Release 2 Report

1. Introduction

During the Release 2 phase, in order to further analyze alumni donation patterns and the impact of various factors on donation behavior, we conducted the following four analyses:

1. **Analyzed the differences in donation behavior between online and on-campus graduate alumni:**

   We compared the differences in average donation frequency and total donation amount per person between online and on-campus graduate alumni, segmented by different age groups and regional categories.

2. **Analyzed the differences in trust participation and donation propensity between the lapsed group and other alumni across different age groups:**

   We conducted a data visualization analysis to compare the likelihood of trust association and donation propensity across different age groups, using bar charts for comparative analysis.

3. **Examined the impact of family affiliation with the university on the likelihood of being a high-donation alumnus:**

   We conducted statistical tests to examine whether family connections—such as parents, children, or the alumni themselves being affiliated with the university—significantly impact the likelihood of becoming a high-donation alumnus. This analysis was performed using chi-square tests and visualized through charts.

4. **Performed a clustering analysis using variables related to alumni engagement and participation in university activities to explore differences in donation behavior across clusters:**

   To analyze overall patterns among different alumni groups and their relationship with donation behavior, we explored clustering using K-Means and DBSCAN methods. We determined the optimal number of clusters and compared the characteristics and donation behavior of different groups.

2. Outcomes in Release 2 Phrase

1. Analysis of Online vs. On-Campus Master's Alumni (Alumni who only hold a master's degree from our university)

Comparison of Average Donation Count per Person by Age Range (Online vs Offline Master Only)
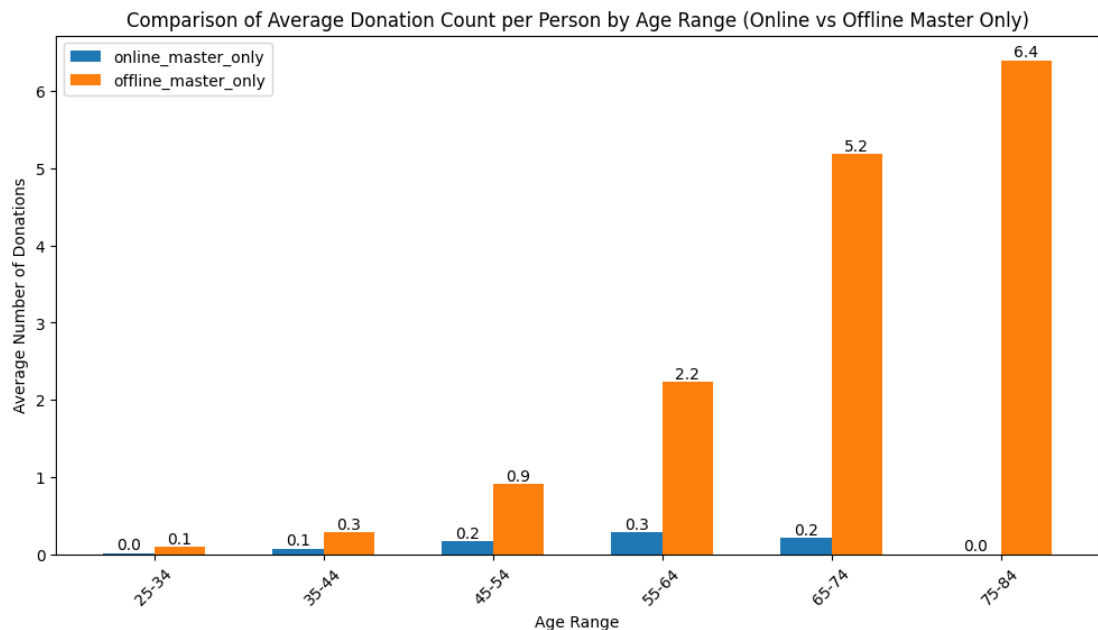
Alumni who took online courses donated significantly fewer times than those who attended in-person courses. Among online course alumni, those around 60 years old had the highest average donation frequency. This may be because online programs were introduced relatively recently, meaning alumni over 70 did not have the opportunity to participate in OLP when they were younger. We may consider targeting alumni aged 55–65 for donation recommendations, as they are more likely to contribute.

Comparison of Average Donation Amount per Person by Age Range (Online vs Offline Master Only)
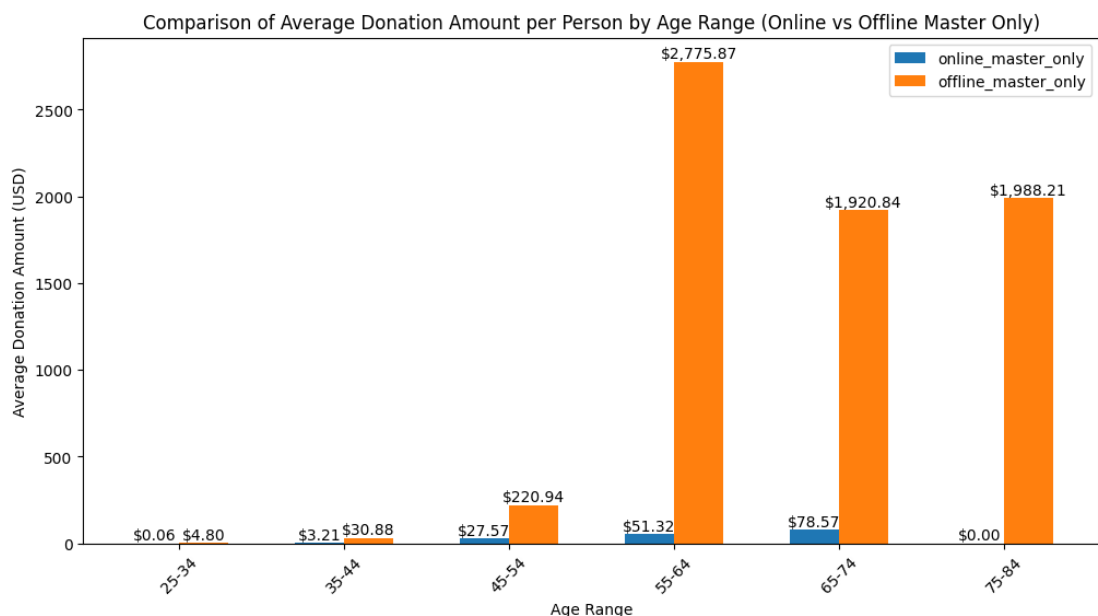
Alumni around 60 years old who attended in-person programs are the main contributors in terms of donations. Among online programs alumni, those around 70 years old have the highest average donation amount per person. We may consider targeting in-person master's program alumni around 60 and online master's program alumni around 70 for donation reminders.
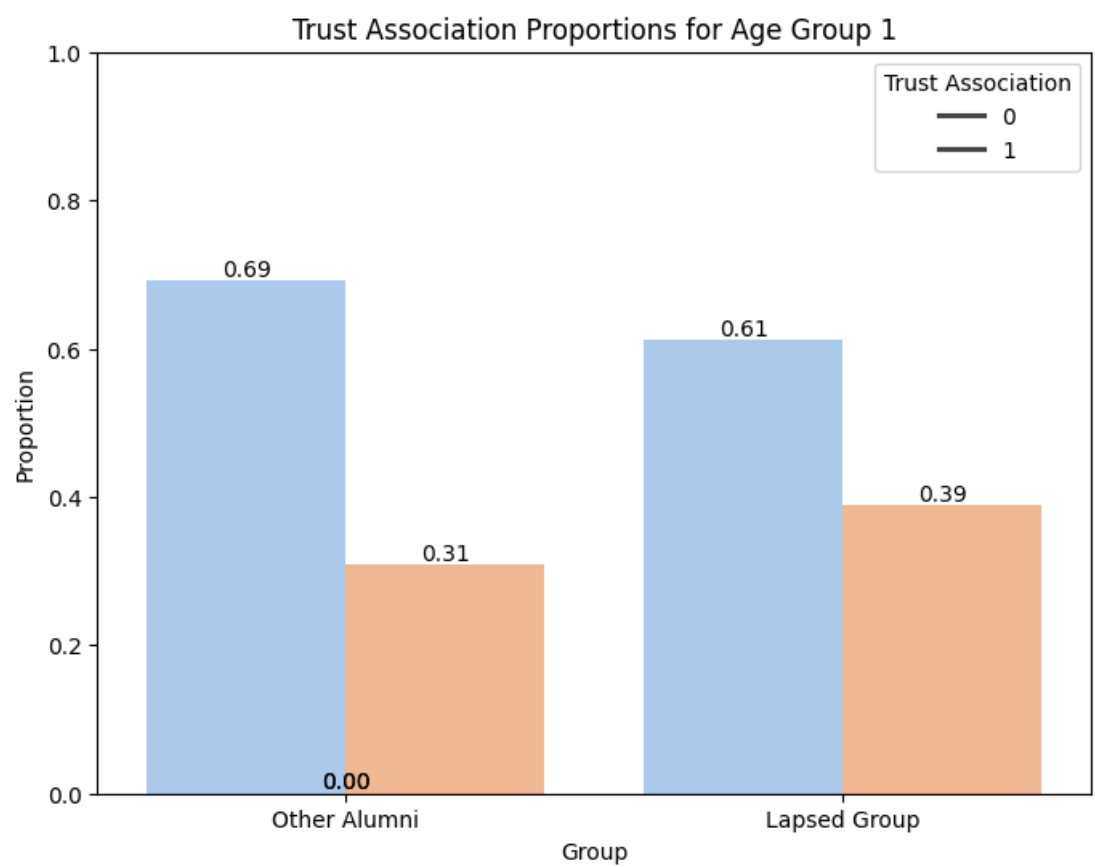
2. **Analysis of Differences in Trust Association and Propensity to Give Between the Lapsed Group and Other Alumni**

Comparison of Propensity to Give: Lapsed vs. Other Alumni
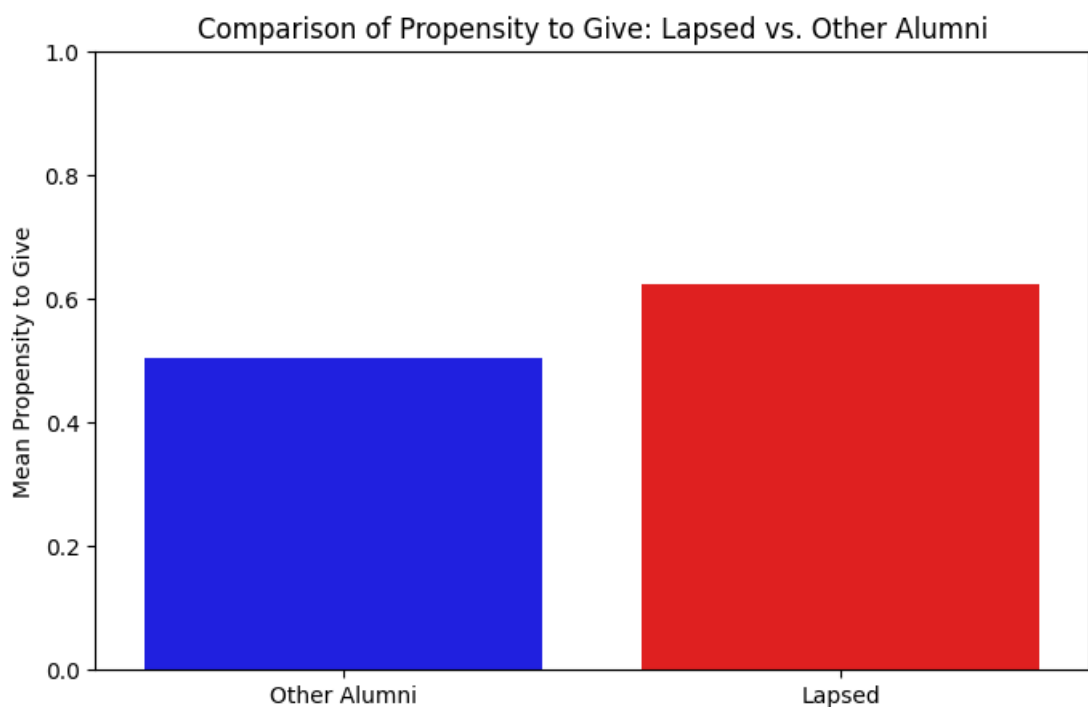
Compared to other alumni, the lapsed group has a slightly higher 'propensity to give', approximately 0.07 higher. This suggests that alumni in the lapsed group have relatively greater donation potential.

Propensity to Give by Age Range: Lapsed vs. Other Alumni
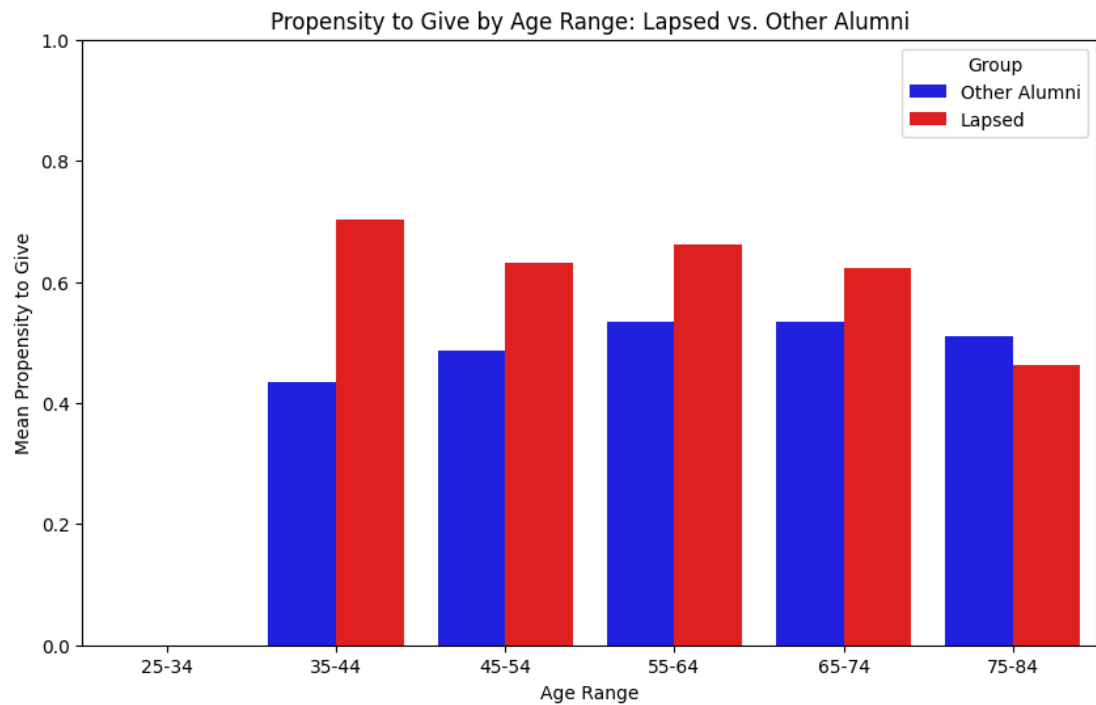
Among other alumni, those aged 60–70 have the highest propensity to give. However, in the lapsed group, alumni around 40 years old show a significantly higher propensity to give, followed by those around 60. Therefore, for the lapsed group, we may prioritize targeting alumni around 40 and 60 years old.

Average Trust Association by Group

Trust Association Proportions for Age Group 1

Trust Association Proportions for Age Group 2

Trust Association Proportions for Age Group 3

Trust Association Proportions for Age Group 4

Trust Association Proportions for Age Group 5
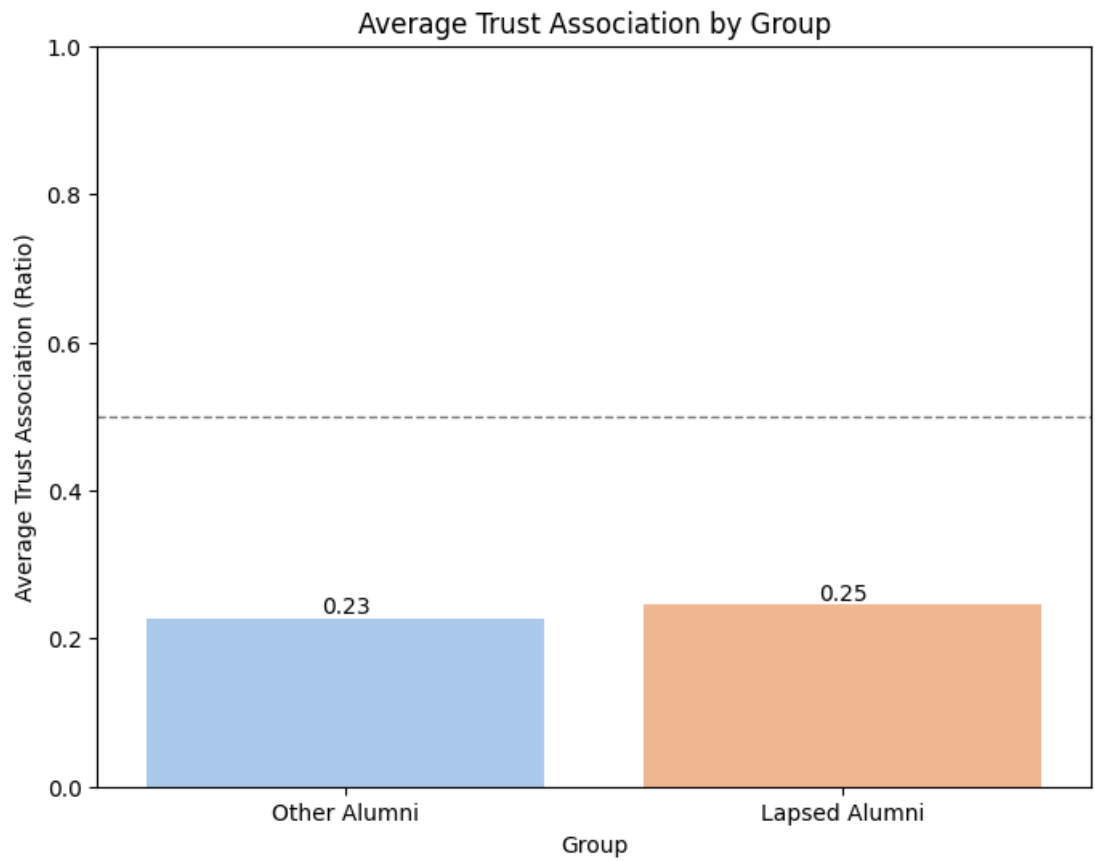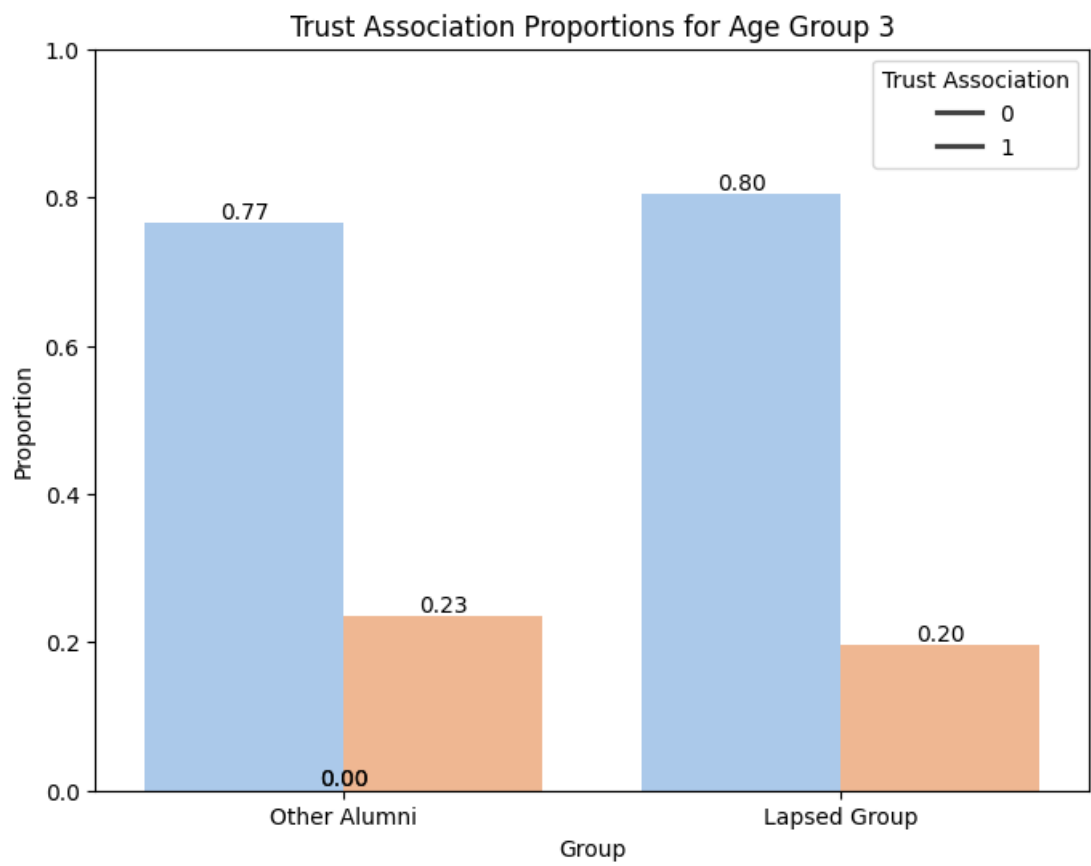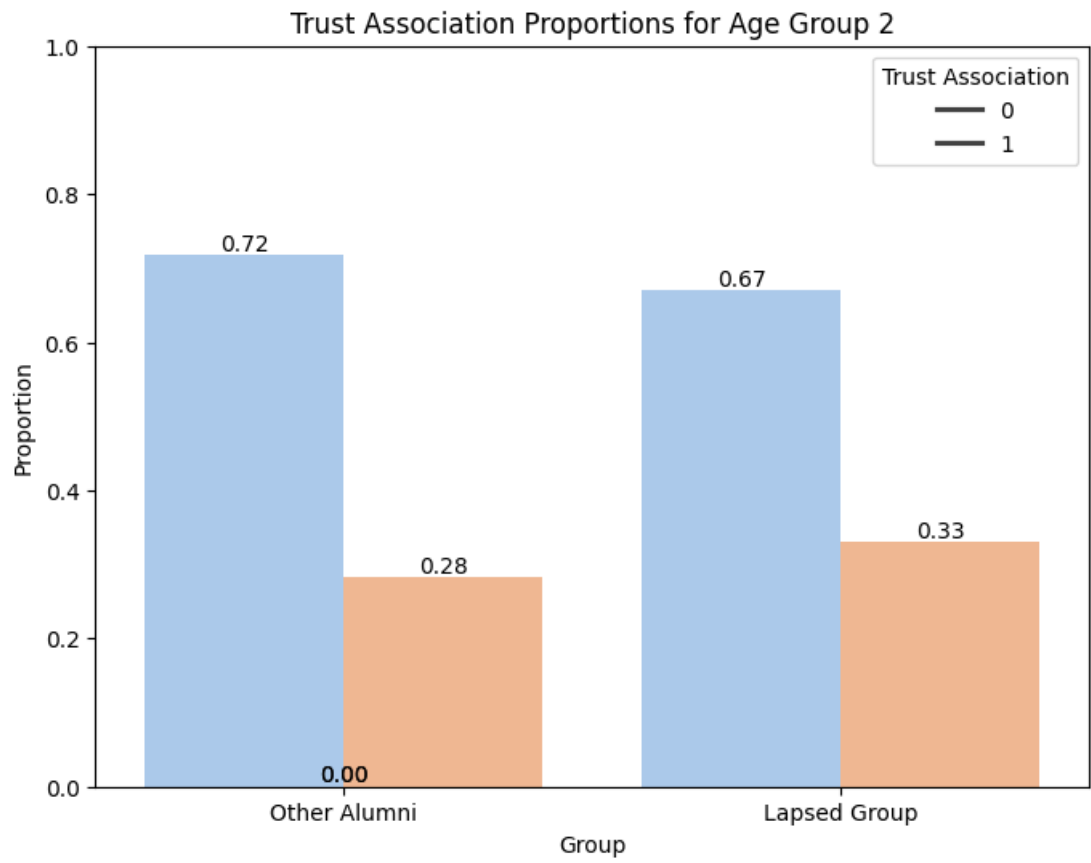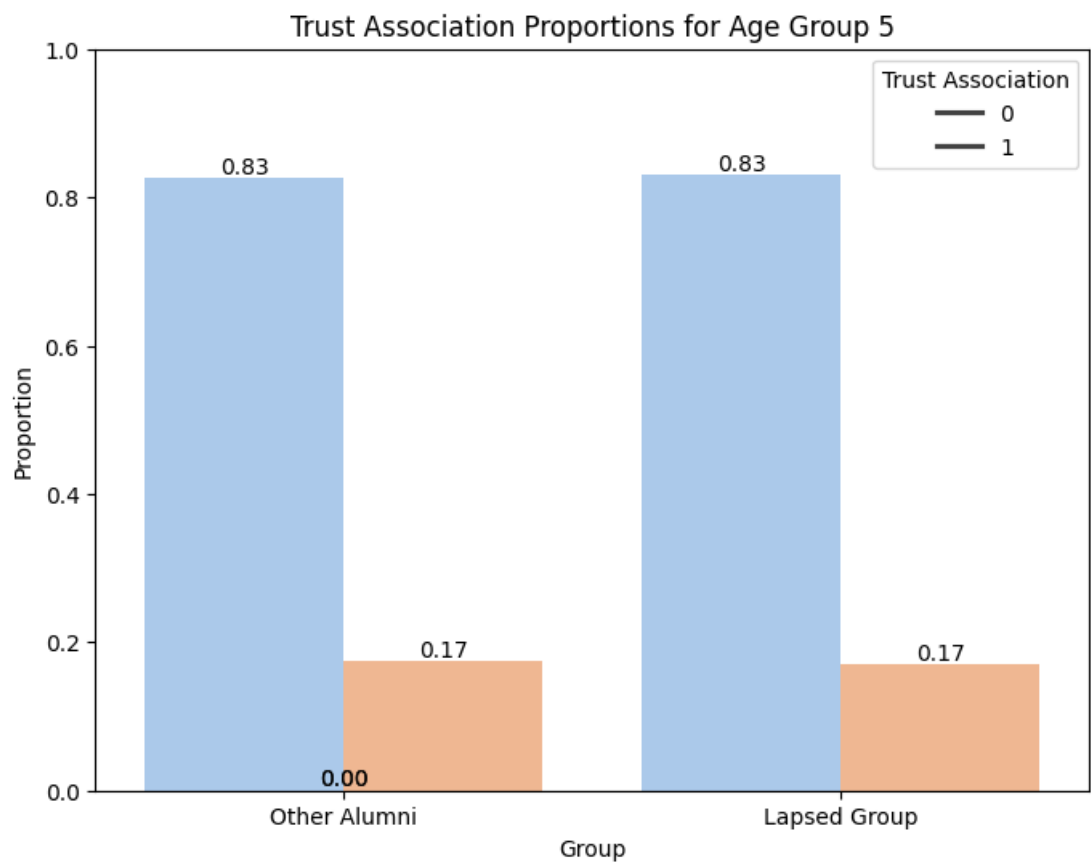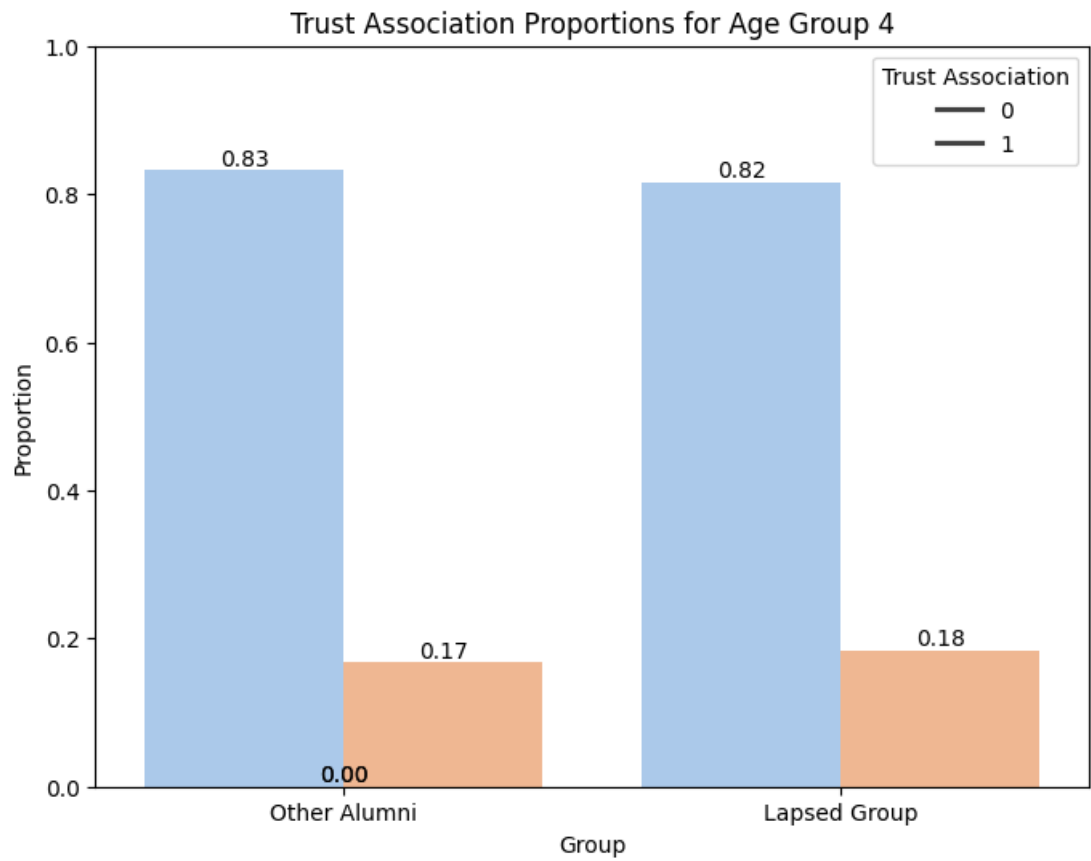
Group 6 consists of alumni around 30 years old. We found no Trust

Associate data for this group, which may be due to their younger age.

We can see that in the 60s and 70s age groups, the proportion of the lapsed group is significantly higher. However, in the younger age groups, the difference between the two groups is not as noticeable, and the proportion of the lapsed group may even be smaller. This suggests that older age groups may have greater donation potential.

3. **Relationship Analysis Between HH Affinity Total and High-Level Giving**

Due to Excel's sluggish overall performance, everyone extracted the data locally and created the 'High Level Giving' variable: Based on the UI Giving Range Code (values of 1, 2, 3, 4, or 5), a new variable was created to identify high-level donors. This variable, named High Level Giving, is represented as 1/0, indicating whether an individual is a high-level donor.

For the entire dataset, we will conduct a chi-square test to examine whether there is a significant relationship between HH Affinity Total and High Level Giving.

```
High Level Giving                              0     1
HH Affinity Total (parent, child, spouse)
0                                          41407   338
1                                           8813   279
2                                           1183   160
3                                             50    15
Chi-Square Statistic: 1461.22
p-value: 0.0000
Degrees of Freedom: 3
Result: There is a significant relationship between HH Affinity Total and High Level Giving (reject null hypothesis).
```


Distribution of HH Affinity Total vs. High Level Giving

High Level Giving Proportion by HH Affinity Total
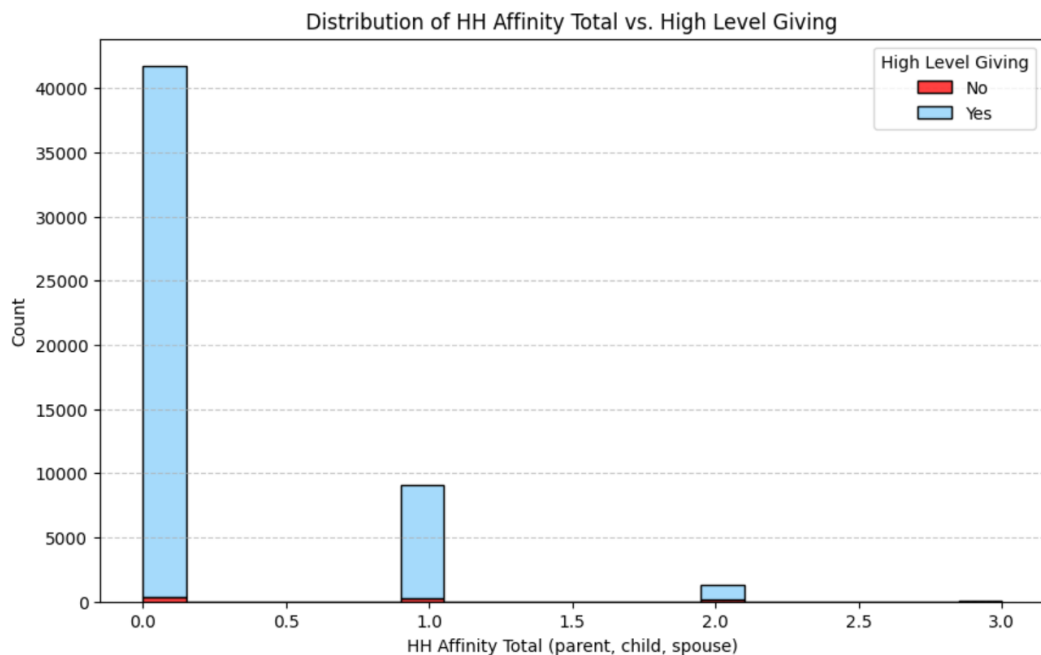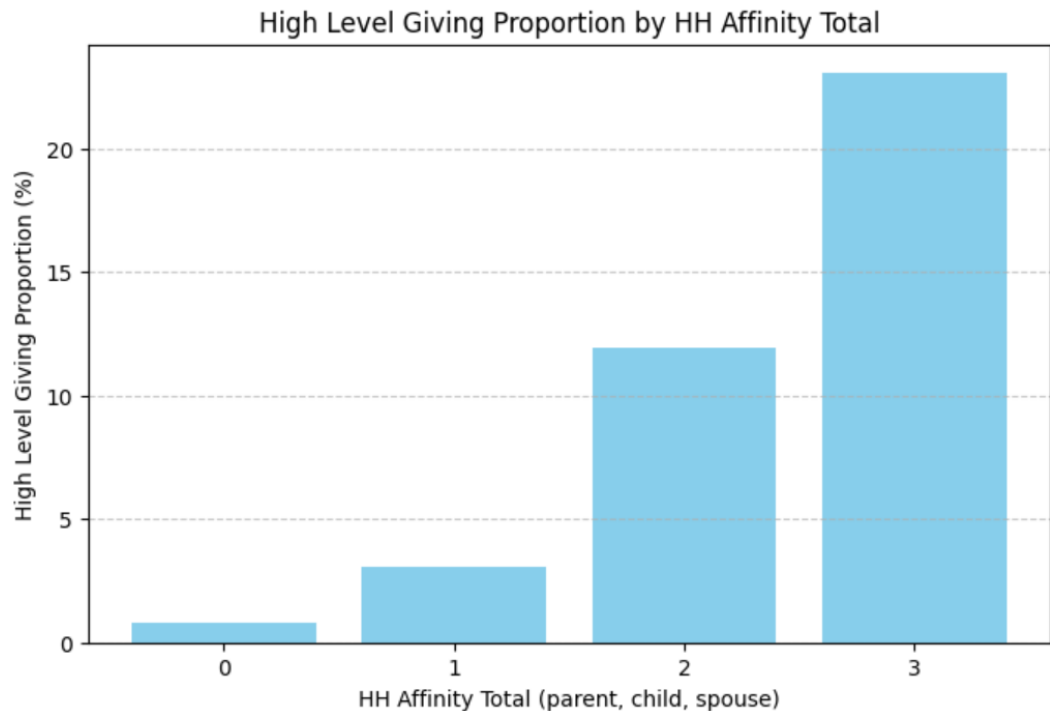
According to the chi-square test, there is a statistically significant relationship between HH Affinity Total and High-Level Giving. As shown in the figure above, alumni with stronger family ties and connections to the university have a significantly higher likelihood of becoming High-Level Givers.

This suggests that we can engage more with alumni who have high HH Affinity scores through outreach and targeted advertisements to encourage potential donations and increase overall contributions.

Next, we conducted separate comparisons and chi-square tests across different age groups and regional categories. The results were all statistically significant, and the overall trend remained consistent.

Different Age Groups:

Age Group: 75-84 (p=0.0000)

Age Group: 65-74 (p=0.0000)

Age Group: 55-64 (p=0.0000)

Age Group: 45-54 (p=0.0000)

Age Group: 35-44 (p=0.0023)

Age Group: 25-34 (p=1.0000)
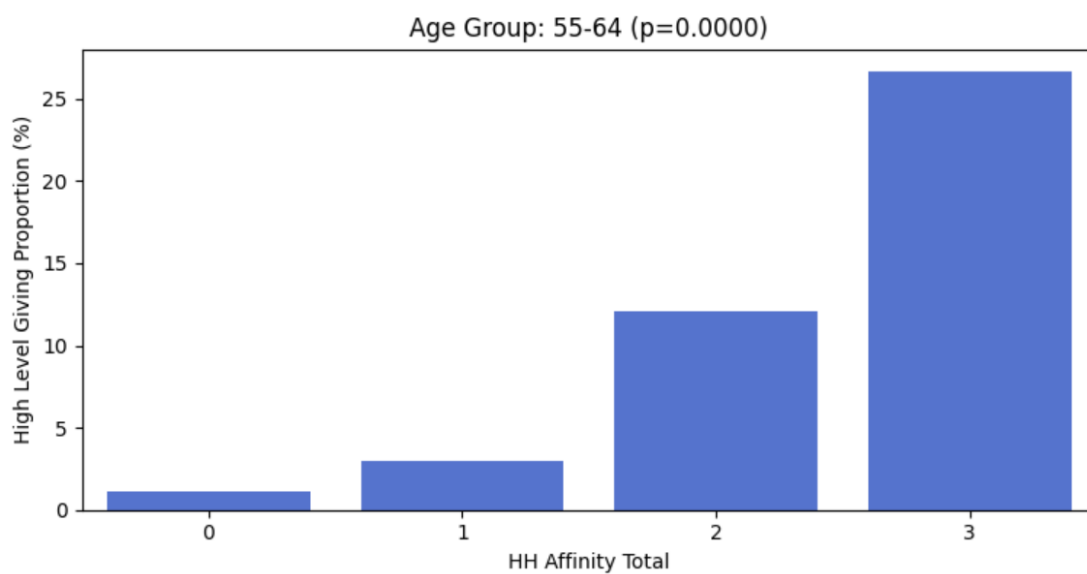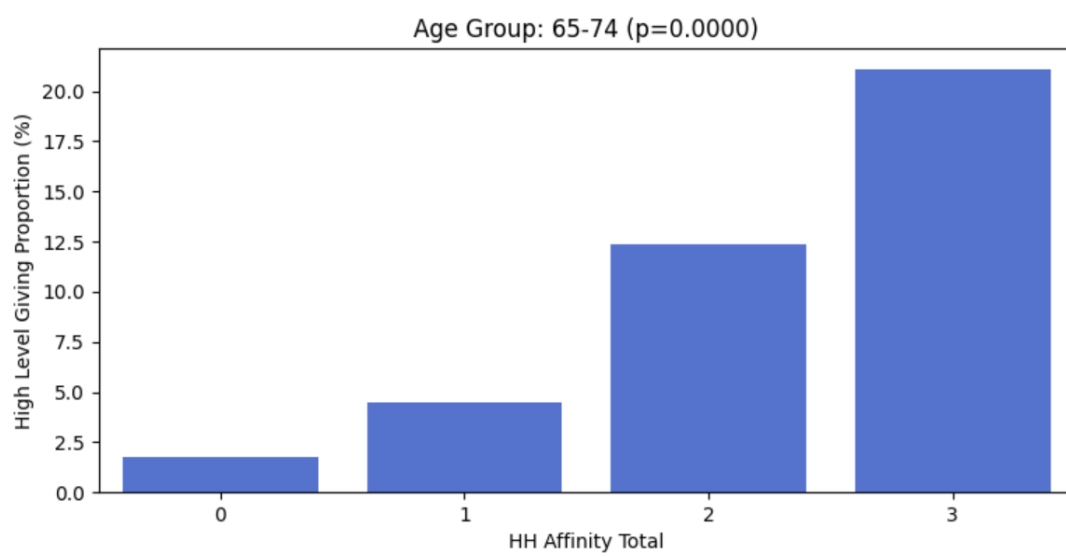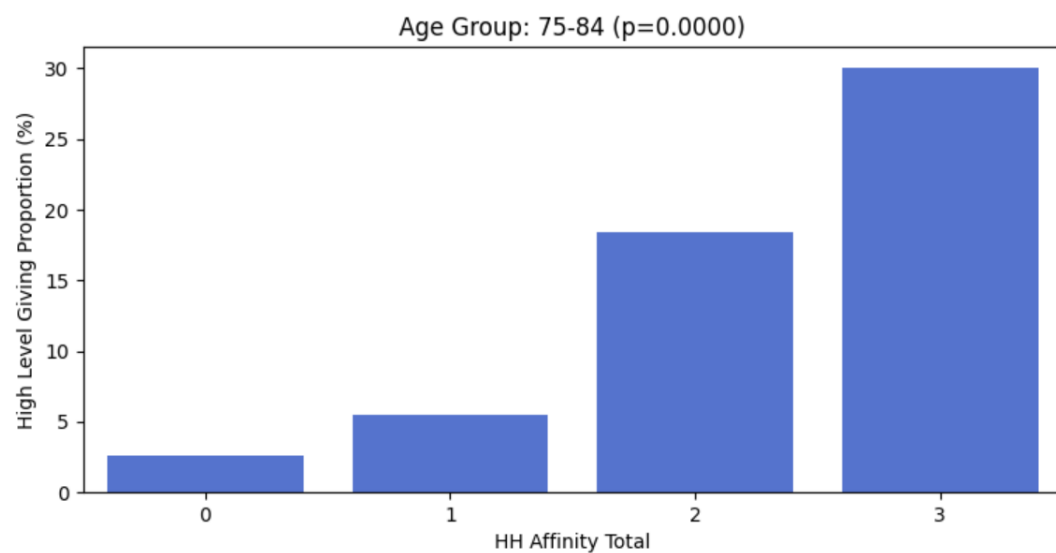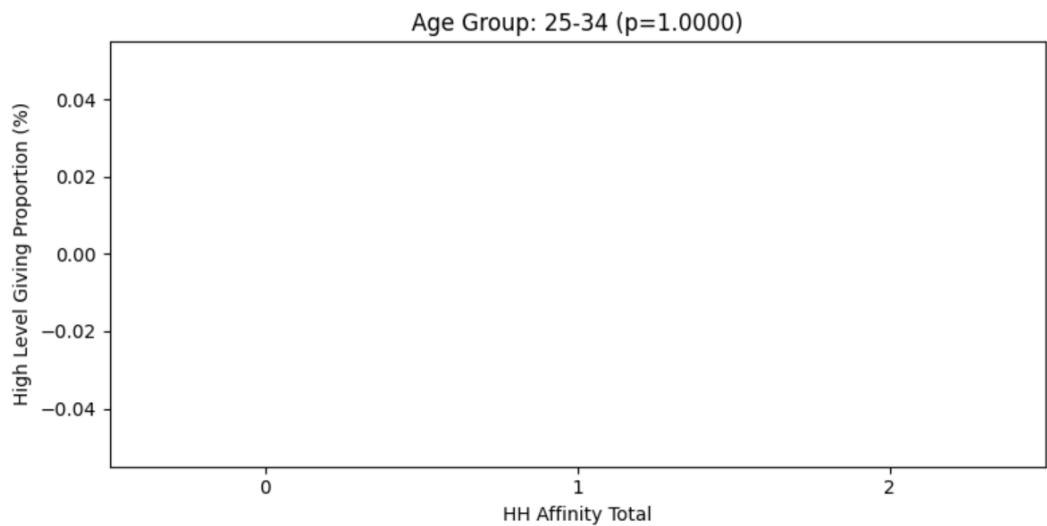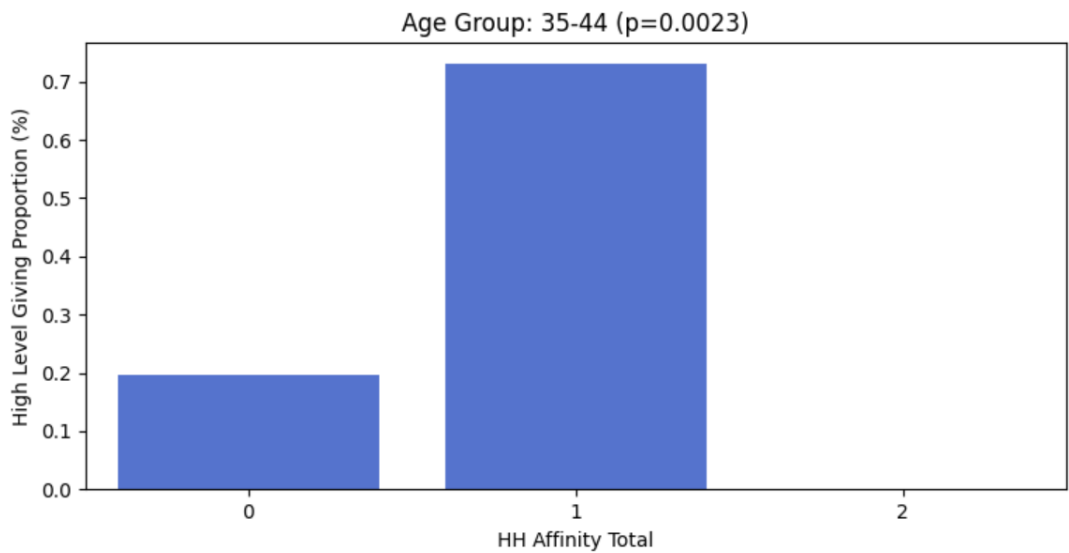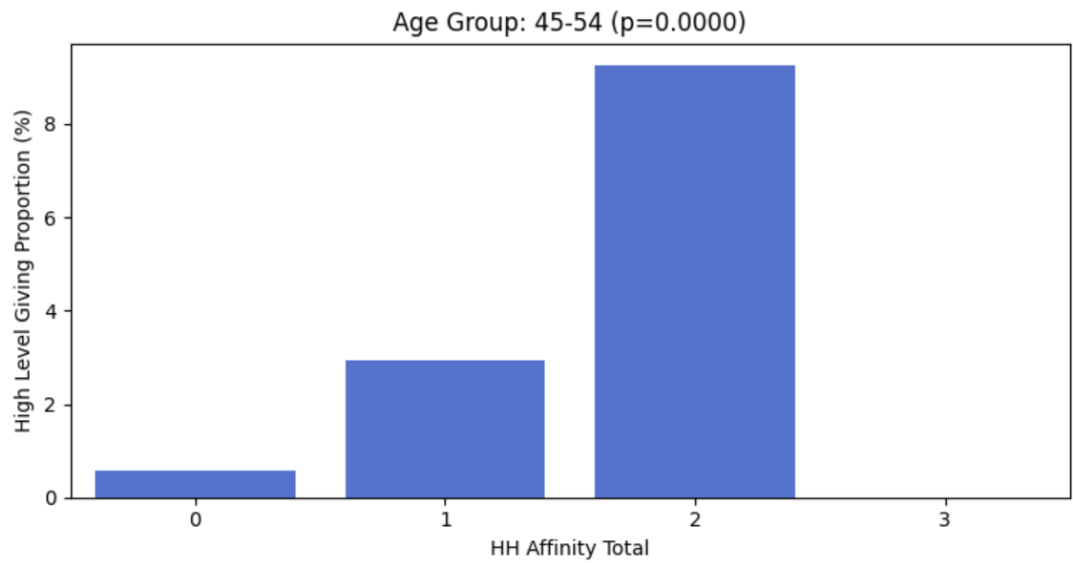
By age group, the overall upward trend remains consistent across different age groups. However, high-level donors are primarily concentrated in the 50+ age group. The 25-34 age group is not displayed because it is too young to have any recorded high-level

donors.

Different Regions:
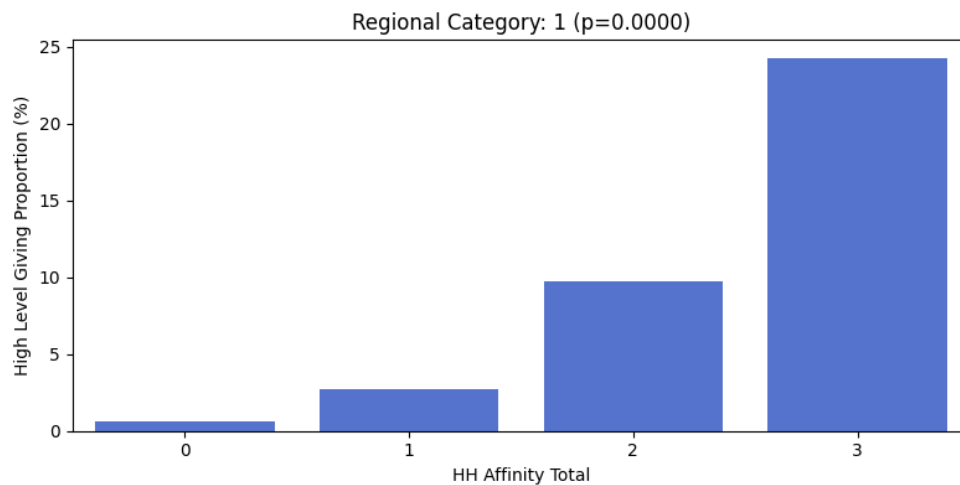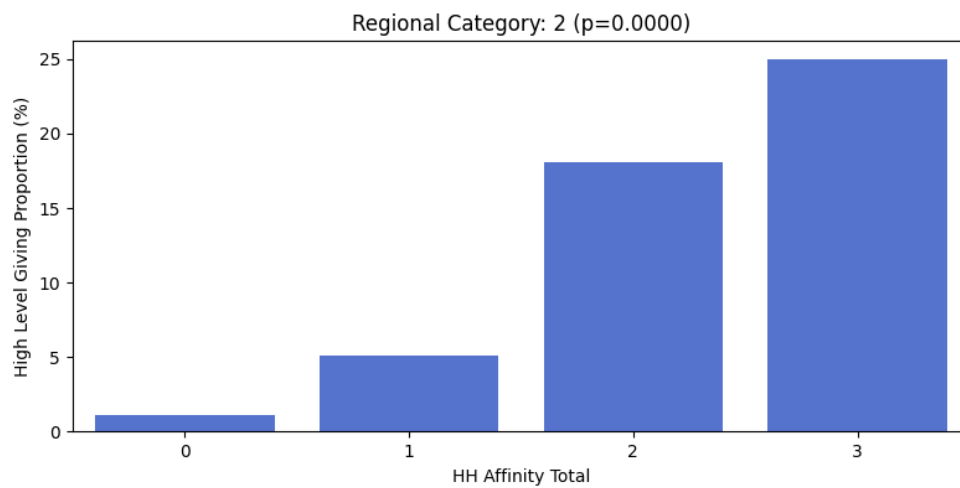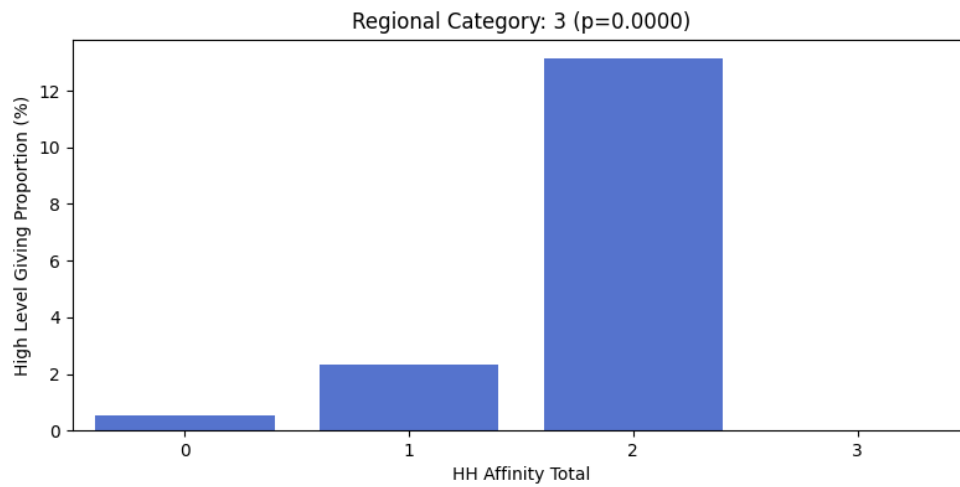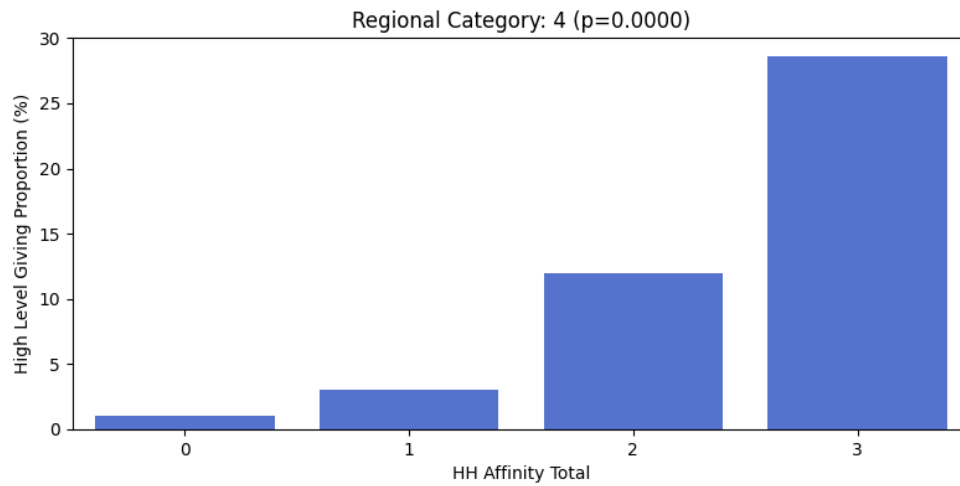
Across different regions, the overall scale is similar, except for Region 3, which has a obvious smaller total number compared to other regions.

4. Clustering Analysis for Pattern Exploration

K-Means:
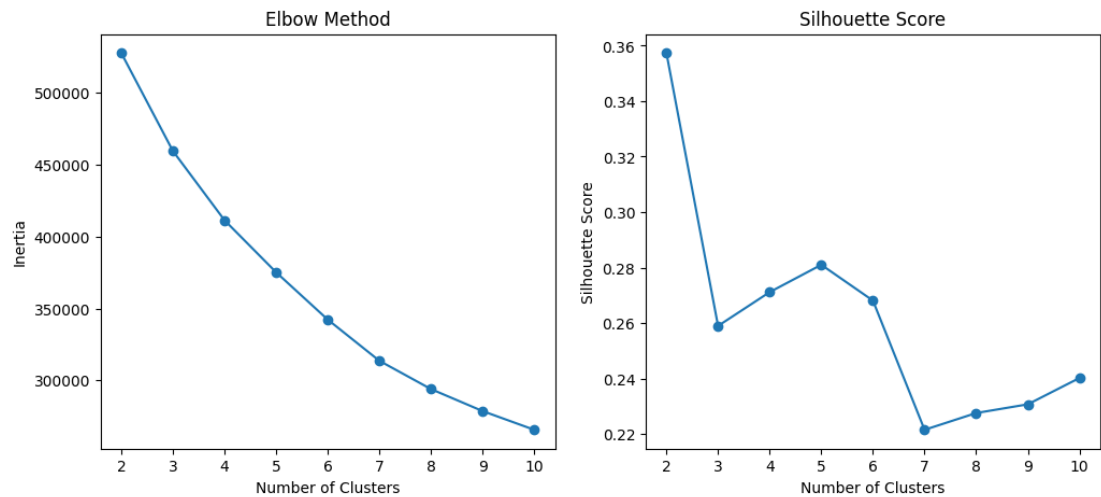
Based on the above chart results, we can see that the classification effectiveness of k=4 and k=5 is similar.

K = 4:

```
        Age Range Category  Regional Category Codes  \
Cluster
0                 5.062371                 1.951160
1                 3.759635                 2.069980
2                 2.713529                 2.413895
3                 3.812003                 2.378311


        BUSN Volunteer (either type) (Y/N)  \
Cluster
0                                      0.0
1                                      1.0
2                                      0.0
3                                      0.0


        HH Affinity Total (parent, child, spouse)  fy23 Engagement excl $$  \
Cluster
0                                        0.180284                 0.325000
1                                        0.421907                 1.136917
2                                        0.437962                 1.281330
3                                        0.147913                 0.127658


        fy24 Engagement excl $$  BUSN Affiliation-Employment  \
Cluster
0                      0.314046                    44.690722
1                      1.119675                    22.870183
2                      1.236467                     0.451749
3                      0.121918                     1.696021


        BUSN Affiliation-Student Involvement  \
Cluster
0                                   5.006443
1                                  15.770791
2                                  15.456811
3                                   2.968037


        Affiliation- BUSN score minus giving  UIUC Affiliation minus giving  \
Cluster
0                                1042.466366                    1042.141366
1                                1017.666329                    1016.529412
2                                 551.803022                     550.521692
3                                 511.648434                     511.520776


        UIUC Affiliation Employment  UIUC Affiliation-Student Involvement
Cluster
0                        113.247423                             31.939433
1                         55.476673                             76.166329
2                          9.767116                             67.569515
3                         15.433790                             23.778539
```

Cluster Count and Average Giving
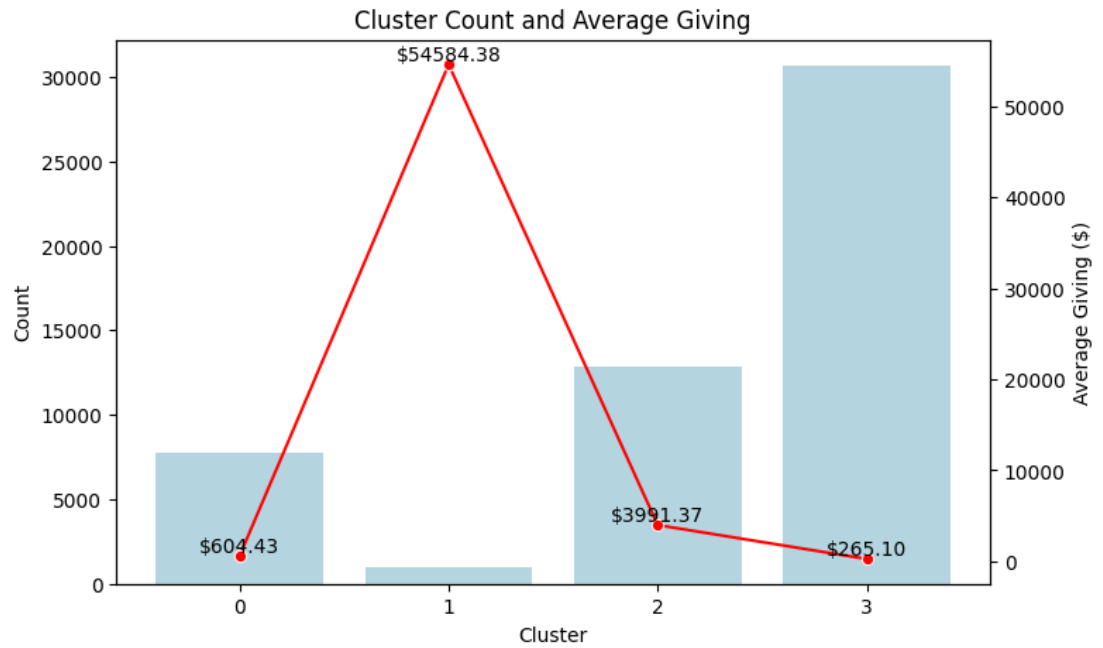
We can see that although Cluster 1 has just over 1,000 individuals, the average donation per person exceeds $50,000, making them our most promising alumni group. Additionally, Cluster 2 has a large population of 13,000, with an average donation of around $4,000, indicating potential as well.

We can identify alumni who share similar characteristics with these two clusters and target them with more outreach efforts to encourage donations.

K = 5:

```
          Age Range Category  Regional Category Codes  \
Cluster
0                  4.983007                 1.934220
1                  3.697743                 2.399489
2                  2.777556                 2.330222
3                  4.406439                 2.547284
4                  3.767656                 2.068577


          BUSN Volunteer (either type) (Y/N)  \
Cluster
0                                   0.000000
1                                   0.000000
2                                   0.000000
3                                   0.018109
4                                   1.000000


          HH Affinity Total (parent, child, spouse)  fy23 Engagement excl $$  \
Cluster
0                                         0.198849                   0.332877
1                                         0.133289                   0.221007
2                                         0.601667                   1.396444
3                                         0.207243                   0.577465
4                                         0.418628                   1.131013


          fy24 Engagement excl $$  BUSN Affiliation-Employment  \
Cluster
0                        0.320680                    22.755927
1                        0.214916                     2.117538
2                        1.335556                     0.755556
3                        0.593561                   328.873239
4                        1.115660                    18.270215


          BUSN Affiliation-Student Involvement  \
Cluster
0                                     5.180211
1                                     2.702036
2                                    21.788889
3                                     3.319920
4                                    15.506653


          Affiliation- BUSN score minus giving  UIUC Affiliation minus giving  \
Cluster
0                               1053.508976                      1053.176100
1                                512.134797                       511.913790
2                                558.958444                       557.562000
3                               1019.637827                      1019.060362
4                               1007.699079                      1006.568066


          UIUC Affiliation Employment  UIUC Affiliation-Student Involvement
Cluster
0                           64.211320                             31.769905
1                           15.605964                             22.151476
2                            9.316667                             92.861111
3                          800.000000                             27.766600
4                           45.138178                             76.253838
```

Cluster Count and Average Giving
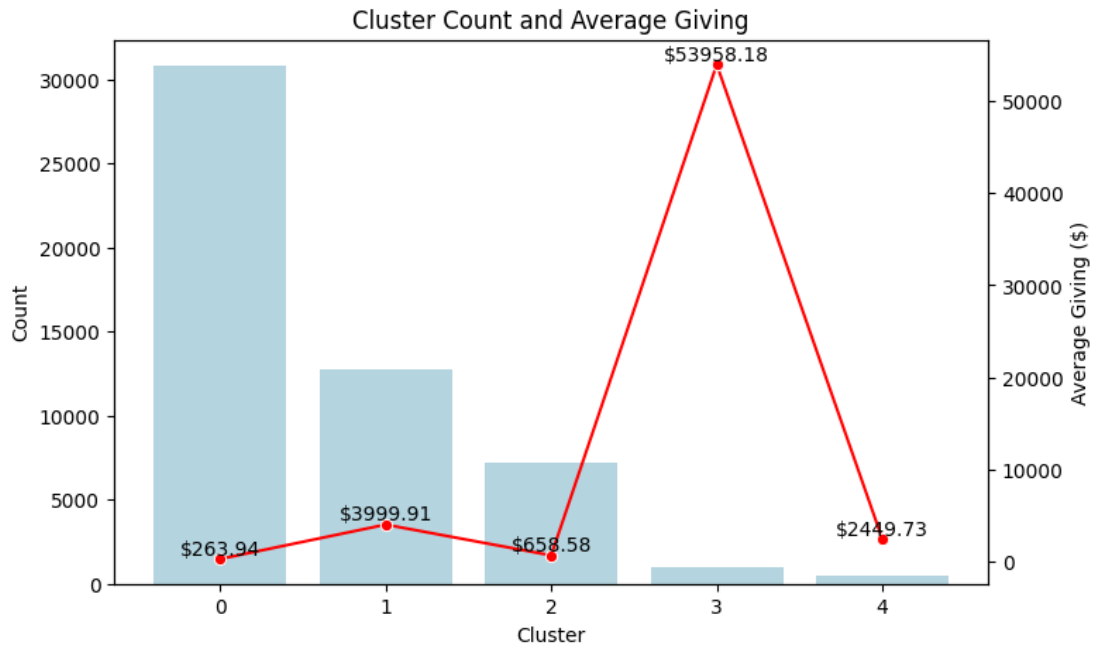
When the number of clusters increases to 5, the effectiveness is not as strong as before. Clusters 3 and 1 can be targeted for increased advertising efforts.

DBSCAN:

DBSCAN Clustering Distribution
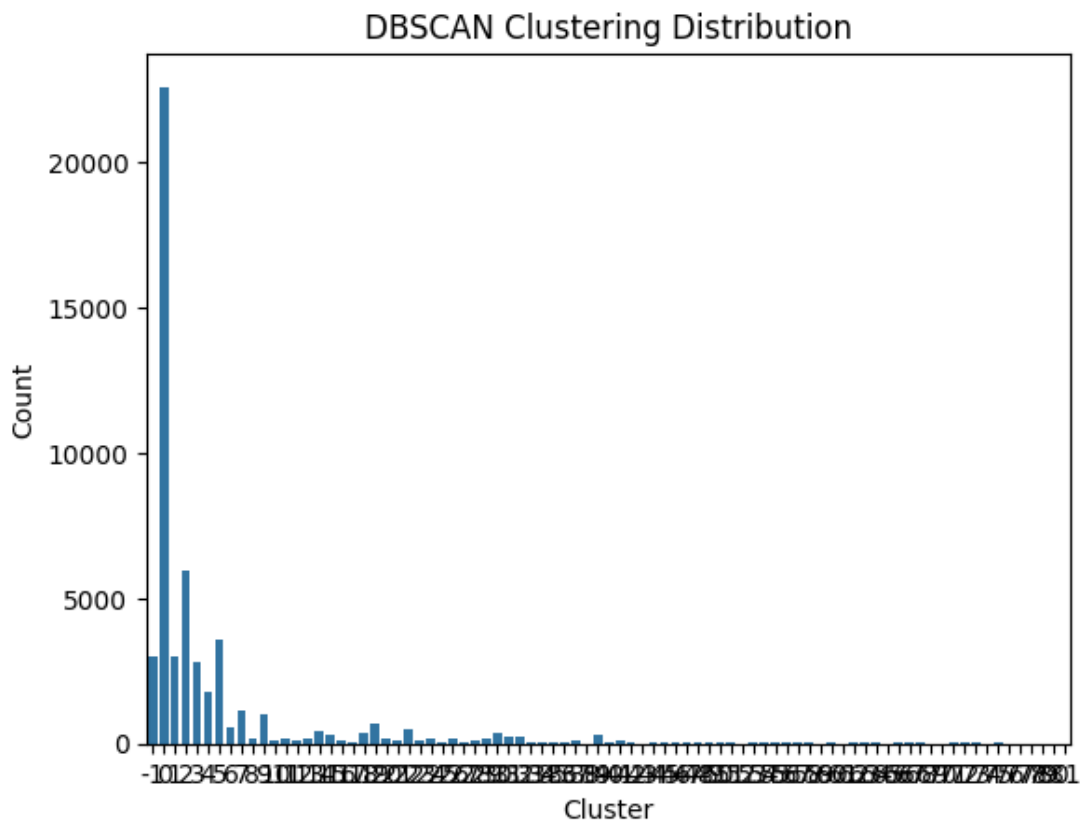
Since density-based clustering does not allow for prior control over

the number of clusters, it instead aggregates all the clusters based on the density of the data points. From the plot, the first column labeled as -1 represents all the outliers, not a cluster. The clusters start from the second column, which is the 0th cluster.

We can see that the first 6-7 clusters are more prominent. We will extract these and examine the specific feature values.

```
Cluster Characteristics (Mean of features for clusters 0-6):
               Age Range Category  Regional Category Codes  \
Cluster_DBSCAN
0                        3.966248                 2.337831
1                        4.368527                 2.342972
2                        3.591466                 2.433563
3                        4.510631                 2.348335
4                        2.828962                 2.307868
5                        3.451308                 2.180022
6                        2.330896                 2.323583

               BUSN Volunteer (either type) (Y/N)  \
Cluster_DBSCAN
0                                             0.0
1                                             0.0
2                                             0.0
3                                             0.0
4                                             0.0
5                                             0.0
6                                             0.0

               HH Affinity Total (parent, child, spouse)  \
Cluster_DBSCAN
0                                                    0.0
1                                                    0.0
2                                                    0.0
3                                                    0.0
4                                                    1.0
5                                                    1.0
6                                                    1.0

               fy23 Engagement excl $$  fy24 Engagement excl $$  \
Cluster_DBSCAN
0                                  0.0                      0.0
1                                  1.0                      0.0
2                                  1.0                      1.0
3                                  0.0                      1.0
4                                  1.0                      1.0
5                                  0.0                      0.0
6                                  2.0                      2.0
```

```
                  BUSN Affiliation-Employment  \
Cluster_DBSCAN
0                            6.115633
1                            7.750504
2                            6.450529
3                           10.878809
4                            1.510832
5                            2.615470
6                            0.091408


                  BUSN Affiliation-Student Involvement  \
Cluster_DBSCAN
0                                              0.0
1                                              0.0
2                                              0.0
3                                              0.0
4                                              0.0
5                                              0.0
6                                              0.0


                  Affiliation- BUSN score minus giving  \
Cluster_DBSCAN
0                                       581.202690
1                                       610.580699
2                                       579.694776
3                                       617.251949
4                                       573.765678
5                                       583.066221
6                                       560.080439


                  UIUC Affiliation minus giving  UIUC Affiliation Employment  \
Cluster_DBSCAN
0                                581.202690                    22.193223
1                                609.580699                    32.178884
2                                578.694776                    23.189988
3                                617.251949                    35.116938
4                                572.765678                     8.608894
5                                583.066221                    13.606010
6                                558.080439                     3.382084


                  UIUC Affiliation-Student Involvement
Cluster_DBSCAN
0                                        20.356321
1                                        19.905851
2                                        25.986897
3                                        18.710135
4                                        37.243444
5                                        28.658876
6                                        50.822669
```
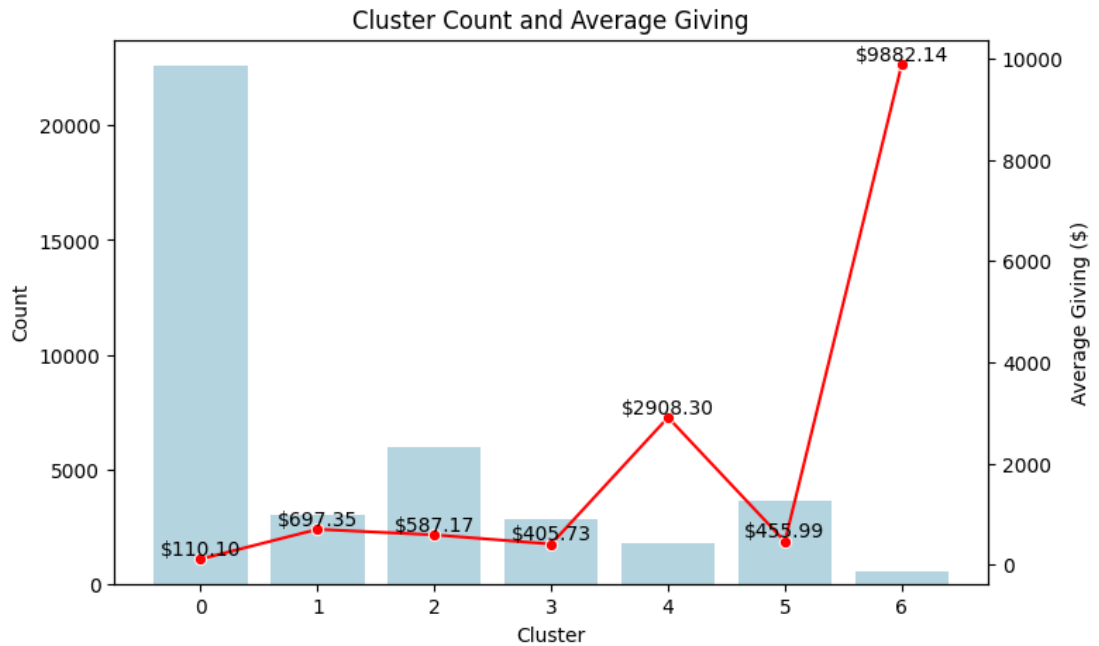
Cluster Count and Average Giving

DBSCAN has provided us with a different perspective. We can first prioritize Cluster 4 and Cluster 6 as the most important targets. Next, Cluster 1 and Cluster 2 can be considered as secondary priorities. Alumni in these smaller groups all show strong donation potential.

3. Reflection and Future Plans

Reflection

1. When comparing different groups or examining the relationships between different variables, simply using data visualization, such as bar charts and histograms, is not sufficient. Statistical tests are also necessary to objectively confirm whether the relationships have statistical significance.

2. When performing clustering analysis, selecting appropriate variables is crucial. There are many raw variables, but including all of them can make the dataset too large, leading to time-consuming and labor-intensive processing. Additionally, choosing the right model is important. A model that is too complex might result in excessively long run times and may not be suitable for practical work environments. In this case, we used K-means and DBSCAN, which took between a few minutes to ten minutes, making them acceptable. However, when using Hierarchical Clustering, the system nearly crashed. Thus, selecting the right model is vital.

3. Finally, it is important to understand the purpose of clustering. Although we did not use donation-related data during the clustering process, after clustering, we need to compare the different groups with donation amounts to understand how their donation behaviors differ. Initially, the clustering only reflected the characteristics of each group but did not compare them with donation behaviors. Only by comparing donation data can we understand how these clusters help us identify alumni with different donation potentials. It is important to always keep the business objective in mind.

Next Plan

1. We will further deepen the clustering analysis by extracting key features from different groups, which will help in identifying alumni belonging to different clusters and their corresponding donation potential.

2. Based on the client's needs, we will continue to conduct significant analysis on the relationships between other variables to determine their specific impact on donation behavior.