The best locations to open a sports bar/pub using UK football stadiums data

Mark Forster

30th June 2020

1   Introduction

1.1   Business description

Football is the UK's national sport and as such it is home to many football (soccer) teams and football leagues. Football generated a turnover of approximately 375 million pounds in 2018. On match days football stadiums around the UK are filled with fans excited to support their teams and celebrate their victories, and pubs and bars located close to football stadiums are able to capitalise on this large customer base.

The aim of this analysis is to predict the best locations to open a new sports bar/pub within close proximity to football stadiums in the UK. It is hypothesised that football stadiums with the highest capacity (most potential customers) and lowest number of existing pubs/bars (least competition) are likely to provide the best locations for a new venue. Stadium location data scraped from the web and Foursquare data of existing venues will be used to determine the best locations.

1.2   Stakeholder interest

This data is of interest to owners of existing sports bar/pub chains who are looking to expand and open new venues.

2   Data acquisition and cleaning

2.1   Data sources

Football stadium data is easily available on many websites, including Wikipedia. Stadium data including stadium name, maximum capacity, longitude, and latitude will be scraped from the web. Foursquare will be used to obtain venue information and the number of nearby pubs/bars based on the stadium location data.

Table 1. Stadium data scraped from the web.

	Name	Team	Capacity	Latitude	Longitude	Easting	Northing	Grid Reference	Also known as
0	Abbey Stadium	Cambridge United	10847	52.212799	0.154298	547284	259362	TL472593	NaN
1	Adams Park	Wycombe Wanderers	10284	51.630600	-0.800299	483134	193102	SU831931	NaN
2	Alexandra Stadium	Crewe Alexandra	10153	53.087502	-2.435690	370916	354565	SJ709545	Gresty Road
3	Almondvale Stadium	Livingston	10122	55.886398	-3.522070	304893	666964	NT048669	NaN
4	Amex Stadium	Brighton and Hove Albion	22374	50.860901	-0.080140	535213	108577	TQ352085	Falmer Stadium

2.2 Data cleaning

Data has been scraped and cleaned as shown below to provide a dataframe containing stadium name, football club, maximum capacity, longitude, and latitude coordinates for all top 5 English leagues and top 4 Scottish leagues. Dataframe contains 141 stadiums.

Table 2. Dataframe containing cleaned stadium data.

	Stadium_name	Team_name	Capacity	Latitude	Longitude
0	Abbey Stadium	Cambridge United	10847	52.212799	0.154298
1	Adams Park	Wycombe Wanderers	10284	51.630600	-0.800299
2	Alexandra Stadium	Crewe Alexandra	10153	53.087502	-2.435690
3	Almondvale Stadium	Livingston	10122	55.886398	-3.522070
4	Amex Stadium	Brighton and Hove Albion	22374	50.860901	-0.080140
5	Anfield	Liverpool	45522	53.430801	-2.960960
6	Ashton Gate	Bristol City	21497	51.439999	-2.620210
7	B2net Stadium	Chesterfield	10400	53.253502	-1.427200
8	Balmoor	Peterhead	4000	57.511200	-1.795990
9	Bayview Stadium	East Fife	2000	56.188900	-2.999140

3   Exploratory data analysis

3.1   Stadiums by capacity

A histogram and histogram with KDE show the number of stadiums as a function of capacity. Most stadiums have a capacity of 20000 or less. There are a small number of very large stadiums with capacities greater than 50000.
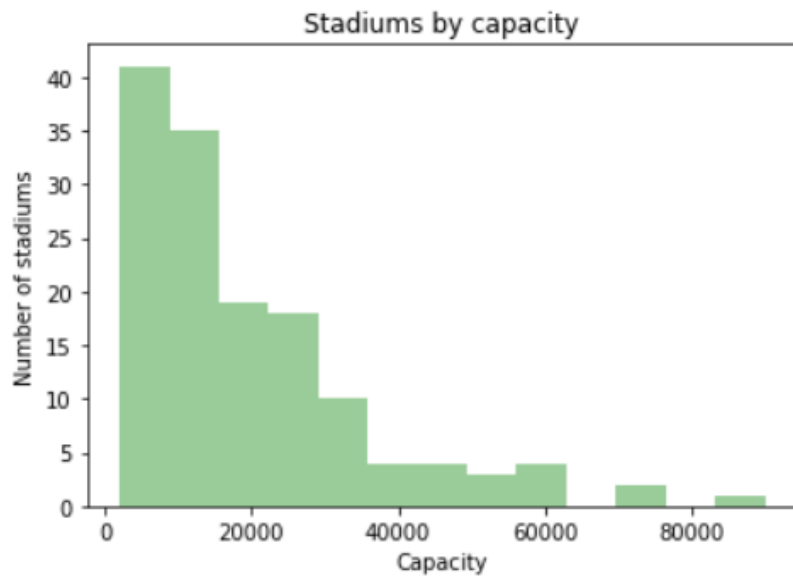

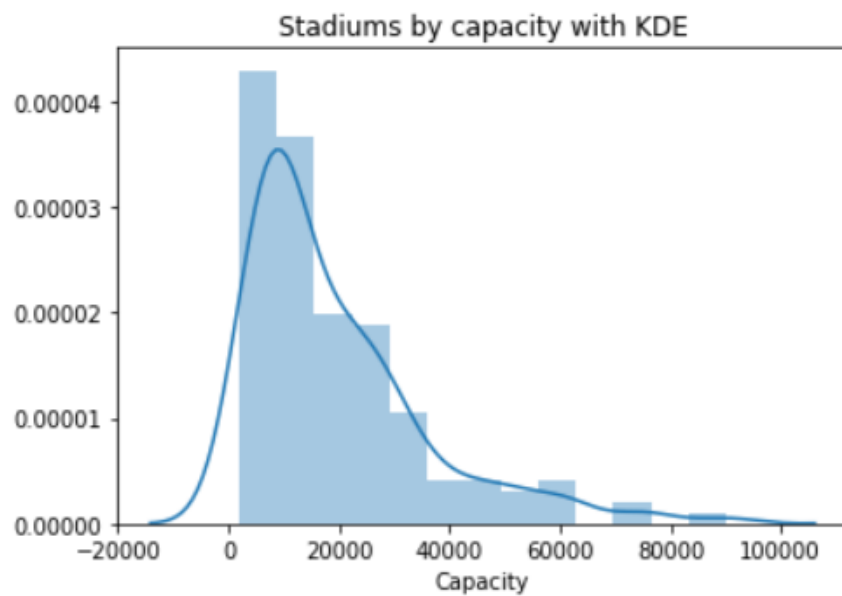*Fig 1. Histogram showing number of stadiums as a function of capacity.*

Fig 2. Histogram with KDE showing number of stadiums as a function of capacity.

3.2   Stadiums north/south divide

Initial visualisation of location data reveals a North/South divide with significantly more stadiums located in the south half of the UK.
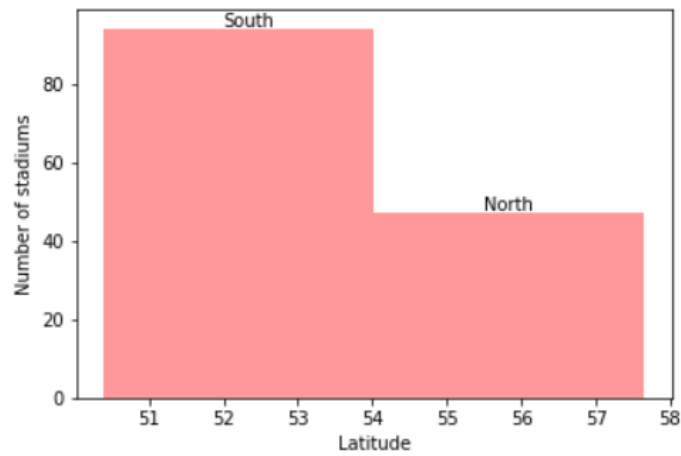

Fig 3. Graph showing the number of stadiums in the south of the UK compared to the north of the UK.

3.3   Stadium locations on UK map

Folium was used to produce a map of the UK showing the stadium locations in addition to the associated team name and stadium capacity. The north/south divide can also be seen here.
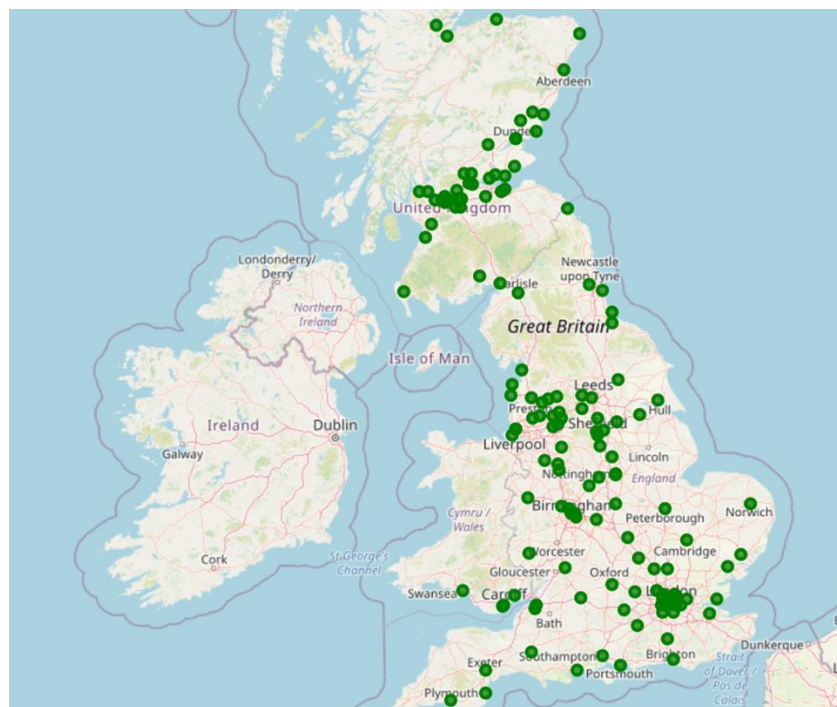

Fig 4. Graph showing the number of stadiums in the south of the UK compared to the north of the UK.

4 Predicting the best locations for new bars/pubs

4.1 Clustering of stadiums using k-means clustering

Foursquare was used to obtain all venues within 500 meters of each football stadium.

Table 3. Dataframe containing stadium data and venues within 500 meters.

```
(1675, 7)
```

	Stadium_name	Stadium Latitude	Stadium Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbey Stadium	52.212799	0.154298	Pipasha	52.213388	0.154665	Indian Restaurant
1	Abbey Stadium	52.212799	0.154298	Abbey Stadium	52.212570	0.154379	Soccer Stadium
2	Abbey Stadium	52.212799	0.154298	Abbey Leisure Complex	52.209225	0.154829	Gym Pool
3	Abbey Stadium	52.212799	0.154298	B&Q	52.211464	0.149038	Hardware Store
4	Abbey Stadium	52.212799	0.154298	Cambridge United FC	52.211833	0.154815	Sports Club

k-means clustering was used to group the stadiums into 5 clusters based on the similarity of their surroundings as shown in figure 5.
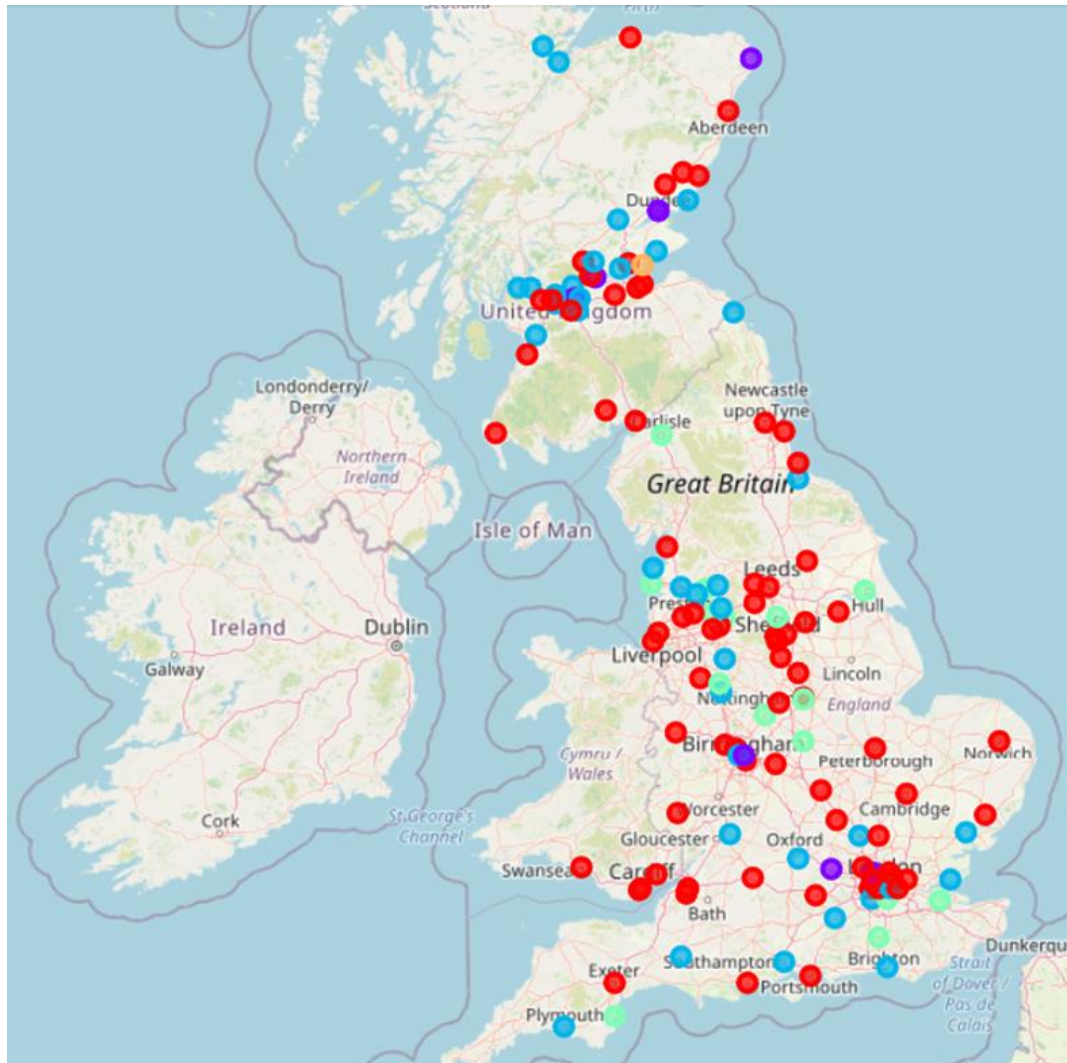

Fig 5. Map showing stadiums split into 5 clusters based on similarity of venues nearby.

4.2   Number of pubs/bars

The dataset was sorted to contain only venues which fell under the categories: Pub, Gastropub or Bar. Figure 6 shows that over 60 stadiums have no pubs/bars within 500 meters and may provide good locations to open a new venue. Figure 7 indicates no significant relationship between number of pubs and stadium capacity within 500 meters.

Table 4. Dataframe containing stadium data and only pubs/bars/gastropubs within 500 meters.

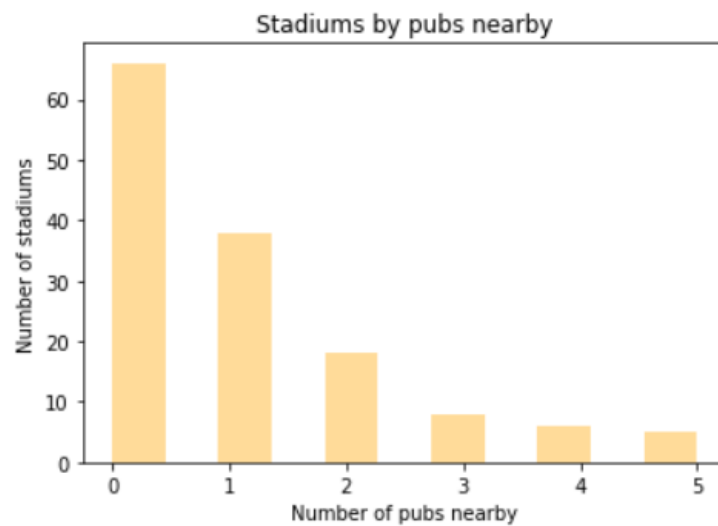|	Stadium_name	Stadium Latitude	Stadium Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
22	Alexandra Stadium	53.087502	-2.43569	The Brockle Bank	53.087550	-2.429237	Gastropub
24	Alexandra Stadium	53.087502	-2.43569	Brocklebank (Brewers Fayre)	53.087405	-2.429179	Pub
27	Alexandra Stadium	53.087502	-2.43569	The Brunswick	53.088780	-2.439393	Pub
44	Almondvale Stadium	55.886398	-3.52207	The Lime Kiln	55.883775	-3.521065	Gastropub
63	Amex Stadium	50.860901	-0.08014	The Swan	50.863851	-0.078795	Gastropub

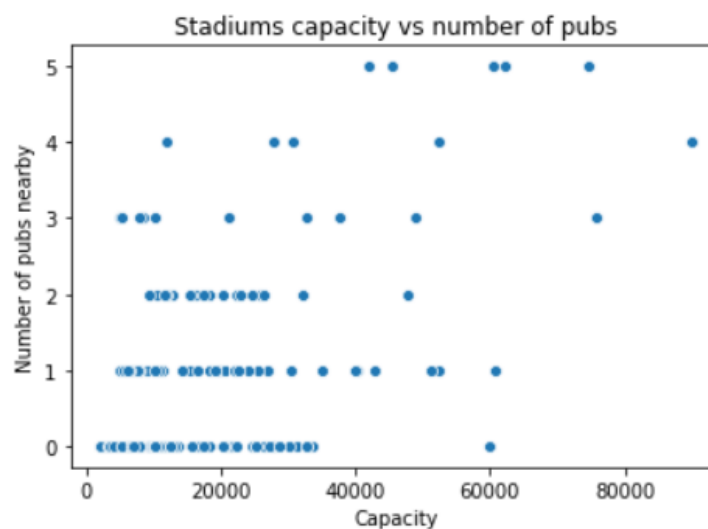*Fig 6. Graph showing the number of stadiums with pubs nearby.*

Fig 7. Graph showing the relationship between stadium capacity and number of pubs.

4.3   Best locations to open venues

Based on the criteria of fewest pubs and largest capacity, 20 stadiums were identified as suitable places to open a new venue. These stadiums are visualised on the UK map in figure 8, and listed in table 5.
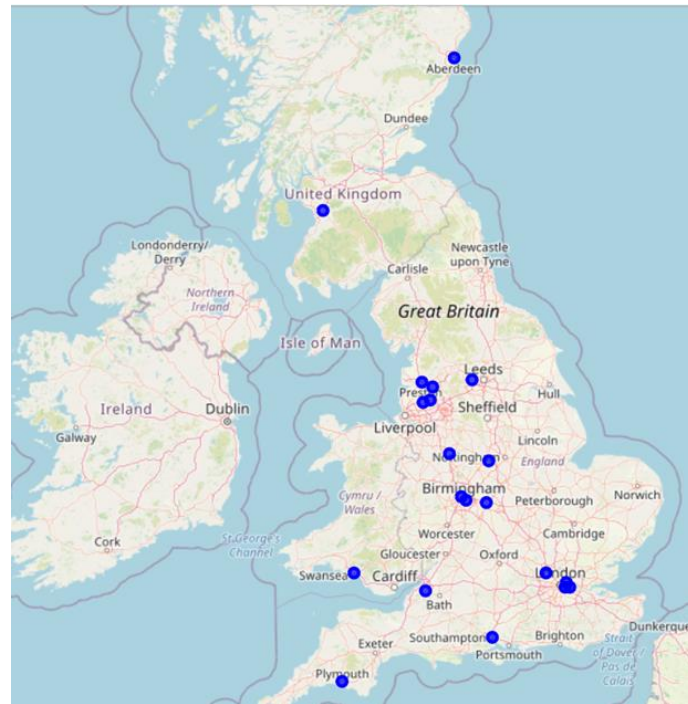

Fig 8. Locations of best stadiums to open a new venue near.

Table 5. List of the 20 best stadiums to open a new bar near.

	Stadium_name	Team_name	Capacity	Latitude	Longitude	Number of pubs nearby
0	Olympic Stadium	West Ham	60000	51.538300	-0.016587	0
1	Pride Park	Derby County	33597	52.914902	-1.447270	0
2	St Mary's Stadium	Southampton	32689	50.905800	-1.391140	0
3	Ricoh Arena	Coventry City	32609	52.448101	-1.495630	0
4	Ewood Park	Blackburn Rovers	31154	53.728600	-2.489370	0
5	St Andrews	Birmingham City	30009	52.475601	-1.868240	0
6	University of Bolton Stadium	Bolton Wanderers	28723	53.580502	-2.535710	0
7	Britannia Stadium	Stoke City	27902	52.988400	-2.175420	0
8	The Valley	Charlton Athletic	27111	51.486500	0.036757	0
9	The Hawthorns	West Bromwich Albion	26272	52.508999	-1.964180	0
10	DW Stadium	Wigan Athletic	25138	53.547699	-2.654150	0
11	Valley Parade	Bradford City	25136	53.804199	-1.759020	0
12	Deepdale	Preston North End	24500	53.771999	-2.688320	0
13	Pittodrie	Aberdeen	22199	57.159302	-2.088720	0
14	Ashton Gate	Bristol City	21497	51.439999	-2.620210	0
15	Liberty Stadium	Swansea City	20532	51.642799	-3.934730	0
16	The Den	Millwall	20146	51.485901	-0.050743	0
17	Rugby Park	Kilmarnock	18128	55.604599	-4.508460	0
18	Vicarage Road	Watford	17477	51.649799	-0.401569	0
19	Home Park	Plymouth Argyle	16388	50.388199	-4.150760	0

5   Conclusions

The best locations to open a new sports bar/pub within close proximity to football stadiums in the UK where identified based on stadium capacity (potential customers) and existing pubs/bars (potential competition) obtained using Foursquare. 20 stadiums with the largest capacities and no existing pubs/bars nearby were identified. K-means clustering was also used to identify stadiums with similar nearby venues, splitting them into 5 clusters.

6   Future directions and criticisms of the model

Current results are based on venues within a 500-meter radius. The model can be expanded to gain a better understanding of the local area close to each stadium and improve the data set. Temporary 'popup' venues such as food trucks which open on football matchdays only do not usually appear in Foursquare data, however they will lead to additional competition.