



**ENTERPRISE BUSINESS ANALYTICS
INTERNSHIP FINAL PHASE REPORT**

Analysis of Donor Behaviour

Name-Shivangi Verma
ID-A0215504R

Table of Contents

1. Executive Summary.....	3
2. Introduction.....	4
3. Business Objectives and Data Analytic Objectives.....	3
4. The plan.....	6
5. Data Acquisition and Processing.....	7
6. Analysis and Modelling.....	10
7. Validation of Results.....	28
8. Project Conclusions and Products.....	29
9. Recommendations for Further Work.....	30
10. References.....	30
11. Appendix.....	31

1.EXECUTIVE SUMMARY

Giving.sg, the giving initiative of NVPC is a one stop online portal where people could donate, raise funds and signup for volunteering activities .It is a Non-profit organization promoting a giving culture in Singapore through volunteerism, philanthropy to build a city of good. NVPC facilitates partnerships with non-profits, companies, public sector bodies and individuals. They also conduct research on giving motivation and behavior, create Roadmaps and landscape in the giving sector.

One of the objectives of NVPC is to create an ecosystem of giving in the community. NVPC created platforms such as Giving.sg, which is an online portal for users to engage in donations and volunteering activities.

The project basically aims to find the attrition for over 5 Years in NVPC in a duration of 2015-2020 ,and observe it in various categories and segments also observe Year 2020 as a special case. In the analysis ,it was observed that the attrition in NVPC is indeed high of around 45-55%,However for Year 2019 it was around 27% ,Three main clusters were identified based on their giving behaviour ,Casual,Regular and Loyal .The Casual cluster shows higher attrition of 69 to 74%.

The metric Donor Lifetime value, which essentially defines the giving power or the quality of donation during there lifetime was analysed based on the attrition and donation amount observed.

The Cluster Loyal has very high giving power or DLTV and lesser attrition but the number of these type of donors is also comparatively less ,which can call for a change or be a signal for NVPC to strategize in the correct direction as the acquisition plan for donors is not very specific or targeted till now.

2.INTRODUCTION

The internship was in Key and Insights Team, It is headed by Fazlin Abdullah and the purpose of the team is to harness Knowledge and influence Sector through insights and also solve Business problems by providing Knowledge and guide NVPC through evidence informed decision making .

The Team works on three main levels namely People of Good, Organisation of Good, Leader of Good.

3.BUSINESS OBJECTIVES

On 11th March, World Health Organization(WHO) announced the COVID-19 outbreak to be a pandemic.

Many efforts and initiatives have come up within Singapore during this crisis period, with both people and corporates doing their parts to give and help others in their own ways.

Giving.sg reached a record \$13.6 million between April 1 and Sunday (April 19), as thousands of people responded to an online campaign encouraging Singaporeans to donate a part of or their entire Solidarity payment. Traffic on the site also surged to a high of 67,000 visitors on April 17 alone, far in excess of the daily average of 3,000 to 5,000. The NVPC said donations in 2020 so far, totaling more than \$22.5 million, have already exceeded that given in the first four months of each of the past four years, with people driven to online giving even as traditional fund-raising and volunteering events are curbed by safe distancing measures

So, Giving on Giving.sg saw a spike during this period. However it also saw a sudden dip as well (Fig 1)

Donations in Year 2020

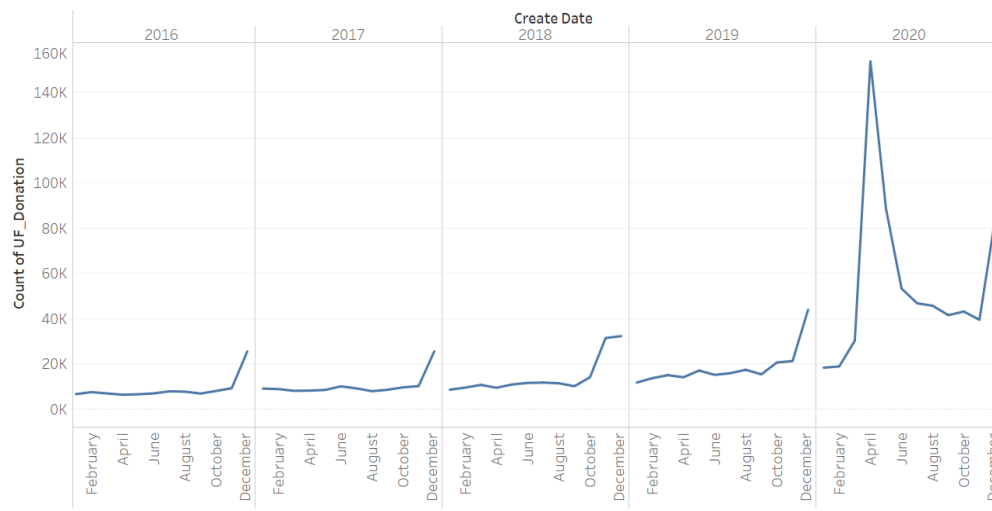


Fig 1

Based on Giving.sg data ,this project aims to investigate more specifically what are the main factors that resulted in this increase also the decrease via Time series or regression model, Including Segmentation analysis across causes or beneficiaries.

The project success criteria is the Validation metrics and the completion of business objectives in time and whether the analysis answers the business problem the company was searching.

ANALYTICAL OBJECTIVES

Preliminary approach

The objective to investigate Year 2020 via Time Series or Regression cannot be done as it is a Rare event with no previous data to leverage. Not many machine learning algorithms will be able to do that. However, there are different approaches of Extreme value theory, Neural Networks, Genetic algorithms which can be applied for detection of rare events.

But application of such algorithms cannot be done in a limited time frame of three months as it is still a new concept and requires research.

Final Approach

To analyze the sudden dip which is attrition of donors a drill down approach will be used where first the metrics will be established and analysis of donor attrition for a period of five years will be done and Year 2020 will be evaluated based on it as a special case, where the dip will be investigated.

Yearly NVPC has shown consistent growth in terms of money but there are no measures or steps taken to track the loss or to check if the donor is satisfied, Is the loyal donor base leaving and in order for sustainable development of NVPC .It should ensure a lasting relationship with donors for dependable source of funding. And Donor attrition is an important indicator of this lasting relationship. In order to do so, the following questions needs to be answered.

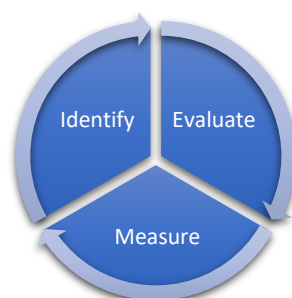
Identify-Which of the donors are leaving?

Evaluate-What is the magnitude of attrition?

Measure-What is the financial impact or lost opportunity?

The main steps involved to answer the above questions are:

- Evaluating attrition in Donor base and Calculating Donor Lifetime Value .
- Using Clustering techniques to divide Donor base into segments and observe metrics
- Reflect upon 2020 as a special case.



4.PLAN

The Project started on Feb 1 ,2021 and major part of time went in Business and Data understanding ,Also there were scope changes in between during the project. The plain laid out is as follows:

Project plan	Start	End	Goal Achieved
Business Understanding	1/2/2021	5/2/2021	This would help to help understand the business problem and find what exactly the company is trying to achieve
Identify scope	8/2/2021	10/2/2021	This would set boundaries on how and the direction in which the project needs to be worked upon
Set Objectives	11/2/2021	15/2/2021	With this a clear set of tasks can be identified and this would help in finding the solutions to business problems through analytical goals
Project plan	16/2/2021	16/2/2021	With a plan ,in place a clear understading of effort estimates and deadline can be placed which can help to track the progress of thr project
Data Understanding			
Data acquisition	18/2/2021	23/2/2021	Data acquisition will help in getting access to required data from source
Data Exploration	24/2/2021	5/3/2021	This is a crucial part and will help in understanding of Data better
Data Preparation			
Data cleaning	8/3/2021	12/3/2021	After the exploration,cleanning will be done to align with the objectives and tasks found earlier
Data Transformation	15/3/2021	19/3/2021	After the exploration,transformation will be done to align with the objectives and tasks found earlier
Feature engineering	22/3/2021	31/3/2021	Feature engineering will help in extracting relevant data
Feature selection	1/4/2021	5/4/2021	Feature selection will help in making decisions on what data to will the analysis be done upon
Modelling and Report submission	6/4/2021	30/4/2021	This will be the final phase which will be utilized to find answers to the goals and business objectives set

The initial Plan included Dashboard development and an in depth report on attrition analysis which was later changed to only an in dept report as a product

5.DATA ACQUISITION AND PREPARATION

Data Source: Giving.sg data warehouse

Giving.sg uses MySQL which is an open source relational database management system.

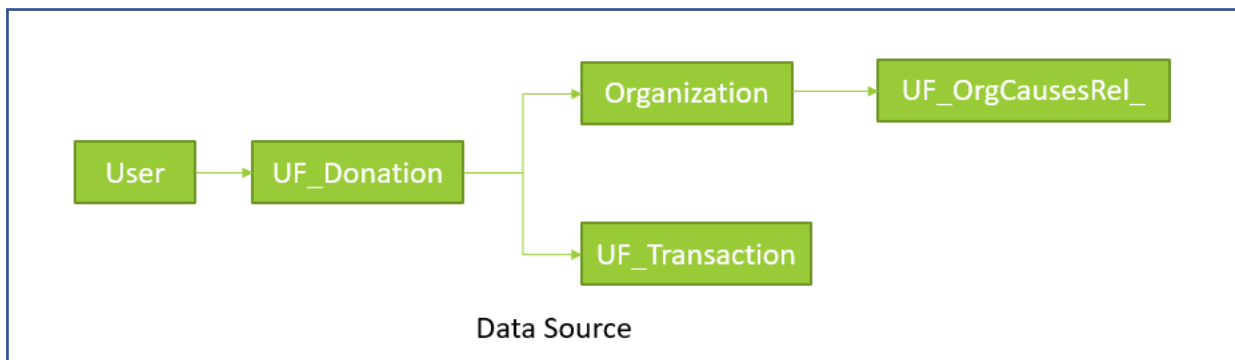


Fig 2

The Data Repository consists of 39 tables. The original data from the giving.sg website gets loaded onto a MySQL repository around two hundred tables. The essential data from these tables have been transformed and stored in the warehouse that is being used for day to day analysis. From the warehouse, Tables used are User_, UF_Donation, Organization_, UF_Transaction, UF_OrgCausesRel_.

Table Name	Description
User_	Details about User like ,Age,NRIC etc
UF_Donation	Details about donation receiving organisation ,donation amount
Organization_	Details about Organization like UEN,Ceo Name
UF_Transaction	User transaction details
UF_OrgCausesRel_	Table with Organization id and the causes it supports

Table 1 (Description of Tables)

These tables are useful for getting the information about the donor and the receiving organization.

The following issues exist in the extracted data: -

- 1.Variables which have more than 50% of values missing.
- 2.Duplicate variables for example User Age, DOB, Year of Birth are present.
- 3.Due to merging of Tables there are variables which are repetitive for example 'causes'.
- 5.Gender is recorded as M, F, Female, Male.

Data Preparation

Data Cleaning areas	Plan
Missing value Treatment	Cannot remove them totally for some important variables or use methods like mean ,median,mode or other ways of

	Machine learning methods as it can make the table loose its value .
Outliers	Treating them as a separate group
Removal of duplicate and unnecessary variables	Transforming it into a single variable 'Age'
	Removing unnecessary variables which are not necessary for project scope
Variable transformation	Variables like Postal Code,NRIC will be further used to transform into sector and Nationality
	Creation of a separate Date variable with a uniform data type
	Changing Gender values to uniform values of Male and Female only

Table 2 : Data Preparation methods

1. User Age ,Year of Birth ,DOB was extracted to form an Age column and values for Gender column is manually changed to Male, Female.
2. Number of Months with at least 1 donation,Number of Donations made, Amount range-These columns were formed from existing columns of Donation amount and donation id , and were used for clustering .

Tableau code for Column analysis for clustering

Number of Donations made

```
IF { FIXED [userId],YEAR([createDate (UF_Transaction))]:COUNTD([donationId])=0 THEN "0"
ELSEIF { FIXED [userId],YEAR([createDate (UF_Transaction))]:COUNTD([donationId])<5 THEN "1 - 4"
ELSEIF { FIXED [userId],YEAR([createDate (UF_Transaction))]:COUNTD([donationId])<10 THEN "5 - 9"
ELSEIF { FIXED [userId],YEAR([createDate (UF_Transaction))]:COUNTD([donationId])<20 THEN "10 - 19"
ELSEIF { FIXED [userId],YEAR([createDate (UF_Transaction))]:COUNTD([donationId])<30 THEN "20 - 29"
ELSEIF { FIXED [userId],YEAR([createDate (UF_Transaction))]:COUNTD([donationId])<40 THEN "30 - 39"
ELSEIF { FIXED [userId],YEAR([createDate (UF_Transaction))]:COUNTD([donationId])<50 THEN "40 - 49"
ELSEIF { FIXED [userId],YEAR([createDate (UF_Transaction))]:COUNTD([donationId])<100 THEN "50 - 99"
ELSE "100 and above"
END
```

Number of Months with at least 1 donation

```
IF NOT ISNULL([createDate (UF_Transaction)]) THEN { FIXED [userId],YEAR([createDate (UF_Transaction))]:COUNTD([z. Months Donations])}
ELSE 0
END
```

Amount range

```
IF { FIXED [userId], YEAR([createDate (UF_Transaction))]:SUM([z. Actual Donation])=0 THEN "0"
ELSEIF { FIXED [userId], YEAR([createDate (UF_Transaction))]:SUM([z. Actual Donation])<100 THEN "$10 - $99"
ELSEIF { FIXED [userId], YEAR([createDate (UF_Transaction))]:SUM([z. Actual Donation])<500 THEN "$100 - $499"
ELSEIF { FIXED [userId], YEAR([createDate (UF_Transaction))]:SUM([z. Actual Donation])<1000 THEN "$500 - $999"
ELSEIF { FIXED [userId], YEAR([createDate (UF_Transaction))]:SUM([z. Actual Donation])<3000 THEN "$1,000 - $2,999"
ELSEIF { FIXED [userId], YEAR([createDate (UF_Transaction))]:SUM([z. Actual Donation])<5000 THEN "$3,000 - $4,999"
ELSEIF { FIXED [userId], YEAR([createDate (UF_Transaction))]:SUM([z. Actual Donation])<10000 THEN "$5,000 - $9,999"
ELSEIF { FIXED [userId], YEAR([createDate (UF_Transaction))]:SUM([z. Actual Donation])<50000 THEN "$10,000 - $49,999"
ELSEIF { FIXED [userId], YEAR([createDate (UF_Transaction))]:SUM([z. Actual Donation])<100000 THEN "$50,000 - $99,999"
ELSE "$100,000 and above"
END
```

6.ANALYSIS AND MODELLING

First step is to identify metrics which will be base for further analysis. Dividing the donor base into broad Donor categories as follows: -

Category	Description
New	donors who never gave before
Recaptured	previously lapsed donors who gave again
Upgraded	donors who gave more than previous year
Downgraded	donors who gave less than previous year
Lapsed	donors who used to give previously but are inactive and are active again
Loyal	donors who have been giving continuously

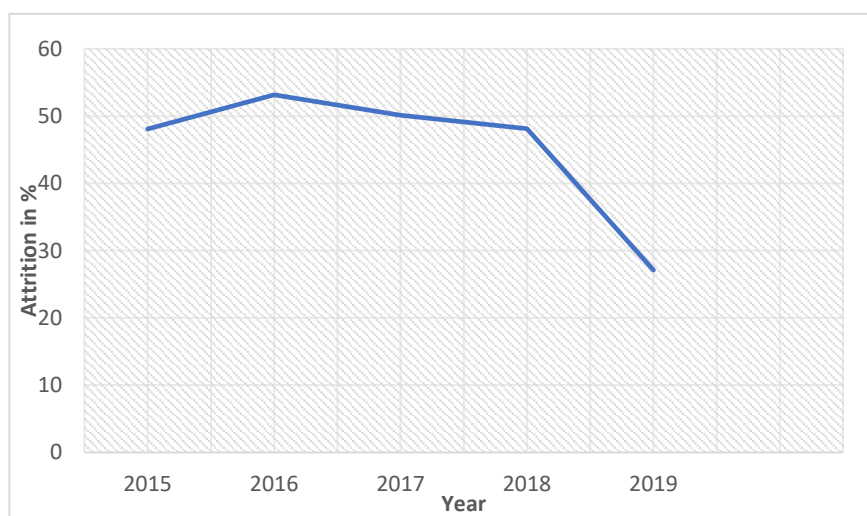
Table 3: Donor Category Description

With donor categorization ,the plan was to dig deeper into and analyse the attrition in various categories, but during ongoing analysis categories of Recaptured ,Lapsed ,Loyal was found not useful for analyse attrition upon and removed from the scope.

Donor attrition rate

To tackle the issue of donor attrition, the first step is to analyze what is the situation and what is the current attrition rate yearly, quarterly and monthly basis. The basic formula lies below,

$$\text{Donor attrition Rate} = \frac{\text{Donors who donated in last year} - \text{Donors who donated this year}}{(\text{Donors who donated last year})}$$



Year	Attrition- percentage
2015	48.07
2016	53.17
2017	50.12
2018	48.12
2019	27.12

Table 4: Overall attrition

Fig 3:Overall attrition

Knowing that ,Giving.sg officially started in 2015 after the closure of “SGCares” the donor counts were not captured correctly in the system.

The attrition observed till 2018 falls in the range of 45 to 55 % .If 2020 had not been an unusual year the attrition would have been around 45% for 2019 .The attrition in 2019 was 27% ,which says that a lot of donors from 2019 gave again in 2020,which will be further confirmed again in the cluster analysis done .

To calculate attrition basic methods of excel was used to compare userids .

Looking at the current scenario ,as COVID has still not been eradicated ,NVPC might expect High donations in 2021,they might also see lesser attrition as compared to other Years this year. The expected value of attrition in 2020 might not be as low as 2019 but will still be lower as compared from 2015 to 2018 as COVID has still not been eradicated.

Expected value metric analysis

- The number of observations is small to build any prediction, still various methods were tried to predict are as follows:
- Auto.arima()-Model output was ARIMA(0,0,0) which indicated this is white noise so cannot build a predictive model using this method.
- Fitting a curve-This resulted in overfitting.
- Fitting a line –This gave intervals which was still not useful to give a comment on expected value.
- Median –This is a Robust way to find expected value but the percentage came around 48% which was still not adequate to finalize a reliable estimation approach.

The expected attrition in 2019 will be around **48%**.

NVPC will save money through lower attrition and raise more money .It costs more time ,effort and resources to attract new donors than it does to retain the current ones.

Donor Lifetime Value

It has become common knowledge that retaining donors is one of the best ways to build and grow your fundraising operation. Acquisition is nice, but more and more development professionals are acknowledging the importance of retaining, stewarding and cultivating relationships with donors.

This is to measure how much revenue a single donor will generate from the moment they first donate, to the time they lapse. Essentially, it is the amount of monetary value a donor brings to the organization in the lifetime of their relationship.

Donor Lifetime Value=Lifetime of Donor * Average Donation Amount

Total Donation amount =DLTV * Number of Donors

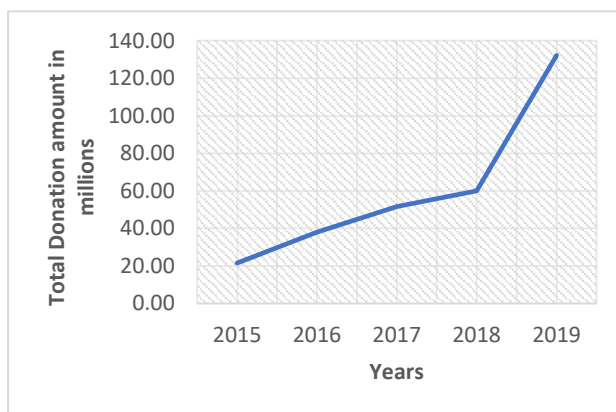


Fig 4:Total Donation Amount

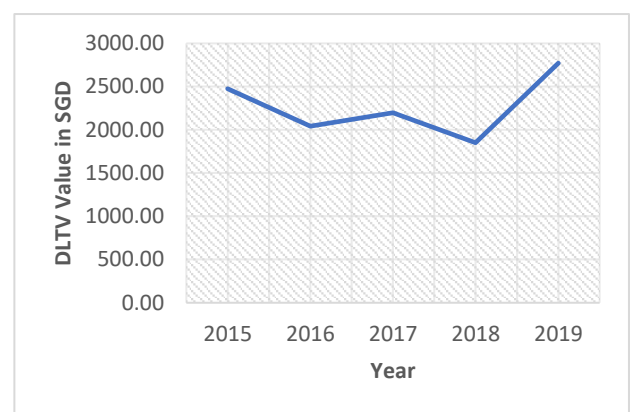


Fig 5:Donor Lifetime Value

It can also be seen as a measure of giving quality or giving power of a donor .Knowing this measure will help NVPC in what they can expect in terms of monetary value on an average at individual level during their lifetime

Years	2015	2016	2017	2018	2019
Attrition Rate	48.05	53.17	50.23	48.12	27.10
Average_donation_amount	1188.77	1086.66	1104.73	890.65	751.01
Lifespan	2.08	1.88	1.99	2.08	3.69
DLTV	2474.03	2043.75	2199.34	1850.89	2771.25
Number of Donors	8717	18610	23470	32471	47696
Total	21566093.84	38034121.87	51618580.73	60100366.06	132177760.00
In Millions	21.57	38.03	51.62	60.10	132.18

Table 5: DLTV and Total Donation amount calculation

Before discussing about DLTV, some important assumption have to be kept in mind , it is assumed that the attrition rate will remain uniform and also the average donation amount will remain uniform during their lifetime.

- The DLTV has a slow decrease since 2015 ,and it stays around a 1000 Singapore dollars mainly, but the number of donors every year also has been increasing, which indicates that even though the number of Donors has been increasing ,the giving quality has not grown.
- In 2019 ,there was a jump in DLTV ,this can be mainly explained by the lower attrition rate observed in that year
- The total donation amount signifies the donation amount in total which that particular cohort would have given before leaving fully.
- The total donation amount has a steady increase till 2018 ,but in 2019 the amount jumps to 132 million, this again can be due to the lower attrition rate.

Expected DLTV:2199.34

New Donor attrition

The New Donors were counted based on the comparison of User id across Years starting from 2010 .

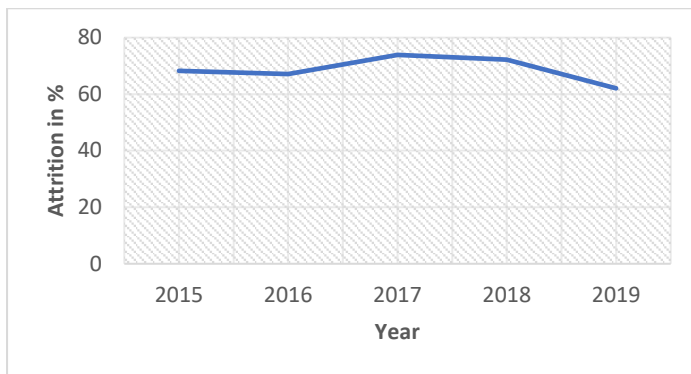


Fig 6 : New Donor Attrition

Year	Attrition Percentage
2015	68.21
2016	67.07
2017	73.88
2018	72.19
2019	62.06

Table 6: New Donor attrition

Years	2015	2016	2017	2018	2019
Attrition Rate	68.21	67.07	73.88	72.19	62.06
Average_donation_amount	681.01	685.49	685.09	427.89	739.254
Lifespan	1.47	1.49	1.35	1.39	1.61
DLTV	998.40	1022.05	927.30	592.73	1191.19
Total Number of Donors	5906	14005	14707	20744	30698
Total Donation	5896562.18	14313832.49	13637816.23	12295539.77	36567224.13
In Millions	5.90	14.31	13.64	12.30	36.57

Table 7: DLTV Calculation for New Donors

New Donor DLTV

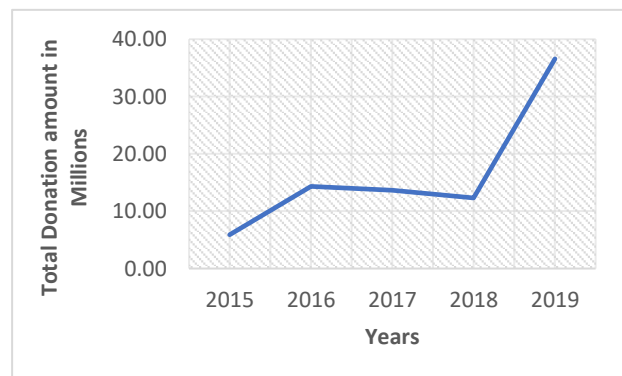
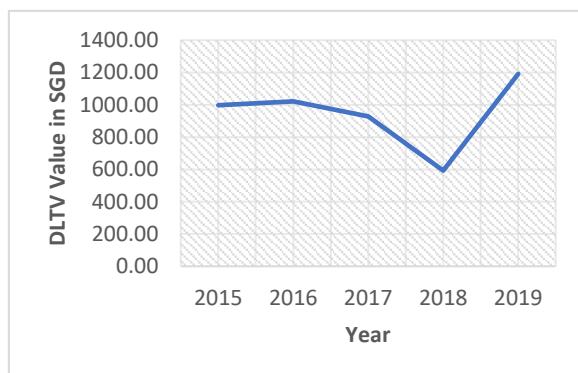


Fig 7: DLTV and Total Donation amount calculation for New Donors

- It can be noted that the attrition in the new donors is very high and lies in the range of 62 to 74% .
- This indicates that the donor acquisition process is possibly good, but the retention is suboptimal.

Looking at the DLTV value for New Donors, it can be said that it typically falls around the range of 1000 Singapore dollars ,however for the Year 2018 it took a plunge ,but then by 2019 it again revived ,The DLTV is mainly influenced with lower attrition rate and high Average donation amount.

The Total Donation amount has also been consistent to 5-15 million, except Year 2019 ,So a new donor has a potential of giving on an average a 1000 Singapore dollars and the total donation amount from new donors for a particular year can be up to 15 million during their lifetime.

For Year 2019 ,by the time ,the 30698 donors would have left NVPC completely they would have given a total of 36 million, The total Donation amount will only be higher for 2020 new donors.

Since the number of new Donors in NVPC is very large , The marketing team can strategize to increase the DLTV value expected from new donors either by placing strategies on retention or by planning to increase there giving quality.

Existing Donors

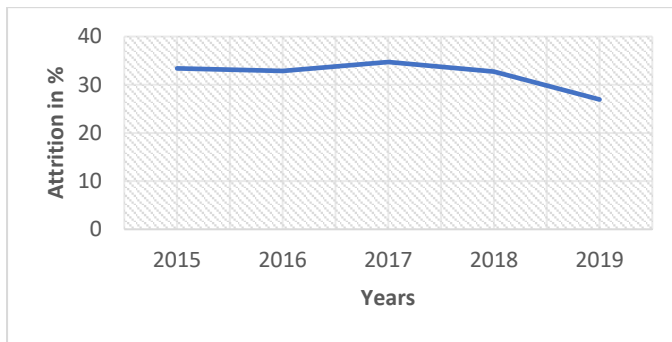


Fig 8: Existing Donor attrition

Year	Attrition-percentage
2015	33.404482
2016	32.884276
2017	34.708458
2018	32.736969
2019	26.933841

Table 8: Existing Donor attrition

Existing Donors DLTV

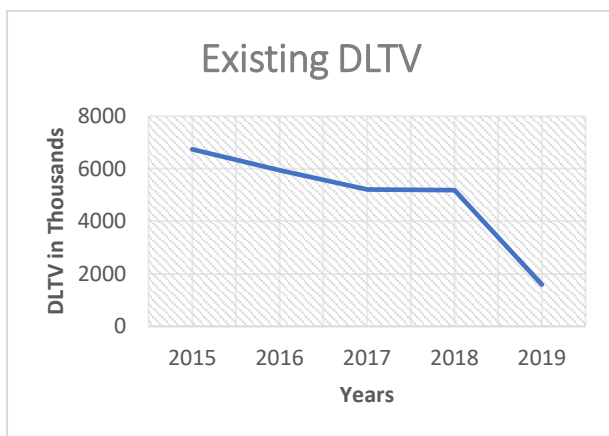


Fig 9: Existing DLTV

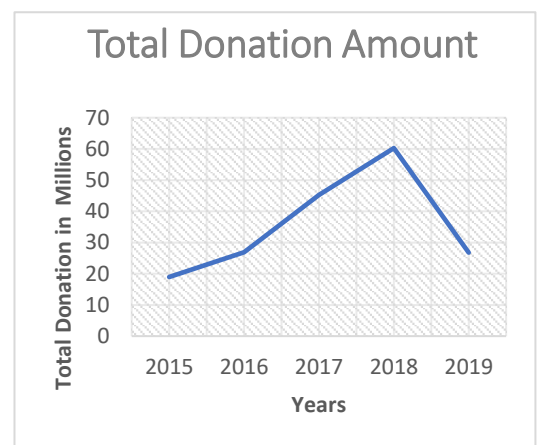


Fig 10: Total Donation

amount

Years	2015	2016	2017	2018	2019
Attrition Rate	33.4044824	32.88427562	34.70845817	32.73696886	26.93384148
Average_donation_amount	2251.62	1955.8	1808.96	1696.82	429.3
Lifespan	2.99361022	3.040967092	2.881142098	3.054650552	3.712801238
DLTV	6740.47265	5947.523439	5211.87081	5183.192149	1593.905572
Number of Donors	2811	4528	8678	11626	16792
Total	18947468.6	26930386.13	45228614.89	60259791.93	26764862.36
In Millions	18.9474686	26.93038613	45.22861489	60.25979193	26.76486236

Table 10: DLTV and Total Donation amount calculation for New Donors

Looking at Existing Donor metrics ,we can conclude that the donor attrition is less compared to what is observed in New Donors and the

DLTV has been decreasing since 2015, which indicates that the quality of giving of Existing donors have been decreasing .

The total Donation amount a donor can give during their lifetime is also increasing but again in 2019 it took a plunge.

In Year 2019 the DLTV Value is the lowest at 1593 SGD ,comparing to DLTV observed in 2019 for new Donors which is 1191 SGD ,indicates that in 2019 ,Existing Donor have lower giving quality than the New donors, also it can be said that in 2019 New Donors had a higher impact on the total donation amount that came in for NVPC.

Clustering

Regardless of the size of Non-Profit, Donor segmentation is an important strategic process, it enables to communicate better in ways that are meaningful to them as individuals.

The Segmentation will be done to further understand profiles on demographic variables like age, Nationality, Causes for a much broader outlook on donor behavior. The plan is to view the donor giving behavior in different level for instance in one segment which category leaves first, does one segment have all new donors etc.

Clustering was initially decided to be done based on RFM analysis. The RFM analysis is majorly done in Profit world, It is a marketing analysis tool to identify a company's or an organization's best customers by using certain measures. It is based on three quantitative factors:-

Recency is How Recent the purchase was ,Frequency is How frequently the purchase is made and monetary is the monetary value for the purchase.

The analysis closely follows the same concept ,except here the clustering is done only on Frequency and monetary values.

The columns used to cluster are Number of donations made, Number of months with at least 1 donation, Amount range,

Number of months with at least 1 donation-Throughout the year ,how many months are there where at least 1 donation is made by donor,

Number of Donations made-Throughout the year ,how many donations are done by a donor in total, the values are binned and converted to ranges as described in the table in appendix.

Amount range-The amount range in which a donor gives donations. The bins for above values are described more in the Table in appendix.(Table 22)

With these three columns K Means Clustering was done ,initial analysis of Clustering was done using Python but the results were not satisfactory to be used for cluster definitions, hence JMP tool was used.

Two approaches: Absolute and relative way.

Absolute-The initial approach decided was to classify all years based on the donor behavior observed in 2019,as 2019 was a good Year for NVPC. However the approach had some disadvantages of not capturing the changing dynamics every Year correctly. The steps taken under this approach are described below.

SMOTE was used for balancing the data as the number of donors in casual cluster is very high.

And classification techniques of Random Forest,KNN,XGboost,Decision Tree and Logistic regression was used . KNN classification technique was found the best .

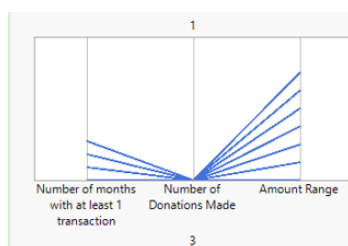
Train ,Test accuracy and confusion matrix values are attached in the appendix. Table no:19

Relative approach -Clustering every Year based on their Giving behavior. observed through the columns of Number of Months with atleast 1 donation, Number of Donations made, Amount range

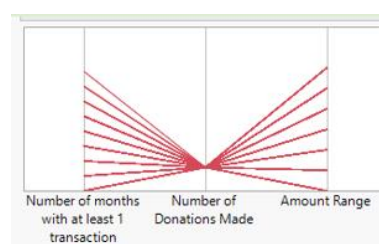
The results of JMP Output for all years are also attached in Appendix :

After clustering three main clusters were found and they are described below.

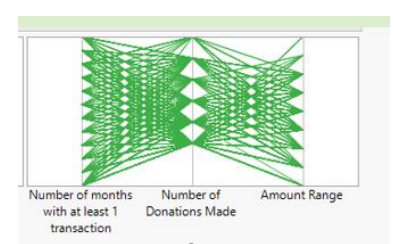
- Casual- Donated for a period of 4 months and the frequency is typically 1-4 donations, and the average donation amount is under 600 SGD
- Regular- Donated for a period of 9 months and the frequency of donation is more than 10-20 donations, and the average donation amount is around 1000-1500SGD
- Loyal- Donated all throughout the year, and the frequency of donation is more than 10-20 donations, with average donated amount ranging from 3000-4000SGD



Casual



Regular



Loyal

- The casual cluster has the largest size among the three every year .

- A Size transition is observed every year in each cluster, the Casual cluster shows highest size transition over the years JMP Output in appendix shows the actual size of Clusters every Year
- The Cluster means show same pattern for year 2015 and 2016 ,for year 2017 the mean for a regular cluster mingles with means of Loyal cluster, However for years after 2017 the means follow same pattern.
- Cluster splitting can be seen every year for all types of clusters
- From a Casual type cluster on an average 10% will convert to Loyal type profile cluster.
- In 2020 many Casual type profile cluster changed to Regular type profile
- Expected change of cluster in 2021 might follow the same pattern.

Cluster attrition over the years

To evaluate the cluster attrition the idea was to use the same formulae ,

Donor cluster attrition Rate= $\frac{\text{Donors who donated in last year in that cluster} - \text{Donors who donated this year}}{(\text{Donors who donated last year in that cluster})}$

	2015	2016	2017	2018	2019
Casual	68.72	69.17	69.7	69.23	59.01
Regular	33.65	35.09	40.9	27.35	21.29
Loyal	20.72	17.33	20.08	12.12	8.571

Table 11: Cluster attrition

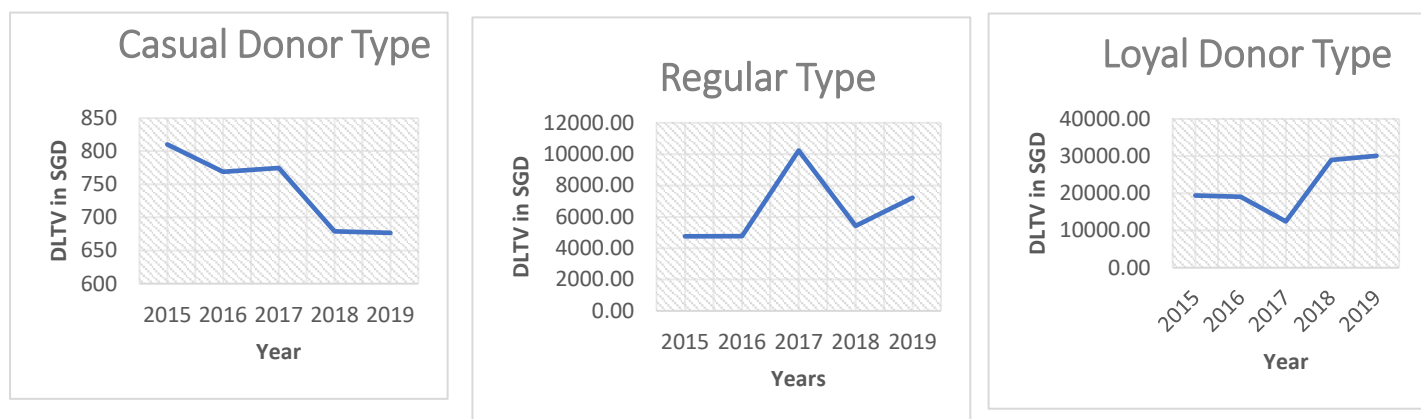


Fig 10: DLTV of Different Clusters

Years	2015	2016	2017	2018	2019
Attrition rate	68.72	69.17	69.70	69.23	59.01
Average_donation_amount	556.72	531.71	539.99	470.01	399.30
Lifespan	1.46	1.45	1.43	1.44	1.69
DLTV	810.13	768.70	774.73	678.91	676.66
Number of Donors	6260	14044	17949	25662	37754
Total	5071419.85	10795627.06	13905659.41	17422210.92	25546745.38
In Millions	5.07	10.80	13.91	17.42	25.55

Table 12: Casual Donor Type

Years	2015	2016	2017	2018	2019
Attrition rate	33.65	35.09	40.90	27.35	21.29
Average_donation_amount	1601.00	1670.25	4185.68	1485.13	1536.81
Lifespan	2.97	2.85	2.44	3.66	4.70
DLTV	4757.80	4759.90	10233.94	5430.09	7218.46
Number of Donor	1257	2408	1457	3542	4691
Total	5980555.72	11461846.68	14910845.38	19233383.77	33861792.91
In Millioons	59.81	11.46	14.91	19.23	33.86

Table 13: Regular Donor Type

Years	2015	2016	2017	2018	2019
Attrition rate	20.72	17.33	20.08	12.12	8.57
Average_donation_amount	4027.53	3299.36	2499.16	3506.56	2578.70
Lifespan	4.83	5.77	4.98	8.25	11.67
DLTV	19437.89	19038.45	12446.02	28932.01	30086.34
Number of Donors	1211	2157	4063	3266	5250
Total	23539280.07	41065931.86	50568182.98	94491955.12	157953272.66
In Millions	23.54	41.07	50.57	94.49	157.95

Table 14: Loyal Donor Type

Casual Expected Attrition	0.6917
Regular Expected Attrition	0.3365
Loyal Expected Attrition	0.1733

Table 15: Expected attrition

Cluster splitting and transition:

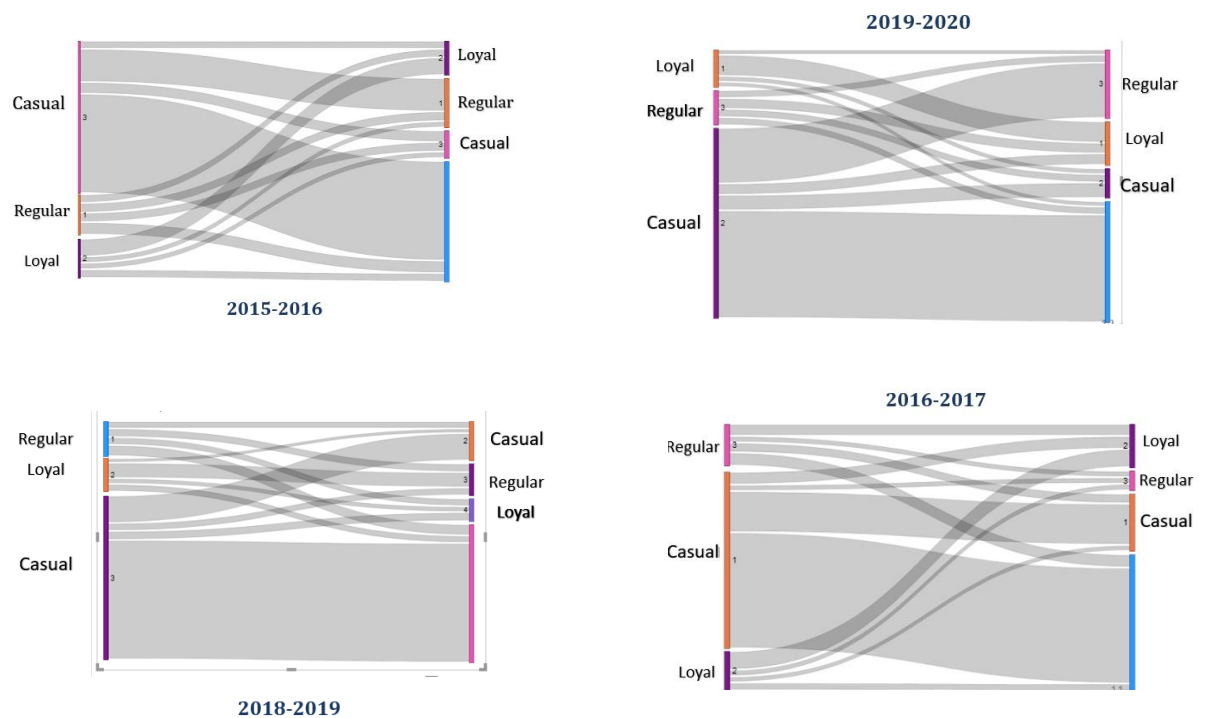


Fig 11: Cluster Transition through Sankey Chart

IMPACT OF REDUCED ATTRITION

The impact of reduced attrition can be an extension to definition of DLTV. It shows that had the attrition been 1% lower, How much extra donation amount could have been saved from being lost .The calculated value of Donation amount that would have been saved from loosing every Year is less but substantial.

Year	Donor	Original_attrition_rate	Avg Donation amount	Total
2015	8736	0.4805	1188.77	10385094.7
2016	4538.352	0.5317	1086.66	4931645.58
2017	2125.310242	0.5028	1104.73	2347893.98
2018	1056.704252	0.4812	890.65	941153.642
2019	548.218166	0.27106	886.16	485809.01
				19091596.9

Table 16: Table with original attrition rate

Year	Donor	Updated_attrition_rate	Avg Gift	Total
2015	8736	0.4705	1188.77	10385095
2016	4625.712	0.5217	1086.66	5026576
2017	2212.47805	0.4928	1104.73	2444191
2018	1122.168867	0.4712	890.65	999459.7
2019	593.4028967	0.26106	886.16	525849.9
				19381171

Table 17: Table with Updated attrition rate

	Yearly_impact
By end of 2016	\$94,930.62
By end of 2017	\$205,590.25
By end of 2018	\$209,035.56
By end of 2019	\$287,751.51
By end of 2020	\$250,342.00

Table 18: Yearly impact

Findings

- After the first year attrition the subsequent year attrition is very low and generally the number of donors after the first year in every cohort does not change much .
- If the initial attrition is contained then possibly overall donors attrition can be reduced significantly.
- Refer to graph below and table 20 in appendix for actual numbers.
- The new donors reduce by more than half of there original numbers and in subsequent Years the attrition is quiet low.

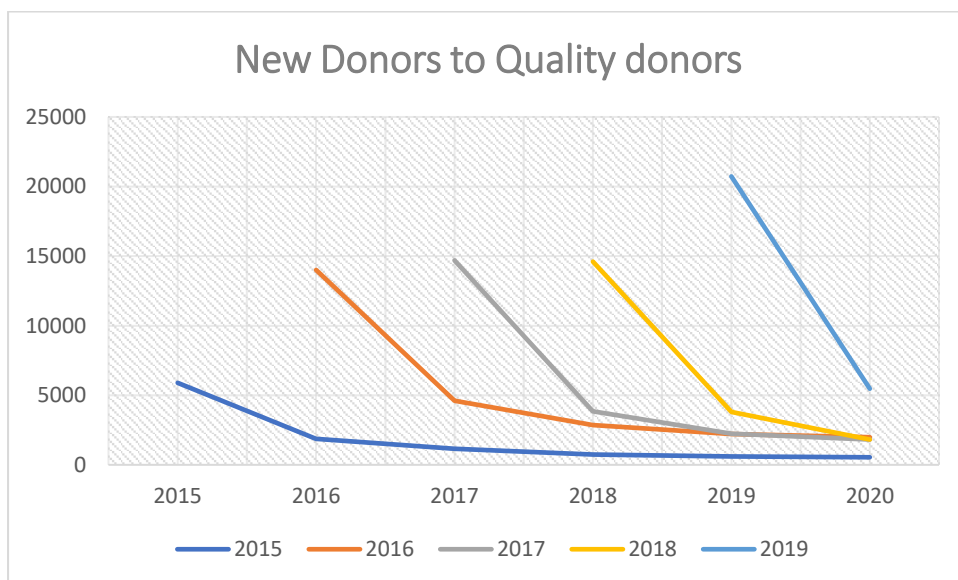


Fig 12: New Quality to Quality Donors

Upgraded /downgraded over the years

The motive of Upgraded and Downgraded indicates donor Health, Most of the times donors wait to be asked or at least have some reason to give more than suggested donations.

Once we know the situation of Upgrade and Downgrade behavior we can use it for forming further strategies .For instance, design a Donor Upgrade Roadmap,

Upgraded -People who gave more than previous Year.

Downgraded-People who gave less than previous Year.

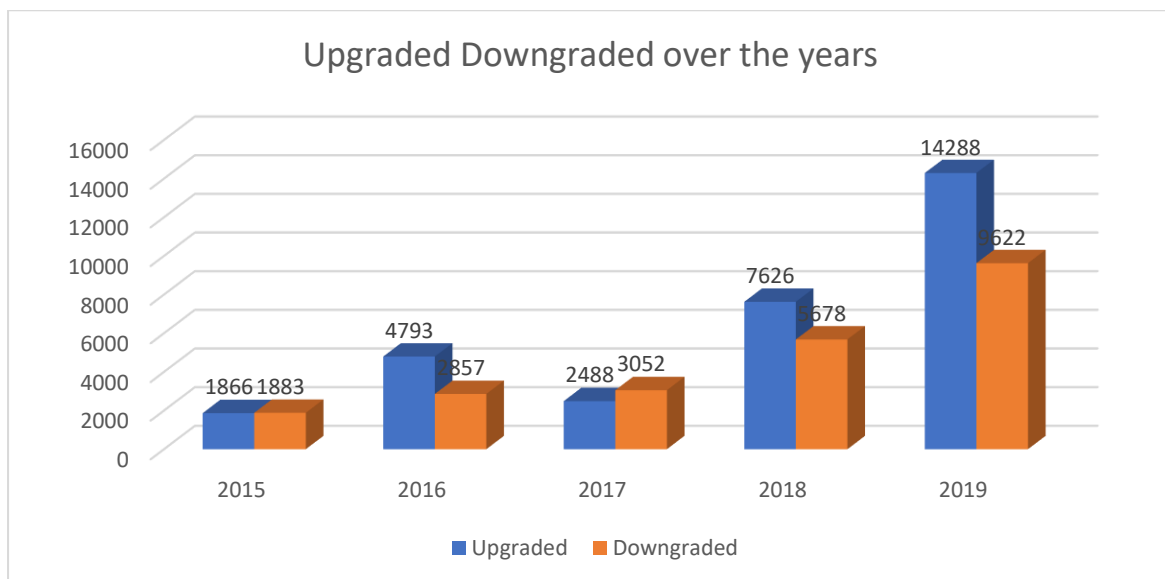
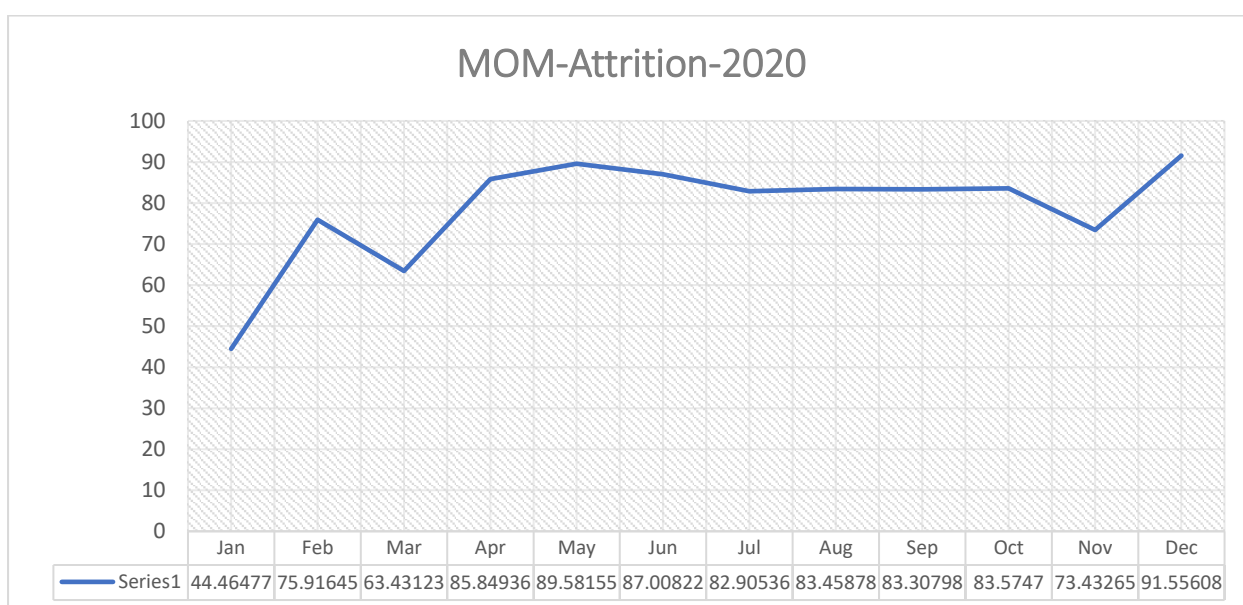


Fig 13: Upgraded and Downgraded over the years

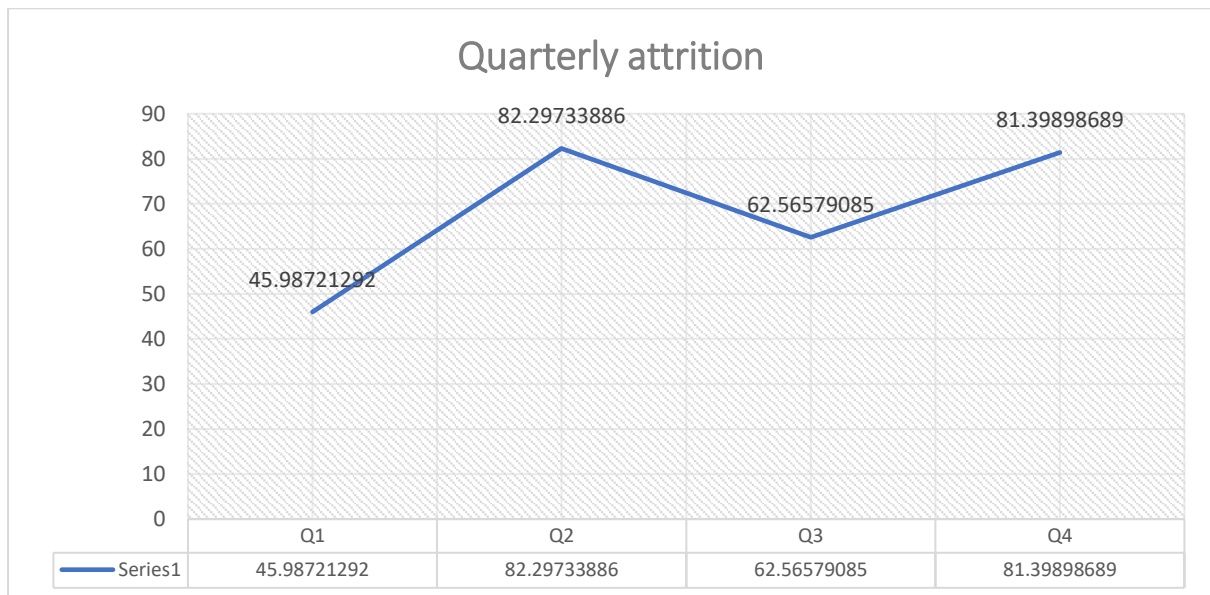
It can be concluded from the above graph that the number of people who upgrade ,or downgraded this year is comparatively similar but for Year 2019 ,it can be seen that a larger number of people upgraded than downgraded.

Year 2020

Monthly attrition



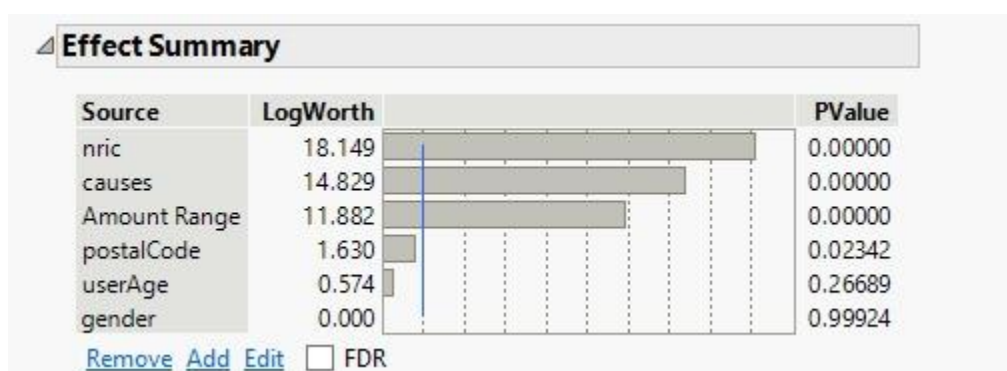
Quarter attrition



The attrition in Year 2020 on a monthly basis was observed highest in May and December and quarterly in Q2 and Q4

Factors that have high possibility to cause the dip

Since Quarter 2 faced high attrition in 2020, A regression analysis was run to find what values effect the dip and as found in effect summary ,Variables of NRIC,Causes,Amount Range have high logworth ,explaining that these might be the reasons for attrition in that quarter.



8.VALDATION OF RESULTS

In Clustering analysis

Initial Clustering was done using KMeans algorithm in python and methods of Silhoeutte analysis,Calinzki Habraz index and Davis Bouldin Index was used respectively in the tables.

Slihhouette Visual analysis is attached in appendix Table 21, A higher silhouette score signifies best cluster number ,similarly a higher Calinzki habraz index and a DB index does not imply that the number of cluster is best for information retrieval.

nclusters	silhouette score	n clusters	Ch index	n_clusters	DB index
3	0.62515	3	201195	3	0.4085317
4	0.696	4	470898	4	0.3332511
5	0.6629	5	547703	5	0.3811915

Fig 14-Silhouette score, Calinski Habraz index,Davis Bouldin Index

Clustering was done using JMP Tool by KMeans method,The clusters were tried for 3,4 and 5 number of clusters and the optimal CCC was observed lowest for 3 clusters ,Optimal ccc is not a reliable approach hence cluster means and distribution was also checked ,which confirmed that 3 clusters are the best to chosen .

In Classification analysis

Accuracy score for Random forest and XGBoost came around 100% both in Train and Test set,which might be due to overfitting.

Accuracy score for algorithms of Logistic regression,Decision Tree resulted in a lesser score than 65% while KNN classifier resulted in

71% accuracy in Train and 60% in Test set. Hence KNN algorithm was used for classification of clusters.

9.PROJECT CONCLUSIONS AND PRODUCTS

- Overall attrition ranges between 45-55%, which indicates that acquisition is working fine but the retention is still suboptimal.
- New Donors type attrition is highest in every year almost 72%, which indicates that acquisition is working fine but the retention is still suboptimal.
- New Donor Lifetime Value is steady and stable over the years of around 1000SGD
- Casual cluster shows highest attrition and highest size transition over the years, but the donor lifetime Value has been decreasing every year
- The other clusters Regular and Loyal type can be a sustainable source of donation amounts, their attrition is less and donor lifetime values is high, means higher quality.
- So for attrition to be evaluated for 2020, 2021 has to be observed completely, but since we have the cluster behaviour for 2020, we know what kind of attrition we can expect in those clusters and can strategize this year to slow down the expected attrition..
- The Expected values is an approximate figure, it is a robust method but still not a reliable approach.
- The first drop can be controlled, possibly the overall attrition can be reduced.
- To Forecast more number of Data points are required.
- NVPC can come up with constraints while capturing data, which can improve the data quality.
- The major learning in this project was Business understanding and Data understanding plays a vital role in the project stages

earlier and has to be given more time than analysis and modelling

- Another major learning was simplification of results while explaining to a non technical audience is very important
- Project management is also very important for a successful project -Changing business scopes and objectives in between leads to loss of time.
- If I had to pursue the project again ,The alternative approach that will be taken is of Project management being done properly and setting a clear understanding of goals and objectives with client.

10.RECOMMENDATIONS

- The Monthly attrition can be analyzed as a time series and forecasting can be done on it .
- The cluster evolution can be studied in more depth as there are various methods available but could not be touched upon because of time limitations.

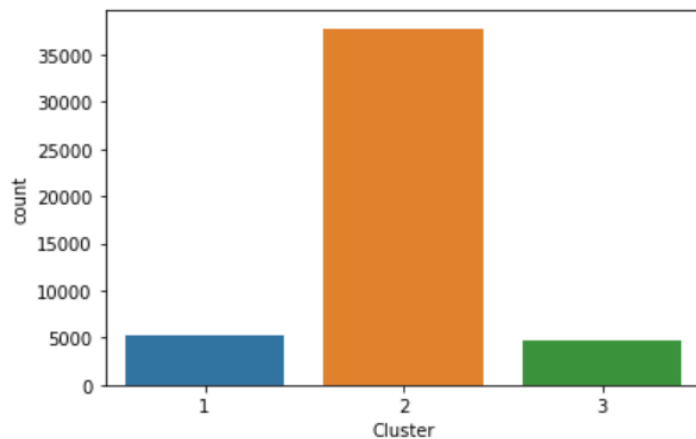
11.REFERENCES

- Reference:<https://www.causeplanet.org/blog/causeplanet-blog/make-a-home-for-root-cause-analysis-in-your-fundraising/>
- <https://fundraisingreportcard.com/donor-lifetime-value/>
- <https://webfiles.blackbaud.com/files/support/helpfiles/rex/content/bb-rfm-analysis.html>
- <https://www.wealthengine.com/donor-retention-what-to-do-to-finish>
- Individual Giving survey and previous reports of interns.

Appendix

Table 19: Classification approach

Segmentation of Donors using Absolute approach



```
[ ] #Balancing
from imblearn.over_sampling import SMOTE
sm = SMOTE()
resampled_training_inputs, resampled_training_outputs_labels = sm.fit_resample(X, Y)
```

```
[25] print('Train Accuracy:', accuracy_score(y_train, knn_classifier.predict(X_train)))
      print('Test Accuracy:', accuracy_score(y_test, knn_predictions))
```

```
Train Accuracy: 0.712789088752718
Test Accuracy: 0.600449474275624
```

```
[ ] #printing the results
print ('Confusion Matrix :')
print(confusion_matrix(y_test, knn_predictions))
print ('Accuracy Score :', accuracy_score(y_test, knn_predictions))
print ('Report : ')
print (classification_report(y_test, knn_predictions))
```

```
Confusion Matrix :
[[ 333 1365   66]
 [ 758 11520  153]
 [ 237 1260   48]]
Accuracy Score : 0.7560991105463787
Report :
```

	precision	recall	f1-score	support
1	0.25	0.19	0.22	1764
2	0.81	0.93	0.87	12431
3	0.18	0.03	0.05	1545
accuracy			0.76	15740
macro avg	0.41	0.38	0.38	15740
weighted avg	0.69	0.76	0.71	15740

	2015	2016	2017	2018	2019	2020
2015 Donors	5891	1869	1157	741	606	543
2016 Donors		14005	4611	2865	2219	1986
2017 Donors			14691	3841	2247	1817
2018 Donors				14604	3806	1817
2019 Donors					20727	5472

Table 20: New To Quality user numbers

2015 JMP Reports

Iterative Clustering

Cluster Comparison

Method	NCluster	CCC	Best
K Means Cluster	3	-58.757	Optimal CCC
K Means Cluster	3	-58.757	
K Means Cluster	4	.	
K Means Cluster	5	-80.163	

Columns Scaled Individually

Control Panel

K Means NCluster=3

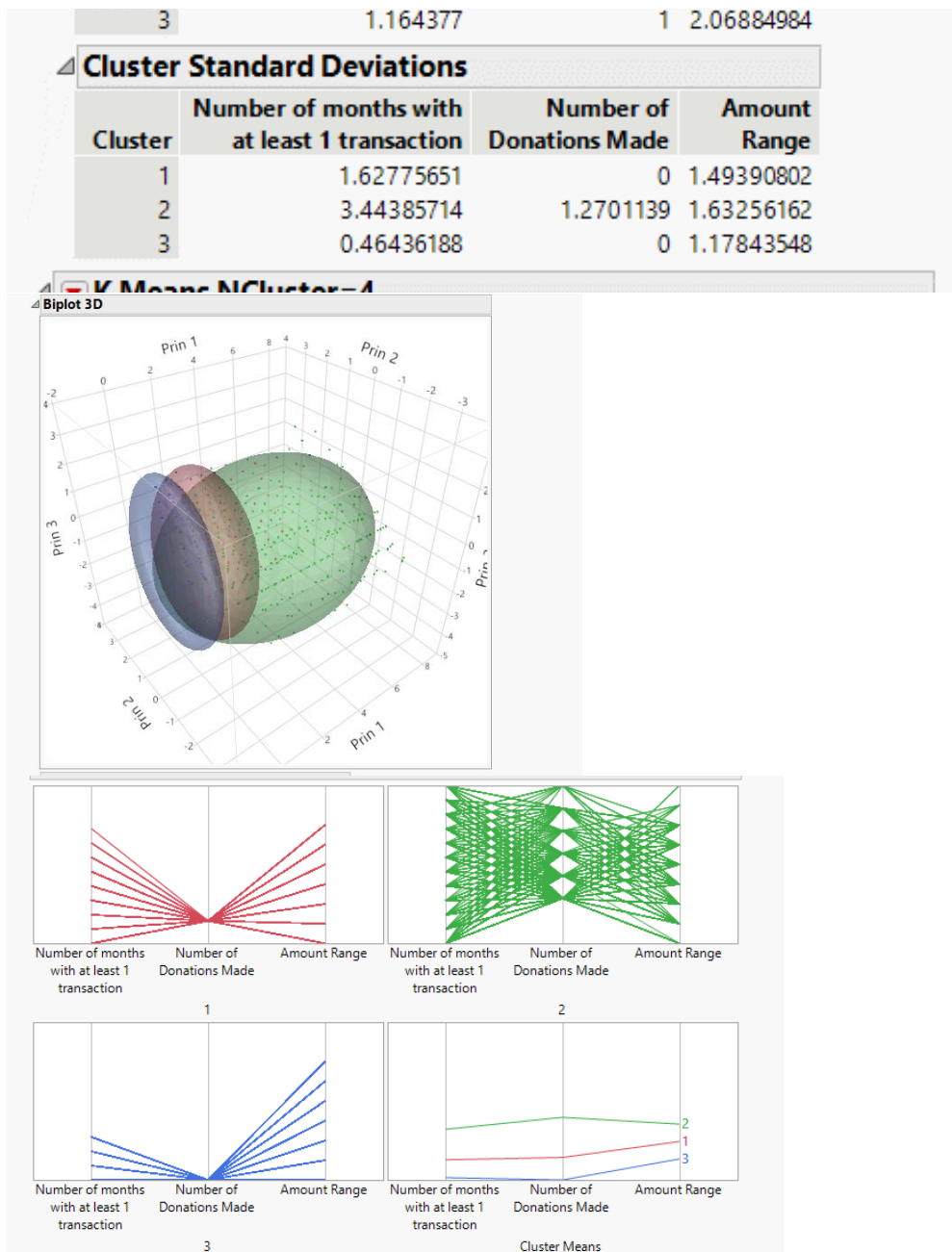
Columns Scaled Individually, Use within-cluster std deviations, Shift distances using sampling rates

Cluster Summary

Cluster	Count	Step	Criterion
1	1257	14	0
2	1211		
3	6260		

Cluster Means

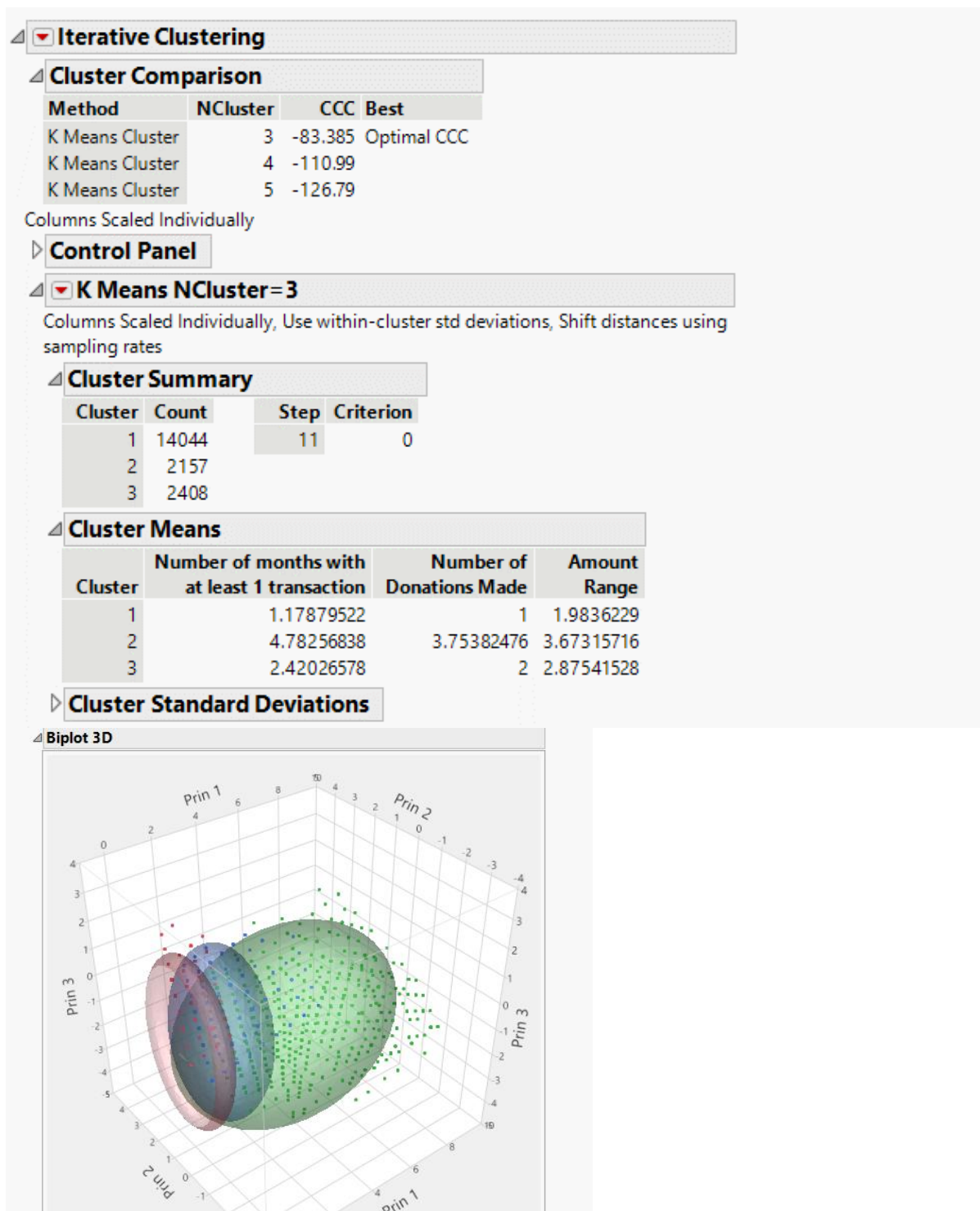
Cluster	Number of months with at least 1 transaction	Number of Donations Made	Amount Range
1	2.41209228	2	2.95385839
2	4.54170107	3.78034682	3.83649876
3	1.164377	1	2.06884984



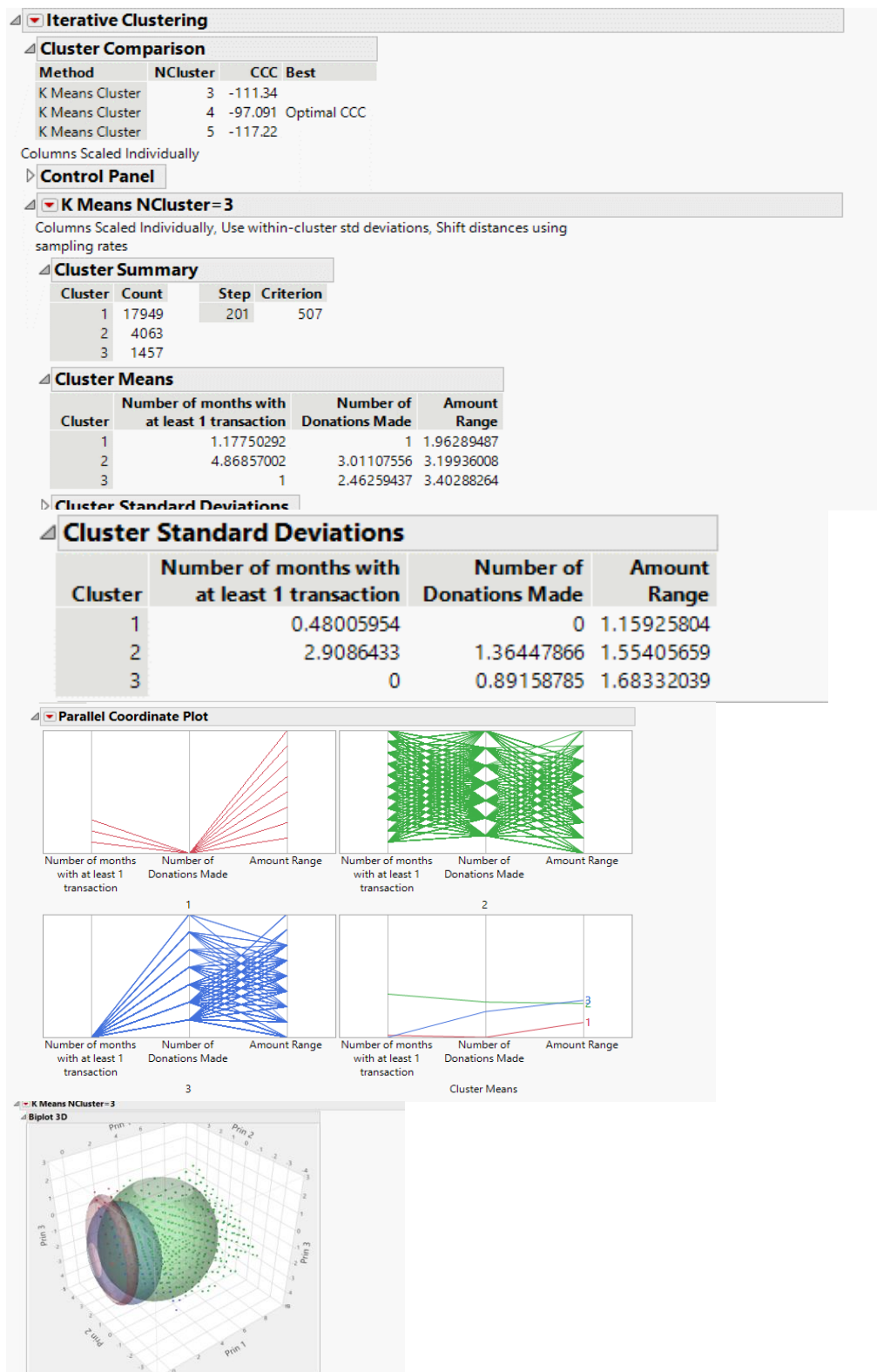
2016 JMP Output

Cluster Standard Deviations

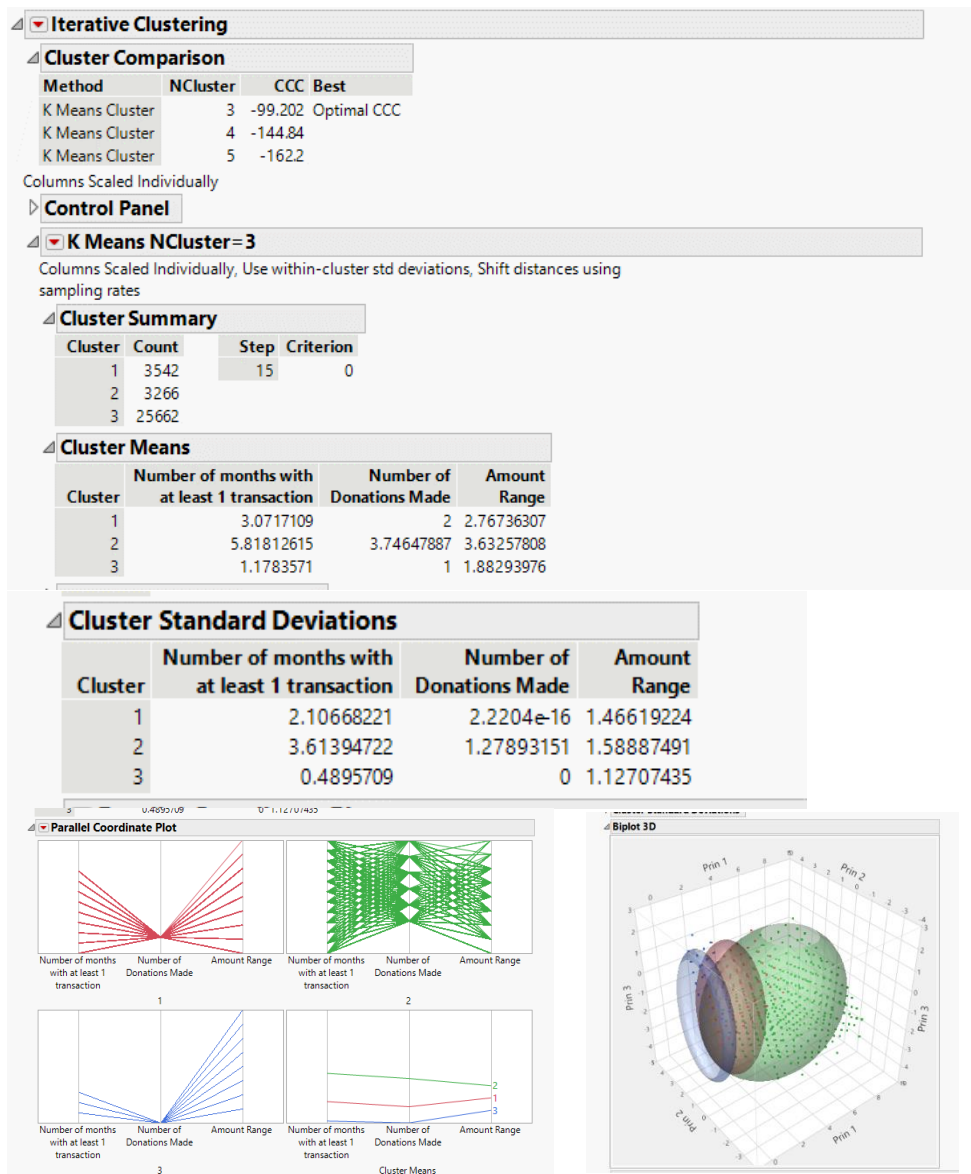
Cluster	Number of months with at least 1 transaction	Number of Donations Made	Amount Range
1	0.48283298	0	1.1606329
2	3.32904392	1.25147682	1.6053018
3	1.66190098	0	1.5108446



2017 JMP Output



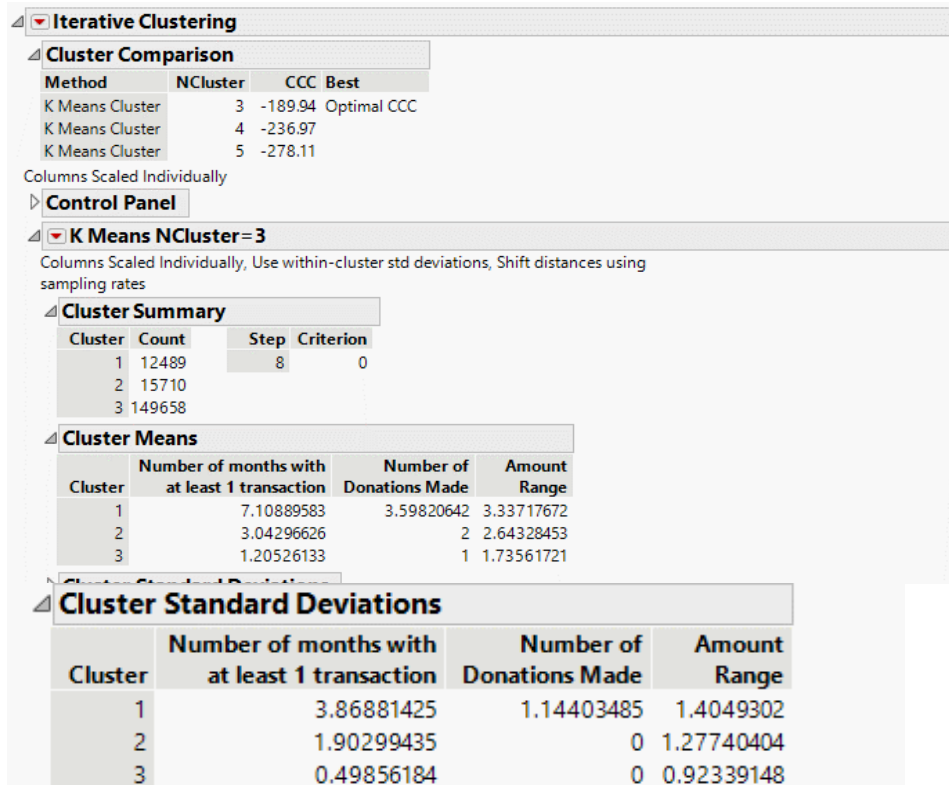
2018 JMP Output



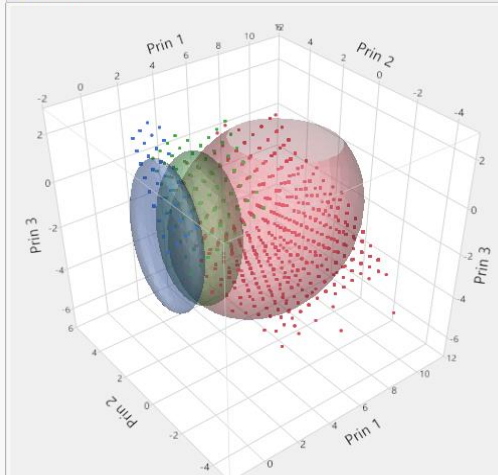
2019 JMP Output



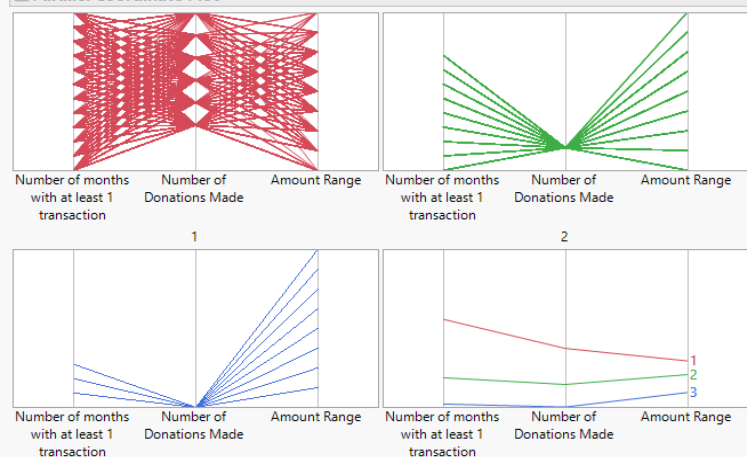
2020 JMP Output



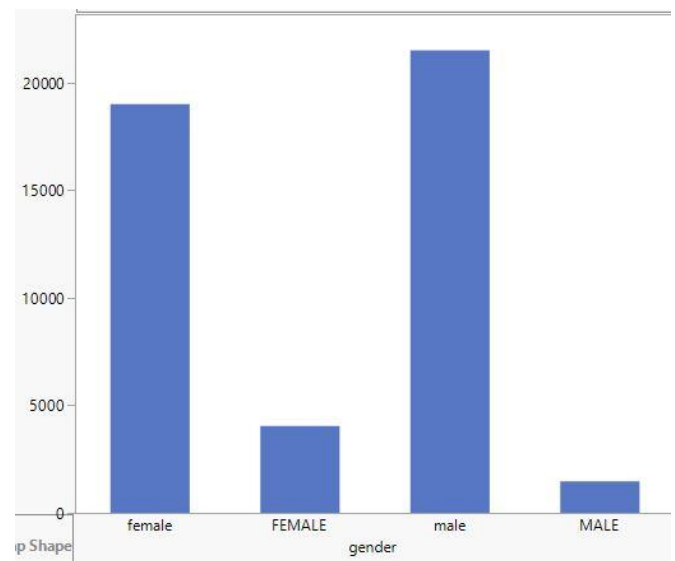
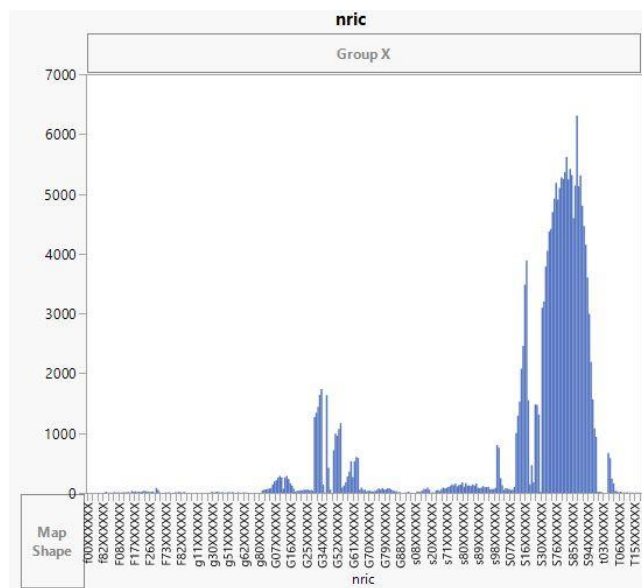
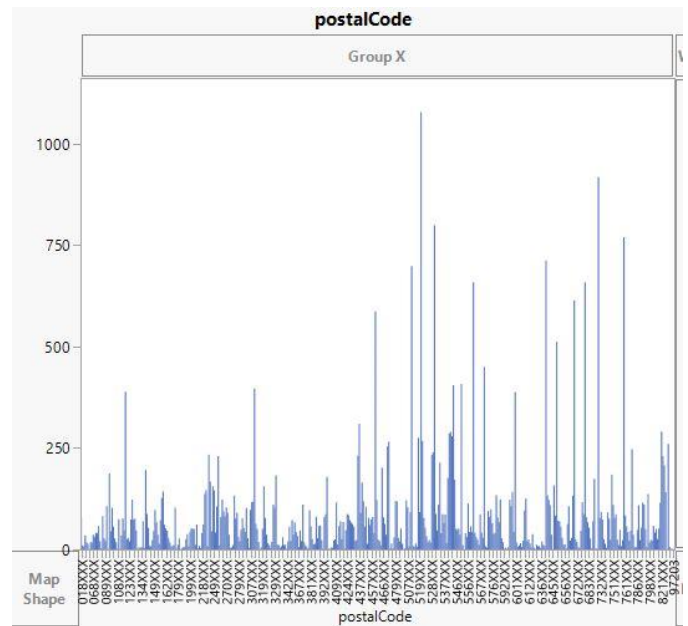
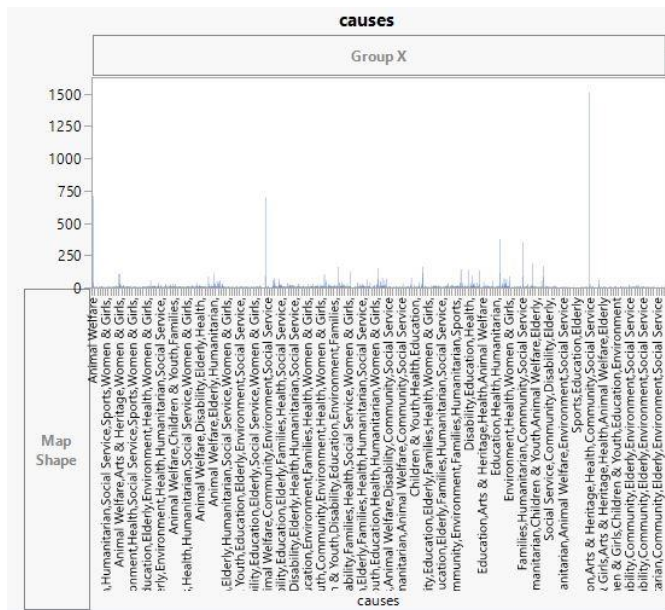
Biplot 3D



Parallel Coordinate Plot



Distribution of Variables



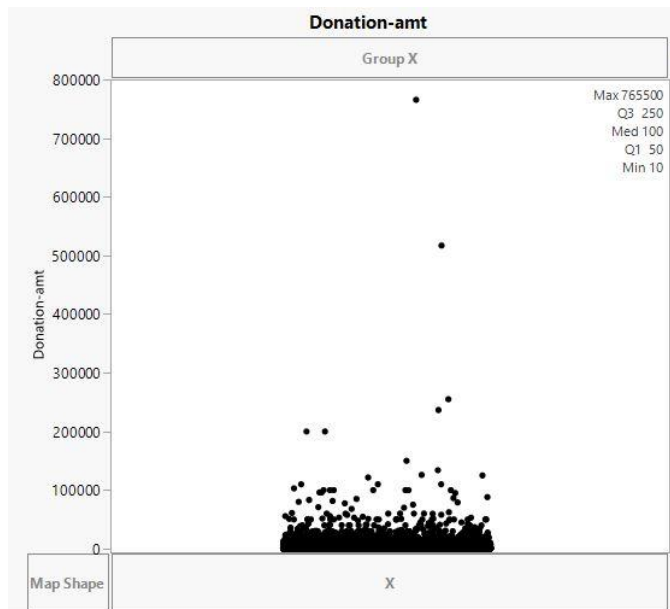
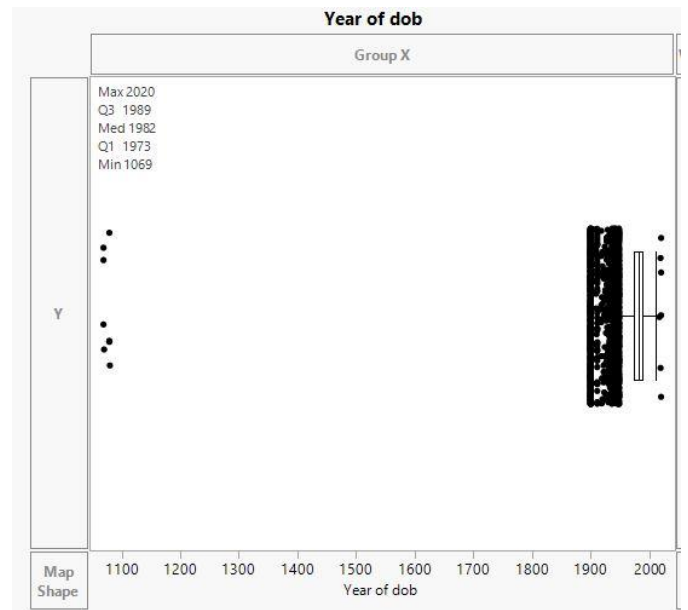
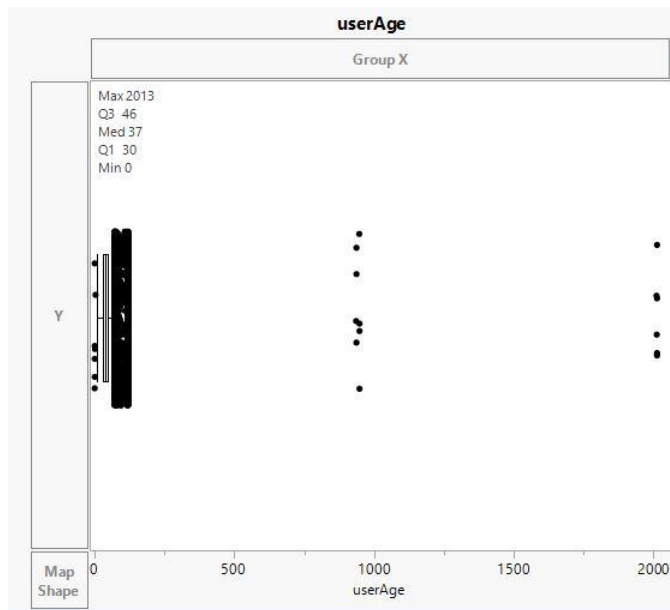


Table 21:Silhouette Visualizer with 3 , 4 , 5 Clusters

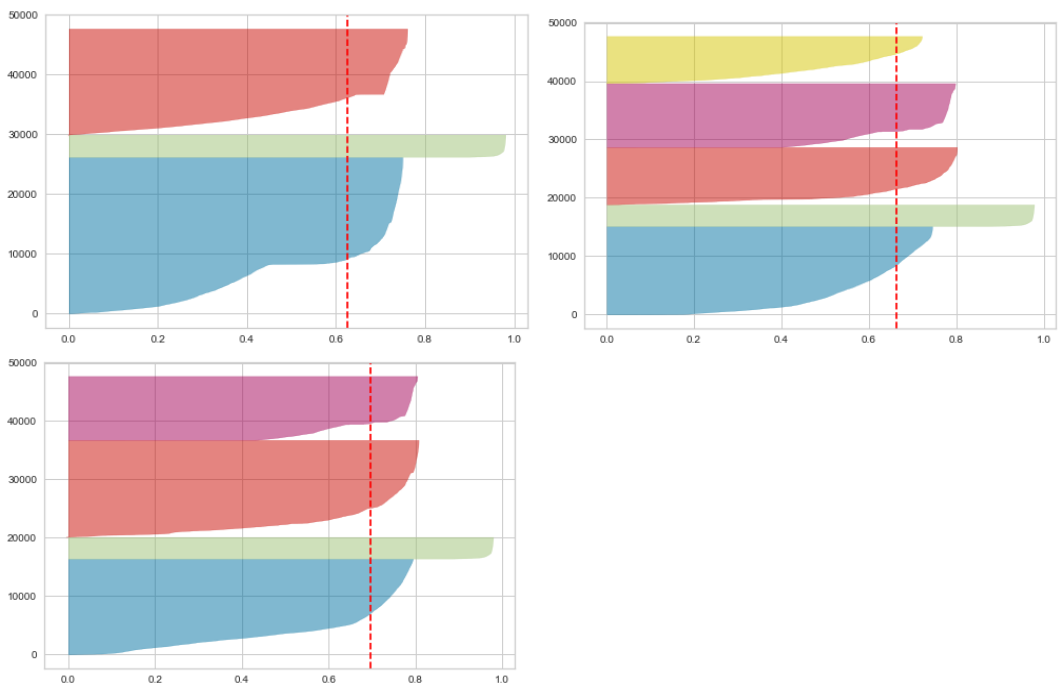
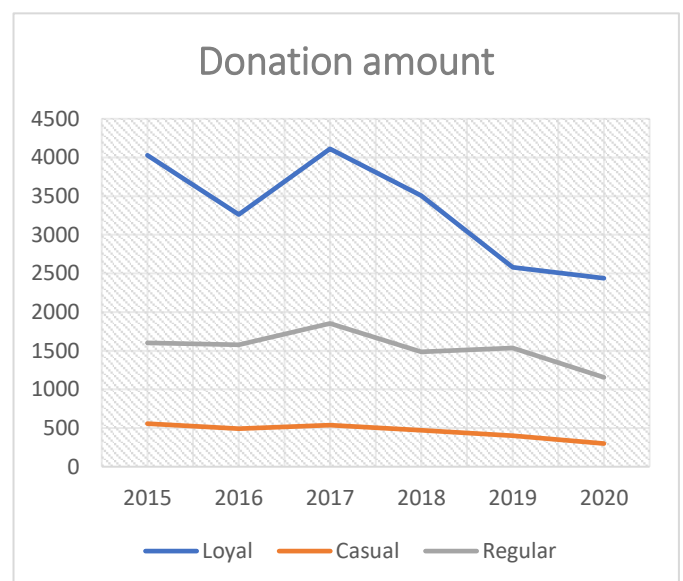
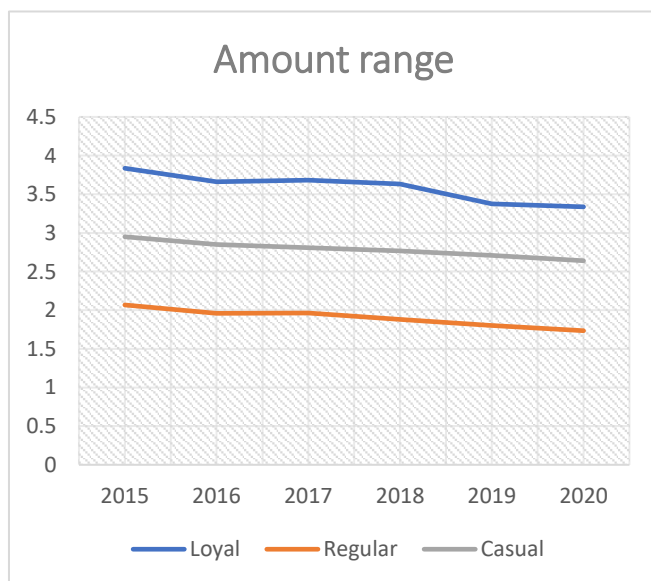
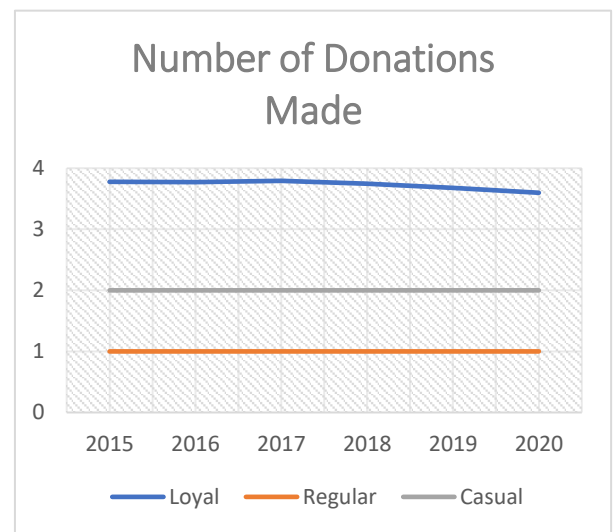
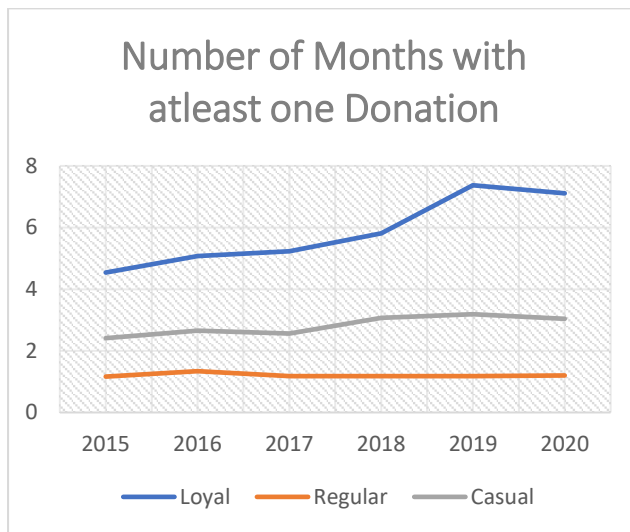


Table 22:Bins of Amount Range and Number of Donatiions

Amount Range	Frequency
10 to 99 SGD	1 to 4
100 to 499 SGD	5 to 9
500 to 999 SGD	10 to 19
1000 to 2999 SGD	20 to 29
3000 to 4999 SGD	30 to 39
5000 to 9999 SGD	40 to 49
10000 to 49999 SGD	50 to 99
	100 and above

Absolute Approach (Cluster Means)



Relative Approach (Cluster Means)

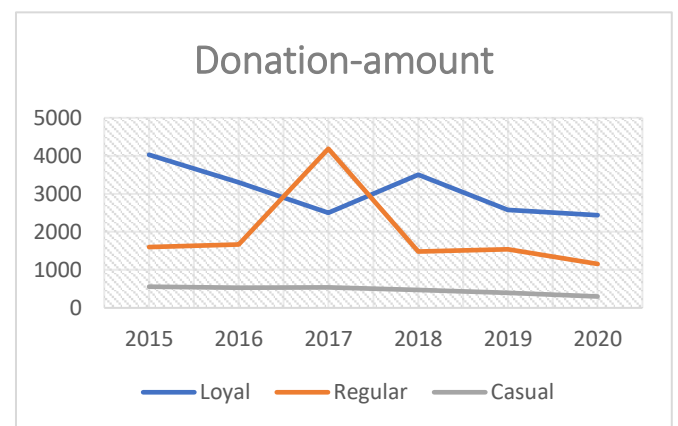
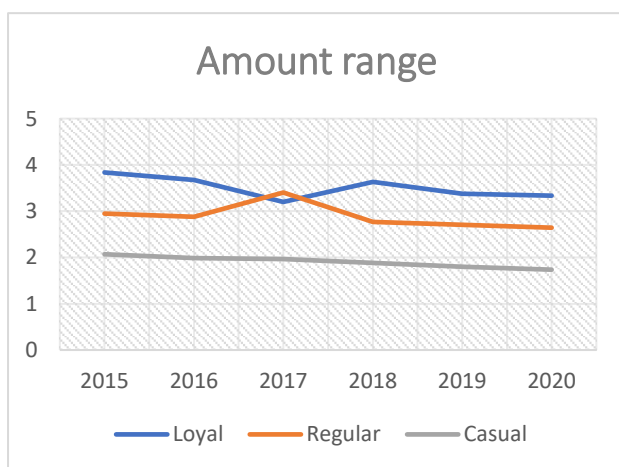
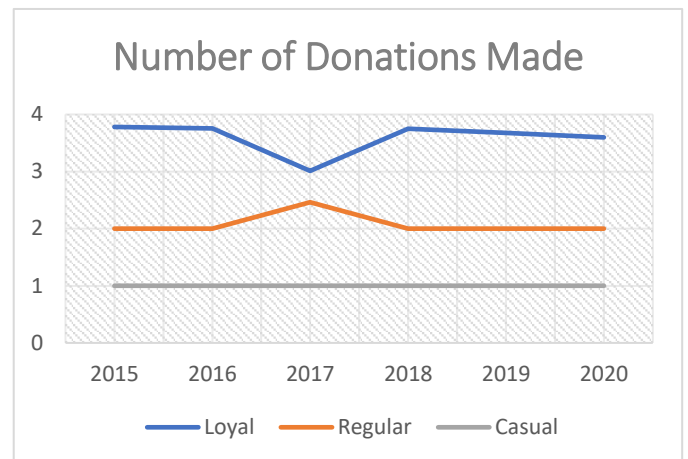
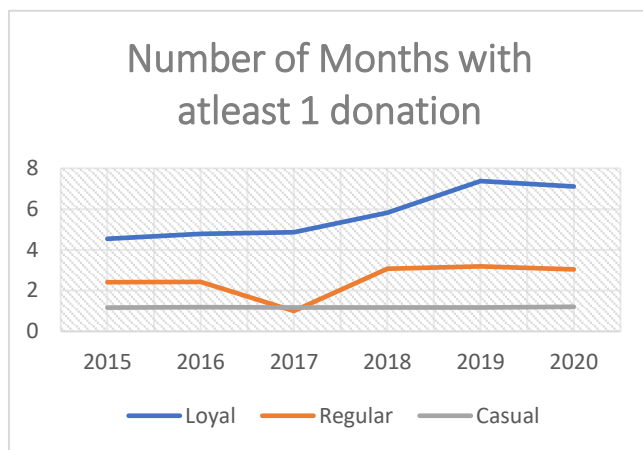


Table:23 Existing and New Donor numbers

Type	2015	2016	2017	2018	2019	2020
Existing	2811	4528	8678	11626	16792	34,617
New	5906	14005	14707	20744	30698	142,624
Total	8717	18533	23385	32370	47490	177241

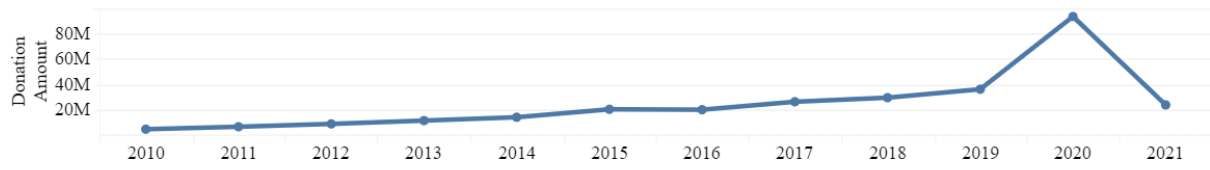


Fig: Donation amount

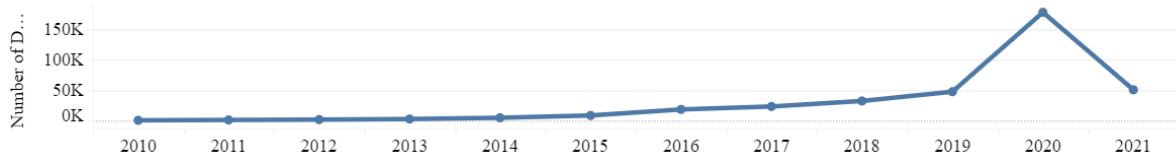


Fig: Number of Donors