Ethan & Minh

Activity 7

12/8/22

<div align="center">Activity 7: Belo Día Hotel</div>

1.

a. 3,000 customers

b. 80 nationalities are represented

c. The most popular market segment for Belo Día customers falls in the other category, accounting for 57.2% (1,716) of customers in the sample. Followed by three segments with similar representations, Travel Agent/Operator 14.4% (432), then Direct 13.03% (391), and Groups 11.57% (347). The least popular market segments are Corporate 3.2% (96), Aviation 0.47% (14), and lastly Complementary 0.13% (4).

Barchart of Market Segment
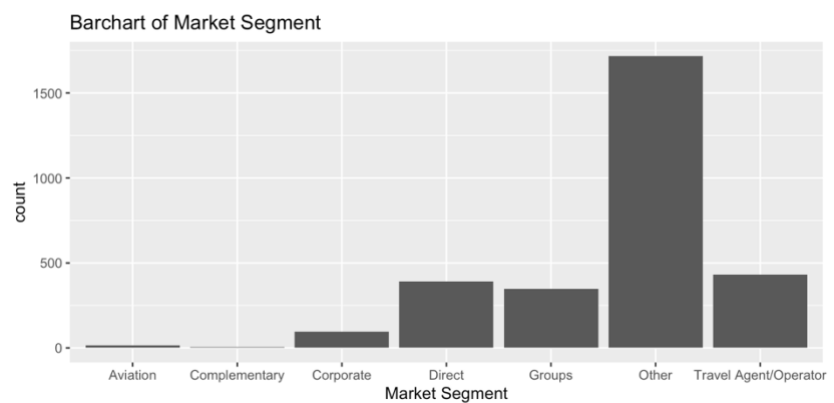
d. The distribution for Average Lead time is unimodal and heavily skewed to the right. The distribution has a median value of 60 days, and an IQR of 120 days. There are no missing values, but there are potential outliers around the 500-550 mark.

Histogram of Average Lead Time
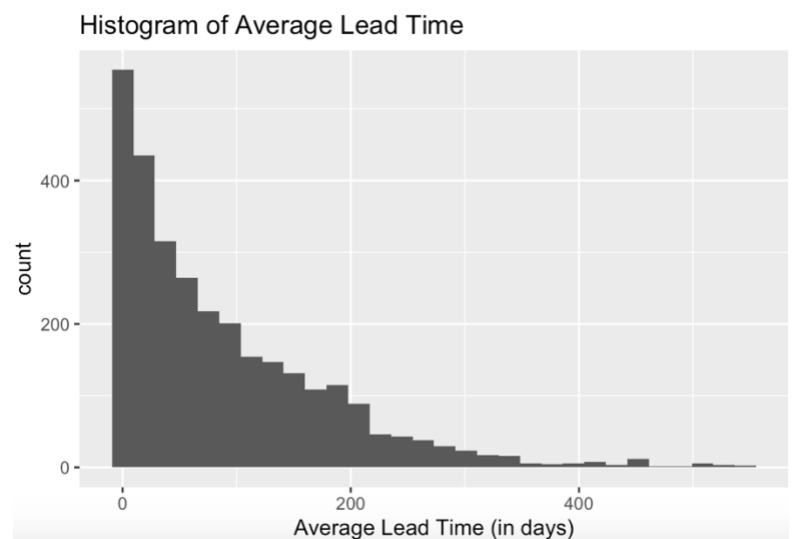
2.

a. We cannot use Euclidean's distance because we have both numeric and categorical variables

b. The basic process of agglomerative clustering involves taking individual observations as their own cluster, then merging clusters that are similar to each other, until there is only one cluster.

c. It would not be reasonable to use 3 clusters based on the dendrogram, as the distance between 4 clusters and 3 on the right side is too short, meaning the clusters are too similar.

d. Based on the dendrogram, 4 clusters is a reasonable amount, as there appears to be enough dissimilarity between clusters, and it gives a closer look at specific information on customer characteristics in each cluster, so the hotel can properly market to those segments.

e.

Cluster 1: Vacationers- This group of customers spent the second most total amount of money, mainly comprised of the other market segment, and relatively evenly split amongst the different nationalities. They also had the highest persons/night.

Cluster 2: Partiers- This group of customers is made of the Groups segment, so they most likely purchase rooms in bulk for events like weddings as such, as they have the second highest lead time, lowest person/night, and room/night, but spend the second least total amount of money.

Cluster 3: Planners- This group is made of only travel agents, so these customers use agents to plan their trip as they have the longest lead time, and spend the least on rooms, but spend the most on other expenses, so they could be vacationers, or business trips, as they also have the second highest person/nights, and highest room/nights. This group comes from a distribution of different nationalities.

Cluster 4: Big Spenders- This group of customers spent the most money, with the highest on lodging and second highest on other, most likely on business, and were mostly represented by Portuguese, and many different market segments (Direct, Corporate, Aviation, Complementary, Other).