TravelTide

Data Segmentation

Data segmentation set up

- TravelTide, a rising e-booking startup, has been gaining traction in the online travel industry since its launch in April 2021.
- However, TravelTide has an intense focus on inventory and searchability and this has led to weaker customer experiences and lower retention rates. They're now committed to resolving this with a robust marketing strategy.

- The goal is to create an amazing personalized rewards program that encourages customers to keep using TravelTide.
- The marketing department has identified perks that are most likely to attract customers as part of the reward program. The identified perks are the following

✓ Free hotel meal

✓ Free checked bags

✓ No cancellation fees

✓ Exclusive discounts

✓ 1 Night free hotel with flight

GOALS

- First, to check if the data backs up the idea that certain customers are extra interested in the perks as the marketing team suggested. After that, we'll figure out which perk is most likely to be a customer’s favorite.

General Summaries of the data

We have a large dataset that includes 4 tables with 1,020,926 users and over 5.4 million sessions ranging from April 2021 to July 2023.

But First...

- New customers have limited interaction data, and this can skew our analysis. Our goal is to create a rewards program based on customer behavior, but it takes time for behavioral data to build up.

So the cohort of the study is

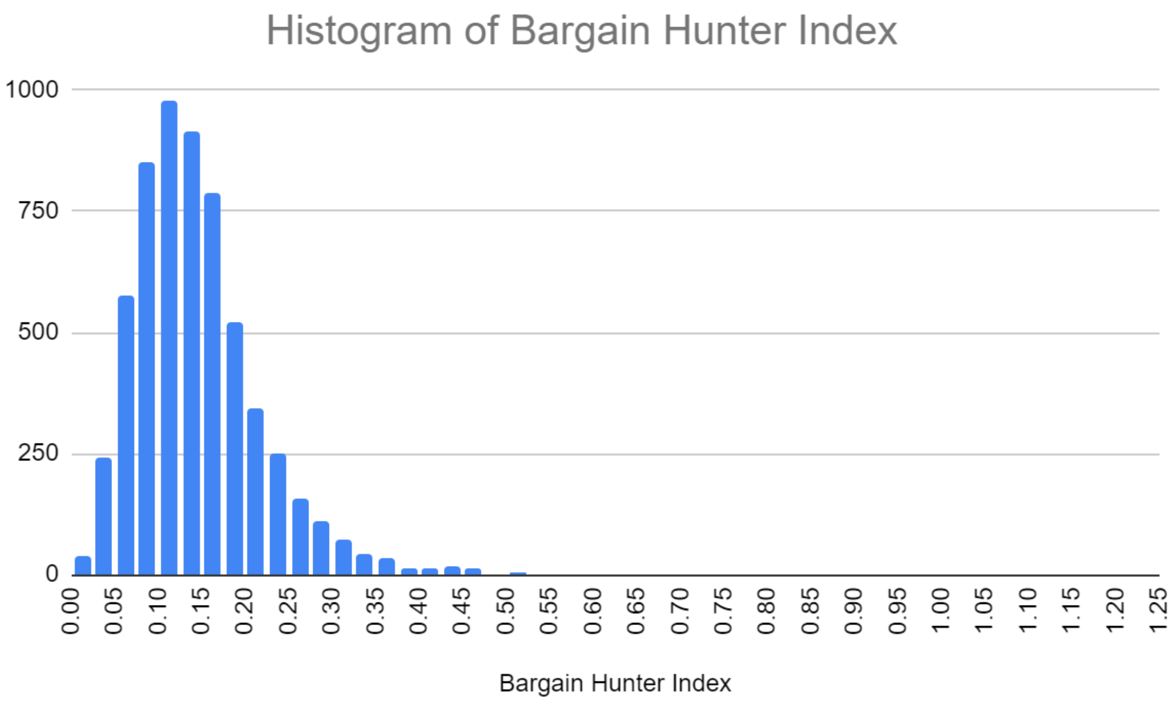
- After discussion with the marketing team, it was decided to only include sessions starting after the New Year's holiday (2023-01-04) until the last available date in the database and to only include users with more than 7 sessions during the same period.

Calculating Metrics

- Based on information from a quick online search on several flight booking websites we have come up with several behavioral metrics for every perk in the reward program to determine the customer's behavior and their interest towards the different perks.
- Each perk with its respective metrics is below.

Exclusive discounts perk

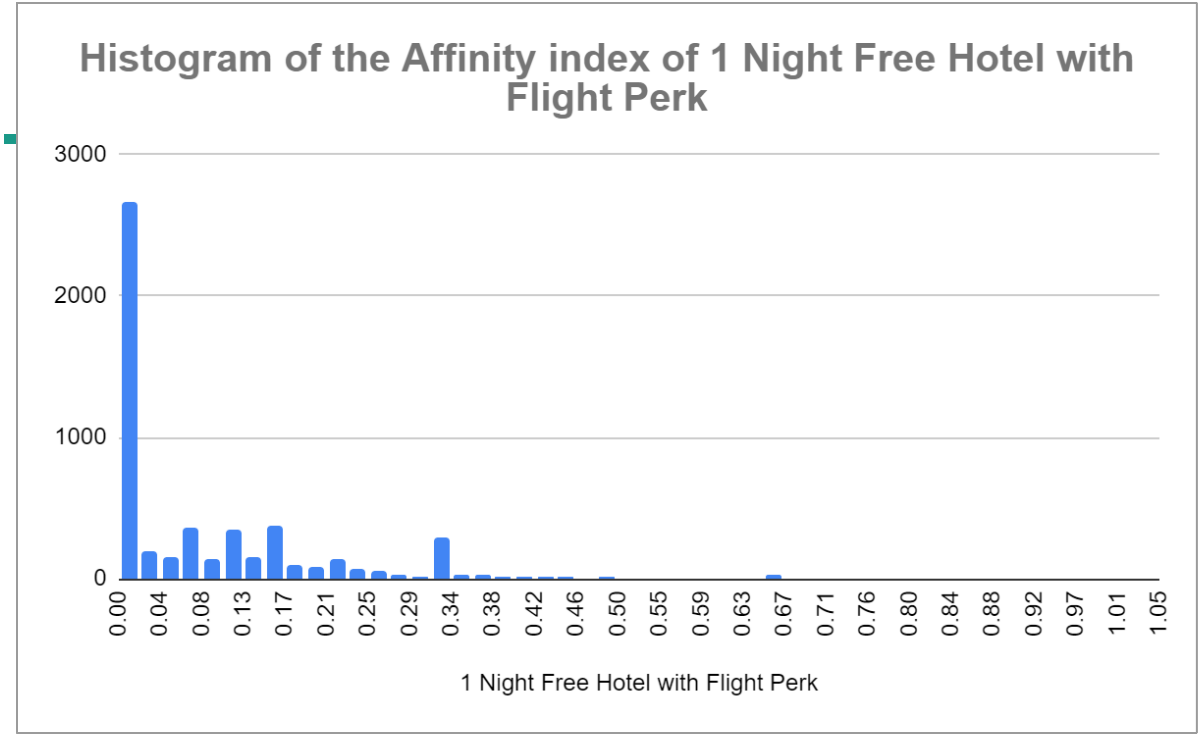
- Identified metrics that are very likely to show customers towards this perk are
  - ❑ The percentage of flight bookings under discount
  - ❑ The average flight discount amount in percentage terms
  - ❑ The scaled Average Dollars Saved per kilometer
  - ❑ Average Browsing duration per session
  - ❑ Average number of clicks per session



Distribution of users on Exclusive discounts perk affinity index (bargain hunter index)

1 Night free hotel with flight

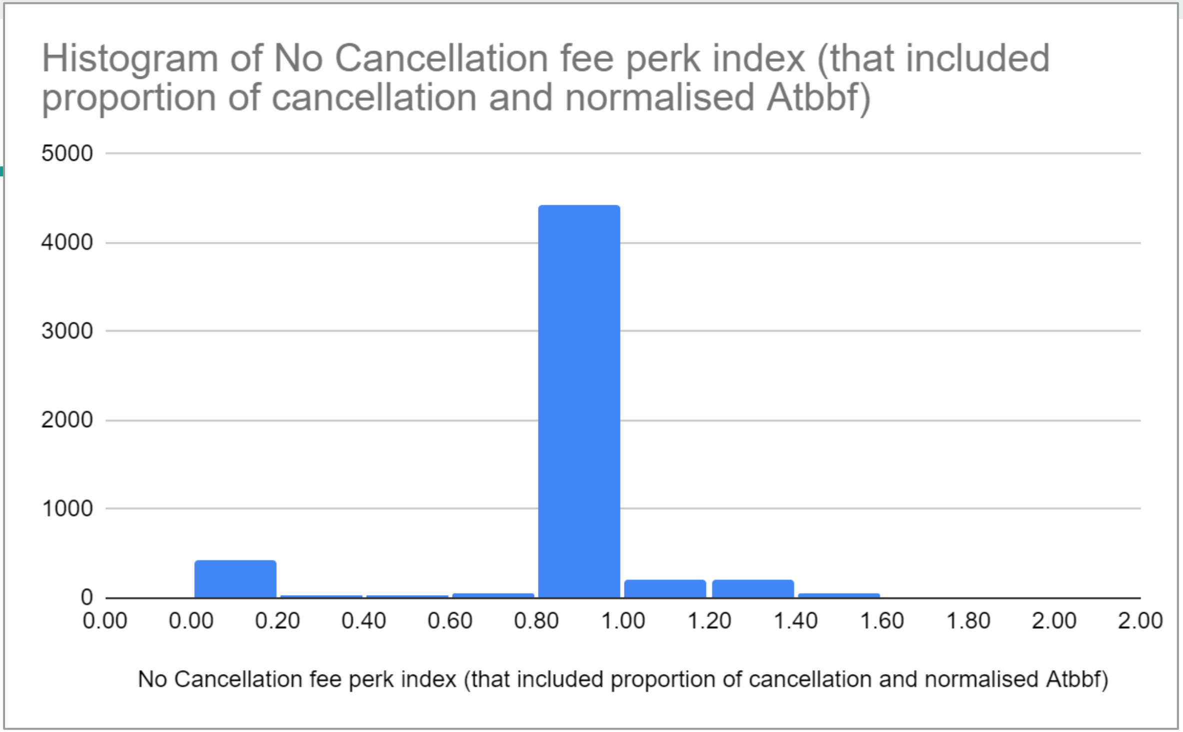
- Identified metrics that are very likely to show customers towards this perk are
- ❑ The percentage of hotel bookings under discount
- ❑ The average hotel discount amount in percentage terms
- ❑ The scaled Average Dollars Saved per booked room
- ❑ Average number of rooms booked per flight



Distribution of users on 1 night free hotel perk index score

No cancellation fees

- Identified metrics that are very likely to show customers towards this perk are
- ❑ The proportion of previous cancellation
- ❑ Customers who book flights last minute who often have unpredictable schedules

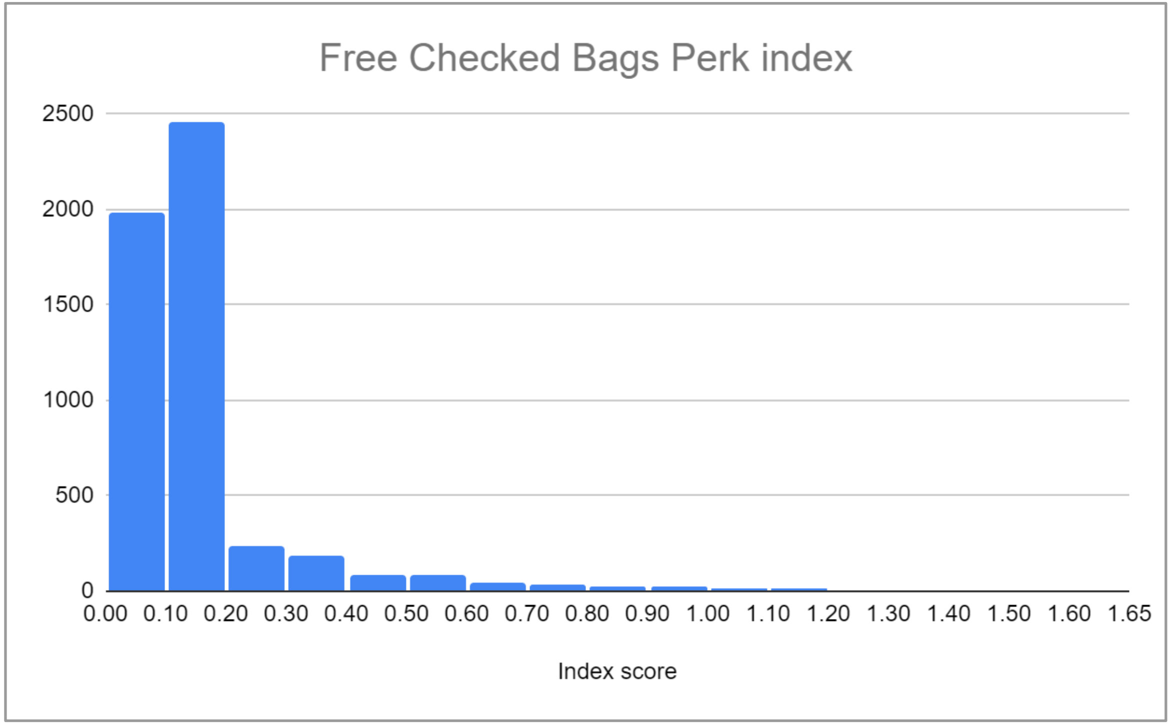


- Distribution of users on No cancellation fee perk index score

Free checked bags

- Identified metrics that are very likely to show customers towards this perk are
  - The Average_number_bags_checked
  - The Average distance travelled
  - Customers who have children
  - Length of stay (Customers with longer travel plans)

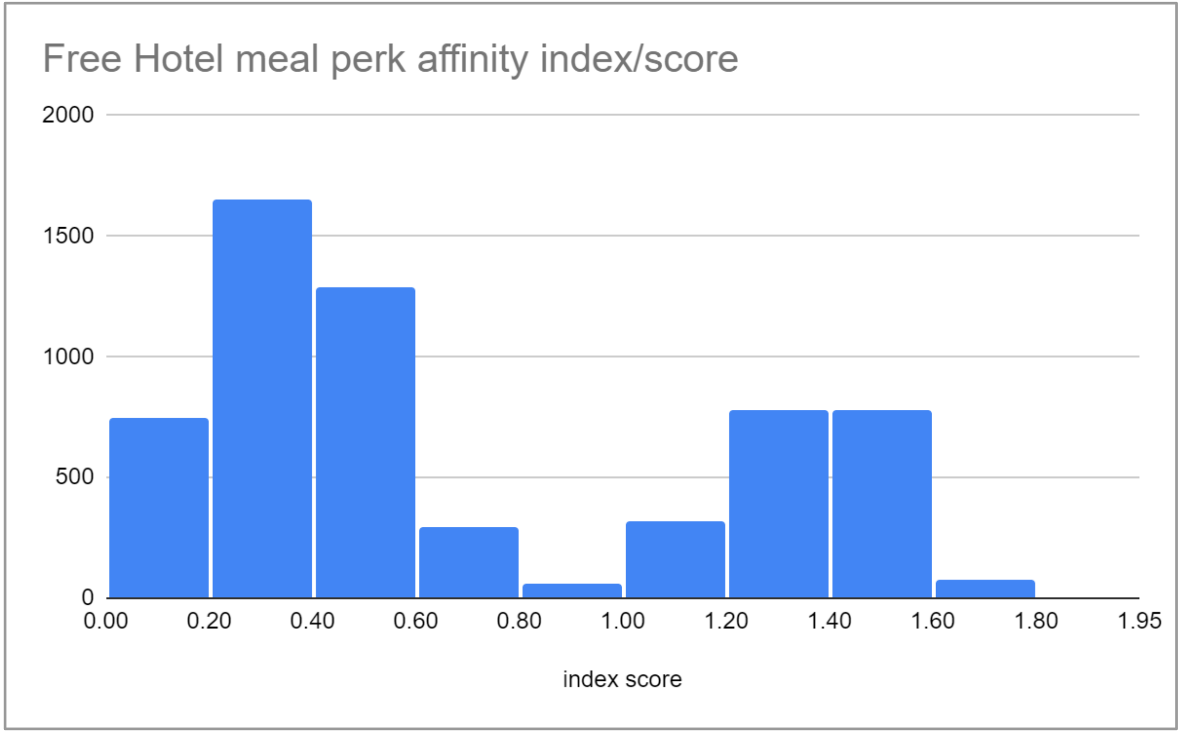
- All those customers are very likely to have more and larger baggage when traveling.



- Distribution of users on free checked bags perk index score

Free hotel meal

- Identified metrics that are very likely to show customers towards this perk are
- ❑ Age – older adults tend to opt to free hotel meal reward programs when booking flight
- ❑ Families with children
- ❑ Price sensitive customers as well (The average flight discount amount in percentage terms was used as a measure to determine this)



- Distribution of users on free checked bags perk index score

Next step?

- All the metrics for each of the perks were checked for any correlation. For those with no correlation with one another, they were scaled to avoid bias and give metrics equal weight when combining them to form the affinity scores.

Affinity score/index

- For all the 5 perks an affinity score was calculated for each of the users.
- **Free Hotel meal index based on** Scaled age in years and has children/not + average discount on booked flights in percentage
- **Free Checked Bags perk index based on** length of stay(scaled) and Average Number of Bags Checked

Affinity score/index

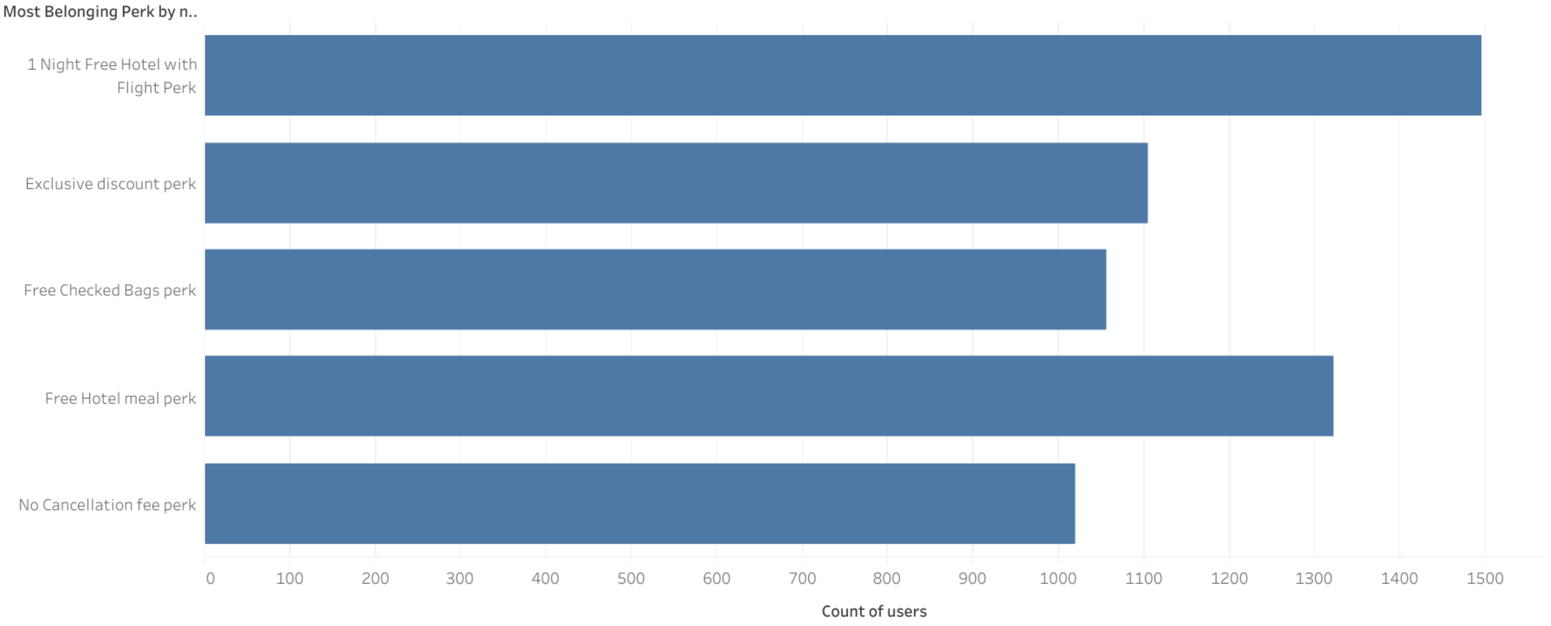
- **No Cancellation fee perk index based on -** the proportion of cancellations and Average time between booking and flight

- **1 Night Free Hotel with Flight Perk index based on -** Scaled average number of rooms booked per flight (scaled_ANRB) and scaled ADS per room booked
- **Exclusive discount perk** (Bargain index) **based on** Scaled ADS per Kilometer and Scaled browsing duration per session

Fuzzy segmentation

- Regarding segmenting customers based on the marketing team's perks for the rewards program, it's essential to assign each customer to one specific perk.
- Then the customers were ranked based on the affinity index score of every perk separately. And customers with the highest score in each perk were given the lowest rank. This gave every customer a ranking in every affinity index for all the 5 perks.
- Then the perk corresponding to this minimum rank is the one the customers most strongly relate to and every user was assigned to this respective perk.

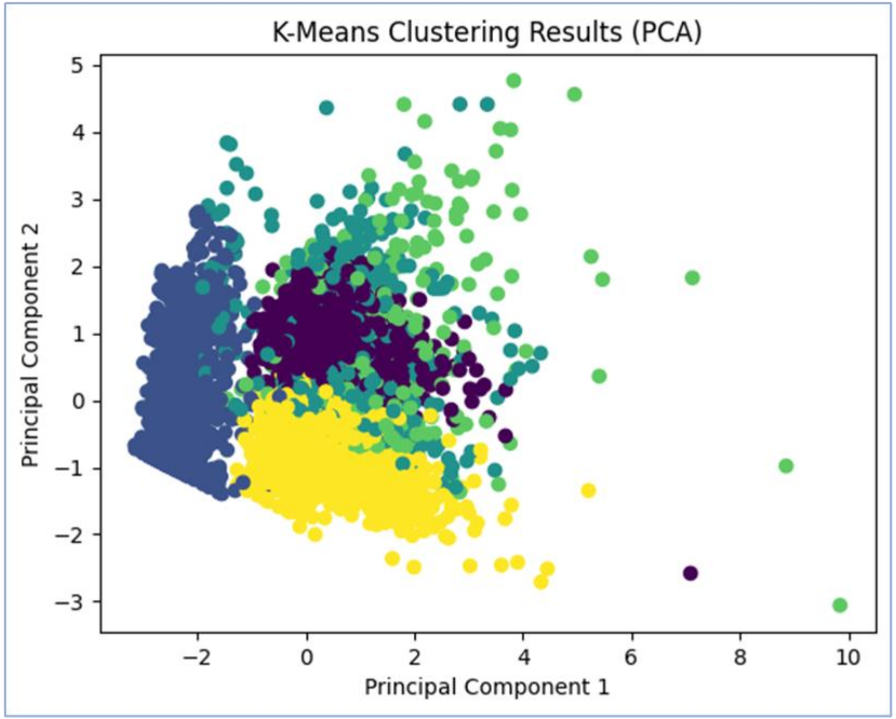
Distribution of users using Fuzzy segmentation



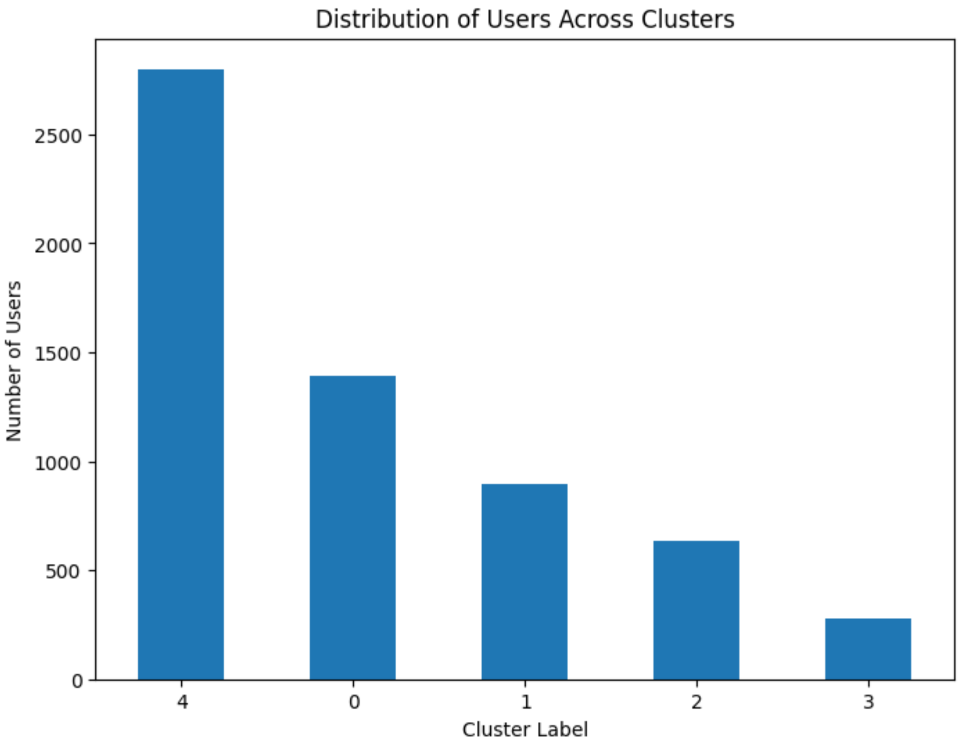
- Distribution of Fuzzy segmentation

Kmeans Clustering

The evaluation of this K-means clustering indicated
that the clusters have some degree of separation and cohesion but may not be perfectly distinct. It indicates a reasonable quality of clustering, with room for improvement.



Distribution of K-means clustering



Both clustering methods were compared

- Several values were used to compare the similarity or agreement between two different clusterings. In this project, we are comparing two clustering methods: one based on the affinity score index and the other based on K-Means.

- The comparison however showed a relatively low level of agreement between the two clustering methods. Overall, the comparison scores indicate that there is some level of similarity between the affinity-based clustering and the K-Means clustering, but it's not a very strong agreement.

Conclusion

- All of the customers have been assigned to a specific group using the Fuzzy segmentation method. However the downside of fuzzy segmentation is that a user with less affinity towards a particular perk can be assigned to that perk merely because the minimum rank they got across all of the 5 perks lay on that perk. This can have an impact on the effectiveness of the reward program. Moreover, Small changes in the ranking can result in different segment assignments.

- The second method of clustering the k-means clustering indicates that the clusters have some degree of separation and cohesion but may not be perfectly distinct. It indicates a reasonable quality of clustering, with room for improvement.
- When both clustering methods were compared there was some level of similarity between the affinity-based clustering and the K-Means clustering, but it's not a very strong agreement.

Recommendation

- Therefore, the customer's assignment to different perks using the fizzy segmentation is provided as a file. The clustering using the K-means clustering is not included as it does not add significant value to the segmentation done using the fizzy segmentation. However further experimentation and reiteration with different clustering algorithms, parameters, or preprocessing steps to see if you can achieve a stronger agreement between the two methods.
- Furthermore, it will be important to identify behavioral metrics from objective and reliable sources to determine customers' affinity toward the different perks.

END