Location Decision of a Museum Service Firm

Introduction

Wave, Inc is a Museum service firm in Seattle, Washington, specializing in exhibit design, display supplies, exhibition services, and restoration. For the past 15 years, Wave has grown from a small service firm offering services in exhibition design to a full-blown museum services firm with an extensive service provision. They've also grown in the number of employees, revenue, and service offering. Wave has been able to differentiate themselves from their competitors and is now a leading museum service firm in Seattle. While their services are sought after by various Museums across the United States, Wave has no businesses in any other state or city apart from Seattle, Washington.

In the last Management meeting, the firm decided to expand its business to 15 more cities in the US. The firm's management is interested in finding their target cities: cities that are more likely to lead to higher revenues for the firm. These cities should have a significant number of top-rated museums. Their most preferred cities should also have a relatively higher population (at least 100,000 people). Further, the museum should have a significant number of people rating or liking the museums in the city.

In this project, we analyze the location decision of Wave Inc. We compare different cities in the US and cluster the cities based on the number of museums in the city, the average rating of museums in each city, the average number of ratings received by museums in the city, the average number of likes received by each museum in the city, and the average number of tips received by museums in each city. This will help Wave to determine the cities with a large number of highly-rated museums to help it to make an optimal location choice decision.

We abstract from an analysis of firm competition or the presence of other museum service firms in the city. That is, while a city may have many several top-rated museums, it may not be the optimal location for a new entrant in the museum service business. This is because there may already be more than enough museum service firms in the city. In this analysis, we assumed that

Wave would be able to compete favorably with the other service firms because of the quality of service it provides its customers.

The results show that New York City is an outlier with the highest number of top-rated museums. A significant number of people have rated the museums in New York City than any other city in the US. The average number of likes received online for museums in New York is higher than the average number of likes received for any other city in the US. However, other cities such as Chicago, Washington, Boston, etc. also have top-rated museums and offers a profitable location for museum service firms.

Data

We use data from different sources. First, since the management of the company is interested in US cities with at least 100, 000 people, we scrapped a table of the list of cities in the US with at least 100,000 inhabitants from Wikipedia (https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population). We extracted the city name and state from the table. We also extracted the population of each city based on the 2018 population estimate. Table 1 shows that a total of 317 cities were obtained from the Wikipedia page. The table also indicates that all the cities have a population of more than 100,000. The average population is about 278000, while the minimum and maximum populations are 100,145 and 8,175,133, respectively.

	2018estimate
count	317.00
mean	277634.42
std	560553.56
min	100145.00
0.25	109565.00
0.5	150441.00
0.75	236123.00
max	8175133.00

Table 1: Summary Statistics of the Population of the selected cities

Next, we did the city search with Foursquare.com API to find all the top museums in each city. The mean number of museums per city is about 13. Surprisingly, the city with the highest number of museums in our dataset is Glendale, Arizona, with 59 museums.

Finally, we did Foursquare venues search to find the rating, number of ratings (ratingCount), number of likes (likesCount), and number of tips (tipsCount) of each museum.

All museums that did not have a rating were dropped from the dataset. After losing the museum with missing ratings, we were left with 3463 museums in 303 cities. Since the goal of the project is to compare the cities based on the average of the variables found for each city, we aggregated the data to the city level. Thus one data point in our final dataset contains the name of the city, the state that the city is in, the average ratings of the top museums in the city, the average ratingCount, the average tipCount, and the number of top museums found from the API search. The data also contains the longitude and latitude coordinates of each city. Table 2 describes our final dataset. While Glendale has the highest number of museums, the museums in Glendale do not have high ratings. The museums in Glendale, Arizona, have an average rating of 7.8. New York City, known for its top rating museums, returned 30 museums with an average rating of 8.97.

	rating	ratingCount	tipCount	likesCount	N
count	303	303	303	303	303
mean	7.75	98.65	16.79	77.77	12.56
std	0.44	161.92	20.04	137.19	10.92
min	6.42	9.33	1	5	1
25%	7.48	31.65	7	21.33	4
50%	7.79	60.72	11.71	45.2	9
75%	8.01	112.85	21.91	88.75	19
max	8.98	2265.83	265.6	1932.6	59

Table 2: Summary Statistics

Methodology

Our primary methodology is K-means clustering. While the datasets are at the museum level, i.e., each data point represents a museum, we collapsed the data to the city level by finding the average rating, average likesCount, average ratingCount, and average tipsCount for each city.

Next, we used the k-means clustering algorithm to cluster the cities into 5 clusters based on the variables listed above. K-means clustering aims to partition all the cities into k clusters (in this project, we choose to partition the data into 5 clusters) in a way that minimizes the within-cluster variation. The resulting clusters were visualized on a map to see the locations of each cluster or city.

Results

The results from the k-means clustering show that New York City is an outlier city with its own cluster. New York has the highest number of high rating museums, with all the museums receiving high ratingsCount, high likesCount, high tipsCount, etc. These results do not change irrespective of the number of clusters we use. The next cluster is cluster two that contains cities with relatively high values for rating, ratingCount, likesCount, and tipsCount. The museums in the cities in cluster 2 have an average rating of 8.68. The next cluster, cluster 3. Table 3 shows the average rating, ratingCount, tipCount, likesCount, and the average number of museums in the cites in each cluster.

Cluster Labels	rating	ratingCount	tipCount	likesCount	N
1	8.976667	2265.83333	265.6	1932.6	30
2	8.685402	588.01757	77.05041	491.48145	32.57143
3	8.165059	235.738679	35.52228	190.43356	23.07143
4	7.971613	115.922321	21.213	90.657796	18.16667
0	7.549628	39.191582	8.224405	28.659822	7.52459

Table 3: Summary Statistics Each Cluster

The results further show that the cities in the top two clusters, clusters one and two are New York, Chicago, Daly City, Los Angeles, Newark, and Richmond, San Francisco, and Washington DC. These cities are, therefore, the top potential profit cities for Waze Inc to locate in. These cities have a relatively high number of top-rated museums with high ratingCounts, tipCounts, and likesCount.

However, there are only eight cities in the top two clusters. Wave Inc aims to move into 15 additional cities (or 16 cities overall including Seattle).

Cluster Labels	City	State	rating	ratingCount	tipCount	likesCount	N
1	New York	New York	8.98	2265.83	265.6	1932.6	30
2	Chicago	Illinois	8.62	542.07	78.17	446.97	30
2	Daly City	California	8.43	441.69	58.59	368.1	29
2	Los Angeles	California	8.85	668.2	88.27	567.37	30
2	Newark	New Jersey	8.58	498.3	66.37	409.3	30
2	Richmond	California	8.66	438.77	56.03	366.6	49
2	San Francisco	California	8.75	640.13	82.93	535.67	30
2	Washington	District of Columbia	8.91	886.97	109	746.37	30

Table 4: Cities in Clusters 1 and 2

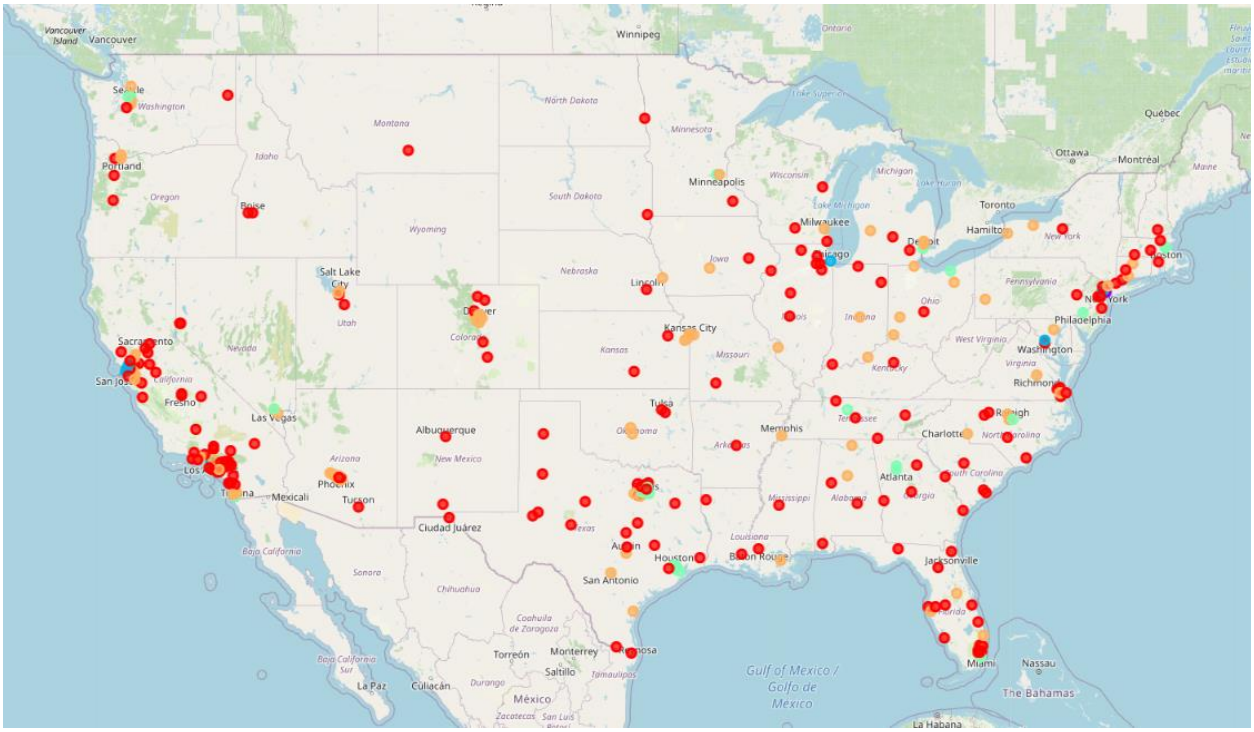
Thus, we will need to supplement our list of cities with some cities in the next rated cluster, Cluster 3. Cluster 3, however, contains 28 cities. Seattle, the home city of Wave Inc, is also in this cluster. There are many ways we could select the seven cities (excluding Seattle) from the 28 cities in cluster 3. We could do a naïve selection and randomly select seven cities from the cluster. We could also rank the cities in the cluster based on one of the variables (say, rating) and select the top 7 cities in the rank. Ranking by rating, for example, puts Raleigh, North Carolina ahead of the list even though the API search output returned only seven high-rated museums for the city. Ranking the cities by any one of the variables leads to the same problem where we select a city very high on the chosen variable but perhaps low on all the other variables. The approach we took in this project is to define a new variable, size, which is the product of all variables: rating, ratingCount, tipCount, likesCount, and N (number of museums returned by the API search for the city). We ranked the cities based on the new variable, size, and selected the top 8 cities in the rank (We select eight cities instead of seven since Seattle is included in the selected cities). Table 5 shows the results of the chosen cities from Cluster 3.

Cluster Labels	City	State	rating	ratingCount	tipCount	likesCount	N
3	Glendale	California	8.38	341.17	44	287.73	30
3	Seattle	Washington	8.31	314.41	49.22	250.48	27
3	Boston	Massachusetts	8.34	292.9	45.23	236.67	30
3	Bellevue	Washington	8.29	288.17	45.47	229.23	30
3	Philadelphia	Pennsylvania	8.5	274.03	44.27	221.17	30
3	Atlanta	Georgia	8.46	247.1	46.4	189.83	30
3	Sandy Springs	Georgia	8.37	240.03	45.37	184.2	30
3	Pasadena	Texas	8.2	215.76	30.1	178.57	36

Table 5: Selected cities in Cluster 3

Cluster Visualization

The map below shows the cites and the clusters to which they belong for the entire dataset. New York City, with its own cluster, is in Cluster 1, shown as a purple circle marker in the map. Cities in cluster two are indicated with the blue circle marker; cities in Cluster 3 are represented with a light green circle marker; cities in cluster 4 are displayed with an orange circle marker. Further, cities in Cluster 0 are shown with a red circle marker. As expected, cluster zero, which is the lowest rank cluster in terms of all of the variables, has the most significant number of cities.



Discussion:

Based on the results, Seattle, where the firm has been quite successful, is not even among the top two clusters. Seattle is in the third-ranked cluster (cluster 3). Thus we expect that if the firm sets businesses in the high clusters (first two), they are more likely to be much more successful than in Seattle. Instead of performing k-means clustering on the cities, one way we could have selected Wave's preferred cities is by naively selecting cities with a high number of museums without regard for the ratings, ratingsCount, or likesCount of the museums in the city. The method would have led to different cities' selection. For example, Glendale Arizona has about 59 high rating museums. Still, the average rating of the museums in the city is 7.8 (far below the average rating of museums in the top three clusters). These museums are more likely to be small and less frequented by people. The rating count for Glendale is 97. To put it into perspective, the cities in cluster 2 has an average rating count of 588 while New York City has a ratingCount of over 2000. The cities in the top two clusters are not only frequented by visitors than the museums in Glendale, but they are also highly rated and liked by the visitors. Museum administrators in these cities are more likely to invest in museum service firms to satisfy their visitors and lead to higher ratings. Therefore, selecting the cities with the highest number of museums may not be an optimal decision for Wave Inc. Thus, the k-means clustering is more likely to lead to top

museums that are more likely to be frequented by museum lovers. And these are the museums that are also more likely to invest in museum services.

Conclusion:

This project compared and clustered different cities in the US-based on the number of factors. The goal is the find cities with high rated museums where a museum service firm can locate to maximize profit. In this project, we assumed the firm's location decision is based only on the presence of high-rated museums that will require their services. While the presence of other museum service firms may alter the clusters and lead to an accurate grouping of high potential revenue cites. We abstracted from the analysis of firm competition on the location decision. The goal was to simplify the analysis and leave the study of firm competition to future work.

The main results show that some cities with a high number of museums may not necessarily be the best profit-maximizing location for the firm since the museums may be less frequented by visitors, and less likely to invest in the services of museum service firms.

Incorporating other variables such as rating, ratingCount, and likesCount into the analysis and performing a k-means clustering is expected to yield the best profit-maximizing city location.