Segmentation Analysis
By Nicholas Jones

Metrics

1. AVG_HOTEL_OCC_DAILY: average amount of days that buyers occupied a hotel room.
2. TOTAL_BAGS: total sum of checked bags.
3. NO_CANCEL_NUM: the sum of the number of book seats on a flight and the number of hotel rooms booked.
4. ALL_DISCOUNTS: the average of the sum of the two discount rates divided by the total number of hotel and flight bookings.
5. ORIGINAL_PAY: the sum of the base fare prices.

Methodology

1. Segmentation strategy: mean +/- 3*Margins of Error using the variable VACATION_DAILY_DIFFERENCE (the average number of vacation days): In this case, the mean (M) is .0452, and an M.E. is .0710.

2. Each group has a significant name: "G1" is the group that is closest to the mean, "G2" is the 2nd closest group, and etc.

Visualizations: Histograms

The Histogram for the segmentation variable: vacation_daily_difference
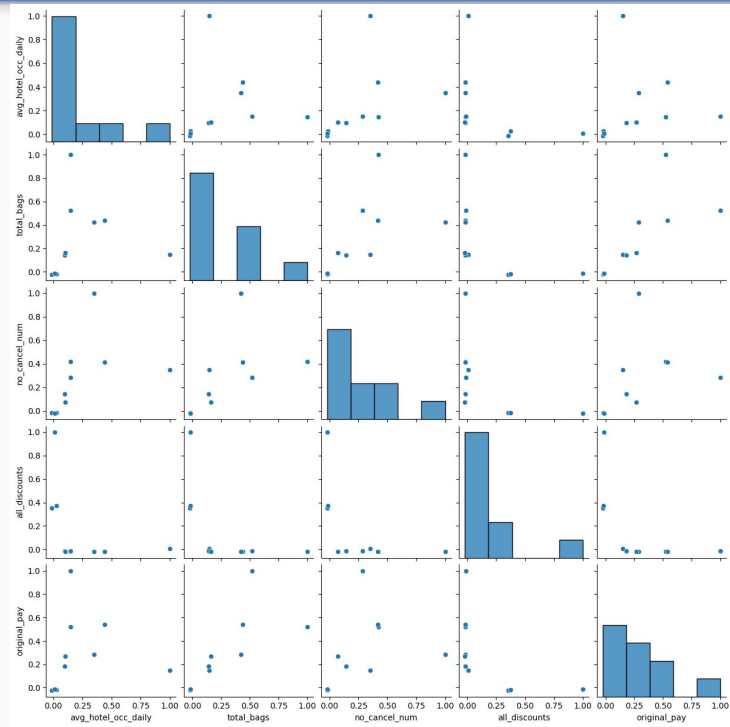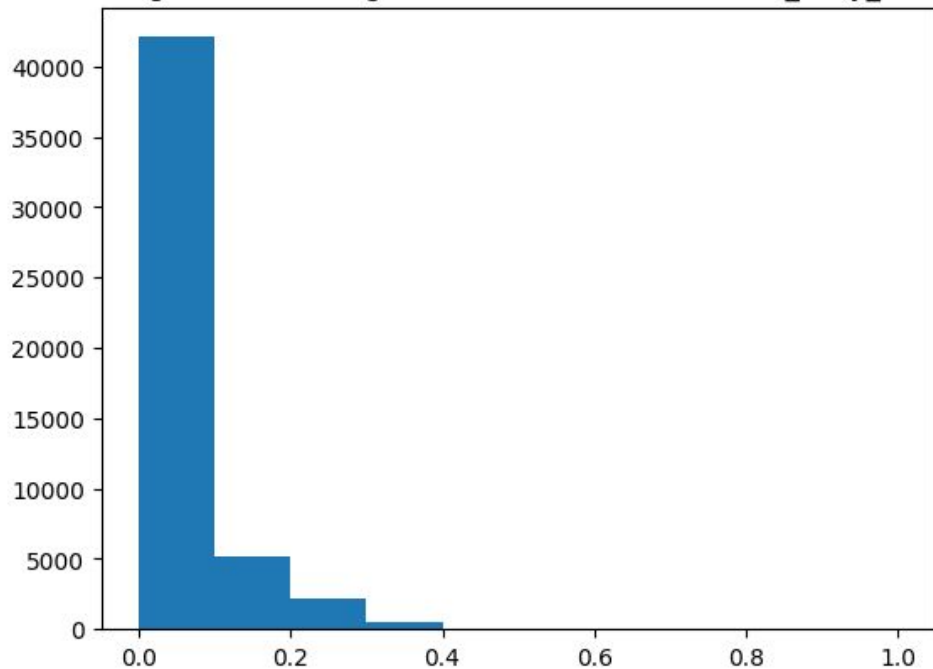
Insights

1. All of the metrics are very similarly distributed relative to the segmenting variable.

2. This means that the mean is very low and most of the customers will be close to the mean.

3. Since the customers are very closely distributed around the mean, this means that not every group will have customers.

4. Customers are mostly in "G1" and "G2."

5. The number of customers in the first 2 groups can be as high as 48,798.

Recommendations

1. For all 5 perks, TravelTide needs to choose customers that are closer to the mean.

2. Elena's theory isn't completely valid.

3. TravelTide needs to segment the range (Mean +/- 3*ME) by segments of the size +/- 1* ME.