Nicholas Jones: TravelTide Segment Analysis
08/24/2023

Introduction/Background

The main idea behind this segmentation analysis was to find an effective way to increase customer retention. An effective framework for this analysis was a perk rewards program; in particular, we chose the following 5 perks that all customers might fall into: Free Hotel Meal, Free Checked Bag, No Cancellation Fees, Exclusive Discounts, and 1 Free Night at Hotel with Flight.

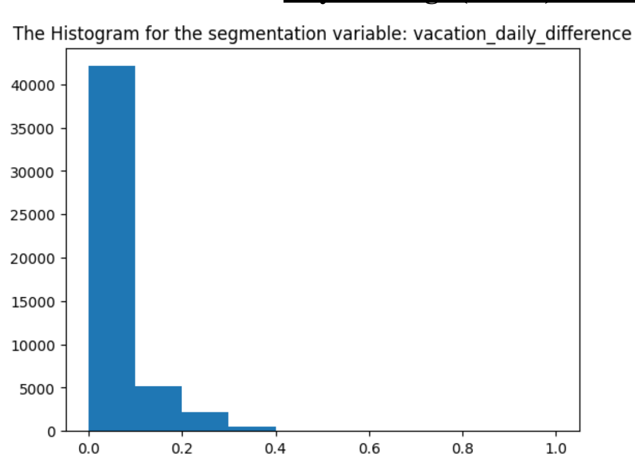
Objective/Summary

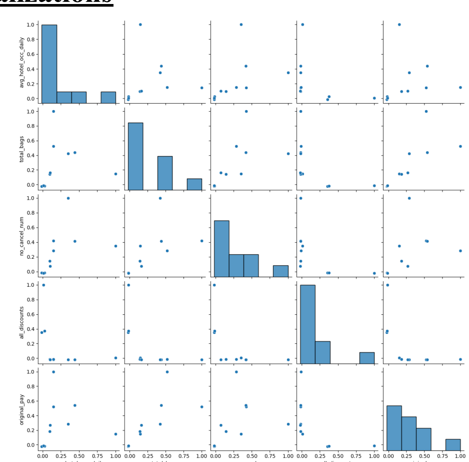
Based on a certain number of customer segments, the main objective is to find the characteristics of the customers that would find the perks very relevant and helpful; so much so, customers will become continuous customers and sign up for the rewards program.

Methodology

The following metrics are the ones that I chose for this analysis: AVG_HOTEL_OCC_DAILY (average amount of days that buyers occupied a hotel room), TOTAL_BAGS (total sum of checked bags), NO_CANCEL_NUM (the sum of the number of book seats on a flight per customer and the number of hotel rooms booked per customer), ALL_DISCOUNTS (the average of the sum of the discount flight rates and the discount hotel rates divided by the total number of hotel and flight bookings), and ORIGINAL_PAY (sum of the base fare prices). As far as the segmentation, I chose to group customers by using the mean and the margin error of the average number of vacation days per customer (VACATION_DAILY_DIFFERENCE): I decided to create 5 equal segments of customers by subdividing the range of the mean +/- 3 margins of error (namely, the range of 0.045237931034482766 days +/- (3)*0.07106363087665117 days). Furthermore,, the segments of the data are denoted as “G1” (the group that is closest to the mean), “G2” (the group that is the second closest group to the mean), and likewise for “G3”, “G4”, “G5”- there is an “Outliers” group as well. Lastly, I tried to confirm my findings with a cluster model as well. The SQL code and Python code can be found in Appendix #1, and there is a Top Perk Dashboard that can visualize different segmentations of the data in Appendix #2.

Key Findings (Part 1): Visualizations





Key Findings (Part 2): Regular Model

- For the Free Hotel Meal perk, I decided to look at the AVG_HOTEL_OCC_DAILY metric: the customers that had vacation durations that are closer to the average should be focused on because both metrics are similarly distributed. Also there are 40,388 customers in the "G1" group, and all customers are in 5 groups;
- For the Free Checked Bags perk, I decided to look at the TOTAL_BAGS metric: Customers with vacation durations that are closer to the mean because both metrics are similarly distributed. Therefore, most of the customers remain in the 2nd lowest group ("G2"), and and all customers are in 2 groups;
- To analyze the No Cancellation Fees perk, I decided to look at the NO_CANCEL_NUM metric: The customers that have vacation durations that are closer to the mean should be considered for the associated perk because both metrics are similarly distributed. Also, there are 46,804 customers in the "G2" group, and all customers are in 3 groups;
- To analyze the Exclusive Discounts perk, I decided to look at a specific rate, ALL_DISCOUNTS metric: The customers with vacation durations that are closer to the mean should be the target group for this metric's associated perk because both metrics are similarly distributed. Lastly there is an extremely high number of people in the "G2" group (namely, 48,798), and all customers are in 3 groups;
- To analyze the 1 Free Night with Flight perk, I decided to look at the ORIGINAL_PAY metric: Customers with vacation durations that are closer to the mean should be considered because both metrics are similarly distributed. Also, there is a sizable number of customers (namely, 21,965) that are in the "G1" group, and every customer is in one of 5 groups;

Key Findings (Part 3): Cluster Model

- Although the Cluster model did not produce a substantial amount of new information, it did yield one important new insight: The number of perks (or clusters) could be reduced to 3 because after that, the amount of variation is meager. Since the K-means measure the variations of data, I interpret each of the three clusters as a margin of error.

Recommendation/Suggestions

- After careful consideration, for each perk, I would recommend that TravelTide choose customers that are very close to the mean of the VACATION_DAILY_DIFFERENCE variable; Since most of the customers are in "G1" and "G2," I have concluded that Elena's model is not accurate.
- The Cluster model suggests that only 3 groups should be used.
- TravelTide needs to find 3 smaller groups. This can be done by equally dividing up the 3 margins of error on both sides of the mean of the variable VACATION_DAILY_DIFFERENCE.

Appendix

(1). SQL code file (file name: sqlcode.txt)/Python script (file name: MS_segment_final.ipynb):
[https://github.com/Ghubby30/MasterSchoolSegmentation]
(2). Top Perk Dashboard:
[https://public.tableau.com/views/TopPerkDash/TopPerk?:language=en-US&:display_count=n&:origin=viz_share_link]