Capstone Project - Battle of the Neighborhoods Coursera/IBM - Applied Data Science Capstone **Table of Contents** 1. Introduction: Problem 2. Background 3. Data 4. Methodology 5. Results 6. Discussion 7. Conclusion 1. Introduction: Problem A movie theater chain is interested in exploring new outdoor business models, such as outdoor showings or drive-in theaters to implement post-COVID-19 in Seattle. The movie theater chain wants to better understand customer segmentation in Seattle based on venues around the city. They are also looking for a location recommendation to implement and test these new outdoor business models in 2021. 2. Background and movie theaters. health of the company and possibly entire movie theater industry. Outdoor showings and drive-in theaters are one possible alternative.

Box office revenues have declined from ~10.3B in 2019 to \$2B in 2020 due to COVID-19 (boxofficemojo.com). Additionally, there has already been an ongoing shift to OTT platforms (such as Netflix and Hulu) over the past decade with consumers increasingly relying on these online streaming services for watching content. This has taken business away from traditional content distribution platforms including linear TV

Given these dual factors, it is an important question for the movie theater chain stakeholder to explore alternative business models and locations to help maintain its revenues/profits and to sustain the future

Finally, it is suitable to have Seattle as a testing ground to find a location as it is less expensive to implement in this city vs. other larger metropolitan areas known for film such as Los Angeles or New York. If the chosen location is successful, then it can be launched in additional locations around the country using similar parameters/location data analysis. 3. Data

## • Approximate latitude, longitude coordinates for neighborhood centers will be found using the **geopy** library Foursquare location data will be used to analyze the following via the Foursquare API:

List of neighborhoods in Seattle:

• Frequency and type of venues in each neighborhood · Density/Number of existing movie theaters in each neighborhood How this data will be used to solve the problem:

• From the Seattle government site (https://www.seattle.gov/neighborhoods/neighborhoods-and-districts)

• Both the foot traffic and density of movie theater and non-movie theater venues will be used to determine the optimal location selection for an outdoor showing/drive-in theater. The number of venues will be a proxy for the foot traffic/popularity of the neighborhood • The recommended area will have at least some nearby outdoor-type venues but not be so densely populated to allow for social distancing

I'll use the list of neighborhoods from the Seattle government site (https://www.seattle.gov/neighborhoods/neighborhoods-and-districts) to create a dataset with neighborhood name information and save as a csv file. I will use the read\_csv function to create an initial pandas dataframe. Neighborhood Latitude Longitude

Jackson 47.597771 -122.329599

Out[2]:

I will then update any neighborhood names that have the special characters "/", "-", or are not recognized by geopy

Indianola

**2** Aurora-Licton Springs 47.695534 -122.337623

Ballard 47.676507 -122.386223 Beacon Hill 47.579258 -122.311598

Geopy is used to fill in the geographical coordinates of each neighborhood center. (43, 3)3.1 Explore Dataset and Visualize Seattle Neighborhoods

Kenmore

Woodinville

Redmond

WA 520

Bellevue

Mercer Island

Mercer

Cottage Lake

Union Hill-Novelty

Sammamish

Issaquah

Carnation

Preston

I will use geopy to find the coordinates of Seattle, Washington and create a map with the neighborhoods superimposed using the folium library The geograpical coordinates of Seattle are 47.6038321, -122.3300624. Out[8]: + WA 307 Poulsbo Naval Base Suquamish Kitsap

Tracyton

Bremerton

Erlands Point-Kitsap Lake

3.2 Foursquare API

4.1 Exploratory Data Analysis

4. Methodology

WA 305 Bainbridge

Island

Manchester

Bainbridge

Port Orchard East Port Orchard WA 160

Vashon

I will use the Foursquare API to gather venue data near each location within 800 meters radius of each neighborhood and a limit of 150 venues returned

0

Burien

I will plot a bar graph of number of venues in each neighborhood to view the density in each neighborhood. Out[16]: Text(0.5, 0, 'Neighborhood') Number of venues in each Seattle neighborhood Venue # of venues 20 Crown Hill Lake City Columbia City Westwood Village Ballard ersity District Capitol Hill oneer Square Central Square Sand Point

Neighborhood

5th Most Common

General College &

Thai Restaurant

Pub

Café

University

4th Most

**Common Venue** 

**Grocery Store** 

**Bus Station** 

Bakery

Mexican

**6th Most Common** 

Pizza Place

Drugstore

**Bowling Alley** 

Sandwich Place

Arts & Crafts Store

Jewelry Store

Mediterranean

Park

7th Most Common 8th Most Common 9th Most Common

**Automotive Shop** 

Gas Station

Food Truck

Spa

Scenic Lookout

Fabric Shop

Taco Place

10th Most

**Common Venue** 

Salon / Barbershop

Martial Arts School

Fried Chicken Joint

Baseball Field

▲Carnation

Preston

We'll also use one hot encoding technique to find the frequency of each venue type in each of the neighborhoods

Next, to further explore the data, I'll create a table showing the top 10 most common venues in each neighborhood:

3rd Most

Pet Store

Brewery

Brewery

**Burger Joint** 

2nd Most

Chinese Restaurant

Park

Park

Bainbridge

Island

Manchester

I'll cluster the neighborhoods to see their similarities/differences using the K Means algorithm

Tracyton

Bremerton

Port Orchard

I'll test out Kmeans for number of clusters ranging from 1 to 15:

And create a plot of the Elbow Method results:

East Port Orchard

Erlands Point-Kitsap Lake

4.2 K Means Testing

3000

2750

2500

2250

• 0: Accomodations and Food

• 2: Asian Restaurants and Bars • 3: Bars and Asian Restaurants

7: Accomodations and Bars

• 9: Bars and General Dining • 10: Bars and Asian Restaurants

• 8: Accomodations and General Dining

First we'll view a **folium map** of the clusters

WA 307

Poulsbo

Tracyton

Tracyton

Bremerton

Port Orchard

outdoor experience opportunities for their audiences.

[NbConvertApp] Converting notebook Coursera-DS-Report.ipynb to html [NbConvertApp] Writing 1281455 bytes to Coursera-DS-Report.html

East Port Orchard

Manchester

WA 160

Erlands Point-Kitsap Lake

Bremerton

Erlands Point-Kitsap Lake

WA 308

Indianola

Suquamish

Bainbridge

Bainbridge Island

• 4: Cafes and Gyms • 5: Asian Restaurants • 6: Zoos and Parks

5. Results

Out[62]:

1: Outdoor Recreation and Transportation

Bainbridge Island

WA 160

Vashon

Since Seattle has the 3rd highest number of coffee shops per capita, I decided to drop the 'Coffee Shop' category to create more distinct clusters

The Elbow Method showing the optimal k

Coffee Shop

1st Most

**Common Venue** 

Coffee Shop

Marijuana

Dispensary

Coffee Shop

Pub

Out[21]:

0

1

2

3

Neighborhood

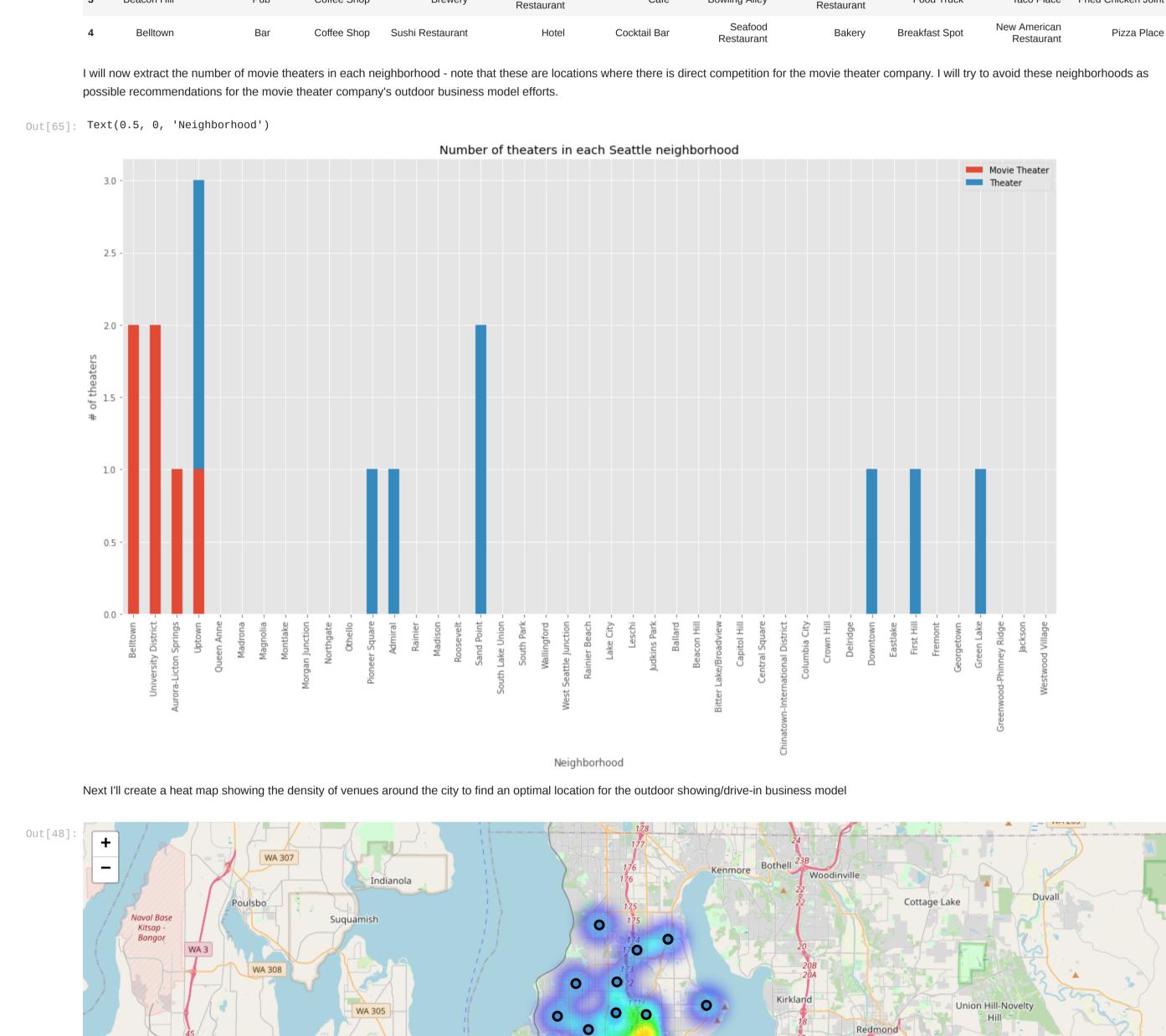
Aurora-Licton

Admiral

**Springs** 

Ballard

Beacon Hill



0

0

Burien

0

WA 520

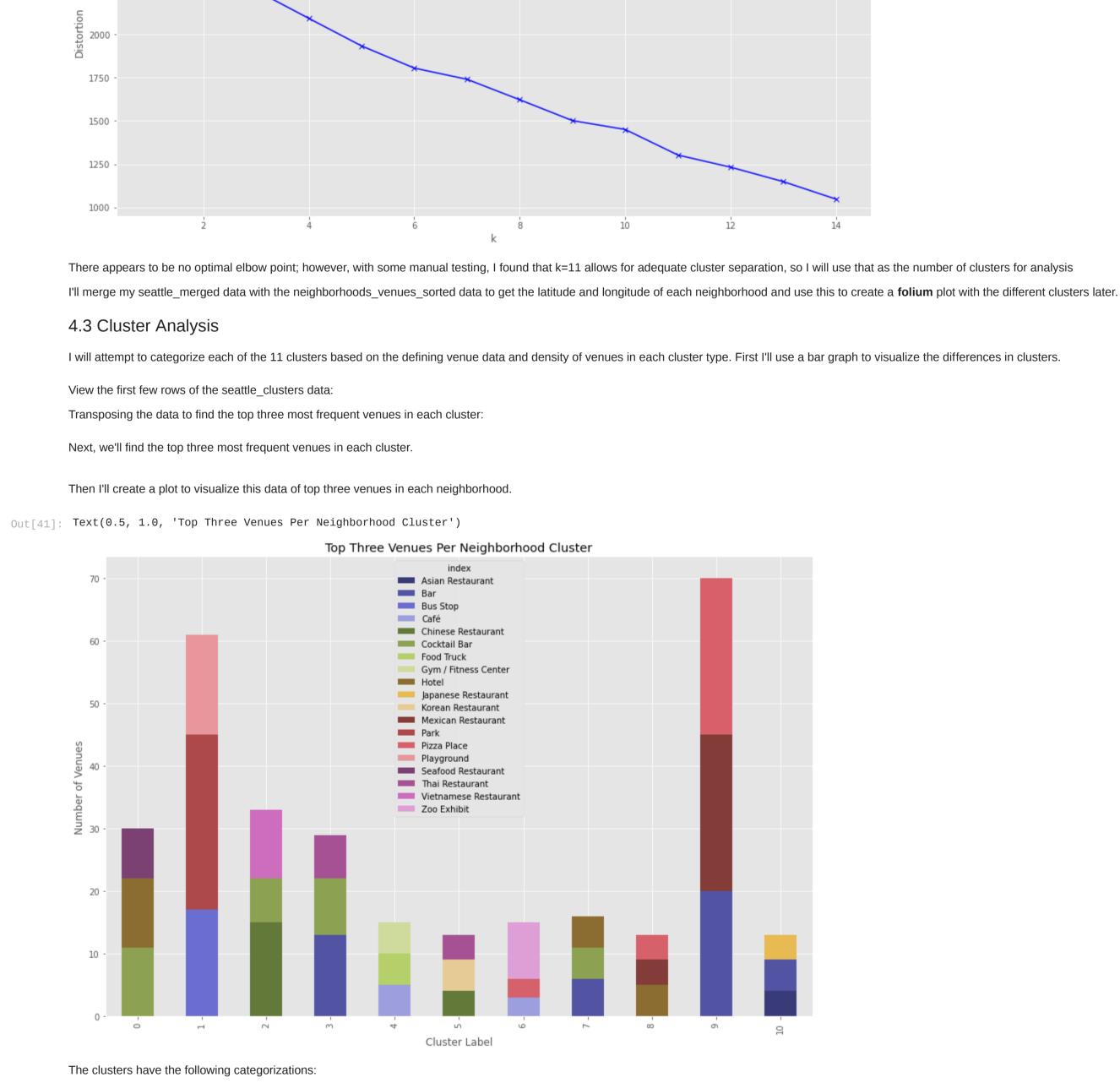
Sammamish

Bellevue

Mercer Island

Mercer Island

Renton



Manchester Issaquah Port Orchard East Port Orchard Preston WA 160 Vashon SeaTac Finally, we'll combine the previous heat map of venue density, along with only the possible location recommendations found in the short list. Out[70]: WA 307 Kenmore Bothell Woodinville Indianola Cottage Lake Poulsbo Naval Base Suquamish 0 Kitsap -WA 308 Union Hill-Novelty WA 305 0 0 Bainbridge Carnation Island 0 WA 520 Bainbridge Island WA 202 Bellevue

We will focus on clusters with access to transportation and/or outdoor areas. In this case, categories 1 and 6. We also see an abundance of categories that are similar, e.g. Bars and Asian Restaurants. Note the

Bothell

Kirkland

Bellevue

Mercer Island

Mercer Island

Mercer Island

Mercer

Woodinville

Redmond

WA 520

Cottage Lake

Union Hill-Novelty

Sammamish

Sammamish

Issaquah

Preston

▲Carnation

WA 202

Kenmore

Lake Washington

distinction is usually in the type of Asian restaurant, e.g. Thai, Chinese, etc. However for simplicity, I have categorized them together as 'Asian restaurants'.

Lets check which neighborhoods belong to cluster 1 and 6 and remove any neighborhoods with existing movie theaters to create a short list of possible recommended locations.

Burien Vashon 6. Discussion Based on the density of venues, I recommend areas around Magnolia, Westwood Village, and Bitterlake/Broadview as potential locations for the movie theater company to create outdoor showing/drive-in theater experiences. These all belong to cluster 1 and have characteristics of being near transportation, making it easily accessible to Seattleites and tourists alike. These areas also have plenty of outdoor areas and open space needed to implement the new outdoor business model in 2021. We also found that there is a large portion of venues dedicated to travel accommodations, outdoors, and asian restaurants, so I can tentatively recommend that movie content chosen can focuse on multiculural, action-adventure type of programming for Seattle viewers. However, further research is needed on customer segmentation, e.g. looking at population demographics data and/or comparing Seattle with other similar cities such as Portland or Vancouver to better understand the distinct viewing audience in Seattle and their differences from those cities (if any). Finally, I also observed in general that many of the clusters only have one or two venues, suggesting that perhaps further data is needed to create more evenly spaced clusters of neighborhoods, such as socioeconomic data, population data, parking data, etc. Carrying out these additional layers of analyses would allow the movie theater company to take into consideration all relevant factors to their decision with regards to finding a location. 7. Conclusion

We found there are several viable candidates for a location to test outdoor screenings/drive-in movies. Based on the densities of venues around Seattle's 43 neighborhoods that were analyzed, we can see this is a good city for the movie theater company to implement their testing of new outdoor business models. This is the perhaps the right time for movie theater companies to pivot to new outdoors business models to

sustain consumer interest at a time when social distancing and streaming movies at home is increasingly rising. Movie theater companies can use this type of data analysis in other cities to identify unique

0