

Release 2 Report

1. Introduction

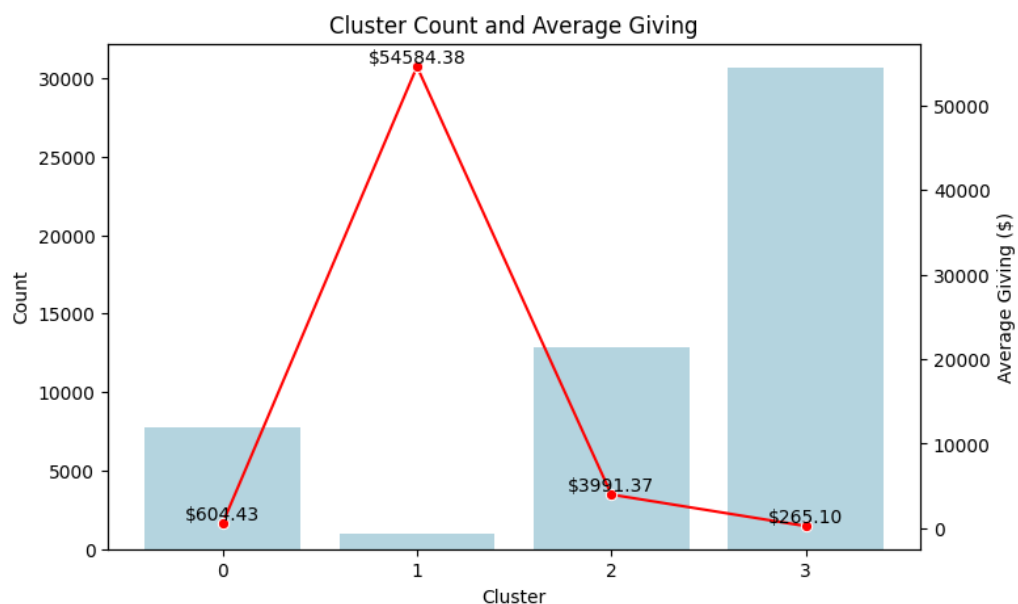
Over the past period of work:

1. We first built on the previous clustering results and conducted a deeper exploration. Specifically, we identified and summarized variables with significant differences in mean values between groups 1 and 2, and between groups 0 and 3 — both of which represent low-potential clusters. We then focused on comparing high-potential group 1 with low-potential group 2, summarizing the key variables where their mean values differ. These insights can help the client better understand and distinguish between these alumni groups.
2. However, comparing mean values alone is not comprehensive. For a more complete understanding of each group, we also examined the distribution of values for the key differentiating variables using histograms. In addition, we included parallel plots for the overall clustering analysis, which allow for a quick overview of how the groups were formed based on variable distributions.

In the following sections, we will walk through the visuals and explain each in detail.

2. Outcomes in Release 3 Phrase

1. Comparison of Variables' Mean Value



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.

Cluster 1: about **1,000 people** with mean **donation about \$55,000**

Mean value of variables (The **green ones** are the values have big difference between cluster 1,2 and cluster 0,4; The **yellow ones** show the difference between cluster 1 and 2):

Age Range Category: about 3.7 (3~4)

Regional Category Code: about 2

BUSN Volunteer: 1

HH Affinity Total: 0.42

fy23 Engagement excl \$\$: 1.14

fy24 Engagement excl \$\$: 1.12

BUSN Affiliation-Employment: 23

BUSN Affiliation-Student Involvement: 16

Affiliation- BUSN score minus giving: 1018

UIUC Affiliation Employment: 55

UIUC Affiliation-Student Involvement: 76

Cluster 2: about **13,000 people** with mean **donation about \$4,000**.

Mean value of variables:

Age Range Category: about 2.7 (2~3)

Regional Category Code: about 2.41(2-3)

BUSN Volunteer: 0

HH Affinity Total: 0.44

fy23 Engagement excl \$\$: 1.28

fy24 Engagement excl \$\$: 1.24

BUSN Affiliation-Employment: 0.45

BUSN Affiliation-Student Involvement: 15

Affiliation- BUSN score minus giving: 552

UIUC Affiliation Employment: 9.77

UIUC Affiliation-Student Involvement: 67.6

We can distinguish the two groups with donation potential from the other two groups based on the following characteristics:

- Their **HH Affinity Total** is around **0.42**.
- Their **FY23 or FY24 engagement** is approximately **1.2**.
- Their **BUSN Affiliation - Student Involvement** is significantly higher than the other two groups (ranging from 2 to 5), at around **16**.
- Their **UIUC Affiliation - Student Involvement** is around **70**, which is noticeably higher than the other two groups (which range between 20 and 30).

Based on these five variables, we can differentiate the high-potential donor clusters (Clusters 1 and 2) from the low-donation clusters (Clusters 0 and 3).

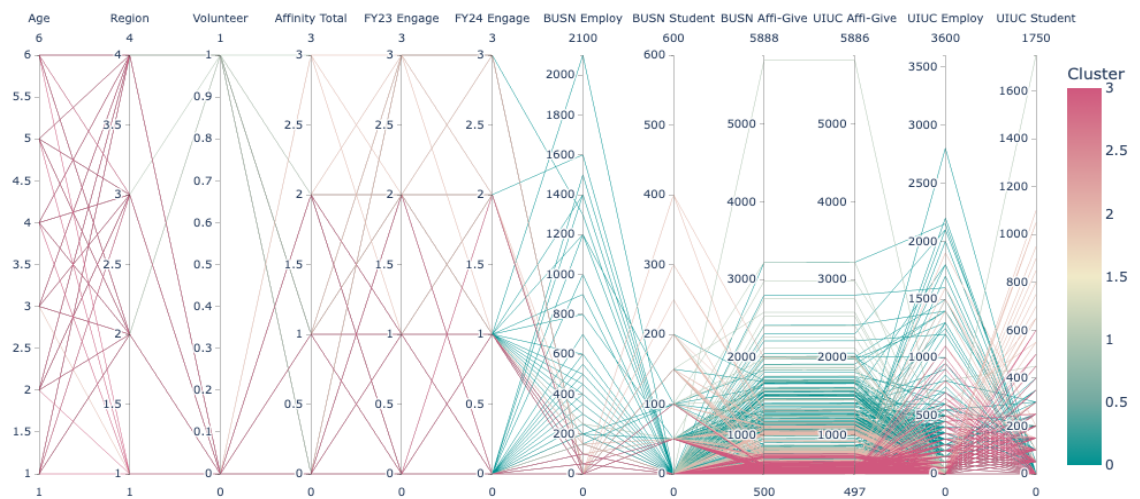
Next, using the variables highlighted in yellow, we can further distinguish **high-donation potential alumni (Cluster 1)** from **moderate-donation potential alumni (Cluster 2)**, achieving a more effective identification.

Variable	Cluster 1 (High Donation Potential)	Cluster 2 (Moderate Donation Potential)
Age Range Category	3.7 (3~4)	2.7 (2~3)
BUSN Volunteer	1	0
BUSN Affiliation – Employment	23	0.45
Affiliation – BUSN Score Minus Giving	1018	552
UIUC Affiliation – Employment	55	9.77

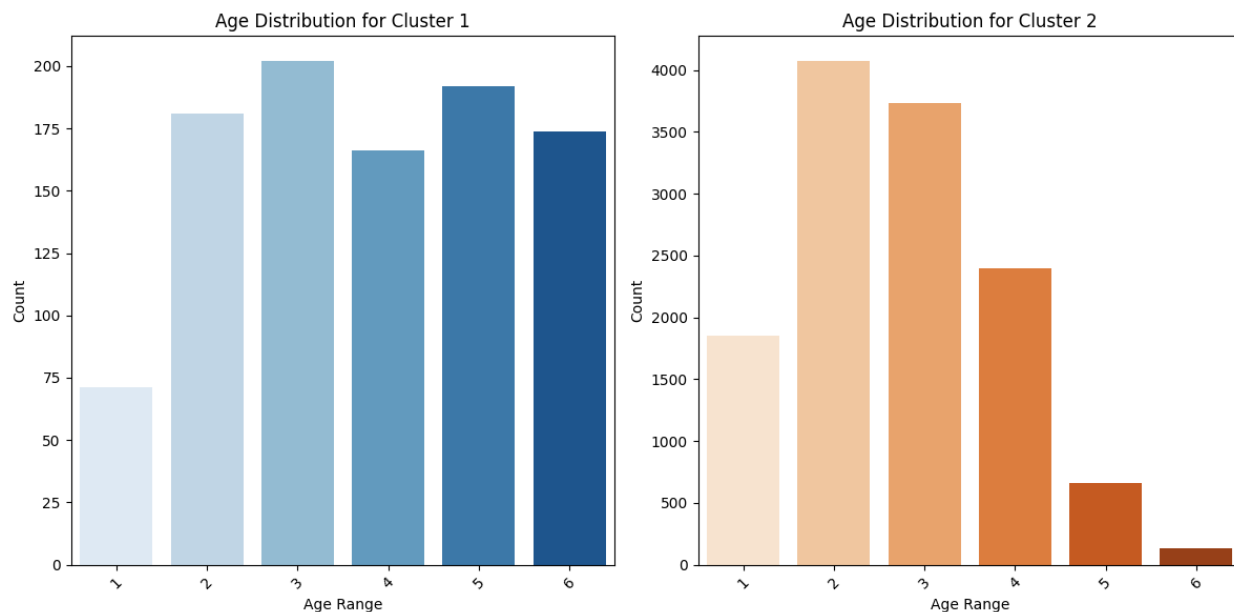
2. Analysis and Significance Testing of Key Variable Distributions Between Cluster 1 and Cluster 2

Overall view for all clusters with **Parallel Plots**

Parallel plots visually display multiple variables at once by representing each variable as a vertical line. Data points are connected across these lines, making it easy to compare patterns and relationships across different features. This is helpful for understanding trends and differences between groups or segments at a glance.



1. Visualize the age distribution separately for Cluster 1 and Cluster 2

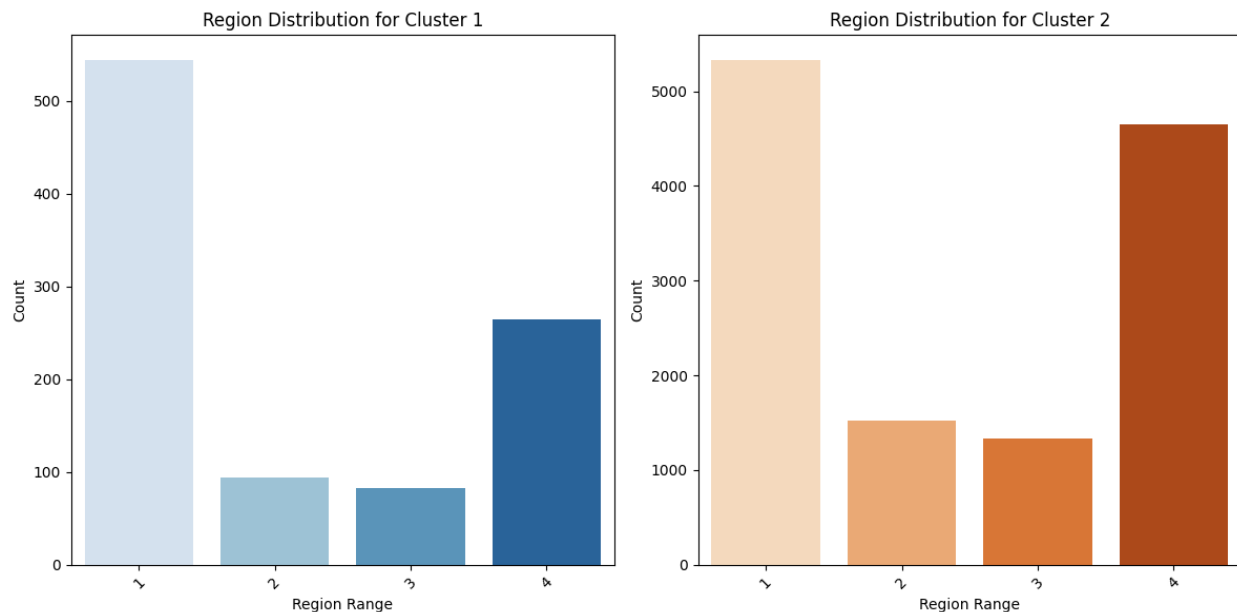


KS-statistic: 0.310

P-value: 0.000

The distributions of 'Age Range Category' between Cluster 1 and Cluster 2 are significantly different.

2. Visualize the region distribution separately for Cluster 1 and Cluster 2

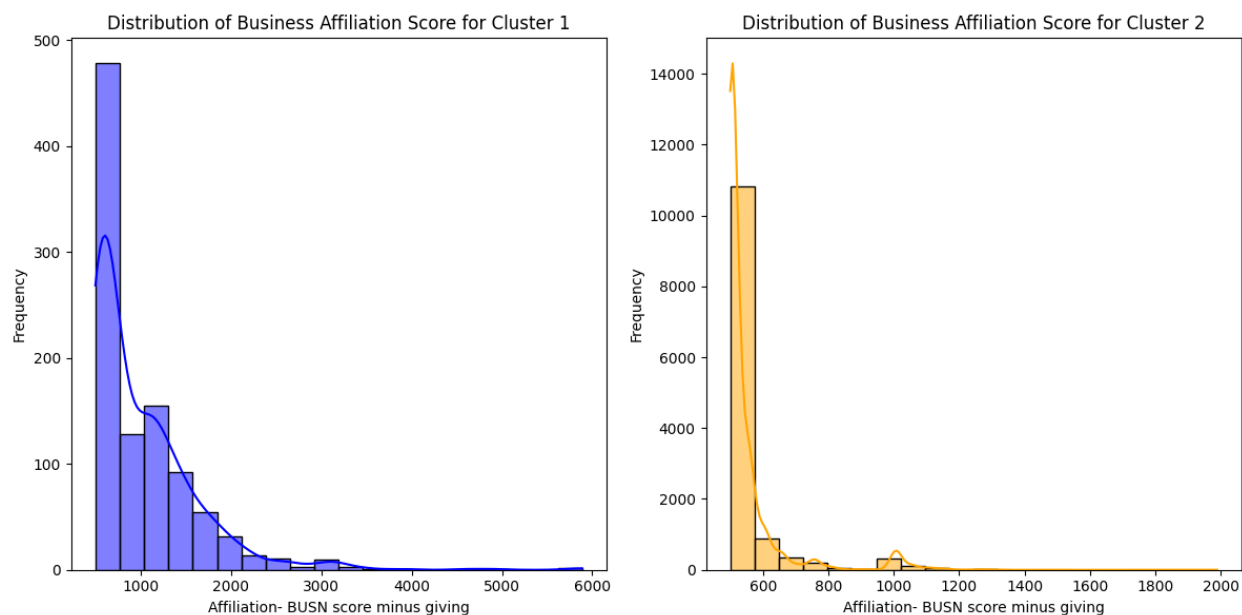


KS-statistic: 0.137

P-value: 0.000

The distributions of 'Regional Category Codes' between Cluster 1 and Cluster 2 are significantly different.

3. Visualize the distribution of the Business Affiliation Score for cluster 1, 2



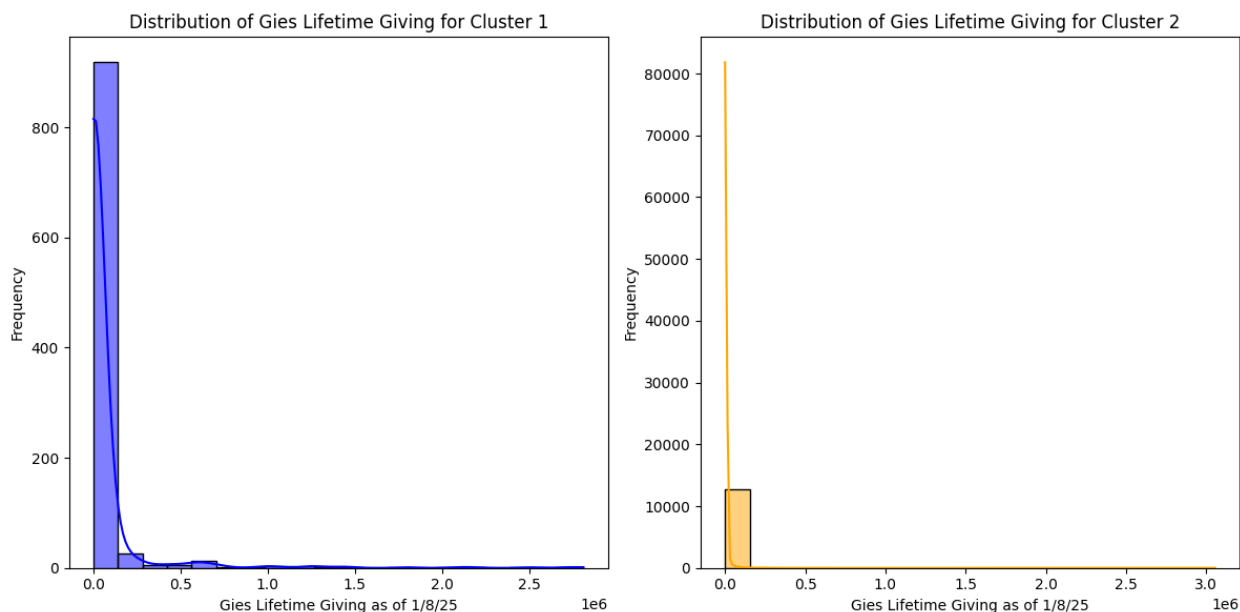
KS-statistic: 0.624

P-value: 0.000

The distributions of 'Business Affiliation Score' between Cluster 1 and Cluster 2 are significantly different.

4. Is the giving distribution also significantly different for cluster 1 and 2?

Gies Lifetime Giving as of 1/8/25



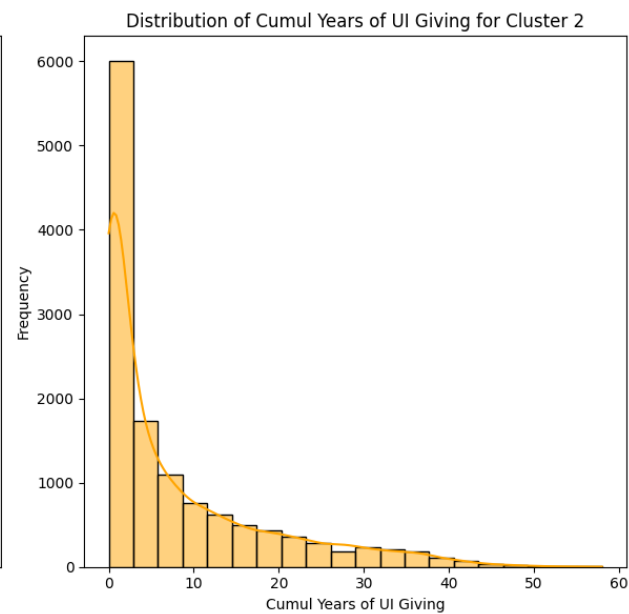
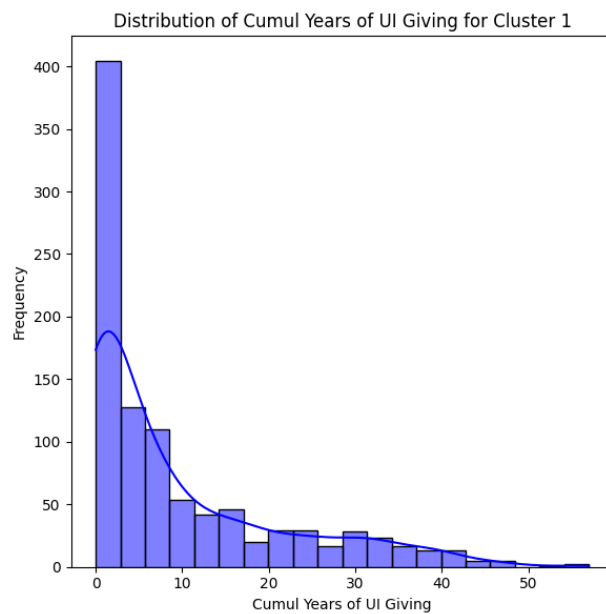
(The smooth curve on top of the histogram represents the estimated **probability density** function of the data. **This curve helps visualize the distribution of the data more smoothly**, in contrast to the stepped appearance of the histogram)

KS-statistic: 0.229

P-value: 0.000

The distributions of 'Gies Lifetime Giving' between Cluster 1 and Cluster 2 are significantly different.

Cumul Years of UI Giving

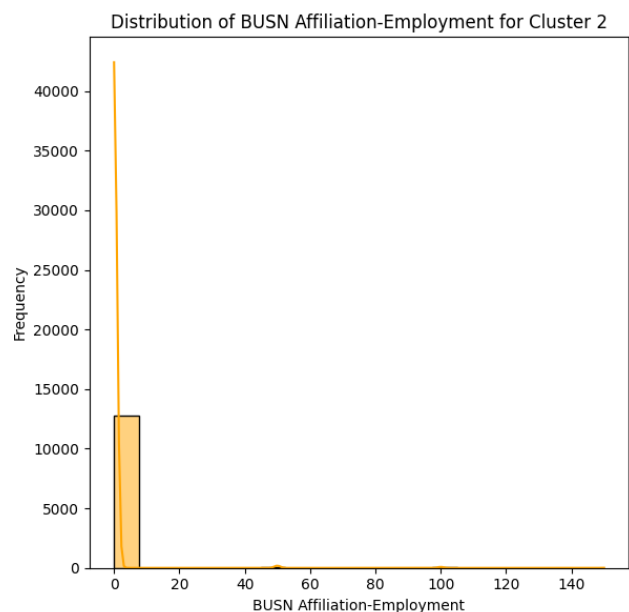
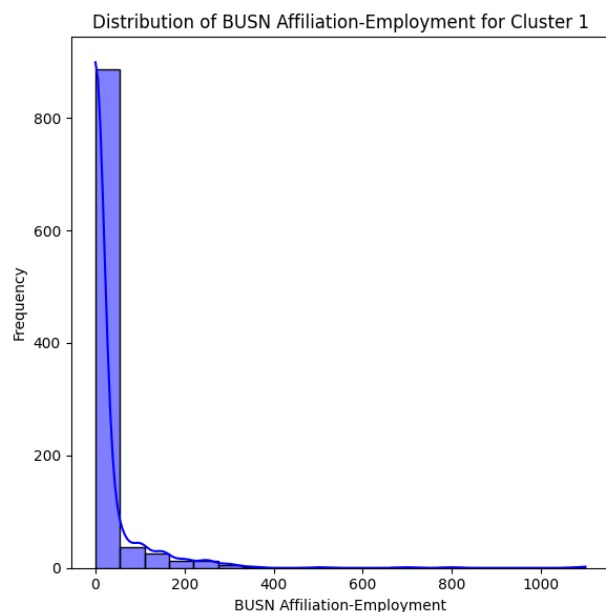


KS-statistic: 0.063

P-value: 0.001

The distributions of 'Cumul Years of UI Giving' between Cluster 1 and Cluster 2 are significantly different.

5. BUSN Affiliation-Employment distribution for cluster 1 and 2

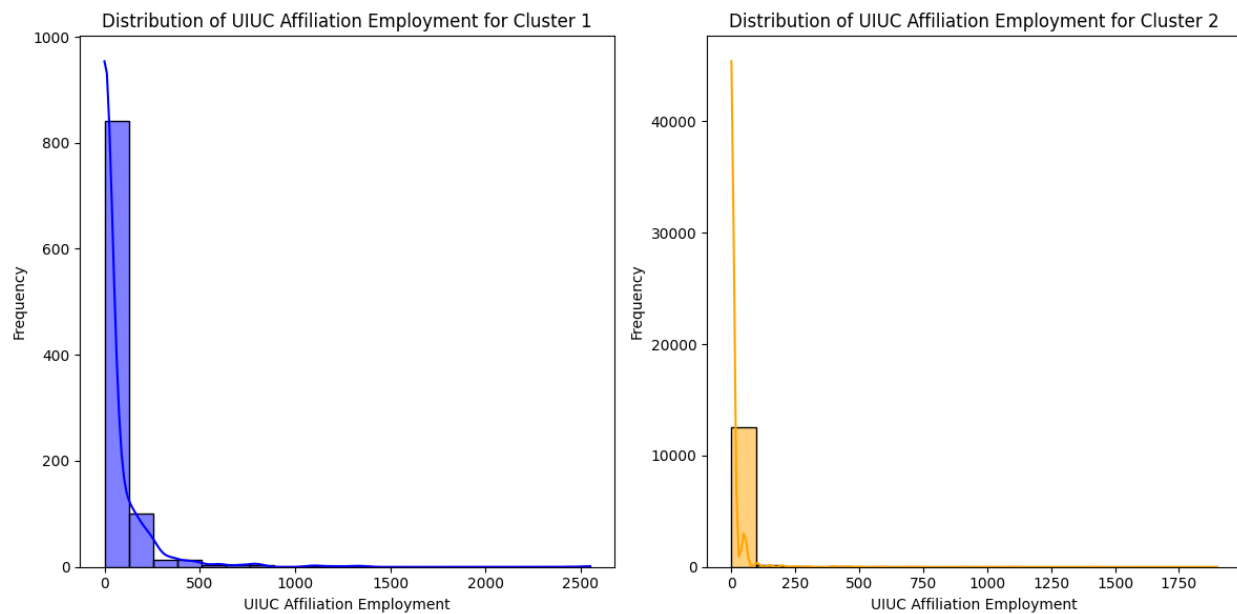


KS-statistic: 0.152

P-value: 0.000

The distributions of 'BUSN Affiliation-Employment' between Cluster 1 and Cluster 2 are significantly different.

6. UIUC Affiliation Employment distribution for cluster 1 and 2



KS-statistic: 0.191

P-value: 0.000

The distributions of 'UIUC Affiliation Employment' between Cluster 1 and Cluster 2 are significantly different.

Finally, we get a excel file with IDs for different clusters-
"Cluster_IDs_by_Group.xlsx"

Like this following:

AutoSave

Cluster_IDs_by_Group

Home

Insert

Draw

Page Layout

Formulas

Data

Review

View

Automate

Paste

Calibri (Body)

11

A

A

B

I

U

General

\$

%

Condition

Format as

Cell Style

N27

	A	B	C	D	E	F	G	H	I	J
1	Individual LookupID	Gies Lifetime Giving as of 1/8/25	Cumul Years of UI Giving							
2	10860126	0	1							
3	13570337	0	0							
4	13710801	0	0							
5	13777302	0	0							
6	11607518	0	0							
7	11766673	0	0							
8	11542811	0	0							
9	11591180	0	0							
10	11625960	50	1							
11	11732161	0	0							
12	13710394	0	0							
13	11069955	0	0							
14	14084151	0	0							
15	11367679	0	0							
16	13517873	0	0							
17	11673416	0	0							
18	11513140	0	0							
19	11547483	0	0							
20	10844227	0	0							
21	11411054	195	9							
22	11526467	0	0							
23	10540702	0	0							
24	10998137	0	0							
25	11103227	0	0							
26	11697725	0	0							
27	11374914	475	18							
28	11101607	0	0							
29	11413521	75	4							
30	11736520	0	0							
31	13778281	0	0							
32	11532484	40	3							
33	13652568	0	0							
34	11599833	0	0							
35	10526057	0	0							
36	11744805	0	0							
37	11467876	0	1							
38	11761831	0	0							
39	11576703	0	0							
40	11133985	0	0							
41	13807379	0	0							

Cluster_0

Cluster_1

Cluster_2

Cluster_3

+

3. Next Plans

1. First, the client was particularly interested in the variable BUSN Affiliation Score minus Giving, which showed significant differences in distribution between cluster 1 and cluster 2. Since this score is composed of several sub-variables, the client would like to understand which sub-variable contributes the most to the difference between these two clusters. This would help them identify which variable to focus on in future monitoring or analysis.
2. Based on this question, we took the following approach: we first joined the Excel file containing the cluster IDs (from our previous clustering results) with the full dataset that includes all sub-variable values. This ensures that for each alumni ID, we have complete information on their corresponding sub-variable scores.
3. Next, we performed linear regression separately on alumni from cluster 1 and cluster 2. We set the total score (BUSN Affiliation minus Giving) as the dependent variable (y), and the sub-variables as independent variables (x). We also included a categorical variable indicating cluster membership (cluster 1 or cluster 2). This regression allows us to assess the contribution of each sub-variable to the overall score and helps identify which one plays the most significant role in differentiating the two clusters.