Optimal Portland Ballpark Location

David Knourek

February 9, 2020

Introduction

The city of Portland, Oregon is a commonly considered location for a new Major League
Baseball (MLB) team should the league decide to expand its number of teams. As with all
business ventures, location is critical to the success of a sports venue and to the financial success
of the team and the league. Although many ballparks are slightly removed from urban centers
and surrounded by large parking lots (Dodger Stadium in Los Angeles, Citi Field in New York),
many beloved stadiums are in the heart of their cities' urban environments (Wrigley Field in
Chicago, Fenway Park in Boston). Other ballparks, such as Coors Field in Denver or Nationals
Park in Washington D.C., have revitalized somewhat forgotten industrial areas in their city with
new businesses conducive to the game-day atmosphere, such as sports bars. This project seeks to
replicate the success of the more urban-centric ballparks by finding the ideal location for a
hypothetical Portland baseball team's ballpark. The primary audience for this analysis is the
MLB league office and the hypothetical Portland team's ownership group, as they would be most
interested in maximizing profits and viability of the team. A secondary audience is baseball fans
interested in learning what sorts of neighborhood characteristics make a ballpark successful, and
comparing the characteristics of different ballpark neighborhoods.

Data

In this project, Foursquare location data is used to characterize the areas around twelve existing
urban-centric MLB ballparks in order to determine what types of venues are successful near
existing ballparks. Foursquare data for the city of Portland is then analyzed in an attempt to find
an existing area in the city of Portland similar to the successful areas around other ballparks. This

ideally determines a location within the city well suited for a ballpark without requiring an entire neighborhood to be built up or converted to suit the enterprise.

Additionally, especially for urban ballparks, public transit is a critical consideration in order to allow fans to efficiently travel to games. Therefore, Portland area mass transit data provided by TriMet (the Portland area transit service) is also analyzed to determine the ideal location for the hypothetical ballpark. The available data include geospatial data for routes, stops, and transit centers, as well as passenger data indicating the popularity of certain routes. The data primarily used for this project are the geospatial data representing rail stop locations within the city of Portland. These data are combined with the Foursquare data in an attempt to find an ideal ballpark location that is well serviced by public transit and surrounded by businesses characteristic of a typical successful ballpark neighborhood.

Secondary data sources are also used to support the analysis performed in this project. Since no existing dataset of existing MLB ballpark locations was found to be available, and the number of locations relevant to this project is small, the latitude and longitude of the 12 ballparks existing ballparks explored in this analysis are manually scraped from the Google Maps application. Additionally, it is necessary to define and separate the neighborhoods of Portland for this analysis. Unfortunately, no dataset was identified that would closely tie the individual neighborhoods of Portland to specific latitude and longitude coordinates. However, an open source dataset of all United States zip codes (US Zip Code Latitude and Longitude, 2020) is used to determine the zip codes which lie within Portland as well as their coordinates. Since the zip codes define geographically contiguous areas, this provides a viable method of separating the city into distinct areas. Within this report, the term "neighborhood," is used to indicate an area corresponding to a distinct zip code.

Methodology

The process of this study includes two primary tasks and one secondary task. The first task to be accomplished is the generic characterization of the ideal ballpark location for an urban-centric ballpark based on the top venues nearby existing stadiums. The second task is to characterize the venues in different neighborhoods within Portland and compare them to the ideal generic location. Finally, the transit stations within each Portland zip code are explored and used to

supplement the neighborhood characterization in order to inform a decision for the optimal ballpark location.

Ideal Neighborhood Characterization

The first step in the analysis is to characterize the neighborhoods around 12 existing urban MLB ballparks and condense these data into a generic assessment of a successful ballpark location. The urban ballparks considered in this analysis are given below. These parks provide a diverse cross section of the current baseball landscape with both high- and low-performing (in terms of on-field success) teams across a geographically diverse spectrum. These ballparks are loaded into a Pandas dataframe with the latitude and longitude of each ballpark.

Ballpark	Location
Wrigley Field	Chicago, Illinois
Fenway Park	Boston, Massachusetts
Coors Field	Denver, Colorado
Nationals Park	Washington, D.C
Oriole Park at Camden Yards	Baltimore, Maryland
PNC Park	Pittsburgh, Pennsylvania
Yankee Stadium	New York, New York
T-Mobile Park	Seattle, Washington
Petco Park	San Diego, California
Minute Maid Park	Houston, Texas
Busch Stadium	St. Louis, Missouri
Chase Field	Phoenix, Arizona

Next, a function is built which utilizes calls to the Foursquare API to obtain information about the venues within 600 meters of each stadium. Most importantly, the call returns the type of venue, such as Sports Bar, Hair Salon, Mexican Restaurant, etc. Next, the resulting venues are grouped by stadium and counted to determine the number of venues nearby each stadium. This is used to determine an appropriate number of nearby venues for the Portland stadium. The

"baseball stadium" venue category is filtered out of the results. Since every existing stadium of course includes a stadium within its neighborhood, and Portland does not currently have a baseball stadium per the motivation of this project, this category does not contribute meaningfully to the results.

After collecting the venue data, the frequency of each venue-type occurrence for each existing ballpark is calculated by using the Pandas get_dummies function, grouping the data by ballpark and taking the average. These data are then combined to provide an overall frequency of the venue-type occurrences across all 12 of the investigated ballparks. This dataset is considered the generic ideal characterization of a ballpark location.

Portland Neighborhood Characterization

The next step in the assessment is to characterize the neighborhoods of Portland and determine which most closely matches the generic ballpark neighborhood characterization found above. This portion of the analysis is similar to the characterization of the existing ballparks' neighborhoods described above, except in this step, the results are separated and grouped by zip code rather than being combined to determine an overall characterization.

After retrieving the venue data from the Foursquare API and grouping the data by neighborhood, the results are filtered to only include neighborhoods which are considered to be "busy" enough to support a ballpark. This determination is made based on the number of venues found nearby the existing urban ballparks. Next, the frequencies of each venue-type occurrence are calculated in the same manner as described above for the existing ballparks.

Finally, a similarity score is calculated to determine how similar each of the Portland neighborhoods are to the generic ideal ballpark neighborhood. For each venue type in each zip code, the "distance" is calculated as the difference between the frequency of that venue type within the current neighborhood and the frequency of that venue type within the generic ideal ballpark neighborhood. The similarity score of each neighborhood is then calculated as the square root of the sum of the squares of these distances. The lower the similarity score, the lower the difference between the neighborhood and the generic ideal location, and therefore the better the location for the Portland ballpark.

Portland Transit Exploration

The final step in the assessment is to explore the accessibility of each neighborhood via public transportation. This analysis focuses on train stops (light rail, commuter train, and street cars), since trains are able to carry many more people much more efficiently than buses in surge situations, such as thousands of people attempting to travel to and from a venue at the same time.

For this portion of the assessment, the rail stop data are loaded into a dataframe, and the rail stop locations are checked against the locations of each Portland neighborhood. A rail stop within a kilometer (approximately a 10 minute walk), is considered sufficiently close to a neighborhood to be a viable transit point for the hypothetical ballpark. The number of rail stops nearby each neighborhood are then counted and used alongside the similarity scores described in the section above to inform the decision of the ideal ballpark location.

Results

Characterization of Generic Ideal Location

In characterizing the generic location of the existing urban baseball stadiums, the results appear to agree with what one might expect. The top venue types (occurring in frequencies of more than 1%), are primarily establishments serving food and especially alcohol, as well as hotels to support those traveling for the games.

```
Bar                      0.0513813
Hotel                    0.0503147
Sports Bar               0.0424983
American Restaurant      0.0378603
Pizza Place              0.0321423
Coffee Shop              0.0302787
Mexican Restaurant       0.0290631
Sandwich Place           0.0228588
Park                     0.0216839
Italian Restaurant       0.0181653
Burger Joint             0.0177956
Lounge                   0.0172618
Steakhouse               0.0168048
Beer Garden              0.0165981
Café                     0.0129147
General Entertainment    0.0127658
Restaurant                0.012007
Food Truck               0.0116598
Seafood Restaurant       0.0110116
Clothing Store           0.0107357
Gym                      0.0106045
Sporting Goods Shop      0.0103253
Brewery                  0.0102797
Cocktail Bar             0.0101666
```

Similarly, many of the lowest frequency venue types (occurring with frequencies less than 1/1000) are retail stores, niche markets, and other forms of entertainment not connected to baseball.

```
Nail Salon                  0.000915751
Bakery                      0.000915751
Yoga Studio                 0.000886525
Food                        0.000886525
Bookstore                   0.000886525
Dog Run                     0.000886525
Piano Bar                   0.000886525
Flea Market                 0.000886525
Karaoke Bar                 0.000886525
Spa                         0.000886525
Neighborhood                0.000886525
Argentinian Restaurant      0.000886525
Comfort Food Restaurant     0.000886525
Vietnamese Restaurant       0.000833333
Big Box Store               0.000833333
Chocolate Shop              0.000833333
Arts & Crafts Store         0.000833333
Wine Shop                   0.000833333
Salad Place                 0.000833333
Cycle Studio                0.000833333
Department Store            0.000833333
Doctor's Office             0.000833333
Garden                      0.000833333
Hostel                      0.000833333
Meze Restaurant             0.000833333
Movie Theater               0.000833333
Speakeasy                   0.000833333
Pedestrian Plaza            0.000833333
Entertainment Service       0.000833333
```

A small cross-section of the stadium-by-stadium results show that although the top venues nearby each stadium are unique, the themes shown above are evident. I.e., each shows a high frequency of bars, restaurants, and hotels.

```
----Sea----
                      venue  freq
0      Sporting Goods Shop   0.08
1              Sports Bar   0.05
2              Beer Garden   0.05
3   General Entertainment   0.05
4            Hot Dog Joint   0.03
5           Clothing Store   0.03
6         Baseball Stadium   0.03
7                      Bar   0.03
8      Fast Food Restaurant  0.03
9                    Hotel   0.02


----Stl----
                      venue  freq
0                      Bar   0.14
1    American Restaurant   0.09
2                    Hotel   0.07
3           Baseball Field   0.05
4              Sports Bar   0.05
5              Beer Garden   0.05
6               Steakhouse   0.05
7               Food Truck   0.05
8           Clothing Store   0.02
9                      Pub   0.02


----Was----
                      venue  freq
0              Beer Garden   0.04
1              Pizza Place   0.04
2               Food Truck   0.04
3                    Hotel   0.04
4          Harbor / Marina   0.04
5             Burger Joint   0.04
6            Souvenir Shop   0.04
7                     Café   0.04
8       Indian Restaurant   0.04
9       Korean Restaurant   0.02
```

Characterization and Scoring of Portland Neighborhoods

In the assessment of the different neighborhoods of Portland, it is found that many of the neighborhoods have very few (approximately five or even fewer) venues within 600 meters. In

order to retain similarity to the existing ballpark locations, which show at least 40 nearby venues, a minimum value of 20 nearby venues is set as the floor for a viable Portland ballpark neighborhood. The following zip codes and their numbers of nearby venues are those considered as potential locations for further assessment.

	Zip	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	97203	27	27	27	27	27	27
3	97204	100	100	100	100	100	100
4	97205	100	100	100	100	100	100
7	97208	25	25	25	25	25	25
8	97209	100	100	100	100	100	100
13	97214	47	47	47	47	47	47
15	97216	37	37	37	37	37	37
16	97217	34	34	34	34	34	34
24	97227	23	23	23	23	23	23
28	97232	38	38	38	38	38	38
42	97266	21	21	21	21	21	21

A small portion of the venue-type frequencies in some of the "busy" neighborhoods is shown below, along with the "distance," or similarity score to the generic ideal location.

```
----97227----
                  venue  freq  Column_Total    dist_sq
0   Mexican Restaurant   0.13      0.029063   0.010188
1             Brewery   0.09      0.010280   0.006355
2                 Bar   0.09      0.051381   0.001491
3  Furniture / Home Store  0.09   0.002246   0.007701
4               Hotel   0.04      0.050315   0.000106
5              Bakery   0.04      0.000916   0.001528
6          Coffee Shop   0.04      0.030279   0.000095
7          Pizza Place   0.04      0.032142   0.000062
8              Lounge   0.04      0.017262   0.000517
9           Food Truck   0.04      0.011660   0.000803
Distance = 0.21


----97232----
                  venue  freq  Column_Total    dist_sq
0          Coffee Shop   0.05      0.030279   0.000389
1         Cocktail Bar   0.05      0.010167   0.001587
2       Sandwich Place   0.05      0.022859   0.000737
3                 Bar   0.05      0.051381   0.000002
4    Italian Restaurant   0.03     0.018165   0.000140
5           Taco Place   0.03      0.003053   0.000726
6   Rental Car Location  0.03      0.004775   0.000636
7           Restaurant   0.03      0.012007   0.000324
8           Food Court   0.03      0.002007   0.000784
9       Clothing Store   0.03      0.010736   0.000371
Distance = 0.14


----97266----
                  venue  freq  Column_Total    dist_sq
0   Mexican Restaurant   0.10      0.029063   0.005032
1                 Bar   0.10      0.051381   0.002364
2          Coffee Shop   0.05      0.030279   0.000389
3    Recreation Center   0.05      0.001701   0.002333
4             Bus Stop   0.05      0.001543   0.002348
5              Brewery   0.05      0.010280   0.001578
6             Dive Bar   0.05      0.001832   0.002320
7       Sandwich Place   0.05      0.022859   0.000737
8        Boat or Ferry   0.05      0.001701   0.002333
9             Pet Store   0.05      0.001832   0.002320
Distance = 0.21
```

The scores of each neighborhood are shown below on an absolute basis on the left, as well as on a normalized basis to the right.

	Score
97203	0.196541
97204	0.140002
97205	0.144396
97208	0.168094
97209	0.120102
97214	0.140936
97216	0.180285
97217	0.198374
97227	0.205616
97232	0.142085
97266	0.210318

	Score
97209	0.000000
97204	0.220589
97214	0.230939
97232	0.243679
97205	0.269293
97208	0.531971
97216	0.667106
97203	0.847294
97217	0.867604
97227	0.947883
97266	1.000000

Transit Data Combined with Scoring

Finally, the number of nearby transit stations (Count) are combined with the similarity score of each neighborhood compared to the ideal neighborhood characterization (Score), to inform a decision on the ideal ballpark location.

	Score	Count
97209	0.000000	30
97204	0.220589	44
97214	0.230939	0
97232	0.243679	1
97205	0.269293	41
97208	0.531971	1
97216	0.667106	2
97203	0.847294	0
97217	0.867604	2
97227	0.947883	2
97266	1.000000	1

Discussion

As discussed in the above results, it appears obvious that certain types of venues are frequently found nearby ballparks. As might be expected, these frequent venues consist primarily of establishments which serve alcohol, hotels, and relatively generic restaurants with familiar cuisine such as pizza places. While serving their purpose at any time, these establishments all additionally cater to sports fans before and after games. Conversely, establishments such as nail salons, yoga studios, and department stores are not frequently found nearby existing baseball stadiums. Although baseball fans may frequent these establishments, they likely do not represent activities fans prefer to participate in immediately before or after a game, therefore locating a ballpark nearby these establishments would represent a lost opportunity cost on gamedays.
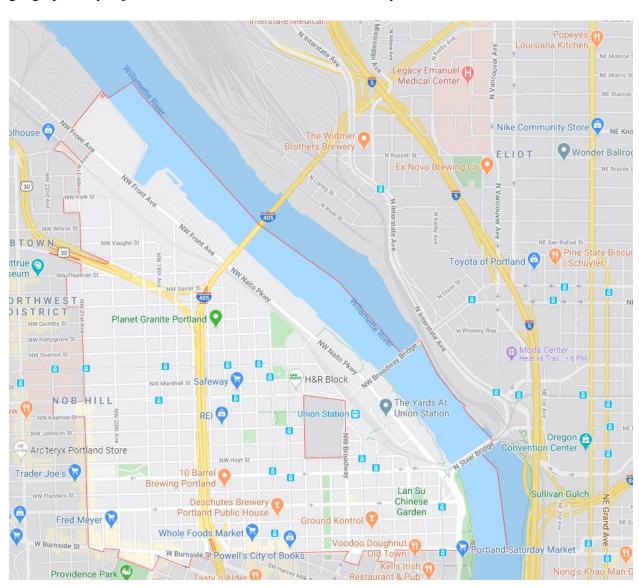
It can also be seen that none of the venue frequencies or their order of precedence matches between each existing ballpark. This indicates that although there is an average or generic neighborhood around each ballpark, there is still room for uniqueness for each individual city.

In exploring the Portland neighborhoods, it is found that many of the zip codes represent areas which only have a few nearby venues, likely representing residential or industrial areas. These areas were filtered out of this study, as the assumption is that the ideal location to locate a ballpark is one in which the neighborhood at least has a start along the path of a busy urban center. Although the floor for this study was set at 20 venues, compared to a minimum of about 40 nearby the existing ballparks, this seems reasonable as additions venues will surely follow in the event of construction of a new ballpark.

In examining the results of the venue frequencies for each Portland neighborhood, it is import to note that at first glance, many of the neighborhoods seem to match the generic ideal ballpark location quite closely, with many restaurants and bars. This highlights the importance of developing an objective metric, the similarity score in this case, which mathematically characterizes the similarity of each neighborhood the generic ideal.

Finally, in examination of the similarity scores alongside the number of nearby transit stops, it appears clear that the 97209 area code is the primary neighborhood for the ideal Portland ballpark location. Although this neighborhood does not have the highest number of rail stops within its radius, the 30 it does have appears to be more than sufficient to support the transit

needs of a baseball game, and the normalized similarity score of 0.00 is significantly better than the next best score of 0.22. These results also highlight the significance of normalizing the scoring, since the un-normalized results appear fairly abstract and difficult to assess in isolation. As shown below from Google Maps, the 97209 area code is located in central downtown Portland, directly on the riverfront, and appears to be a prime location for a ballpark. It is important to note that the other apparently viable neighborhoods (97204 and 97205) are geographically adjacent to the 97209 area, and contain many of the same characteristics.

Conclusion

Overall, this study determines that the 97209 area code of Portland, Oregon represents the ideal location for a hypothetical Major League Baseball stadium within the city. This conclusion is based on the assumption that a new ownership group would seek to mimic the success of existing urban MLB stadiums without requiring building an entire neighborhood around the new park. This conclusion also takes into consideration the accessibility of the stadium by public transit.

It should be noted that there are, of course, other considerations involved in where to build a new ballpark. For example, one may want to consider the availability of a sufficiently-sized plot of empty or available land within the area, as well as the cost of real estate within the neighborhood.

Future work on this subject could include further cleaning of the data and additional venue information. For example, many of the venues found nearby the existing ballparks primarily serve alcohol, but fall into at least a dozen different categories within the Foursquare system (i.e. a bar can be noted as a bar, a pub, a lounge, a dive bar, a sports bar, etc.) Further studies could consider whether it is appropriate to keep these groupings separate, or to combine all types of bars into a single category. Additionally, this study presupposes that the venues nearby existing ballparks are successful. One may want to consider using venue reviews and ratings to actually determine the most successful venues nearby ballparks, rather than just the most frequent.

References

US Zip Code Latitude and Longitude. (2020). Retrieved from
https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/