

“Where to open an Indian restaurant in Ireland?”

Capstone Project for the IBM Data Science course

Tamela Maciel

June 2020



Figure 1: Credit: Joe mon bkk, Wikimedia Commons, CC-BY-SA-4.0

1 Introduction and Business Problem

The UK is famous for its Indian restaurants, and perhaps no where more so than in Leicester, a small city in the heart of the country. Leicester has a large and vibrant Asian community and prides itself on the quality of its authentic curries.

Our client runs several successful Indian restaurants in Leicester and throughout the UK, and is now seeking opportunities to expand into the Republic of Ireland. Ireland’s economy is rapidly growing, especially in the tourism and food industry, and our client sees a gap in the Irish restaurant market for high-quality Indian food.

The question is, where in Ireland should he open his new Indian restaurant? Where will the demand and relative scarcity of existing Indian restaurants be best served by a new venue?

Our client has three criteria which he believes will lead to a successful new Indian restaurant:

1. A city in Ireland that is similar or larger in relative size compared to Leicester.
2. A relative lack of Indian restaurants in that city.
3. A relatively high proportion of residents who are Asian or Irish Asian.

Guided by these criteria, this report analyses a combination of census and venue location data in order to recommend the best location in Ireland for a new Indian restaurant. It clusters and profiles Irish cities based on types of restaurants, and uses this in combination with census data on population and demographics to identify the ideal new location for our client.

2 Data - Acquisition and Cleaning

In order to analyse and profile Irish cities based on the client's criteria for a successful Indian restaurant, data is drawn from a mixture of UK and Irish census data as well as Foursquare venue location data.

2.1 Urban population size

Rank 2016	Change since 2011	Urban Area	County	Population 2016	Increase since 2011
1	-0	Dublin ^[a]	County Dublin	1,173,179 ^[2]	5.6%
2	-0	Cork ^[a]	County Cork	208,669 ^[3]	5.1%
3	-0	Limerick ^[a]	County Limerick & County Clare	94,192 ^[4]	3.0%
4	-0	Galway ^[a]	County Galway	79,934 ^[5]	4.1%
5	-0	Waterford	County Waterford & County Kilkenny	53,504 ^[6]	3.9%
6	-0	Drogheda	County Louth & County Meath	40,956 ^[7]	6.2%
7	▲1	Swords ^{[a][b]}	Fingal	39,248 ^[8]	6.3%
8	▼1	Dundalk ^[a]	County Louth	39,004 ^[9]	3.1%
9	-0	Bray	County Wicklow & County Dublin	32,600 ^[10]	2.3%
10	-0	Navan ^[a]	County Meath	30,173 ^[11]	5.7%
11	▲1	Kilkenny ^{[a][c]}	County Kilkenny	26,512 ^[12]	8.6%
12	▼1	Ennis ^[a]	County Clare	25,276 ^[13]	-0.3%
13	▲1	Carlow ^[a]	County Carlow & County Laois	24,272 ^[14]	5.4%

Figure 2: Top 13 most populous urban areas in Ireland, from the 2016 Irish census. Credit: Wikipedia¹

In order to identify all Irish urban centres that are the same relative size as Leicester or larger, we use population statistics from the **2016 Irish census**, scraped from the Wikipedia table: ‘List of urban areas in the Republic of Ireland by population’¹, and the **2011 UK census**, from the Wikipedia table: ‘List of urban areas in the United Kingdom’². The python package ‘BeautifulSoup’ was used to scrap and filter this data. According to the UK

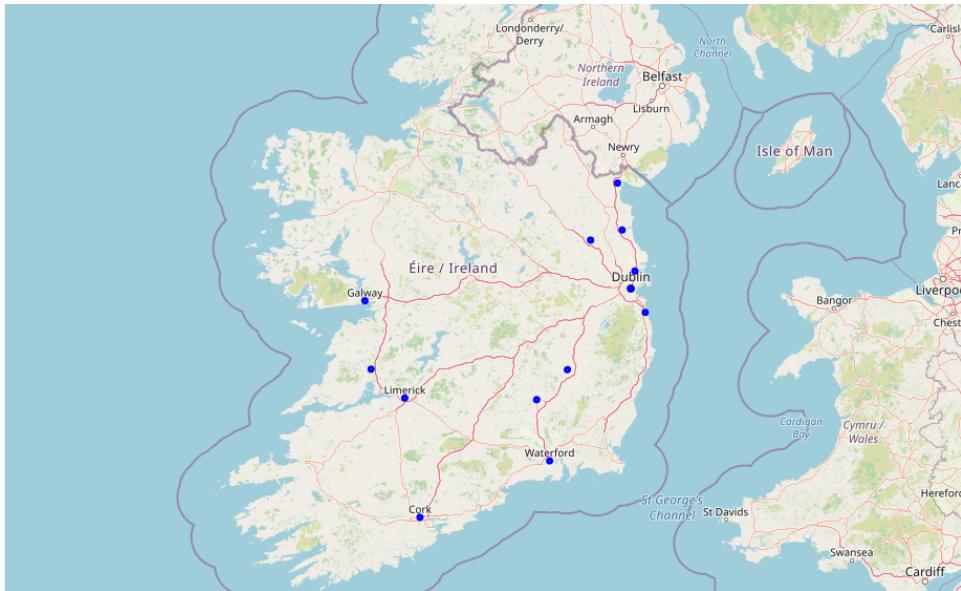


Figure 3: Locations of the top 13 most populous cities in Ireland, from the 2016 Irish census.

census, Leicester is the 13th largest built-up area in the UK, and so in order to match the client’s criteria of a similarly large urban location, we will analyse and profile the 13 largest urban areas in Ireland (shown in Figure 2) in order to determine the best new Indian restaurant location.

The location of each city centre (latitude and longitude) is derived using the python package ‘geopy’ and Nominatum. The locations of the 13 most populous Irish cities are shown in Figure 3.

2.2 Types of restaurants

In order to capture the restaurant profile of a city, we need a large, up-to-date database of location-based venue data.

¹https://en.wikipedia.org/wiki/List_of_urban_areas_in_the_Republic_of_Ireland_by_population

²https://en.wikipedia.org/wiki/List_of_urban_areas_in_the_United_Kingdom

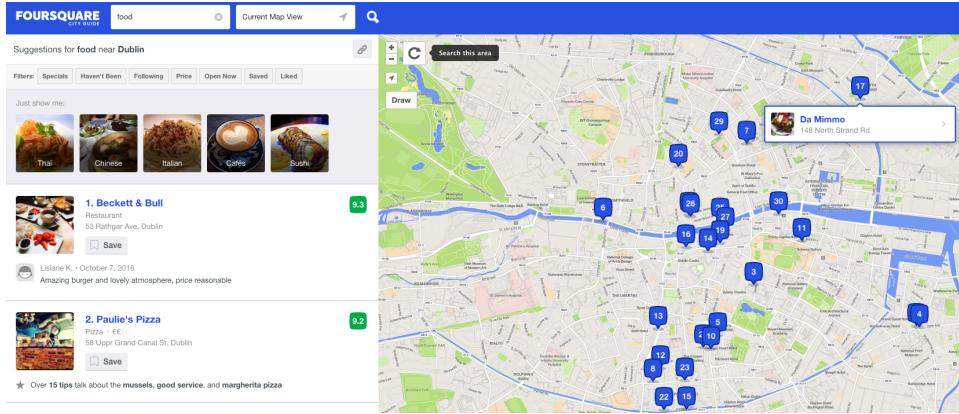


Figure 4: Top restaurants in Dublin. Credit: Foursquare.com

In this analysis, we use the **Foursquare API** to return up to 100 food venues within a 1.5 kilometre radius of the geographical centre of the 13 largest Irish cities.

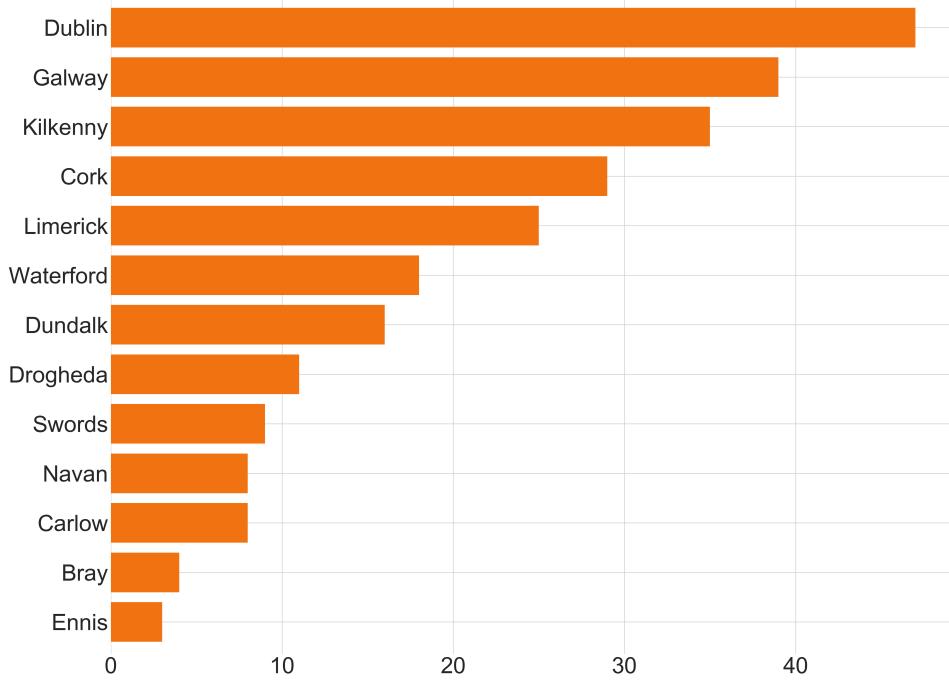


Figure 5: Number of restaurants within 1.5 kilometres of the centres of the 13 largest Irish cities.

We use the ‘explore’ option for Foursquare queries rather than ‘search’,

and return results sorted by popularity, as this returns a better picture of the top restaurants in any given city. We avoid biasing the results by time of day by setting the time and day parameters to ‘any’.

Many of the food venues in a given city are cafes, pizza places, and fast food, and not direct competitors to restaurants. Hence we use the ‘venue category’ parameter from Foursquare to filter out non-restaurants and distinguish between the type of cuisine served (e.g. Indian).

After filtering out non-restaurants, our restaurant data consists of 252 of the most popular restaurants, spread over 13 Irish cities, as shown in Figure 5. This data does not represent the total number of all restaurants in any city, especially for a larger city such as Dublin where many restaurants exist outside the 1.5 kilometre city centre radius. However, because they are ranked by popularity (according to Foursquare), and because only the most popular restaurants within a given radius are returned, this data can be considered representative of the restaurant competition that our client will face in each city.

The number of restaurants per city varies widely between 3 and 47. The cities with the lowest number of restaurants are Ennis (3 restaurants) and Bray (4 restaurants). After manually inspecting both cities via the Foursquare app, it’s clear that these are in fact accurate numbers of restaurants. The vast majority of food venues in both small cities fall into the fast food or cafe category, which are not considered in this analysis. This is true even if the search radius is substantially increased.

While it is not possible to simply gather more restaurants in order to more accurately profile the smaller cities, we bear in mind that it is difficult to directly compare a city the size of Dublin with the much smaller urban areas such as Ennis and Bray. For most of the analysis that follows, we keep the city restaurant data as is. However, for the purposes of clustering (Section 3.2), we’ll drop the cities that have less than 5 restaurants (e.g. Ennis and Bray) in order to have enough venues to build an accurate profile of each city.

Of the 252 restaurants, 7.1% are Indian restaurants. The purpose of this study to identify the regional variations in this percentage and where the market gaps are.

2.3 Asian residents

For a new restaurant to be a success, it should ideally locate itself in a place where there is a market gap, but also an existing base of future customers. In this analysis, we assume that areas with higher populations of Asian residents will be the most receptive to the authentic Indian cuisine that our client prides himself on.

The **2016 Irish census** includes ethnicity demographics grouped into nine categories. While ethnicity data is not available on the city level, eth-

nicity by the 31 administrative counties is recorded by the Central Statistics Office in Table E8001: Irish Travellers Ethnicity and Religion³. Overall, Ireland has a predominately white population, however the most common non-white ethnicity is Asian at 1.7% as shown in Figure 6.

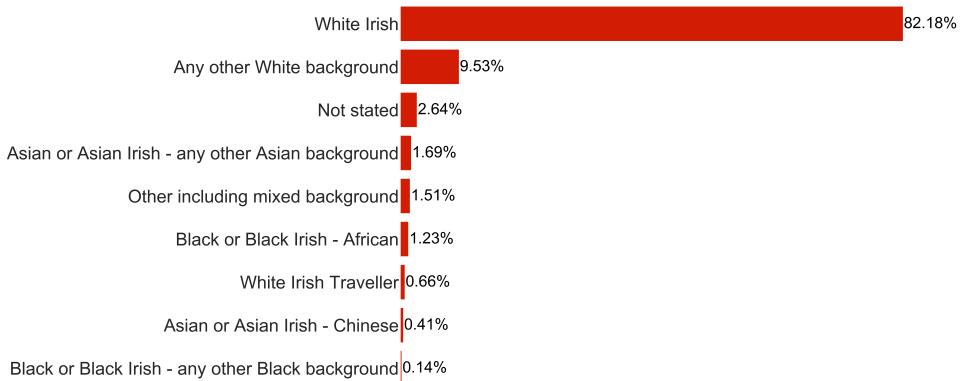


Figure 6: Percentage of Irish residents by ethnic or cultural background, from the 2016 Irish census. Data from the Central Statistics Office.³

In order to determine the percentage of Asian residents by county, we also download the total population per administrative county from the 2016 Irish census, from Central Statistics Table E2004: Population Distribution and Movement⁴.

This county-level data is used to rank counties by their relative proportion of ‘Asian or Asian Irish - any other Asian background’ communities (see Figure 8 in the next section). Given the fact that Irish counties are relatively small, this ethnicity data will be used to distinguish between and recommend cities that may otherwise have similar restaurant profiles.

In order to accurately link our 13 cities to the administrative county in which they reside, we also need the county boundaries as a geographic file. Boundaries of the 31 administrative counties used in the 2016 Irish census can be downloaded from data.gov.ie⁵. Here we use the 20 metre resolution geojson file for county boundaries, as shown in Figure 7.



Figure 7: The boundaries of the administrative counties used in the 2016 Irish census are shown in dark green, with the cities shown as blue circles. As an example, Swords lies in the geographic county of Dublin, but in the administrative county of Fingal County Council.

3 Methodology - Data Analysis and Clustering

In this study we want to identify the best place in Ireland in which to open an Indian restaurant, by maximising the urban population, especially with regards to Asian residents, and minimising the existing number of Indian restaurants. Our analysis is limited to the Irish cities that have the same relative population as Leicester or larger, and to an area within 1.5 kilometres radius of each city centre.

The first step was data acquisition and cleaning, as described in Section 2. Using the most recent UK and Irish census data, the 13 largest Irish cities by population were selected for this study, in order to match the relative population of Leicester in UK. We then collected **Irish population and ethnicity data** from the 2016 Irish census, using the data portal on the Central Statistics Office webpage.³ Two tables on county population and ethnicity were combined and cleaned, and matched to the city locations using a geographic file of the boundaries of counties. We also collected **restaurant data (frequency, location, and type)** for each of the 13 Irish cities using the Foursquare API. All of this data was cleaned and wrangled into a final single dataset of the 13 most populous Irish cities, with 17 key features ranging from population, demographics, and restaurant data.

The second step is exploratory data analysis, in which we analyse and visualise trends in both the demographic and restaurant raw data. This is described further in Section 3.1 below. This allows us to short-list the cities

³<https://statbank.cso.ie/px/pxeirestat/Statire>SelectVarVal/Define.asp?maintable=E8001&PLanguage=0>

⁴<https://statbank.cso.ie/px/pxeirestat/Statire>SelectVarVal/Define.asp?Maintable=E2004&PLanguage=0>

⁵<https://data.gov.ie/dataset/administrative-areas-osi-national-statutory-newlineboundaries-generalised-20m>

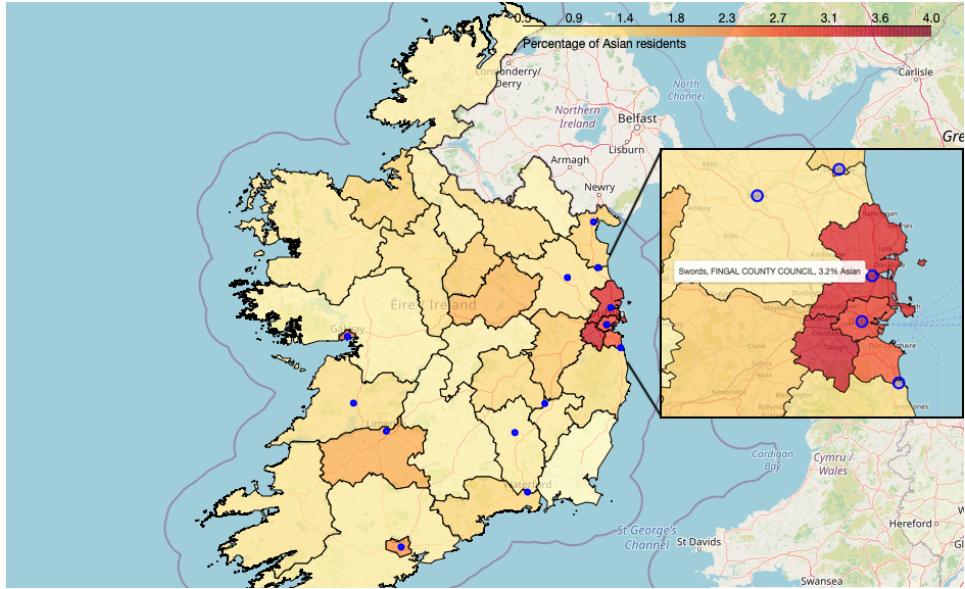


Figure 8: The percentage of Asian residents by administrative county, with the 13 most populous cities shown as blue circles.

that are the most promising locations for an Indian restaurant, according to the client’s criteria.

In the third and final step, we use the frequency of different types of restaurants per city in order to cluster these cities into distinct groups (using k-means clustering). This allows us to build up a profile of each Irish city, and further narrow our recommendations of the ideal city location for an Indian restaurant. Further details are described in Section 3.2 below.

The results of this analysis are summarised and discussed in Sections 4 and 5.

3.1 Exploratory data analysis

As a first step, we matched county-level ethnicity data to the locations of the 13 Irish cities in this analysis, using the county boundaries and the geographic python package ‘Shapely’. This allows us to visualise both the city locations and the proportion of Asian (non-Chinese) residents per county, as a choropleth map shown in Figure 8. From this map, it is clear that the highest percentage of Asian residents are located around Dublin city, shown in dark red. This is coloured relative to other counties; the actual percentages (2.5-3.5%) are quite low.

When we only consider the counties where our 13 cities are located, **Swords, Dublin, Galway, Cork, and Limerick have the top five**

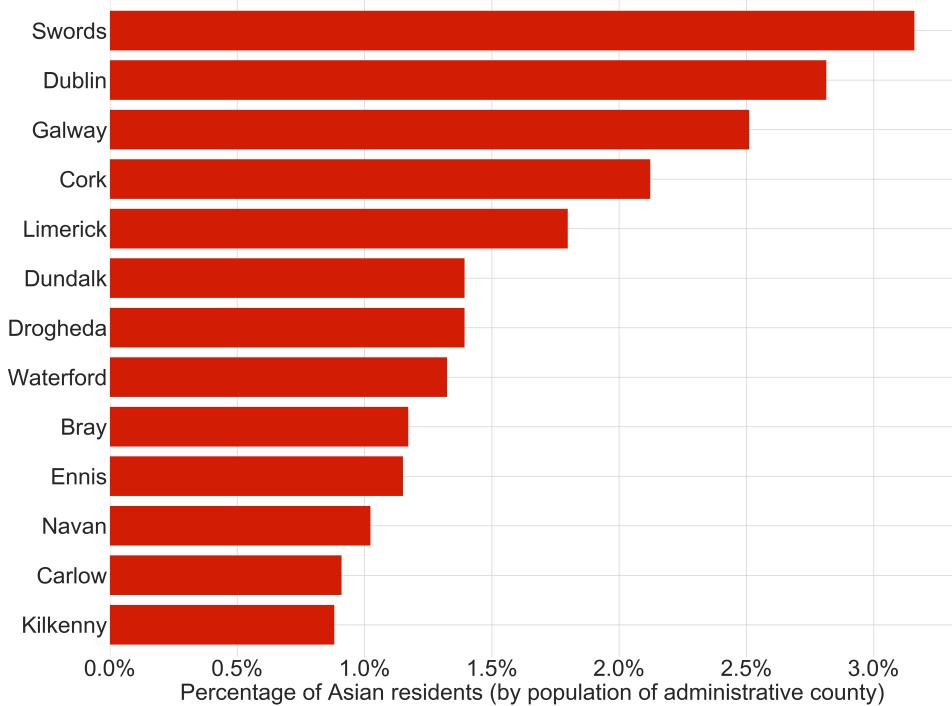


Figure 9: The 13 most populous cities, ranked by the percentage of Asian residents within their administrative county.

highest percentages of Asian residents, as shown in Figure 9. As an initial recommendation, ignoring the restaurant data, these five cities are strong contenders for a new Indian restaurant location, as they satisfy both criteria of a high urban population and a high proportion of Asian residents.

Restaurant data for each city comes from the Foursquare API, as described in Section 2.2. This data includes the frequency, location, and type (category) of restaurants within 1.5 kilometres of each city centre. Food venues that are not restaurants (e.g. fast food and cafe) are filtered out as they are not direct competitors to a new Indian restaurant. The category of restaurants will be used to cluster and profile each city in the next section, but first we explore some basic trends in the data.

Let's consider the percentage of Indian restaurants in each city, as this represents the direct competition for our client's cuisine and customer market. Figure 10 ranks the 13 cities by the percentage of Indian restaurants within 1.5 kilometres of their city centres.

Our client hopes to identify a gap in the Indian restaurant market, so lower percentages of Indian restaurants are more favourable. Ennis has no Indian restaurants, but this is perhaps unsurprising given that it has only

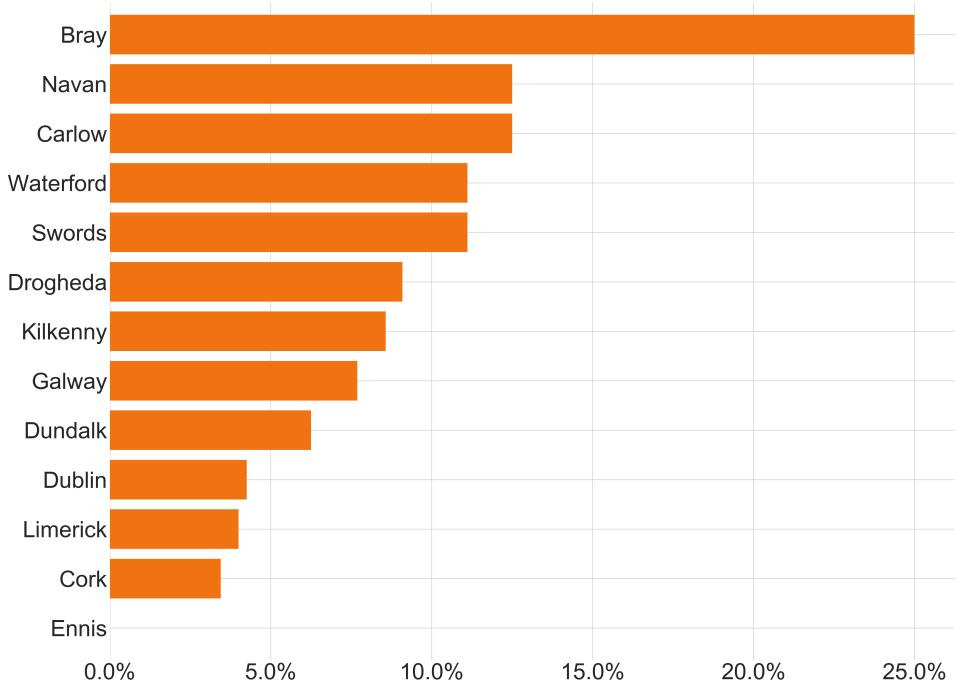


Figure 10: The 13 most populous cities, ranked by the percentage of Indian restaurants within 1.5 kilometres of their city centres.

3 restaurants in total. However **Cork, Limerick, and Dublin also have very few Indian restaurants and they are much larger cities**. These are potential good candidates for a new Indian restaurant.

On the opposite end of the spectrum, Bray has the largest percentage of Indian restaurants, at 25%. However this should be taken with a pinch of salt given that Bray only has 4 restaurants in total.

Now let's combine the data on Asian residents and the percentage of Indian restaurants in order to identify the cities that have a high proportion of Asian residents *and* low proportion of Indian restaurants. In Figure 11, the 13 cities are plotted according to their percentage of Asian residents and Indian restaurants.

The data is divided into quarters and the cities within the quarter that represents a high proportion of Asian residents for a low proportion of Indian restaurants are highlighted in the red dashed box.

We now begin to have a **shortlist of cities that meets the client's criteria for a relatively high population of Asian residents, and relatively low proportion of Indian restaurants**.

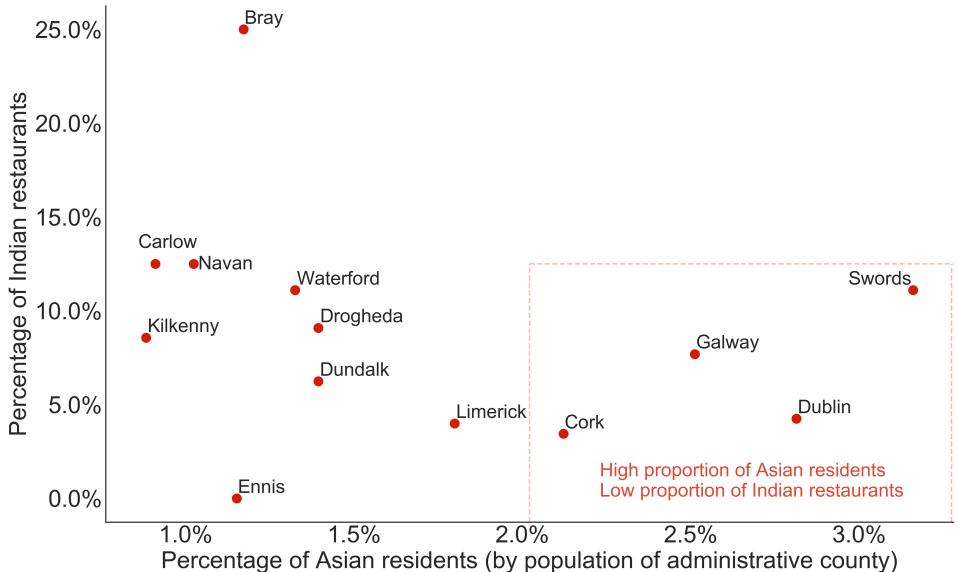


Figure 11: The 13 most populous cities, ranked by their percentage of Asian residents and Indian restaurants. Cities that match the client's criteria are highlighted in the dashed box.

These cities are (by decreasing proportion of Asian residents):

- Swords
- Dublin
- Galway
- Cork

While this shortlist is certainly indicative, we are still lacking knowledge of the restaurant profile of each city, in order to determine what type of restaurants are most popular and successful. This is achieved by clustering the cities according to their restaurants, as described in the next section.

3.2 City clustering by restaurants

In order to get a better sense of the restaurant environment of each city, we can compute the frequency of each type of restaurant in a city (e.g. Chinese, Indian, Italian, Burger restaurants), and use a clustering algorithm to group similar cities together.

Clustering algorithms are a type of unsupervised machine learning, where the data is unlabelled, and the algorithm groups data points based on their

similarity, or proximity, to other data points. It can work across many features (dimensions), and helps provide insight into data that is otherwise difficult to visualise or categorise. The final labels for each cluster require manual inspection for trends and similarities.

For each of the 13 most populous cities, we have the frequencies of 30 different types of restaurants, from ‘American’ to ‘Vietnamese’ restaurants. This unlabelled, high-dimension data is ideal for a clustering algorithm in order to group similar restaurant environments together and build up a profile of each city for our client. However, as discussed in Section 2.2, there is a big range in the total number of restaurants per city. In order to profile cities appropriately, we should have minimum number of features, and hence a minimum number of restaurants, per city. As a result, for the clustering analysis, we drop the cities with less than 5 restaurants (i.e. Bray and Ennis), leaving 11 cities for clustering. The five most common types of restaurants for each of these 11 cities are shown in Figure 12.

city	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0 Carlow	Asian Restaurant	Thai Restaurant	Tapas Restaurant	Chinese Restaurant	Restaurant
1 Cork	Restaurant	Burger Joint	Chinese Restaurant	Mediterranean Restaurant	Mexican Restaurant
2 Drogheda	Restaurant	Italian Restaurant	Mediterranean Restaurant	Diner	Indian Restaurant
3 Dublin	Restaurant	Italian Restaurant	Sushi Restaurant	Burger Joint	French Restaurant
4 Dundalk	Chinese Restaurant	Diner	Italian Restaurant	Restaurant	Malay Restaurant
5 Galway	Restaurant	French Restaurant	Indian Restaurant	American Restaurant	Seafood Restaurant
6 Kilkenny	Restaurant	Italian Restaurant	Gastropub	Steakhouse	Chinese Restaurant
7 Limerick	Restaurant	Italian Restaurant	Steakhouse	Asian Restaurant	Seafood Restaurant
8 Navan	Chinese Restaurant	Restaurant	Asian Restaurant	Diner	Gastropub
9 Swords	Italian Restaurant	French Restaurant	Tapas Restaurant	Sushi Restaurant	Steakhouse
10 Waterford	Restaurant	Chinese Restaurant	Steakhouse	Italian Restaurant	Indian Restaurant

Figure 12: The top five most common types of restaurants for 11 Irish cities.

K-means is a simple clustering algorithm which groups data according to a pre-set number of clusters (k), and minimises the total sum of squared distances from each data point to the cluster centre. In this analysis, we use the K-means algorithm⁶ from the python ‘scikit-learn’ package, and cycle through a range of k values in order to determine the most appropriate number of clusters. This is done by plotting a measure of cluster density (‘inertia’) against k , and selecting the k value where the graph ‘bends’ (known as the elbow-method). We use the ‘full’ algorithm and the ‘k-means++’ method to initialise our cluster centres.

Using this approach, we clustered the 11 cities by their frequencies of restaurant types, and determined that 4 clusters was the most appropriate grouping (due to a subtle but distinct bend in the slope of cluster inertia at $k=4$). After inspecting the cities and types of restaurants in each cluster,

⁶<https://scikit-learn.org/stable/modules/clustering.html>

distinct trends make it possible to give these clusters profile descriptions as shown in Figure 13. The 11 cities, coloured by cluster type, are mapped in Figure 14.

Cluster	No. of cities	Cities	Profile description
1	1	Drogheda	Drogheda, popular generic and Mediterranean restaurants
2	1	Swords	Swords, popular Italian restaurants
3	3	Carlow, Navan, Dundalk	Mid-size cities, popular Chinese-Asian restaurants
4	6	Dublin, Cork, Galway, Limerick, Waterford, Kilkenny	Larger cities, popular generic 'restaurants'

Figure 13: Table and description of the 4 clusters of Irish city restaurants.



Figure 14: 11 cities, clustered by frequency of restaurant type, and coloured by cluster label (1-4).

These clusters make some sense. The majority of cities (6) have larger populations with popular generic 'restaurants'. Three cities, shown in dark green in Figure 14 are smaller cities located in the rural areas around Dublin with fewer restaurants and popular Chinese-Asian restaurants (perhaps the

takeaway market). The final two clusters contain one city each. Swords is unique for its prevalence of Italian restaurants and Drogheda has a prevalence of both generic and Mediterranean restaurants. Both are located just north of Dublin.

These cluster profiles, combined with our previous analysis of city ethnicity and prevalence of Indian restaurants, allow us to recommend with some certainty the top Irish cities for a new Indian restaurant. This final result is described and discussed in the next section.

4 Results

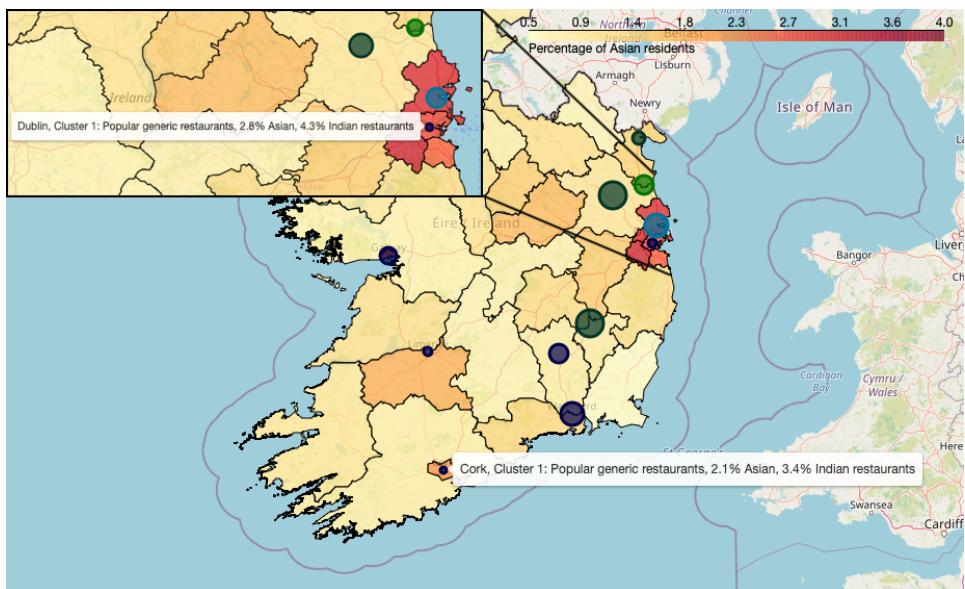


Figure 15: Choropleth of the percentage of Asian per administrative county. 11 cities are coloured by restaurant cluster and the size of the city circle indicates the percentage of Indian restaurants.

Figure 15 summarises the result of our analysis and clustering. It shows a choropleth of the administrative counties of Ireland, coloured by the percentage of Asian (non-Chinese) residents in that county. It also shows the final 11 cities included for analysis as coloured circles, where the size of the circle indicates the percentage of Indian restaurants in that city and the colour of the circle indicates which cluster the city belongs to, based on the types of restaurants that are most common. An interactive version of this map is available online on the project’s Github repository⁷.

⁷<https://github.com/tamelamaciel/capstone-IBM-Data-Science>

The cluster profiles combined with knowledge of the proportion of Asian residents in each city allows us to clearly identify those cities with a current lack of Indian restaurants but a relatively high Asian population.

Our top three recommendations for where to locate a new Indian restaurant in Ireland are:

- **Dublin:** 2.8% Asian population, only 2 Indian restaurants out of 47 (4.3%), and no Indian restaurants in its most common 5 restaurants
- **Swords:** 3.2% Asian population, only 1 Indian restaurant out of 9 (11.1%), and no Indian restaurants in its most common 5 restaurants
- **Cork:** 2.1% Asian population, only 1 Indian restaurant out of 29 (3.4%), and no Indian restaurants in its most common 5 restaurants

In addition, there are two other cities that have good potential and would be worth considering by the client:

- **Galway:** 2.5% Asian population and only 3 Indian restaurants out of 39 (7.7%). However, its 3rd most common restaurant is an Indian restaurant, unlike our top three cities.
- **Limerick:** 1.8% Asian population (which is low but not far off of Cork), only 1 Indian restaurant out of 25 (4%), and no Indian restaurants in its most common 5 restaurants

5 Discussion

In this study we analysed population, ethnicity, and restaurant location data in order to shortlist the best locations in Ireland for an Indian restaurant. The result of both traditional data analysis and machine learning clustering leads to three clear top recommendations for where to locate a new Indian restaurant: Dublin, Swords, and Cork. There were two strong runners-up: Galway and Limerick.

5.1 Data completeness and sensitivity

While we believe that both the data and conclusions are robust, there are some potential uncertainties to acknowledge. The Foursquare restaurant data was necessarily limited by number and search radius, in order to return a consistent result for each city. However, this means that for the larger cities (such as Dublin and Cork), many restaurants will have been missed either because they're outside the search radius of 1.5 kilometres, or because they are not as popular, and so were not included in the top Foursquare results. These missing restaurants are unlikely to skew the overall results, however

they may be potential competition depending on the client's choice of city and neighbourhood, and should be analysed further before a final decision is made.

The restaurant data for the smaller cities such as Carlow, Navan, and Bray is sensitive to low statistics. They have high percentages of Indian restaurants but this due to a single Indian restaurant each, in cities that have less than 10 restaurants in total. Should the client wish to locate in a smaller Irish city, we recommend that the percentage of Asian residents be the deciding factor, and not the percentage of Indian restaurants, which can be misleading.

Finally, clustering can be sensitive to the number and type of features included in the algorithm. In this analysis, we only clustered cities based on the frequency of their different types of restaurants. However, in further investigation it would be interesting to include population and ethnicity demographics in the clustering algorithm. This has the potential to further differentiate between cities in our largest cluster of popular 'generic restaurants', and may lead to clearer clustering with lower errors.

5.2 Further investigations

Our top recommendations should be viewed as the starting point for further investigation by the client.

An additional key factor to investigate further is the price of commercial properties (either sale or lease). This may well differentiate our recommended cities and help the client decide, based on their budget. Dublin and Swords are well-known to have much more expensive property prices, compared to anywhere else in the country. However, on the flip-side, Dublin and Swords have higher densities of potential customers with larger disposable incomes.

Finally, once an ideal city is selected, further analysis must be done to select the ideal neighbourhood, and even street address, based on footfall, property prices, and demographics of the area. This is outside the scope of this work, but a similar approach can be applied to neighbourhoods just as easily as to cities.

For a given neighbourhood, it would also be helpful to know if previous Indian restaurants have ever opened and then failed, and why. This data would require records of historic businesses, or strong local knowledge, and would be more difficult to extract, but could prove to be quite indicative of the best area (or not) to locate a new Indian restaurant in Ireland.

6 Conclusion

This study using a combination of demographic and venue location data in order to recommend where in Ireland to open a new Indian restaurant. Based

on the client's criteria of what makes a successful restaurant, we maximised both population and proportion of Asian (non-Chinese) residents in each Irish city, in order to guarantee a large existing base of future customers, and we minimised the proportion of existing Indian restaurants, in order to identify a gap in the restaurant market.

Using population and demographic data from the 2016 Irish census, we analysed and ranked the 13 largest Irish cities by percentage of Asian residents. Using restaurant location data from Foursquare for each city, we clustered and profiled each city based on their most common types of restaurants. **The combination of these two analyses led to three clear top recommendations for where to locate a new Indian restaurant: Dublin, Swords, and Cork.** There were two strong runners-up: Galway and Limerick.

Our analysis and recommendations only include demographics and a profile of restaurants for each Irish city and should be considered as a starting point for further investigation by the client. Property prices and neighbourhood-level demographics and restaurant profiles should also be included in a final decision of where to locate a new Indian restaurant in Ireland.