Brian Rabinowitz - Data Preprocessing, Visualizations, Analysis
Sudiksha Kalepu - Clustering, Silhouette Scoring, Analysis/Proofreading

<u>Nightclub Recommendations Based on Agglomerative Clustering</u>

In this report, we will attempt to derive meaningful analysis and actionable insights through applying clustering on the given nightclub customer dataset. We will discuss the methodology of how we arrived at our clusters, and how we can interpret said clusters into a plan of action moving forward.

<u>Methodology</u>

We began by preprocessing our data by dropping entries with missing values (because there were almost none) and encoding categorical data in different ways. Categorical variables with meaningful ranking relationships such as education and nightclub frequency were given numeric rankings, and income was specifically encoded by taking the lower bound of the given bracket (i.e $20,000-$50,000 becomes 20,000). Other categorical variables without meaningful ordinal relationships between categories were left alone when performing K-Prototypes clustering, and were encoded using one-hot encoding when performing Agglomerative clustering. All encoded and numeric variables were normalized using z-scores prior to clustering.

We then performed both K-Prototypes and Agglomerative clustering on our data with varying numbers of clusters, and performed silhouette analysis to determine the best possible clustering for interpretation. This resulted in using our two cluster, Agglomerative based labels.

<u>Results/Visualizations</u>

Our clustering analysis ended up revealing two clusters, one larger cluster, around ⅘ of the data (Cluster P), that largely reflected the makeup of the overall data, reflecting a generally positive stance on the club's operations, and one smaller cluster that was largely dissatisfied with the club (Cluster N).
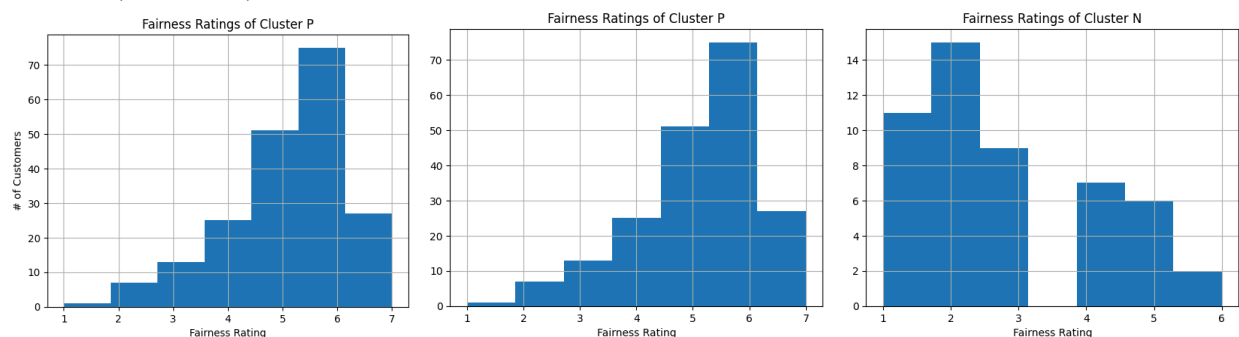

Above is the fairness rating of the full data and each cluster, and below is the intent to revisit ratings of the same, reflecting Cluster P's positivity and Cluster N's negativity.

When looking at Cluster N's defining features, we found it consisted largely of non-students, unlike the rest of our sample and Cluster P, and also consisted of largely female customers. Along with this, they are also older and have higher income on average than the overall sample, however these are reasonably logical consequences of them being non-students so we will not visualize them below.
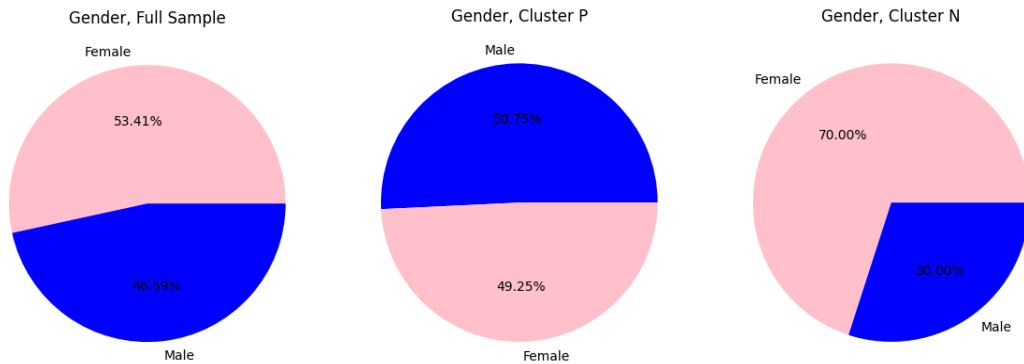
Above we have visualized gender, and below we have visualized employment using pie charts.
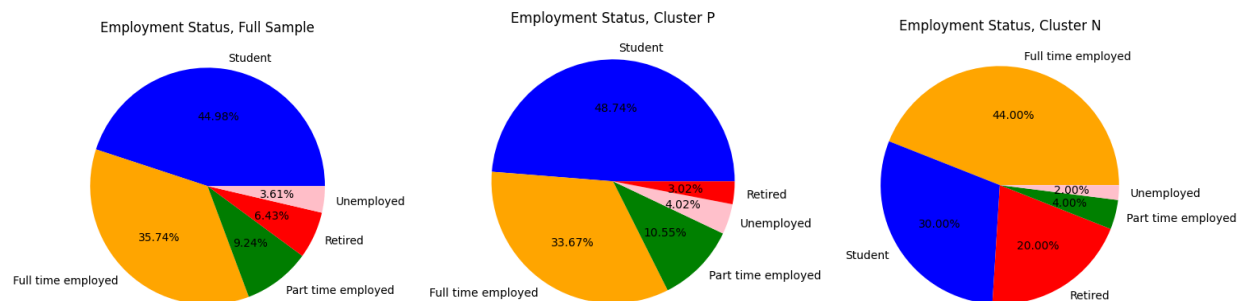
Potential Plans Of Action and Recommendations

This cluster of dissatisfied, non-student women and their potential to leave the nightclub's customer base leads us to recommend two potential plans of action.

One is to completely ignore this dissatisfied group and maintain current pricing strategies, while looking to appeal more to students and the younger demographic. This could be through student discounts, decor, music selection, loosening dress code, etc. The goal of this plan is to expand the total number of people coming into the club, especially due to the higher rates of word-of-mouth advertising indicated by Cluster P. However, implementing this strategy should only be considered after considering both whether the nightclub may run the risk of hitting capacity too often with this strategy, and whether the nightclub's profits come mainly from entry fees or from purchases made while in the club. If the club is already consistently close to or hitting capacity, then increasing demand this way may not lead to increased profits, and if the students we're targeting do not purchase enough once in the club to make them reasonable to target, then this strategy of increased demand may not be the correct choice.

The other is to shift our strategy towards this smaller population of higher income, working women. This may be possible through selling long-term memberships to our club, adding more high-end catering to the menu, potentially decreasing capacity to help them feel more comfortable, and through keeping pricing consistent. Once again, this strategy should only be implemented after considering the club's current capacity and revenue streams.