Martin Liwosz

Coursera.org  IBM Data Science Professional CAPSTONE

Franchise Competitive Analysis Your VISUAL COMPANION to Competition USING MACHINE LEARNING

Contents

[1. Introduction 2](#_Toc14776677)

[1.1 Business problem 2](#_Toc14776678)

[1.2 Business interest 2](#_Toc14776679)

[1.3 Background 3](#_Toc14776680)

[2. Data 3](#_Toc14776681)

[2.1 Description 3](#_Toc14776682)

[2.2 Source 4](#_Toc14776683)

[3. Methodology 5](#_Toc14776684)

[3.1 Exploratory data analysis 5](#_Toc14776685)

[3.2 Inferential statistical testing 6](#_Toc14776686)

[3.3 Machine learning 7](#_Toc14776687)

[4. Results 8](#_Toc14776688)

[4.1 Outcome 8](#_Toc14776689)

[5. Discussion 10](#_Toc14776690)

[5.1 Observations 10](#_Toc14776691)

[5.2 Recommendation 10](#_Toc14776692)

[6. Conclusion 10](#_Toc14776693)

## Introduction

### 1.1 Business problem

Should you start a business form scratch or simply buy a franchise? If you opt to NOT start a business from scratch and choose a franchise;

What’s involved?

Where do you start?

What’s the competitive landscape?

That’s the biggest problem when choosing this career path. Evaluating the competitive landscape and it’s this problem that we will look to address in more detail; as well as provide you further insights into more of the background of exactly where you should setup your franchise.

### 1.2 Business interest

Anyone who might be considering going into the franchise model would find this of interest. This report outlines the 7 key areas to be particularly engaged in If you're looking to start a business, buying into a franchise because purchasing a franchise can be a good alternative to starting a unique venture.

Alternatively if you're planning to expand your business franchising could also be an effective way of moving into new markets as the key to franchising is that it allows one business to operate under the trading name of another business' established brand and sell its products and/or services for a specified period.

Therefore in knowing this the next step and where this report particularly focuses on is the ability to search for yourself and particular business within a given area to see your competitors on a geographic map.

Once that’s done we will sue machine learning algorithms to clearly illustrate for you which of these businesses are of particular competitive concern due to proximity of the search area and relation of one another.

### 1.3 Background

Let’s start at the beginning by discussing the background and specifically answer What’s involved?

Well these are the 7 things you need to consider when choosing to franchise.

1. Start with learning everything you can and get professional advice before you sign up to buy a franchise. Begin with free online courses, and it’s important to ask questions and make sure you understand how the franchise model could affect the way you do business.
2. Agreements. The first could be the franchise agreement which is the contract you agree to for a set amount of time, which on average can be five years. It covers exactly where and how you will run your franchise and worth consulting a professional to make sure you understand your rights and responsibilities under every clause.
3. Disclosure statement. Franchising in Australia is regulated by the Franchising Code of Conduct. Before entering a franchising arrangement, you should be given certain documents including your franchise agreement, code of conduct and disclosure statement. It’s worth contacting more than one franchisee and even past franchisees for their experiences.
4. Finances. Running a business will always come with risks, especially in relation to factors out of your control, such as competition and the economy.
5. Restraint of trade clause. This stops you competing with the franchise itself during the term of your franchise agreement and after the agreement ends.
6. Fees. In most franchise business arrangements, the franchisee could have to pay a royalty fee to the franchisor on a weekly, monthly or yearly basis. You need to know how this royalty fee works as well as other fees on top of these.
7. Understanding your territory. Is last but not least important because your territory is the area you’re allowed to serve customers within. Increased competition might result if you haven’t selected your target territory well.

## Data

### 2.1 Description

Addressing the competitive landscape proves to be the hardest thing to address because there isn’t the ability to simply search for the 50 of a whatever it is you want to search for in a given area. Even Google maps can’t easily achieve this. However, with this tool you can define the area, search criteria, and radius in meters around the given area.

The data requirements to achieve this include:

Inputs from the user:

1. Area anywhere on the map
2. Search term in the case of the default notebook it uses Mc Donald’s
3. Radius in meters around the area up to 100, 000 meters

### 2.2 Source

The outputs as a result are those the Foursquare API which include:

{ "meta":

{ "code": 200, "requestId": "5ac51d7e6a607143d811cecb" },

"response":

{ "venues": [

{ "id": "5642aef9498e51025cf4a7a5",

"name": "Mr. Purple",

"location":

{ "address": "180 Orchard St",

"crossStreet": "btwn Houston & Stanton St",

"lat": 40.72173744277209,

"lng": -73.98800687282996,

"labeledLatLngs": [

{ "label": "display",

"lat": 40.72173744277209,

"lng": -73.98800687282996 } ],

"distance": 8,

"postalCode": "10002",

"cc": "US",

"city": "New York",

"state": "NY",

"country": "United States",

"formattedAddress": [ "180 Orchard St (btwn Houston & Stanton St)", "New York, NY 10002", "United States" ] },

"categories": [ {

"id": "4bf58dd8d48988d1d5941735",

"name": "Hotel Bar", "pluralName": "Hotel Bars",

"shortName": "Hotel Bar",

"icon": { "prefix": "https://ss3.4sqi.net/img/categories\_v2/travel/hotel\_bar\_", "suffix": ".png" }, "primary": **true** } ], "venuePage": { "id": "150747252" } } ] } }

The results of which are combined into producing the final interactive maps and machine learning analytics.

## Methodology

### 3.1 Exploratory data analysis

There are several requirements that are used as part of the delivery of this analysis. Namely, I start with several libraries which include: requests, pandas, numpy, random geopy.geocoders, Ipython.display and .core.display, we also use the json\_normaise function from pandas, io.json, folium, matplotlib and lastly but not least I’ve used sklearn.cluster for using Kmeans.

In terms of the methodology another requirement I’ve used is a GitHub repository to ensure that I’m able to share said results with you and others in similar positions or using this notebook for what it was intended. Upon the establishment of the above libraries then the exploratory data analysis can take place.

The original concept of this report was to find and recommend where next to put a Mc Donald’s franchise, however it quickly became clear that the functionality available in this notebook allowed the search of any term given the input. Therefore, the decision was made to extend the core of the notebook to capture inputs form the user to best server their needs and target a larger user base.

Given that the following inputs are correctly entered and knowing that there is an external limitation then the results allow any user the ability to find 50 of any search term around the target area.

print ('What version of the API would you like to run in YYYYMMDD format?')

VERSION = input ()

print ("Where would you like to search?")

address = input ()

print('What would you like to search for')

search\_query = input ()

print('How far away in meters woud you like the radius to be?')

radius = input ()

From there the next step in the methodology is to take the result of the Foursquare API to explore the area and search term with the above given search parameters.

url = 'https://api.foursquare.com/v2/venues/search?client\_id=**{}**&client\_secret=**{}**&ll=**{}**,**{}**&v=**{}**&query=**{}**&limit=**{}**&radius=**{}**'.format(CLIENT\_ID, CLIENT\_SECRET, latitude, longitude, VERSION, search\_query,LIMIT,radius)

The results of the above API call takes the relevant part of the JSON response and using the json\_normalise function transfers the data into a Pandas data frame.

results = requests.get(url).json()

*#print (results)*

*# assign relevant part of JSON to venues*

venues = results['response']['venues']

*# tranform venues into a dataframe*

dataframe = json\_normalize(venues)

print (dataframe)

Once these results are within a data frame then we are able to use the following code to transpose the datapoints onto a map.

maccas\_map = folium.Map(location=[latitude, longitude], zoom\_start=13) *# generate map centred around the centre location*

*# add a red circle marker to represent the Conrad Hotel*

folium.features.CircleMarker(

[latitude, longitude],

radius=10,

color='blue',

popup='Centre',

fill = **True**,

fill\_color = 'blue',

fill\_opacity = 0.6

).add\_to(maccas\_map)

*# add the Maccas restaurants as blue circle markers*

dataframeSummary = dataframe[['location.lat','location.lng', 'location.city']]

*#print(dataframeSummary.dtypes)*

*#print (dataframeSummary.head())*

**for** index, row **in** dataframeSummary.iterrows():

folium.features.CircleMarker(

[row['location.lat'], row['location.lng']],

radius=5,

color='red',

fill = **True**,

fill\_color='red',

fill\_opacity=0.6

).add\_to(maccas\_map)

*# display map*

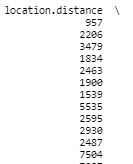
maccas\_map

### 3.2 Inferential statistical testing

Inferential statistical testing refers to a random sample of data taken from a population to describe and make inferences about the population. In the case of this report we took a random sample and with the use of pandas we determined two key items.

1. Within the API response itself there is a field that’s retrieved namely location.distance that when analysed with the random sample could be further analysed.

Example of data set



1. Point two will be discussed in the subsequent section.

### 3.3 Machine learning

Machine learning (ML) is defined as an application of artificial intelligence that provides systems the ability to automatically lean and improve without being explicitly programmed to do so. Therefore, in having defined what ML is I’m able to begin to discuss what method of ML was used and why.

I used a type of ML that is unsupervised and called K-means clustering. This requires a set of unlabelled data and the goal of the algorithm is to find groups in the data.

Therefore with the use of ML and in particular K-means we’ll be able to determine the closest competitors. From there you could then define your territory and the boundaries of that to a level where you are able to determine your outer limits.

*# set number of clusters*

kclusters = 3

*# run k-means clustering*

kmeans = KMeans(n\_clusters=kclusters, random\_state=0).fit(dataframeSummary)

*# check cluster labels generated for each row in the dataframe*

kmeans.labels\_[0:10]

*# add clustering labels*

*#dataframeSummary.insert(0, 'Cluster Labels', kmeans.labels\_)*

dataframe\_merged = dataframe[['location.city']]

*#print (dataframe.head())*

print (dataframeSummary.head())

print (kmeans)

*# merge grouped with data to add latitude/longitude for each*

dataframe\_merged = pd.concat([dataframe\_merged, dataframeSummary], axis=1, ignore\_index=**False**)

dataframe\_merged.head() *# check the last columns!*

## Results

### 4.1 Outcome

Your business will always have competition and if you don’t know where your competition is, what the competition is up to, you could be missing out on huge opportunities that’s why competitive analysis is so crucial to your success as a business.

It gives you the tools to quickly adapt to any changes in the competitive landscape and potentially capitalise on industry trends that your competitors haven’t even noticed hence the need of this tool and others like it.

So given the below code:

print ('What version of the API would you like to run in YYYYMMDD format?')

VERSION = input ()

print ("Where would you like to search?")

address = input ()

print('What would you like to search for')

search\_query = input ()

print('How far away in meters woud you like the radius to be?')

radius = input ()

The following was input:

What version of the API would you like to run in YYYYMMDD format?

20190701

Where would you like to search?

Sydney, Australia

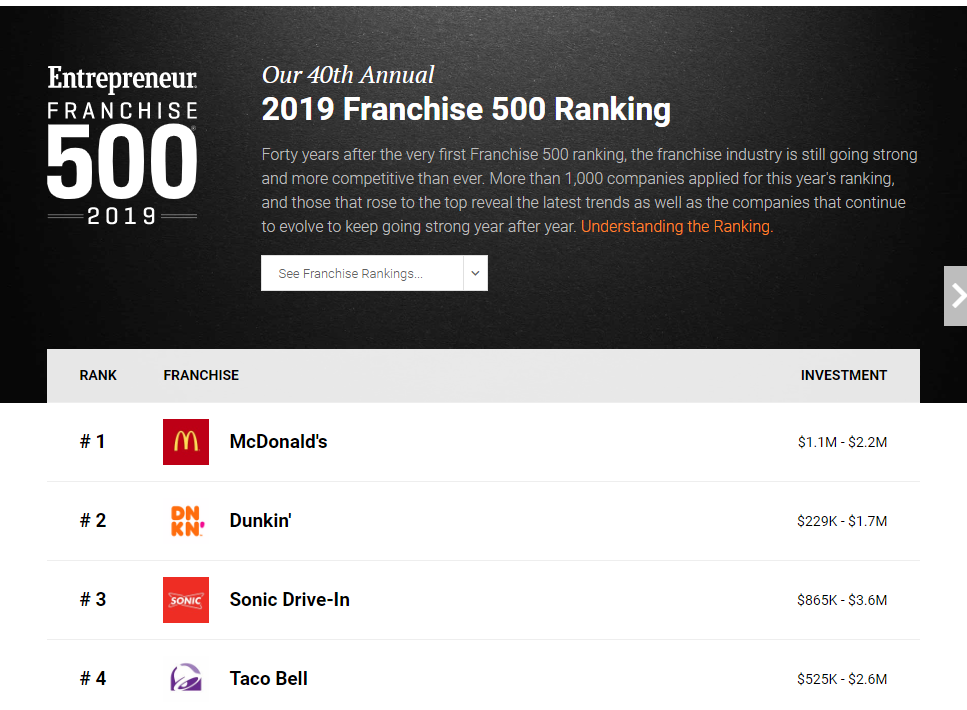
What would you like to search for

Mc Donald's

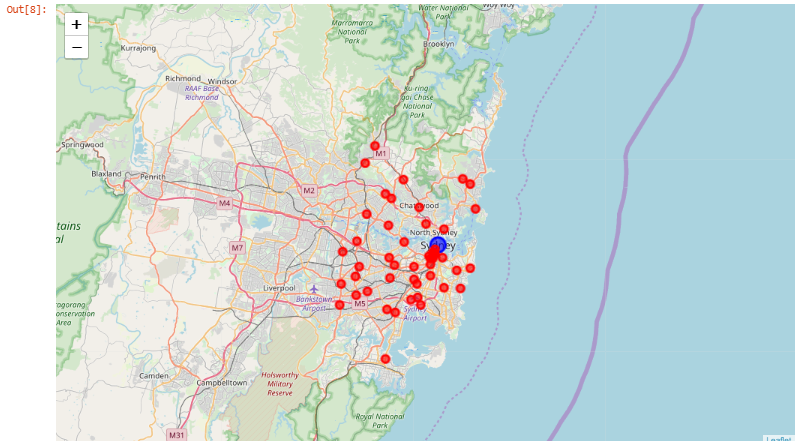
How far away in meters woud you like the radius to be?

100000

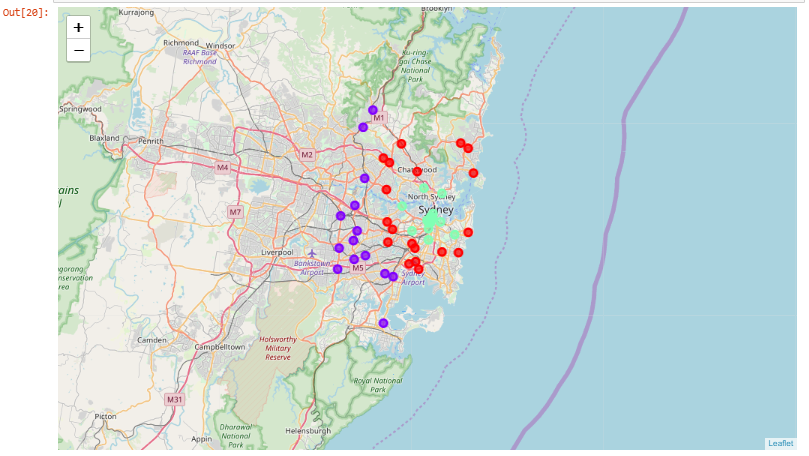
This produces the franchise analysis of 50 nearest Mc Donald’s stores within a 100,000m radius. Why was the outcome focused on Mc Donald’s you might ask? Well it is most probably the worlds most prominent franchise, the most easily found franchise the list could go on but a quick internet search found that it’s the top franchise according to the 40th Annual 2019 Franchise 500 ranking.

[](https://www.entrepreneur.com/franchise500)

The results of which are 2 key things the first of which being the direct competition on an interactive map for you to search and



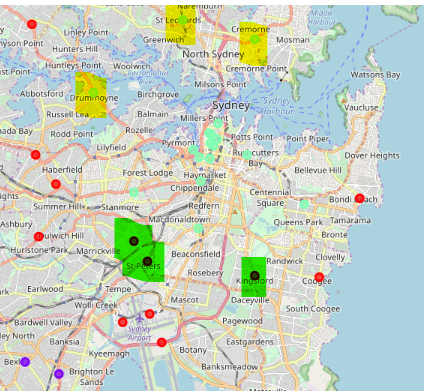
The other key outcome is the result of the ML k-means algorithm. Your direct competition. Their competition and the out reaches of your territory.



## 5. Discussion

### 5.1 Observations

* Given the proximity to the ocean further analysis over an inland based city or target location such as a street address theoretically will produce closer results within the 50 records retrieved.
* Upon review of the analysis. Given I know the area well I was shocked to find that the k-means algorithm produced competition I didn’t consider those across the bridges around Sydney, this is reflected on the below image as the yellow competitors. This is in stark contrast to the green black combination I naturally would have considered but shouldn’t have.



* 1. Recommendations

1. Competitive analysis is that the combination of business strategy and the way it relates to the competition.

It’s aim is to see the strengths and weaknesses of the competitors among your market, ways which will give you with a definite advantage, the barriers that may be developed so as to forestall competition from getting into your territory and any weaknesses that can be exploited among the merchandise development cycle.

The first step in a very contestant analysis is to spot the present and potential competition that is what's being tired the notebook here.

1. Once you have got sorted your competitors using this notebook, you'll begin to analyse their ways and determine the areas wherever they're vulnerable. this will be done through an examination of your contestant analysis through weaknesses and strength analysis.
2. There are basically 2 ways in which you'll determine competitors.
   1. The primary is to appear at the market from the customer's viewpoint and cluster all of your competitors by the degree to that they contend for the buyer's dollar.
   2. The second methodology is to cluster competitors consistent with their varied competitive ways therefore you perceive what motivates them.
3. Research the franchise so that there isn’t anything you don’t know about it.
4. Use this tool over and over until you’ve found exactly where you want to locate your ideal franchise location.

## Conclusion

I hope you’ve valued the information and the angle in which I aim to address the requirement of the coursea.org IBM Data Science Professional course.