Project Goal

- The goal of the project is to determine the most appealing rewards for individual customers based on their behavior and preferences, subsequently categorizing customers into groups according to their preferred perks
- This groups will be used in personalized rewards program that keeps customers returning to the TravelTide platform

Project Workflow

Data Collection: Gather customer data, including behavioral metrics and demographic information, using SQL queries. Filter the data to extract information only from the targeted cohort.

Index Calculation: Calculate various indexes, such as Hotel Hunter Index, Average Bags Index, Cancellation Rate, Bargain Hunter Index, Combined Booking Index, and Session Intensity Index, for each customer using SQL.

Ranking Customers: Rank customers based on their index values in descending order using Python.

Perk Assignment: Determine the most attractive perk for each customer based on their minimum rank in the indexes. Segment customers into groups based on their preferred perks using Python.

Preprocessing users data: checking missing values, converting data to appropriate formats, replacing incorrect values, additional necessary calculations using Python.

Demographic Analysis: Analyze customer segments by demographic characteristics, such as age, gender, marital status, and parental status, using Tableau.

KMeans Analysis: Conduct unsupervised customer segmentation using the KMeans approach. Compare results with manual segmentation and draw conclusions about the more effective approach.
Visualization: Create visualizations in Tableau to present the findings and insights from the analysis.

Conclusions and Recommendations: Summarize the project's goals, findings, and provide recommendations based on customer segmentation. Create slides and a video presentation in canves.com

Approach to Customer Segmentation and Indexes Calculations

Customer Segmentation

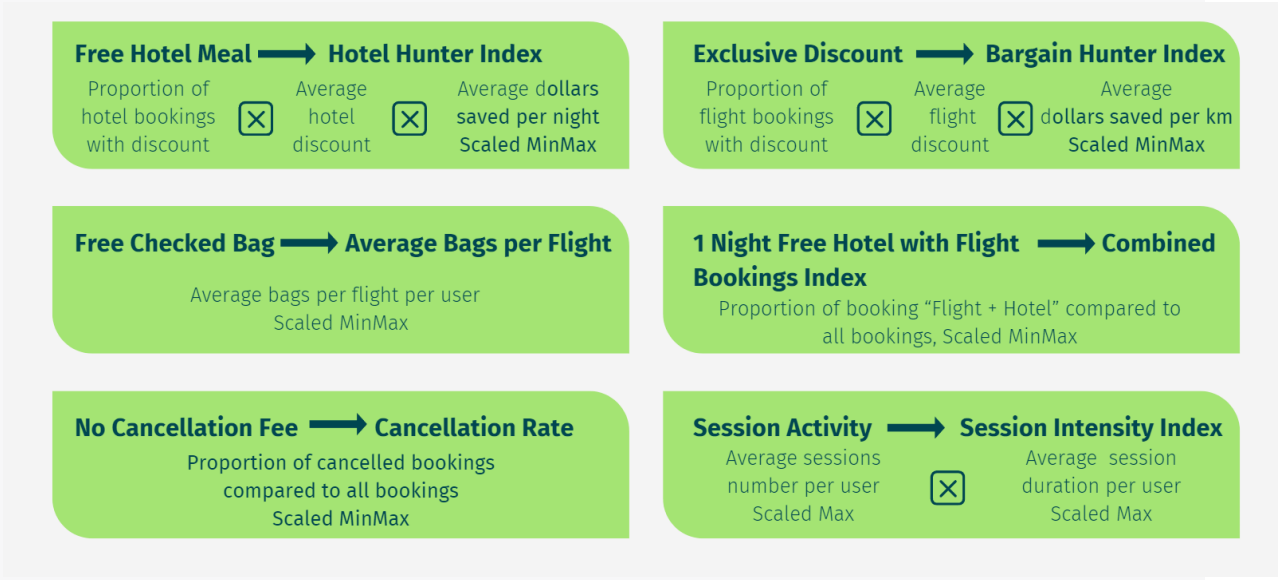
To categorize customers for each perk, it's crucial to determine the specific characteristics that customers targeted for each perk should possess and identify the relevant indicators for customer division.

- **Free Hotel Meal**
  Target audience: Customers who actively seek and use hotel discounts.
  Metrics: Proportion of booking with hotel discounts, average saving per night, average discount on hotel bookings
- **Free Checked Bags**
  Target audience: Customers who typically book flights with multiple checked bags.
  Metrics: The average number of checked bags per flight
- **No Cancellation Fee**
  Target audience: Mainly customers who frequently cancel trips.

Metrics: Percentage of booked flights or hotels that were cancelled

- **Exclusive Discount**
  Target audience: Customers who actively seek and use flight discounts
  Metrics: Proportion of booking with flight discounts, average saving per km, average percentage discount
- **Free Hotel Night with Flight**
  Target audience: Customers who often book both flights and hotels together
  Metrics: Proportion of booking combination of flight and hotel compared to all bookings

Indexes Calculation



Here is more detailed information about the index formulas used for calculations. To facilitate comparisons based on these metrics, we will apply scaling techniques. Scaling allows us to standardize the values across different metrics, ensuring a fair and accurate comparison.

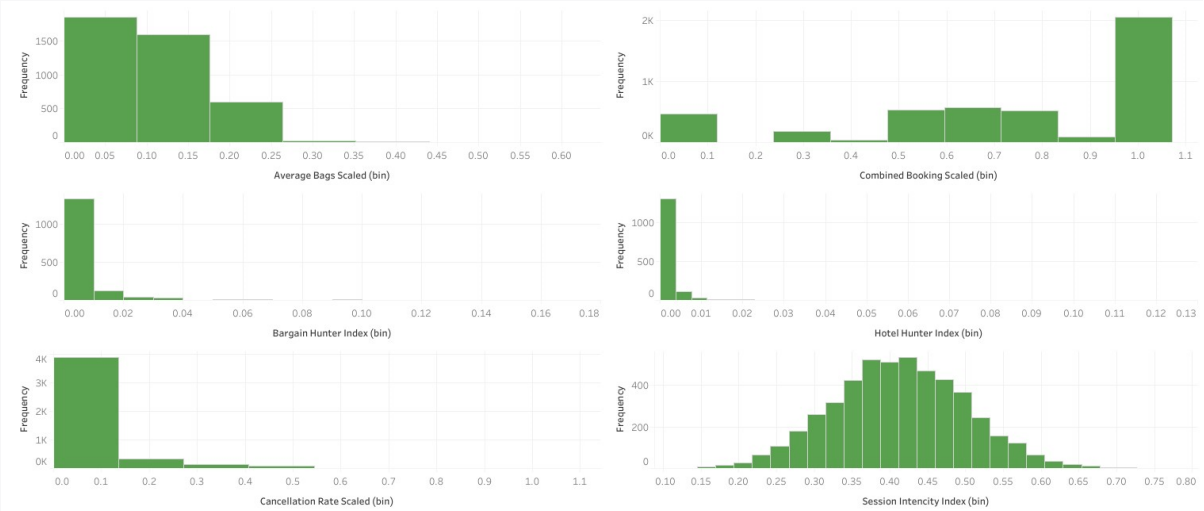
Approach to Customer Classification

In our approach to customer segmentation, we followed a systematic process:

1. **Ranking Customers:** ranking all customers for each perk based on their respective index values in descending order.
2. **Determining Minimum Rank:** identifying the minimum rank for each customer and assigning them the label associated with the perk linked to this minimum rank.
3. **Creating Rank Determination Function:** To facilitate the ranking process, we created a custom function to determine the rank based on the minimum rank encountered for each customer.
4. **Handling Multiple Minimum Ranks:** In cases where customers exhibited multiple minimum ranks, we assigned them the label corresponding to the first minimum rank encountered during the analysis.
5. **Active Searchers Segment:** Customers who had displayed no booking activity and possessed all NaN values for indexes were categorized as "Active Searchers." We retained these customers within this segment for an initial period to better comprehend their preferences before making specific perk recommendations. Additionally, we included customers in this group whose minimum rank was determined by their Session Activity, ensuring that we had customers with non-null index values. This inclusion contributed to a more comprehensive understanding of the preferences within this segment.

Indexes Calculation Results

Main Indexes Distribution



As observed in the histograms, the Average Bags Index, Hotel Hunter Index, Bargain Hunter Index, and Cancellation Rate exhibit right-skewed distributions, which suggests that most customers have lower levels of engagement in activities such as checking bags, hunting for hotel deals, seeking discounts, or experiencing cancellations.

In contrast, the Combined Booking Index tends to display a left-skewed distribution that indicates that a lot of customers make booking hotel and flight together. The Session Intensity Index closely approximates a normal distribution.

6.

Customers' Perks Analysis

In the treemaps for each index, we can observe that customers with matching perks exhibit the highest index indicators. This reinforces the validity of our segmentation approach.

Among customers with high session activity, a significant portion falls into the "Active Searcher" category. Interestingly, many of them also display non-null index values, particularly in the Hotel Hunter Index and Bargain Hunter Index. This suggests that offering them one of these perks could be a promising strategy.

Session Activity Analysis

There is not a significant variance in the number of sessions per user. Since we included only users with more than 7 sessions starting from 2023-01-04, all user groups have a median of 8 sessions. However, users labeled as "Active Searcher" exhibit a higher upper whisker in the boxplot.

After labeling users as "Active Searcher," the highest average and median session durations are observed among users labeled with "Free Hotel Meal" and "Exclusive Discount" perks, followed by “Free Checked Bag” and “Exclusive Discount” perks users.

Analysis by Demographic Characteristics

Age groups

Age Groups

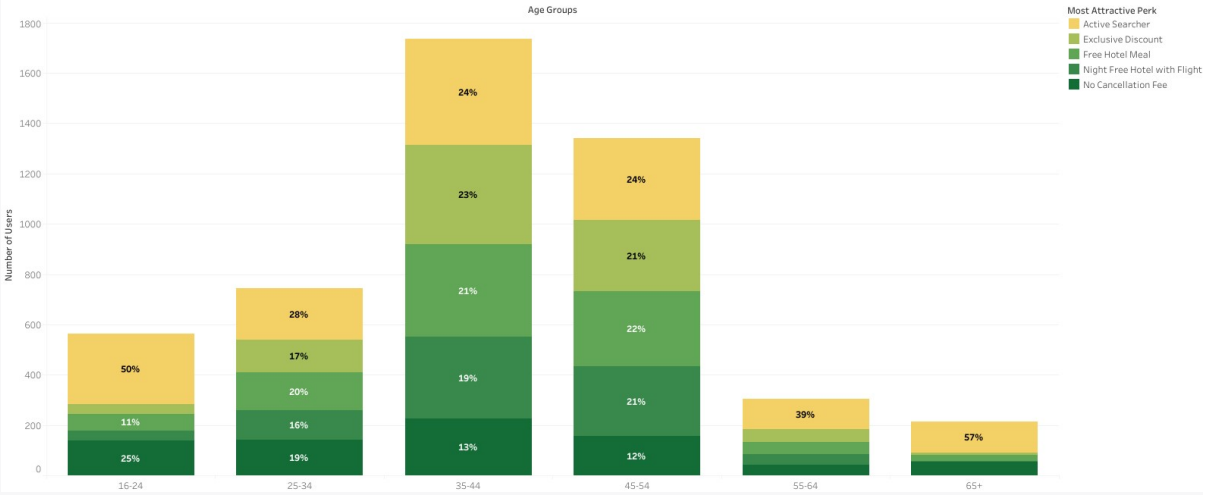
Most Attractive Perk	16-24	25-34	35-44	45-54	55-64	65+
Active Searcher	18.9%	13.9%	28.6%	22.1%	8.1%	8.3%
Exclusive Discount	4.4%	14.1%	43.4%	31.3%	5.6%	1.1%
Free Hotel Meal	6.7%	15.9%	38.5%	31.2%	5.2%	2.4%
Night Free Hotel with Flig..	5.0%	14.5%	40.3%	34.4%	5.3%	0.5%
No Cancellation Fee	18.3%	18.6%	29.9%	20.6%	5.4%	7.1%

Age Groups

% of Total Count of Dat..

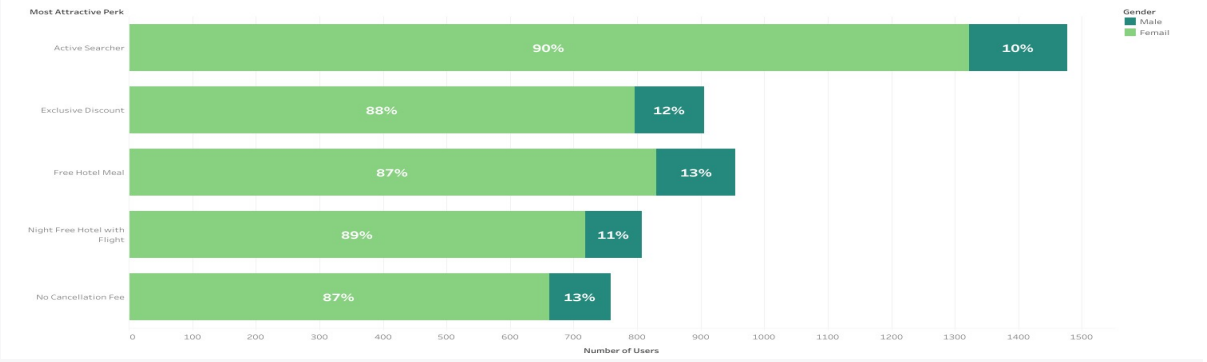
0.5% 43.4%

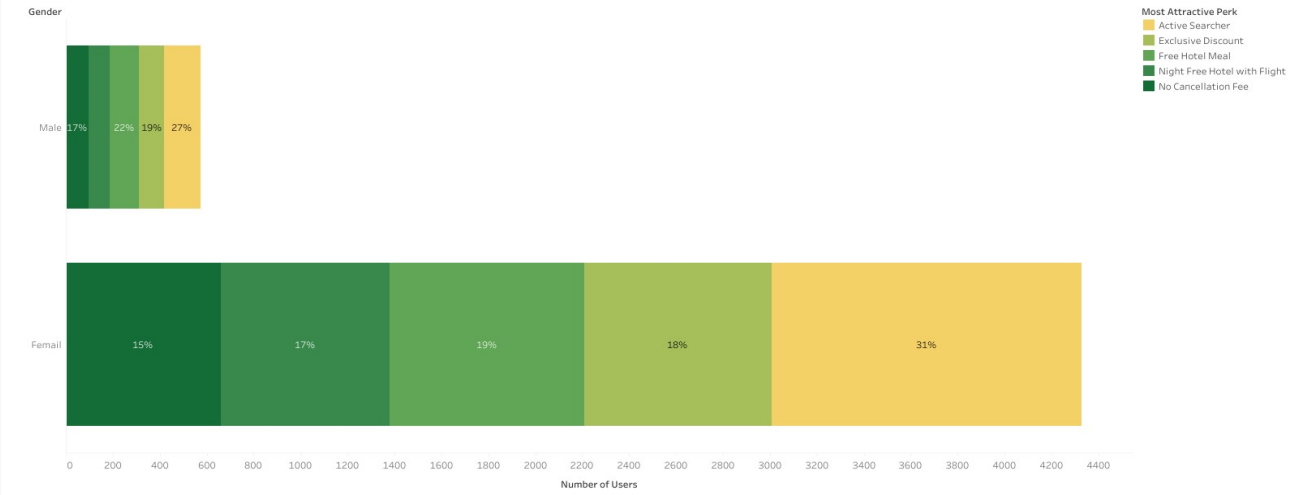
Bar Chart Age Groups



It's evident that the most active users fall within two age groups: 35-44 and 45-54. Additionally, when compared to other age groups, these two cohorts exhibit a lower prevalence of the "Active Searcher" label.
Within the 35-44 age group, there's a pronounced proclivity for seeking discounts on both flights and hotels. Conversely, the 45-54 age group demonstrates the highest level of engagement in hunting for hotel discounts, flight discounts, and combined booking opportunities.

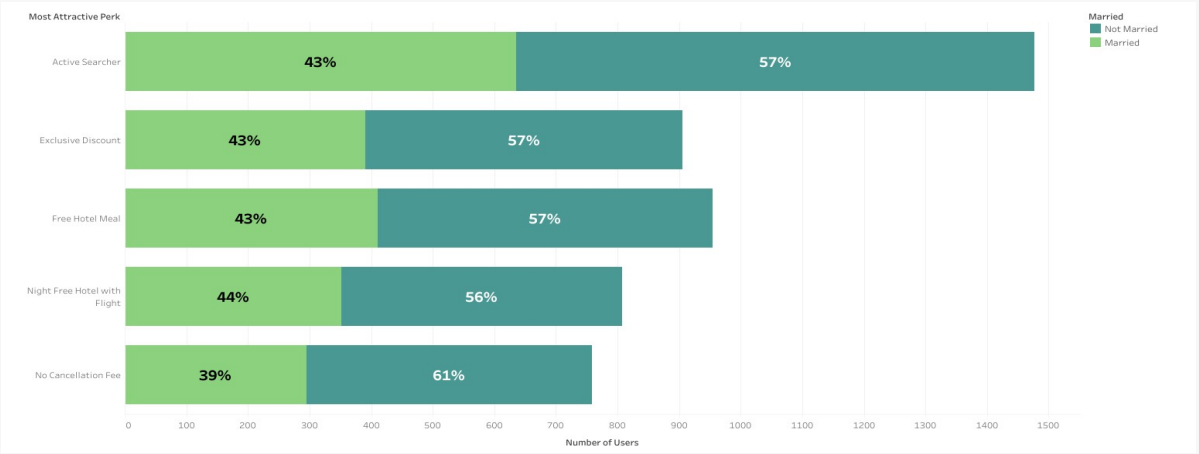
Gender





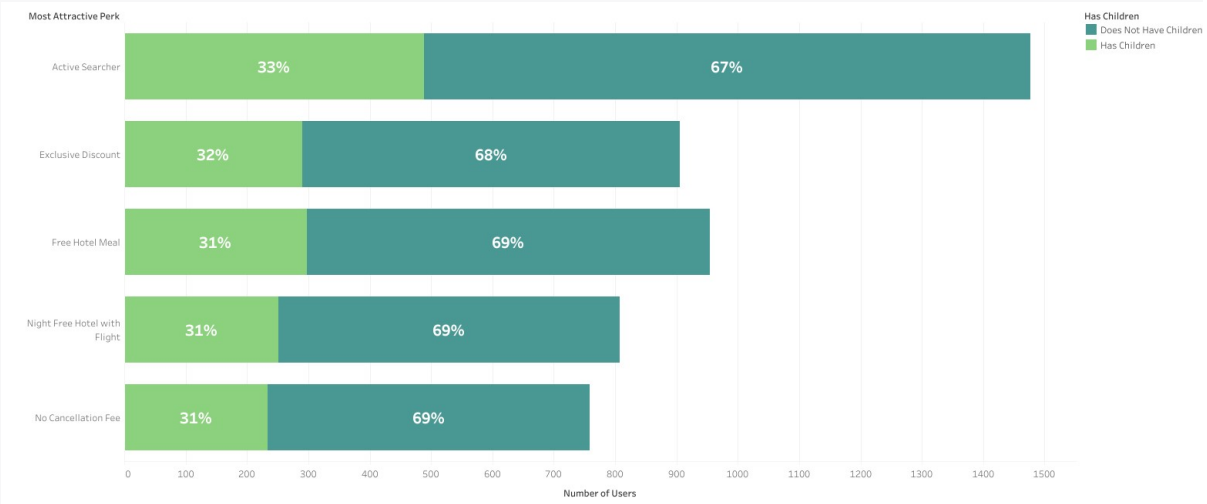
Our primary audience comprises females, accounting for a significant 88% of all users
Among male users, there is a relatively higher rate of booking cancellations and an increased interest in hotel discounts. Conversely, female users display a greater inclination towards combined bookings and a higher proportion of "Active Searcher" labels

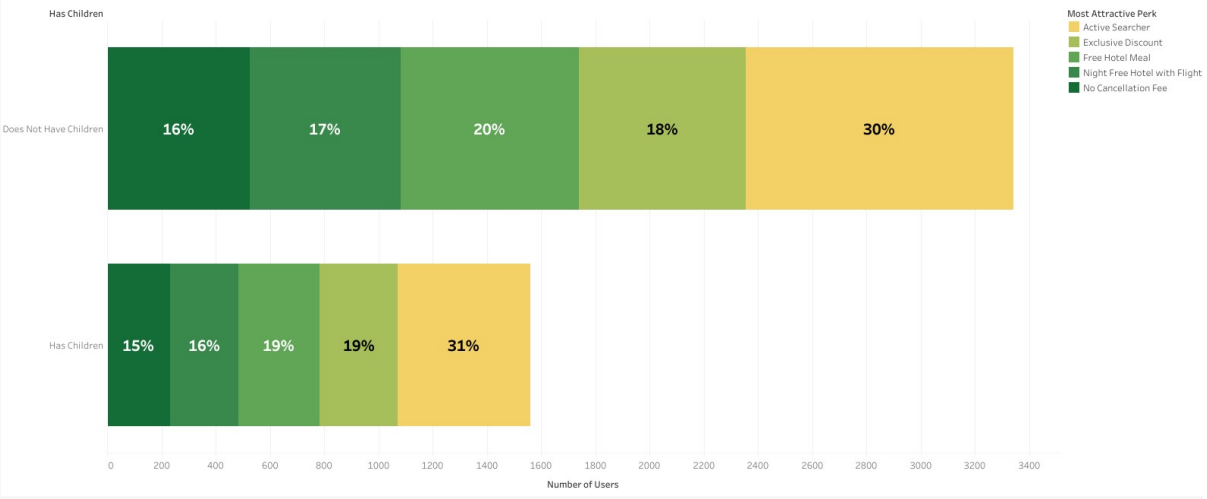
Married Status



Unmarried users exhibit higher levels of activity compared to their married counterparts.
Unmarried users display a greater preference for the "No Cancellation Fee" perk compared to married users, whereas the latter group shows a stronger interest in hotel and flight discounts

Has Children Status





Users without children exhibit higher activity levels compared to users with children
Users with children tend to show relatively more interest in the "No Cancellation Fee" perk

KMeans Analysis

In addition to calculating indices and assigning groups based on the best results from one of the indices in comparison to others, client segmentation was also conducted using unsupervised learning through K-Means analysis.

We've gotten 5 clusters:

Cluster 0: This cluster does not have a leading position in any of the indexes. They perform well in Cancellation Rate (2nd place) and Bargain Hunter Index (3rd place), indicating an interest in perks such as 'No Cancellation Fees' or 'Exclusive Discounts.'

Cluster 1: Cluster 1 holds leading positions in multiple indexes, including Average Bags, Cancellation Rate, and Bargain Hunter Index. This suggests their interest lies in both flight discounts and accompanying services.

Cluster 2: Cluster 2 excels in the Hotel Hunter Index and ranks 2nd in Bargain Hunter Index, indicating a strong interest in hotel and flight discounts.

Cluster 3: This cluster ranks last in most indexes but holds the leading position in Session Activity, showing a keen interest in hunting for discounts.

Cluster 4: Cluster 5 leads in the Combined Booking Index, indicating a preference for perks related to combined bookings, such as '1 Free Night Hotel with Flight.'

In summary, we can label our clusters based on their interests in the following perks:

Cluster 0: 'No Cancellation Fees' Cluster 1: 'Free Checked Bag' Cluster 2: 'Free Hotel Meal' Cluster 3: 'Exclusive Discount' Cluster 4: '1 Free Night Hotel with Flight'

In our specific case, calculating indexes and labeling customers based on index ratings can be regarded as a more effective and precise approach. The results obtained through the first approach exhibit greater homogeneity in terms of demographic characteristics within groups. This approach has proven to be more effective in addressing customers who have not made any bookings yet. However, K-Means analysis should undoubtedly be employed in future work to uncover any hidden patterns and insights in customer segmentation.

Conclusions

Customer Segmentation: We successfully segmented our customers into distinct groups based on their preferences for different perks, including Free Hotel Meal, Free Checked Bags, No Cancellation Fee, Exclusive Discount, and Free Hotel Night with Flight. This segmentation will allow us to target each group with personalized rewards.

Index Metrics: We calculated and analyzed various index metrics, such as Average Bags, Discount Flight Proportion, Discount Hotel Proportion, Booking Cancellation, Combined Booking, Session Number, and Session Duration. These metrics helped us understand customer behavior and identify those most likely to be interested in specific perks.

Active Searches Group: Customers labeled as 'Active Searchers' display a strong inclination towards hotel and flight discounts. Therefore, it would be advisable to offer them perks related to these categories. However, it's essential to consider their demographic characteristics for more targeted recommendations. For instance, users in the age group 35-44 appear to be more interested in flight discounts, while users aged 45-54 show a greater preference for hotel discounts. As a result, offering 'Free Hotel Meal' may be a more attractive option for the latter group.

Session Duration: After labeling users as "Active Searcher," we observed the highest average and median session durations among users with "Free Hotel Meal" and "1 Free Night Hotel with Flight" perks. At the same time, some of them also display non-null index values, particularly in the Hotel Hunter Index and Bargain Hunter Index.

Demographic Analysis: We explored the influence of demographic characteristics, including age, gender, marital status, and parental status, on customer preferences. This analysis highlighted distinct patterns in perk preferences among different demographic groups.

KMeans Analysis: We conducted unsupervised customer segmentation using the KMeans approach and compared the results with our manual segmentation. This allowed us to assess the effectiveness of different segmentation methods and finalize our customer groups.

Recommendations

A/B Testing: Conduct A/B testing of marketing strategies to evaluate the effectiveness of different approaches in driving customer engagement and retention within each segment.

Continuous Monitoring: Regularly monitor customer behavior and adjust the rewards program based on evolving preferences. Be agile in adapting to changing market conditions and customer needs.

Further Data Analysis: Continue to explore data patterns and hidden insights using advanced analytics techniques. Incorporate K-Means clustering or other machine learning algorithms to uncover additional segmentation nuances and refine targeting strategies.

Engage "Active Searcher" Users: Develop and implement a system of key demographic and session behavior characteristics that can be used to label by their most attractive perk even if they haven't made any bookings yet or don't have a sufficient number of sessions. This proactive approach will enable us to identify and engage potential "Active Searcher" customers early in their journey on our platform

Feedback Mechanism: Implement a feedback mechanism within the rewards program to gather insights directly from customers. Use this feedback to make real-time adjustments and continuously improve the program.

***Slides can be accessed via the following link:
https://www.canva.com/design/DAFtYDpuErE/Y0ZKZGRWpbUigcby1MRhdw/edit?utm_content=DAFtYDpuErE&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

***Video presentation:
https://www.canva.com/design/DAFtYDpuErE/TiREg8ncRvrsOX7qcu45zA/view?utm_content=DAFtYDpuErE&utm_campaign=designshare&utm_medium=link&utm_source=editor

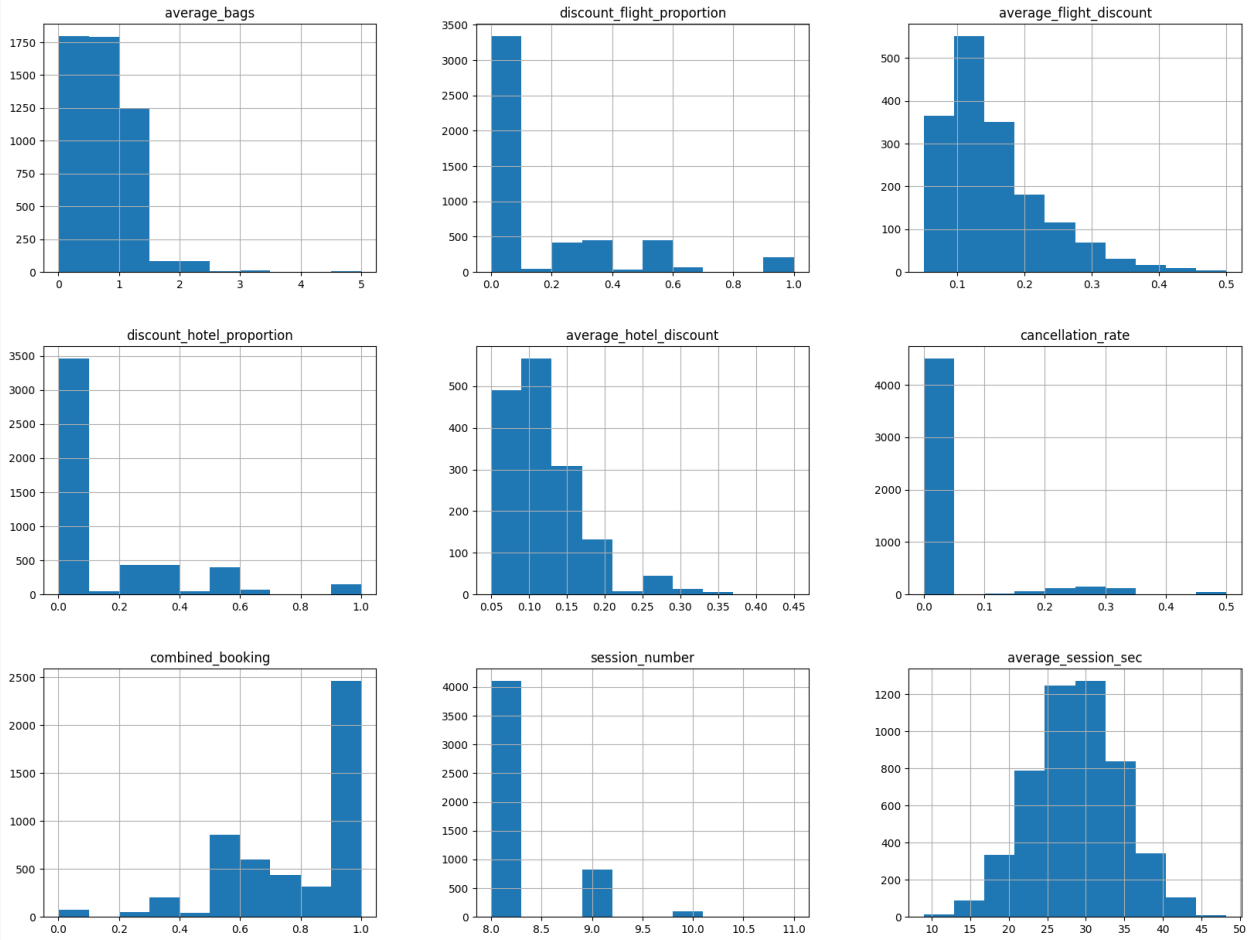
***Interactive Dashboards can be accessed via the following link:
https://public.tableau.com/app/profile/alexandra.meshi/viz/Metrocar_16888430060580/Story1?publish=yes

In the interactive dashboards, you can explore the detailed analysis and visualizations of our findings. Dive deeper into the data, interact with the charts and graphs, and uncover valuable insights to drive informed decision-making and optimize Metrocar's performance.

APPENDIX

Exploratory data analysis

Before starting indexes calculation we will explore intermediate results and check them for outliers:



Average Bags: The majority of customers have one or fewer average bags per flight. The maximum amount is 5 bags, which could be considered an outlier. However, in this case, having 5 average bags per flight is a relatively high amount, but it doesn't appear unrealistic. Since we are searching for customers who are most interested in the 'Free Checked Bag' perk, it's reasonable to retain the data without trimming the maximum value.

Discount Flight Proportion & Average Flight Discount: The majority of our customers either didn't use flight discounts or book flights. Among those who did, some have a notably high proportion of flights with discounts (60% or more) and an average discount of about 50%. However, this only indicates their preference for booking flights with discounts.

Discount Hotel Proportion & Average Hotel Discount: The majority of our customers either didn't use hotel discounts or book hotels. Among those who did, some have a relatively high proportion of hotel bookings with discounts (60% or more) and an average discount of about 45%. Again, this primarily suggests their inclination to book hotels with discounts.

Booking Cancellation: The vast majority of our clients either didn't cancel their bookings or didn't have any bookings at all. Therefore, all customers who did cancel could be considered outliers. However, in our case, we are interested in these outliers, so we will retain the data without making changes.

Combined Booking: This indicator exhibits significant fluctuations. Customers either don't book flights and hotels together at all or always combine flights with hotels.

Session Number & Session Duration: Since we filter only customers with more than 7 sessions, the most common session number is 8, with a maximum of 11, which is within a reasonable range and doesn't appear as an outlier. The mean and median session durations are quite close (28.7 and 28.75), and the distribution is approximately normal

In essence, these outliers offer valuable opportunities for TravelTide to design a rewards program that caters to a diverse range of customer behaviors and preferences. Instead of discarding outlier data, we should leverage it to develop targeted perks that resonate with specific customer segments, ultimately enhancing customer satisfaction and loyalty.