Release 4 Report

1. Introduction

Over the past period of work:

As we already found several important features whose distributions are significantly different between cluster 1 and cluster 2, such as the features about ages, regions, business school's affiliation, UIUC's affiliation, lifetime giving, employment and so on.

Then, after seeing these variables' distribution difference, the client is more interested in the difference of BUSN affiliation total score (the total is a score combination with lots of sub-variables) between cluster 1 and cluster 2 – which sub-variables make the biggest contribution into this difference of BUSN affiliation total score between cluster 1 and cluster 2. Therefore, we used Attribution of Group Mean Differences, KS-Test and Wasserstein Distance methods to figure out the sub-variables which give the biggest contribution to this.

2. Outcomes in Release 4 Phrase

Attribution of Group Mean Differences:

This method helps identify which factors contribute the most to the overall mean difference between two groups, making it easier to prioritize areas for targeted action or strategic focus.

As we can see from the graph below, Volunteer, Degree and Events give the top 3 contribution to the mean difference of BUSN affiliation between cluster 1 and cluster 2. However, seeing the mean difference is not enough, we also want to analyze the difference contribution in the whole distribution.

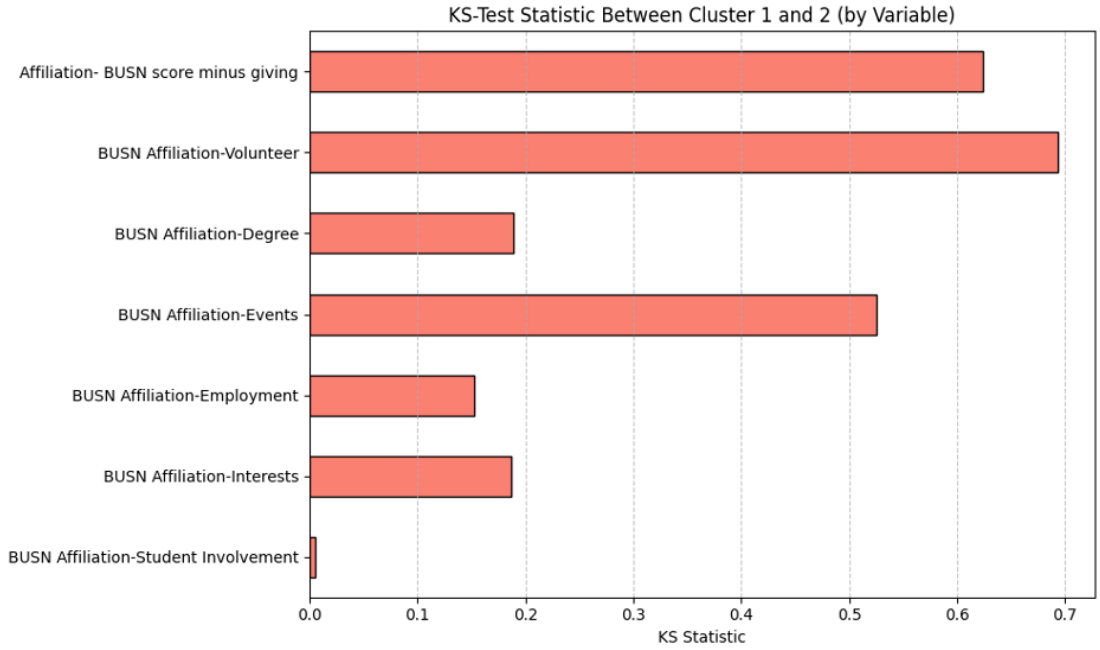
	Mean Difference	Contribution to Total Diff (%)
BUSN Affiliation-Volunteer	277.381350	59.541360
BUSN Affiliation-Degree	97.879322	21.010309
BUSN Affiliation-Events	54.737175	11.749621
BUSN Affiliation-Employment	22.418434	4.812234
BUSN Affiliation-Interests	13.171354	2.827300
BUSN Affiliation-Student Involvement	0.313980	0.067397

KS-Test:

The KS-test helps us measure whether two groups differ significantly in how

their affiliation scores are distributed, highlighting where the biggest behavioral gaps lie.

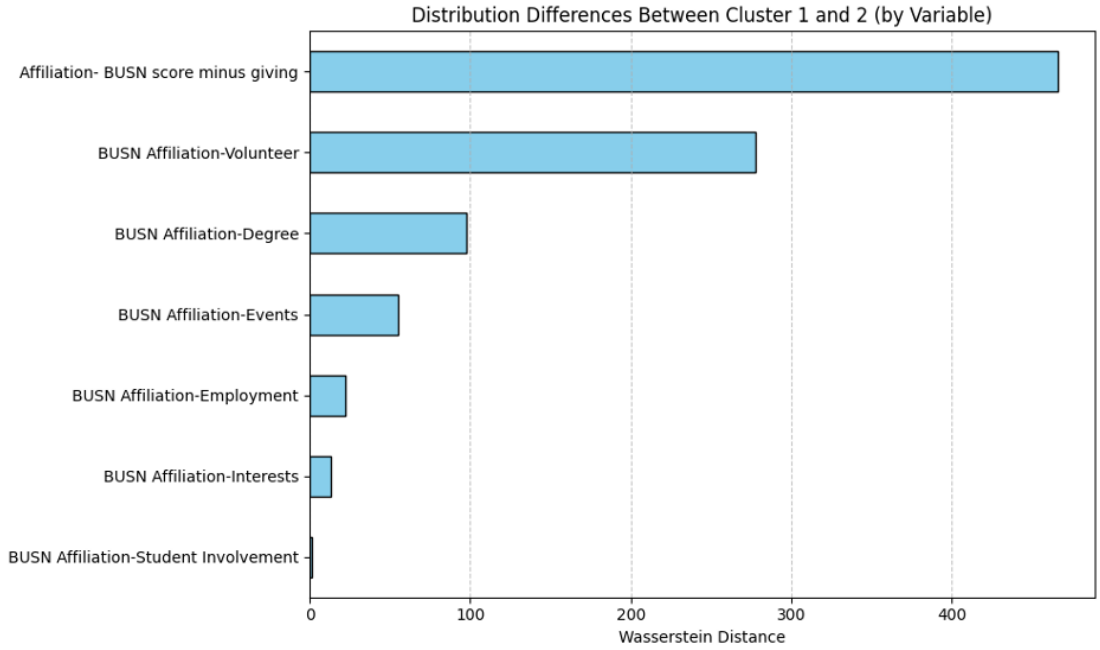
This illustration shows that main contribution comes from Volunteer, Degree, and Events.




```
Affiliation- BUSN score minus giving: KS-stat=0.624, p-value=0.000
BUSN Affiliation-Degree: KS-stat=0.189, p-value=0.000
BUSN Affiliation-Employment: KS-stat=0.152, p-value=0.000
BUSN Affiliation-Events: KS-stat=0.525, p-value=0.000
BUSN Affiliation-Interests: KS-stat=0.187, p-value=0.000
BUSN Affiliation-Student Involvement: KS-stat=0.005, p-value=1.000
BUSN Affiliation-Volunteer: KS-stat=0.693, p-value=0.000
```

Wasserstein Distance:

The Wasserstein Distance quantifies how far two groups' distributions are from each other, helping us understand which factors show the biggest overall shifts in behavior.




```
Affiliation- BUSN score minus giving: Wasserstein Distance = 465.863
BUSN Affiliation-Degree: Wasserstein Distance = 97.879
BUSN Affiliation-Employment: Wasserstein Distance = 22.418
BUSN Affiliation-Events: Wasserstein Distance = 54.737
BUSN Affiliation-Interests: Wasserstein Distance = 13.171
BUSN Affiliation-Student Involvement: Wasserstein Distance = 1.014
BUSN Affiliation-Volunteer: Wasserstein Distance = 277.381
```

According to the three methods outcomes above, we can make a conclusion that the big difference of business school's affiliation total score between cluster 1 and cluster 2 is mainly due to Volunteer, Degree and Events, which means the alumni with higher score about Volunteer, Degree and Events could be potential people to make a consideration.

3. Reflection and Next Plans

Challenges and Future Recommendations

In the past weeks, we encountered two major challenges in our project and identified one critical area for future improvement.

Challenges:

Our first significant challenge was accurately identifying different donation groups among alumni. To address this, we applied clustering analysis methods. This approach successfully grouped alumni based on their characteristics and donation behaviors, effectively resolving this identification issue.

The second challenge was understanding which specific variables had the greatest impact on alumni donations. Initially, we attempted to analyze these

contributions using linear regression. However, we found this method was not suitable for clearly distinguishing key variables. We then adopted the KS-test and Wasserstein Distance methods, which provided clearer insights into the factors driving differences in donation behaviors between alumni groups.

Future Improvement:

Moving forward, we have identified a crucial improvement area: conducting a detailed analysis of affiliation scores. Our client specifically required us to clarify differences in affiliation scores between UIUC and BUSN alumni. Thus, our next step will involve breaking down the overall affiliation score into its individual components. By analyzing these sub-factors in detail, we aim to clearly identify which specific factors contribute most significantly to the higher donation rates observed among BUSN alumni.

Through addressing these challenges and implementing our recommended future improvement, we anticipate enhancing the effectiveness and precision of our alumni donation strategies.