The Starbucks Exploration



Contents

[1. Introduction 2](#_Toc42013065)

[1.1. Context 2](#_Toc42013066)

[1.2. Business Problem 2](#_Toc42013067)

[2. Data 3](#_Toc42013068)

[2.1. City Populations 3](#_Toc42013069)

[2.2. Starbucks and Neighbourhood Venues from Foursquare API 3](#_Toc42013070)

[3. Methodology 4](#_Toc42013071)

[3.1. Finding the City with the Fewest Stores 4](#_Toc42013072)

[3.2. Finding the City with the Most Stores 6](#_Toc42013073)

[3.3. Finding the Most Popular Starbucks & Reference Neighbourhood 6](#_Toc42013074)

[3.4. Finding the Best Rated Store in London and Study its Neighbourhood 6](#_Toc42013075)

[3.5. Studying Neighbourhoods in Rome 7](#_Toc42013076)

[3.6. Similarity between Neighbourhoods 9](#_Toc42013077)

[4. Results 10](#_Toc42013078)

[5. Discussion 11](#_Toc42013079)

[6. Conclusion 11](#_Toc42013080)

# Introduction

## Context

Thanks to Coursera and IBM, we are certified in Python and machine learning, and it did not take long for coffee company Starbucks to hire us as data scientists! We will work for the European Division!

Starbucks has only been growing bigger[[1]](#footnote-1) since its creation, and continuously open more stores across the world.

| **Year** | **Revenue in mil. US$** | **Net income in mil. US$** | **Total Assets in mil. US$** | **Average Price per Share in US$** | **Employees** |
| --- | --- | --- | --- | --- | --- |
| 2013 | 14867 | 8 | 11517 | 33.71 | 182000 |
| 2014 | 16448 | 2068 | 10753 | 37.78 | 191000 |
| 2015 | 19163 | 2757 | 12416 | 53.25 | 238000 |
| 2016 | 21316 | 2818 | 14313 | 56.59 | 254000 |
| 2017 | 22387 | 2885 | 14366 | 57.27 | 277000 |
| 2018 | 24720 | 4518 | 24156 | 57.50 | 291000 |
| 2019 | 26509 | 3599 | 19220 | 81.44 | 346000 |

According to this same Wikipedia page, as of May 2020, Starbucks is present in over 30,000 locations, on 6 continents and 79 countries.

## Business Problem

Our mission is to keep this global expansion going by opening a new store in Europe, but the location must be carefully chosen to guarantee success.

Our problem will be solved by studying the current stores locations. We will then choose a highly populous big city where Starbucks is not yet too present.

We will then try to find a more precise location within this city. In order to do so, we will select several successful Starbucks coffees and use Foursquare API to characterize their neighbourhood and try to find a similar location in our target city where there is no store yet!

# Data

## City Populations

We will need some population data to be able to find out where Starbucks is not yet heavily present.  
The table from Wikipedia[[2]](#footnote-2) also contains GPS coordinates, which will be useful later, therefore I put this under usable form.

| **City** | **Country** | **Population** | | | **Latitude** | **Longitude** |
| --- | --- | --- | --- | --- | --- | --- |
| Istanbul | Turkey | | 15519267 | 41.013611 | | 28.955000 |
| Moscow | Russia | | 12615279 | 55.750000 | | 37.616669 |
| London | United Kingdom | | 9126366 | 51.507221 | | 0.127500 |
| Saint Petersburg | Russia | | 5383890 | 59.950001 | | -30.299999 |
| Berlin | Germany | | 3748148 | 52.516666 | | 13.383333 |
| Madrid | Spain | | 3223334 | 40.383331 | | -3.716667 |
| Kiev | Ukraine | | 2950800 | 50.450001 | | 30.523333 |
| Rome | Italy | | 2844750 | 41.900002 | | 12.500000 |
| Paris | France | | 2140526 | 48.856701 | | 2.350800 |
| Bucharest | Romania | | 2106144 | 44.432499 | | 26.103889 |

## Starbucks and Neighbourhood Venues from Foursquare API

Foursquare API will be used to find the number of Starbucks store for each city and which stores have the most reviews, and hence are likely to be top spots in their respective cities!

We will also use the Foursquare API again to characterize the surroundings and try to find a similar neighbourhood in our target city.

# Methodology

## Finding the City with the Fewest Stores

A big limitation of Foursquare API is that the maximum number of results for a venue search is 50.

It is OK for our application because we are interested in cities with a low number of stores!

The first step of our study is to group Starbucks stores by city and count the number of occurrences.

Let us query Foursquare API to find Starbucks Stores for each city above.

| **City** | **Country** | **Population** | **Starbucks Stores**  **Count** |
| --- | --- | --- | --- |
| Istanbul | Turkey | 15519267 | 50 |
| Moscow | Russia | 12615279 | 50 |
| London | United Kingdom | 9126366 | 50 |
| Saint Petersburg | Russia | 5383890 | 0 |
| Berlin | Germany | 3748148 | 22 |
| Madrid | Spain | 3223334 | 50 |
| Kiev | Ukraine | 2950800 | 1 |
| Rome | Italy | 2844750 | 2 |
| Paris | France | 2140526 | 50 |
| Bucharest | Romania | 2106144 | 32 |
| Minsk | Belarus | 1982444 | 1 |
| Hamburg | Germany | 1930996 | 14 |
| Vienna | Austria | 1899055 | 22 |
| Warsaw | Poland | 1802237 | 30 |
| Budapest | Hungary | 1768073 | 32 |

Let us now add a column SB\_Density as Starbucks Density.  
This will calculate the number of people per Starbucks. The higher, the better for our study!

| **City** | **Country** | **Officialpopulation** | **StarbucksStoresCount** | **SB\_Density** |
| --- | --- | --- | --- | --- |
| Nizhny Novgorod | Russia | 1259013 | 0 | inf |
| Saint Petersburg | Russia | 5383890 | 0 | inf |
| Perm | Russia | 1051583 | 0 | inf |
| Ufa | Russia | 1121429 | 0 | inf |
| Kiev | Ukraine | 2950800 | 1 | 2.950800e+06 |
| Minsk | Belarus | 1982444 | 1 | 1.982444e+06 |
| Kharkiv | Ukraine | 1451132 | 1 | 1.451132e+06 |
| Rome | Italy | 2844750 | 2 | 1.422375e+06 |
| Tekirdağ | Turkey | 1055412 | 1 | 1.055412e+06 |
| Volgograd | Russia | 1013533 | 1 | 1.013533e+06 |
| Voronezh | Russia | 1054537 | 2 | 5.272685e+05 |
| Odessa | Ukraine | 1011494 | 2 | 5.057470e+05 |
| Belgrade | Serbia | 1397939 | 3 | 4.659797e+05 |
| Samara | Russia | 1170910 | 3 | 3.903033e+05 |
| Istanbul | Turkey | 15519267 | 50 | 3.103853e+05 |
| Moscow | Russia | 12615279 | 50 | 2.523056e+05 |
| Kazan | Russia | 1243500 | 5 | 2.487000e+05 |
| London | United Kingdom | 9126366 | 50 | 1.825273e+05 |
| Milan | Italy | 1390434 | 8 | 1.738042e+05 |
| Berlin | Germany | 3748148 | 22 | 1.703704e+05 |
| Rostov-on-Don | Russia | 1119875 | 7 | 1.599821e+05 |
| Sofia | Bulgaria | 1238438 | 8 | 1.548048e+05 |
| Hamburg | Germany | 1930996 | 14 | 1.379283e+05 |
| Cologne | Germany | 1085664 | 8 | 1.357080e+05 |
| Munich | Germany | 1471508 | 17 | 8.655929e+04 |

A few observations on these results:

* Russia, Ukraine and Belarus all look like promising markets. Starbucks are already successful in Moscow and many other populous cities could be good locations for a new store. However, the alphabet being different, it would be difficult to do the next part of this project (with some neighbourhoods spelt in Cyrillic!).
* Italy is a peculiar case, with 2 cities (Roma, Milan) in our top 25. It may sound astonishing that Starbucks has not already taken over this market, but in fact the coffee culture is very traditional and deeply rooted in Italy. This Forbes Article[[3]](#footnote-3) is a good read. Still, Starbucks is not only about coffee, the lifestyle experience is equally important, as proves the recent store addition in Milan!
* Germany is another country with a reasonably small density of Starbucks stores!
* SB\_Density for cities with more than 50 stores are naturally wrong, because of the API results limit.

Rome therefore looks like a good place to build a new store!

## Finding the City with the Most Stores

The number of results limit from the API is quite annoying for this point, but a quick Google Search[[4]](#footnote-4) shows that London has the most Starbucks stores.

The best method would have been to simply count the number of stores in the results of the API query.

## Finding the Most Popular Starbucks & Reference Neighbourhood

Let us explore the centre of London and try to find the most popular Starbucks!

| **Name** | **Address** | **Postcode** | **City** |
| --- | --- | --- | --- |
| Starbucks | 1-3 Villiers street | WC2N 6NN | London |
| Starbucks | 10 Kingsway, Unit B2; St Catherines House | WC2B 6LH | London |
| Starbucks | Charing Cross Road, Unit 1 129-133 | WC2H 0EA | London |
| Starbucks | 10 Russell Street | WC2B 5HZ | London |
| Starbucks | 112 - 116 New Oxford Street | WC1A 1HH | Bloomsbury |

## Finding the Best Rated Store in London and Study its Neighbourhood

We can get the rating of each store in Central London by using Foursquare API and sort the results by descending order.

| **Name** | **Address** | **Rating** | **Likes** |
| --- | --- | --- | --- |
| Starbucks | 52 Berkeley St. | 7.3 | 186 |
| Starbucks | 10 Kingsway, Unit B2; St Catherines House | 7.3 | 70 |
| Starbucks | 27 Berkeley St | 7.1 | 96 |
| Starbucks | 34 Great Marlborough St, (Carnaby Street) | 7.0 | 108 |
| Starbucks | 6A Vigo Street, London | 7.0 | 380 |

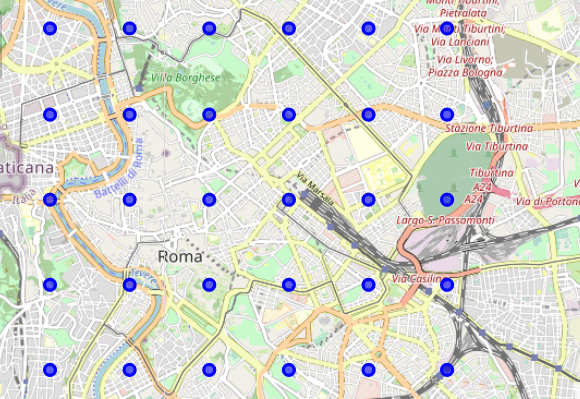
We have 2 winners! However, the first one has more likes, so it sounds like a good place to start from.

52 Berkeley St. is the reference location for the remainder of this battle.

## Studying Neighbourhoods in Rome

I could not find a list of Rome neighbourhoods along with GPS coordinates, so I created my own using Folium.

I fine-tuned my grid until I was visually happy with the point positions. I wanted them to cover most of the city centre.



Each of these neighbourhoods is given a name from R0 to R29.

At this latitude and for 1/110 deg of latitude and 1/80 deg of longitude, 1 grid step is approximately 1000m.[[5]](#footnote-5)

A function was defined to extract all venues within 700m radius of each of these neighbourhood centres, including the reference London neighbourhood.

The first elements are shown below:

| **Neigh.** | **Venue** | **Venue Latitude** | **Venue Longitude** | **Venue Category** |
| --- | --- | --- | --- | --- |
| London | The Ritz London | 51.507078 | -0.141627 | Hotel |
| London | Novikov | 51.507767 | -0.142850 | Asian Restaurant |
| London | Brown's Hotel | 51.509127 | -0.142077 | Hotel |
| London | Burger & Lobster | 51.507118 | -0.145477 | Seafood Restaurant |
| London | Prada | 51.508998 | -0.140959 | Boutique |

To characterize each neighbourhood, we used one-hot encoding and calculated the frequency of each venue category.

Ideally, I would need to spend a lot of time post-treating this data in order to have better groups of venues. I have only so much time, so I simply grouped together restaurants, bars, etc. There are still 114 categories at the end, so quite a lot...

The top 5 venues for the first 5 neighbourhoods of our list:

----London----

venue freq

0 Store 0.17

1 Art 0.09

2 Boutique 0.05

3 Hotel 0.05

4 Lounge 0.04

----R0----

venue freq

0 Hotel 0.09

1 Plaza 0.06

2 Winery 0.06

3 Pizza Place 0.06

4 Café 0.06

----R1----

venue freq

0 Plaza 0.09

1 Ice Cream Shop 0.07

2 Hotel 0.07

3 Café 0.07

4 Trattoria/Osteria 0.05

----R10----

venue freq

0 Hotel 0.10

1 Ice Cream Shop 0.07

2 Pizza Place 0.07

3 Bed & Breakfast 0.05

4 Trattoria/Osteria 0.05

----R11----

venue freq

0 Pizza Place 0.19

1 Café 0.07

2 Bistro 0.05

3 Plaza 0.05

4 Ice Cream Shop 0.05

A proper classification of the venues would be necessary to improve the accuracy of the correlation between neighbourhoods.

I have done a first step but for example perhaps not all Restaurants are equal! For now, I am happy with this.

## Similarity between Neighbourhoods

Let us assume that we did all the necessary work to properly classify venues of the same type.

We can now calculate the correlation between our London reference neighbourhood and each Rome neighbourhood. I will use Pearson correlation for this task.

The results are sorted by descending correlation score:

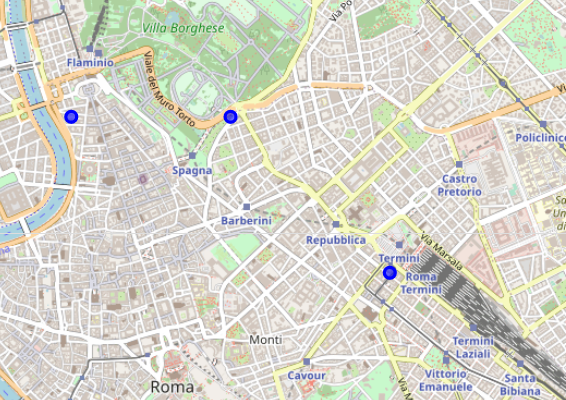
| **Neighbourhood** | **Correlation** | **Latitude** | **Longitude** |
| --- | --- | --- | --- |
| London | 1.000000 | 51.507442 | -0.142527 |
| R20 | 0.448627 | 41.909093 | 12.487500 |
| R19 | 0.435693 | 41.909093 | 12.475000 |
| R15 | 0.408138 | 41.900002 | 12.500000 |
| R28 | 0.326147 | 41.918184 | 12.512500 |

Naturally, correlation of London with itself is 1. We also found that R20, R19 and R15 are all plausible location for a new Starbucks store!

Let us map these places in Rome.

# Results

Let us plot the location of the 3 suggested locations for a new Starbucks Store in Rome.



Looking back at the typical venues of each neighbourhood, the main common point between our London "reference" neighbourhood is the high presence of hotels, restaurant and cafes.

Interestingly, Termini Station comes as a potential location (R15), which is a major transportation hub of the city. In fact, the only "Starbucks" location that the Foursquare API search query returned for Rome is located inside Termini station. It is not a licensed Starbucks store but very much looks like it. This is a good sign that this algorithm has not completely lost it!

The other 2 are essentially next to Villa Borghese park. R19 is next to the river. I would certainly enjoy a coffee in either place!

# Discussion

There are several ways this analysis could be improved.

Because of time and resources limitations, I have taken certain shortcuts, but I believe the method would still apply.

Major improvements would consist in:

* An up to date list of current Starbucks locations, including coordinates and sales volume. This would certainly be available as an employee of the company.
* Not limiting the study to European countries. In a situation where Starbucks wanted to expand to areas where alphabet/culture can be challenging to non-locals, the best would be to have local offices able to carry out this job.
* Improving the queries to the API and sorting the request results better. This is as time consuming as necessary to get good quality data. Still the results are far from illogical.

# Conclusion

As a new data scientist for Starbucks, the mission was to find the best location to open a new store. To do so, I have:

* Determined which cities in Europe have the fewest stores per inhabitant. I decided to focus on Rome.
* Chosen a neighbourhood where Starbucks is highly present, with a high user rating. The store is in London.
* Searched venues in Rome and near the reference store in London and worked out the similarity between all those neighbourhoods.

I have determined that 3 specific locations in Rome were similar to the neighbourhood in London where one of the most popular Starbucks stores in Europe is located. This result is a good point to start from.

Next step would be to verify if customers would be likely to visit these new locations.

1. https://en.wikipedia.org/wiki/Starbucks#Locations [↑](#footnote-ref-1)
2. https://en.wikipedia.org/wiki/List\_of\_European\_cities\_by\_population\_within\_city\_limits [↑](#footnote-ref-2)
3. https://www.forbes.com/sites/jennawang/2018/09/13/why-it-took-starbucks-47-years-to-open-a-store-in-italy/ [↑](#footnote-ref-3)
4. https://www.newstatesman.com/jonn-elledge/2014/05/london-has-more-branches-starbucks-any-eu-country [↑](#footnote-ref-4)
5. <http://www.csgnetwork.com/degreelenllavcalc.html> [↑](#footnote-ref-5)