Daniel Kim, Tae Yong Ahn, Taewook Ha, Kevin Liao

Ticketmaster: Dude, Where Is My True Fan?

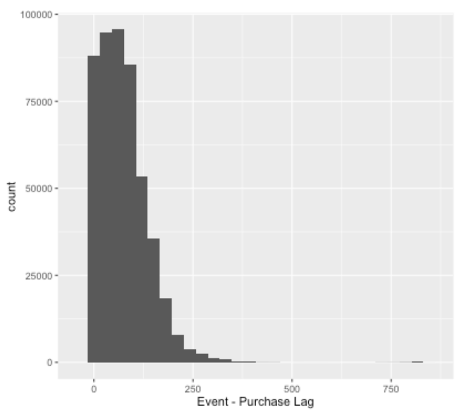
Objective: Our team's objective is to identify Ticketmaster's distinct customer segments for marketing optimization.

Hypothesis: To achieve this objective, we have formulated the assumption that each customer segment differs in the following:

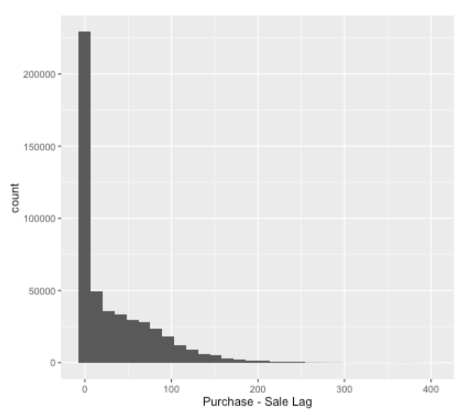
1. The average distance between the attended events
2. The SD of the distances of the attended events
3. Average price of tickets purchased
4. Average number of tickets purchased at once
5. Total number of purchases
6. Time lag between sale start date and purchase date

Methodology: Our very first action was to conduct an exploratory analysis of the given data. Since we were given real-life data, there were a lot of unprocessed parts, and hence, we spent a sizable amount of time data-cleaning. As this project has the characteristic of unsupervised learning, our next move was to acquire domain knowledge through outside research. After learning about the dynamics about the primary and secondary online ticket sales market, we extracted various features that we thought were relevant to customer segmentation. The features were then used to group the different clusters into a solid customer segment.

The analytical methods we have used to prove our hypothesis are as follows: PCA, K-Means, research papers, geographical resources, and other external resources.



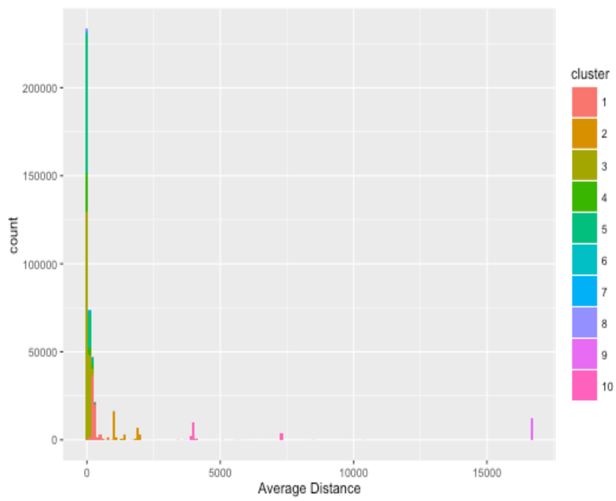
[Figure 1]



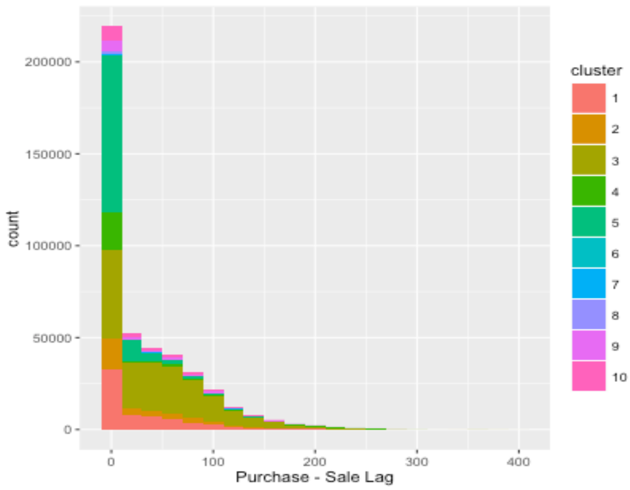
[Figure 2]

Analysis: [Figure 1] and [Figure 2], relatively, are event-purchase lag and purchase-sale lag, meaning [Figure 1], for example, shows the time lapse between the actual date of the event and the date of purchase. One of the assumptions we had about scalpers was that since profitability is their priority, they will have a tendency to buy as early as possible when the tickets are relatively cheap. Hence, we used [Figure 1] and [Figure 2] to verify the validity of our assumption.

Based on the extracted features, we did k-means clustering in order to observe any detectable common user behavior within each clusters.

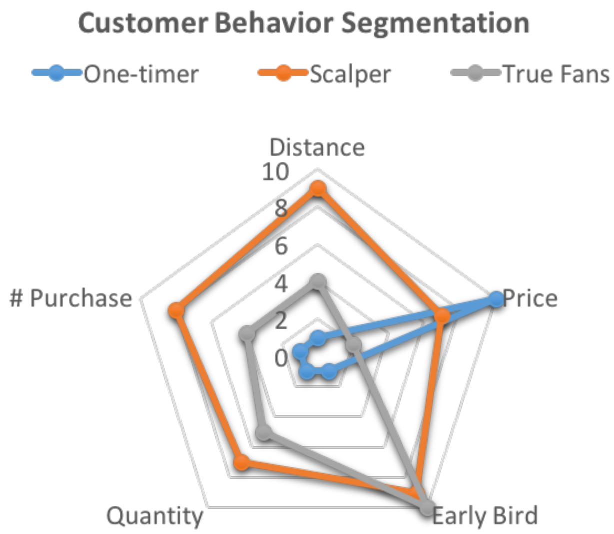


[Figure 3]

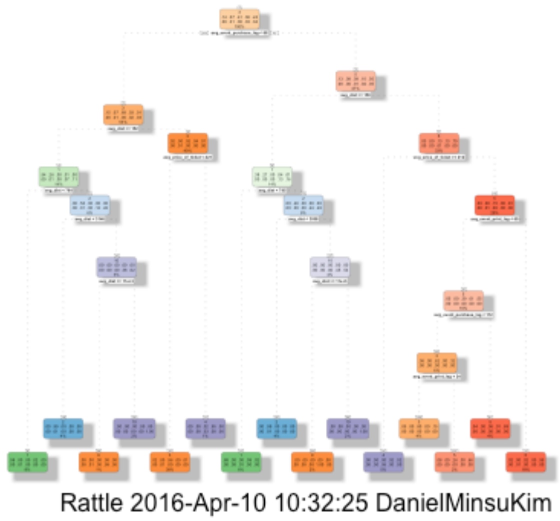


[Figure 4]

We used a total of 10 clusters because we wanted to avoid broad generalizations about common behavior, and make our groupings as precise as possible. Then, we manually merged certain clusters with similar characteristics, ending up with three final characteristics. The manual merging process was inevitable for this was an unsupervised learning process. [Figure 3] shows the k-means clustering result for average distances between attended events. [Figure 4] shows the time lapse between the sales start date and the purchase date. In the two aforementioned figures, we observed both distinct behaviors and common trends, through which we grouped the clusters.



[Figure 5]



[Figure 6]

[Figure 5] shows the distinct behaviors of the three final clusters based on a pentagonal diagram. The scalper (orange), for example, has a tendency to purchase tickets for events further away from his/her location, purchase large quantities repetitively at a lower price, and purchase rather early on in the sales process.

Conclusion:

Recommendation: For future operations purposes, we have automated the customer identification process for Ticketmaster. [Figure 6] shows a decision tree model Ticketmaster can prospectively use in order to identify which segment a new customer falls into. From our decision tree model, a new customer will be identified with our features, leading him/her with a suitable classification label we provided.

Further Study: Due to time constriction, we were only able to analyze Purchase data. We can further inspect GA data to capture more specific behaviors leading to their ticket purchases. Since the interest of each segment of customer is different from one another, there must be distinct patterns between their behaviors prior to their purchases. By incorporating this data to purchase history data, we can classify customers into distinct three categories, fans, scalpers and one-timers.