The Battle of Neighborhoods (Week 1)

1. Description of problem and background

Many people associate horse racing and breeding with the state of Kentucky, and Central Kentucky in particular. Lexington, Kentucky's official slogan is "Horse Capitol of the World," and it's impossible to travel backroads in the bluegrass without driving by miles and miles of pristine horse farms. Although Lexington's economy has expanded beyond horse racing alone, owing to a large university and decades of relatively rapid development, the city is still highly dependent on the industry and has established an urban/rural boundary to prevent over-development of farmland surrounding the city.

The center of horse racing in the city is the racetrack Keeneland. Hosting 18 stakes races in its fall 2019 meet with six grade I races and purses topping out at $1 million (Staff, 2019), the track attracts spectators both from the US and globally. The 2019 fall meet has attendance of 262,630 people (Keeneland, 2019). Considering the size of the city is around 330,000 people (US Census Bureau, 2019), the track drives a large part of the local economy during this period. Another meet is scheduled every year in April, with the 2019 spring having an attendance of 242,547 people (Keeneland, 2019). The 2020 Breeders' Cup, one of the premier racing events in the US and in the world, is also on the horizon in November. When Keeneland last hosted this 2-day event last in 2015 it attracted record attendance with 95,102 attendees (Breeders' Cup, 2015).

Because Keeneland is very important to the city and clearly attracts a great number of visitors, the indirect benefits of racing in the area could influence decision making and new development. As such, the central problem proposed is how to quantify the indirect impact of Keeneland on different neighborhoods and venues. Also, how can this information be leveraged by someone looking to start a business in the area that caters to racetrack attendees? Developing a business themed around horse racing could be a promising venture if placed in the correct location and in the right category.

2. Description of available data and usage to solve problem

To approach the problem of developing a business themed around horse racing first requires an understanding of which areas are most visited by attendees and which categories of venue are most frequented. The Foursquare API includes an option to search for "Next Venues". This search requires a venue id and returns the top five, or fewer if not available, venues that users who checked into the searched venue have checked into next. Several levels can be created with the next venues of Keeneland, followed by the next venues of these venues, and so on. If a certain fraction is assigned to each next venue based on the order returned, this fraction can be taken of the whole, the original input of Keeneland, at level 1, and then multiplied by the fraction of the prior venues for each subsequent level. What results is the fraction of the whole population that visits each venue in that level, with the fractions of all venues totaling to 1 for each level. In addition to determining the fraction of the population that visits each venue at each level, the program can also match each venue to a census tract. A geojson file of census boundaries is available through

the Kentucky Geographical Network (KYGovMaps, 2020). The geojson file can be used to determine the tract that each venue belongs to.

After determining the proportion of the original visitors that each venue has at each level, those levels are first concatenated and then grouped by venue to determine the total impact. The total number of visitors can be determined by multiplying by the number of visitors in the fall 2019 meet. After a data frame is created with the total number of visitors per venue, the venues can be clustered together using a K Means algorithm. A grid of points can then be constructed at constant intervals away from each of the cluster centers. Using the zoning shapefile for Lexington (LFUCG, 2020), which can be converted to a geojson file, the points can be filtered to include only the areas zoned for business. Next, the distance of the points to each venue with Keeneland visitors can be calculated, and the distances, along with a target radius, can be used to include only venues nearby to the studied points.

After creating a grid of points to study and determining venues with Keeneland visitors nearby, the venue data frame can be grouped by category and the number of visitors can be summed. After being sorted this yields the ranking of categories by visitors. The top categories can be chosen based on total visitors.

Next the top categories that are not within the certain radius of each point are determined using the top categories and the categories that are within the radius. The categories are then matched up with category id using the venue json from the Foursquare API. These category ids along with the coordinates, radius, and limit on number of results are used with the API to determine the nearby venues in the chosen categories for each point. Next the results are grouped by coordinates of each point and the categories. Using this grouping the number of venues in each category within a given radius and the minimum distance to each category were determined for each point.

With this information, four criteria are included for each point: number of venues in each category with a given radius, minimum distance to a venue of each category, Keeneland visitors within a given radius of each point, and Keeneland visitors of each category. A K Means clustering algorithm is then used with these four criteria for each point. A certain number of clusters is chosen based on the elbow method and the top clusters are chosen based on their cluster centers. The ideal cluster will have few nearby venues of the given category, a large minimum distance to a venue of the given category, and a large number of visitors both in the vicinity and in the category. Finally, the venues in the top clusters are filtered based on desired values for these characteristics and the top location-category pairs are determined.

Bibliography

Breeders' Cup. (2015, November 17). *Press Releases*. Retrieved from breederscup.com: https://www.breederscup.com/media-center/press-releases/2015-11-17

Keeneland. (2019, April 26). *Keeneland Closes 2019 Spring Meet With Strong Attendance And Wagering Results*. Retrieved from keeneland.com: https://www.keeneland.com/media/news/keeneland-closes-2019-spring-meet-strong-attendance-and-wagering-results

Keeneland. (2019, October 26). *Strong Attendance And Community Events Highlight Keeneland Fall Meet*. Retrieved from keeneland.com: https://keeneland.com/media/news/strong-attendance-and-community-events-highlight-keeneland-fall-meet

KYGovMaps. (2020, June 5). *Ky Census Tracts 2010 WGS84WM*. Retrieved from opendata-kygeonet: https://opendata-kygeonet.opendata.arcgis.com/datasets/ky-census-tracts-2010-wgs84wm/geoservice

LFUCG. (2020, January 29). *Zoning*. Retrieved from data.lexingtonky.gov: https://data.lexingtonky.gov/dataset/zoning

Staff, H. L. (2019, June 26). Record purses, 10 Breeders' Cup berths await Keeneland Fall Meet competitors. *Lexington Herald Leader*.

US Census Bureau. (2019, July 1). *QuickFacts Lexington-Fayette, Kentucky*. Retrieved from census.gov: https://www.census.gov/quickfacts/lexingtonfayettekentucky