letters by whether they include a discussion of their program's impact as measured by scientific research. The base script, used for both treatment and control, included a standard qualitative story about an individual beneficiary. Adding scientific impact information has no effect on average likelihood of giving or average gift amount. However, we find important heterogeneity: large prior donors both are more likely to give and also give more, whereas small prior donors are less likely to give. This pattern is consistent with two different types of donors: warm glow donors who respond negatively to analytical effectiveness information, and altruism donors who respond positively to such information.

- We collaborated with Freedom from Hunger (FFH) to conduct two waves of direct-mail marketing to prior donors. FFH is a US- based nonprofit organization that provides technical advisory services to microfinance institutions (MFIs) in developing countries.
- In the first wave, the control group received an emotional appeal focused on a specific beneficiary, along with a narrative explaining how FFH ultimately helped the individual. The treatment group received a similar emotional appeal (trimmed by one paragraph), with an added paragraph about scientific research on FFH's impact. The second wave was similar in design, except that the treatment group narrative included more details on the research, including a brief discussion about the benefits of randomized trials.
- We find that average donation behavior does not change when previous donors are presented with evidence of the charity's effectiveness in achieving its goals. However, we find that the aggregate effect masks different responses by small and large prior donors: larger prior donors, as measured by

[5] Karlan, Dean & Wood, Daniel H., 2017. "The effect of effectiveness: Donor response to aid effectiveness in a direct mail fundraising experiment," Journal of Behavioral and Experimental Economics (formerly The Journal of Socio-Economics), Elsevier, vol. 66(C), pages 1-8.

the amount given in the last donation before the experiment, donate more and small prior donors donate less in response to being told about the scientifically-measured effectiveness of the charity.

- Following a paradigm put forward by Kahneman (2003) we explore a model of giving that incorporates two motivations for giving, **altruism** (akin to Kahneman's System II decisions, which are deliberate, effortful, reasoned and focused on impact) and **warm glow** (akin to Kahneman's System I decisions, which are intuitive, effortless and reactive). The model makes an important prediction: **that individuals driven by altruism, holding all else equal (such as wealth and education), will respond favorably to information about the effectiveness of a particular charity, whereas those driven by more emotionally-based triggers may actually reduce giving.**
- We find that presenting positive information about charitable effectiveness increases the likelihood of giving to a major U.S. charity for large prior donors, but turned off small prior donors.

Building Donor Relationships: An Investigation into the Use of Relationship and Database Marketing by Charity Fundraisers - LAURA WEIR atid SALLY HIBBERT, 2000.[6]

The findings suggest that using a fund-raising database is not enough, by and of itself, to improve effectiveness in building donor loyalty. Rather, relationship and database marketing principles need to be applied in the use of the database to achieve greater loyalty from the charity's donors.

- It has been demonstrated that it takes 12-18 months before a newly acquired donor will become an asset to the charity in terms of net income [Gaffney, 1996] which suggests that charities should aim to extend relationships with donors beyond an 18 month period.
- More general explorations of why people donate to charity have produced various taxonomies of motives for giving. Among motives that have been identified are: self-esteem, recognition from others, habit, peer pressure, concern, self-preservation/fear, guilt and because the person was asked [see Bruce, 1994; Burnett, 1992; Kotler, 1982]. In addition to appreciating these motives, fundraisers will increasingly need to understand what encourages donors to maintain and increase their commitment to a charity over the long term and, equally important, the factors that inhibit the development of donor commitment [see Alorbi, Hibbert, Home and Tagg, 1998].
- With regard to communication appropriate for appealing to donor motives, it was traditionally suggested that the benefits of supporting a particular charity should be selected and promoted according to the traditional marketing tool AIDA: attention, interest, desire, action [Strong, 1925]. However, this approach implies a transactional view of marketing exchanges. It does not account for the development of individuals' commitment to a charity and the evolution of motivation that accompanies that development. More recently, Shultz et al [1994] have suggested that organisations

[6] Weir Laura & HIBBERT SALLY (2000) Building Donor Relationships: An Investigation into the Use of Relationship and Database Marketing by Charity Fundraisers, The Service Industries Journal, 20:2, 114-132, DOI: 10.1080/02642060000000023

adopt an integrated marketing communications approach. This approach proposes that organisations use their database to distinguish among customers on the basis of their existing level of commitment to the organisation and that separate communications objectives and strategies are developed for each of the groups. The role of the database in enabling integrated marketing communications in the charity context is stressed by Burnett [1992] who claims that it is at the core of relationship

- Among the articles advising on effective database use, Rodd [1996] emphasises that Pareto's '80:20' rule applies to fundraising, that is, that a small proportion of donors normally account for a large share of donated income. He, therefore, advocates that top donors should be managed with care and receive a larger proportion of fundraising attention. This said, there may be potential to build up the commitment on individuals who currently donate small amounts, and they should not be neglected.