

# SHOE STORE

## Finding the best location

---

January 14, 2020

---

IBM Coursera Capstone Project  
Liubomir Hristov



---

# SHOE STORE

## Finding the best location

### 1. Introduction

#### 1.1. Background

Tim Petrov is a young and talented designer producing handmade shoes under his own brand “TP collection” in Russia. His collections are targeted at young professionals aged 25 – 35, both male and female. Business is going well and last year he opened shops at almost every big city in Russia with population over one million.

#### 1.2. Problem

Now he is planning to open his first shop abroad in 2020 and must choose in which European city it will be. He is very familiar with Russian business reality but has no firm understanding of market situation in European cities. Tim Petrov hired consultancy company “ABC Wise” to get help with identifying which city and location there he should choose for the first store and get the ordered list of other preferable locations.

During a discussion with “ABC Wise” specialists, Tim Petrov outlined his main criteria for selecting the city – population aged 25-35 with big purchasing power.

Regarding location in the city:

- high concentration of existing shoe shops. It’s a little bit strange, but the location with more shops is better as the brand is new and yet unknown, so few customers will travel specially to visit the shop. More likely they will look to buy shoes from an area with a lot of other shoe shops.
- venues often visited by population aged 25-35 and tourists.

#### 1.3. Metrics

There are no available datasets for purchasing power of population aged 25-35 by European cities, so we must estimate it using publicly available data.

We can start city selection with metrics like:

- City with high GDP: obviously in city with high GDP the people earns more and can afford to spend more.
- Big population: a city with the bigger population is expected to have more potential customers
- High percentage of people aged 25-35: they are our most prospective target group

---

number of tourists visiting the city: tourists are our second target group

- Doing business ranking.

Location selection:

- Total number of shoe shops in the same area.
- Concentration of different kind of venues preferred by young professional like: trending café, fashion restaurants, art-galleries
- University campuses
- Tourist attractions

## Data

### 2.1 Data source used for city selection

In order to find which city suits our better-defined criteria we will use the following public available data from Eurostat <https://ec.europa.eu/eurostat>.

Eurostat is the statistical office of the European Union situated in Luxembourg. Its mission is to provide high quality statistics for Europe, offering a whole range of important and interesting data that governments, businesses, the education sector, journalists and the public can use for their work and daily life.

As our main target group consist of young people 25-35 years old, so we'd like to rank the major European cities by purchasing power of this group. There is no such data at Eurostat site ready to be used, but we can easily get it using available data, Excel files:

- met\_10r\_3gdp.xls - Gross domestic product (GDP) at current market prices by metropolitan regions. These regions include countries and regions, so we have to filter the data.
- urb\_cpop1.xls - Population aged 25 – 35 cities and greater cities
- urb\_cpop1 total.xls - City population total
- Urb\_ctour.xls - Total nights spent in tourist accommodation establishments per city
- Rankings.xlsx - A World Bank Group flagship publication measuring the regulations that enhance business activity and those that constrain it. Doing Business presents quantitative indicators on business regulations and the protection of property rights that can be compared across 190 economies.

Doing Business covers 12 areas of business regulation. Ten of these areas—starting a business, dealing with construction permits, getting electricity, registering property, getting credit, protecting minority investors, paying taxes, trading across borders, enforcing contracts, and resolving insolvency—are included in the ease of doing business score and ease of doing business ranking.

---

## 2.2 Data source used for location selection

Foursquare - The Places API offers real-time access to Foursquare's global database of rich venue data and user content for over 105MM places available across 190 countries and 50 territories.

# Methodology

## 3.1 Estimate purchasing power of population aged 25-35 per each European city.

Ideally it will be very suitable for our research to find data about purchasing power of population 25-35 per each European city. Unfortunately, it's not possible to find such data, so we must try to estimate it using publicly available datasets from Eurostat.

There is a dataset called "Population on 1 January by age groups and sex - cities and greater cities" from [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=urb\\_cpop1&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=urb_cpop1&lang=en)

We can make queries for all major European cities regarding either total population or select specific age and sex group. Let's request and download the data for total population – file urb\_cpop1 total.xls and the same data for Population aged 25 – 34 – file urb\_cpop1.xls.

Another useful dataset from Eurostat is Gross domestic product (GDP) at current market prices by metropolitan regions - file met\_10r\_3gdp.xls

If we assume that proportion of GDP produced by population 25-35 to the GDP produced by total population

$$C = \text{GDP}_{25-35} / \text{GDP}_{\text{total}}$$

is the same for different European cities, then we can calculate the GDP of population 25-35 for every city.

We just must multiply city GDP by population aged 25 – 35 and divide it by city population total.

$$\text{GDP}_{25-35} = \text{GDP}_{\text{total}} * (\text{POP}_{25-35} / \text{POP}_{\text{total}})$$

Thus, we can sort all major European cities by GDP of population aged 25 – 35, which we assume is proportional to its purchasing power.

## 3.2. Estimate purchasing power of tourists per each European city.

Our second target group consist of international and national tourists. At Eurostat web page we can find a lot of different data regarding tourism. Ideally, we'd like to see data about how many pairs of shoes tourists buy from every city, or how much money they spend there for shoes, but obviously it's not possible to get such data.

---

In order to find the best place it is not necessary to have the exact figures, but to rank the cities by purchasing power of tourists, so suitable metric for us will be the “Total nights spent in tourist accommodation establishments per city”, which we assume proportional to tourist purchasing power. This is the file Urb\_ctour.xls

### 3.3. Economy’s ease of doing Business

File – **Rankings.xlsx** – Ranking of economy’s ease of doing business for 192 countries. There is an overall ranking and rankings for specific parts of doing business. We’ll use the rank for Starting a Business.

Now, we have three independent lists with EU major cities ordered by purchasing power, tourists and easiness of doing business in respective country. We’d like to calculate the overall ranking based on importance of criteria. It seems reasonable to set weight coefficients of 0.7 for purchasing power of population 25-35 and 0.15 for both tourists ranking and Doing Business and find which city will get maximum score.

### 3.4. Finding the best location in selected city

Now, after we’ve selected the city the next step is to find the best neighborhood. We are going to start utilizing the Foursquare API to explore the neighborhoods and find which one will get maximum score according our criteria.

- we will make a list of neighborhoods ordered by number of shoe shops located there
- we will make 5 lists of neighborhoods ordered by number of venues preferred by young professional: trending café, fashion restaurants, art-galleries, university campuses, tourist attractions

Now we have 6 lists of neighborhoods ordered by different criteria. It seems reasonable to set weight coefficients of 0,5 for the list with shoe shops and 0,1 for others lists.

The neighborhood with maximum score will be the best place for opening new shop of “TP collection”

## 4. Results

### 4.1. Selecting the best city

After clearing the data for cities GDP from file **met\_10r\_3gdp.xls** and arranging cities by GDP we can

see that two cities London and Paris are at the top with very close results and the next three Madrid, Milano and Munich have more than three times smaller GDP.

	City	GDP
0	London	763494.39
1	Paris	683962.00
2	Madrid	211782.00
3	Milano	200116.42
4	München	178945.00

After clearing the data for Population aged 25 – 34 from file **urb\_cpop1.xls** and arranging cities by population 25-34 we can see almost the same proportions: London and Paris are very close at the top and Madrid, Milano and Munich have about three times smaller population

	City	POP_25-35
0	London	1644020.0
1	Paris	1497168.0
2	Madrid	628193.0
3	Milano	443827.0
4	München	267864.0

After clearing the data for Total Population from file **urb\_cpop1 total.xls** and arranging cities by total population we can see different ranking – this time Paris is at the top followed by London and the rest.

	City	POP_total
0	Paris	9803494.0
1	London	8866541.0
2	Madrid	4955432.0
3	Milano	4106356.0
4	München	1456039.0

After a simple arithmetic operation, we have a list with top five European cities, ordered by purchasing power of population 25-35. At the top is London followed by Paris.

	City	YP	GDP	gdp_25_35
0	London	0.167697	763494.39	128035.988704
1	Paris	0.168856	683962.00	115491.037555
2	München	0.183968	178945.00	32920.082141
3	Madrid	0.126769	211782.00	26847.300079
4	Milano	0.108083	200116.42	21629.169594

Rank the top 5 cities by purchasing power of tourists.

We used the metric “**Total nights spent in tourist accommodation establishments per city**” from file **Urb\_ctour.xls** from Eurostat, which we assume proportional to tourist purchasing power. Due to some unknown reason, there is no data for tourists visiting London, so we replaced missing data with the mean from other 5 top European cities. Data is normalized and the top city is Madrid followed by London, Munich and Paris.

	City	Tourists
1	Madrid	1.000000
4	London	0.729684
0	München	0.707566
2	Paris	0.658051
3	Milano	0.553120



## Starting a business

There is no available data for ease of doing business for any specific city. The metric by its nature refers to category “country” so we just use it from the respective country.

	City	Starting a Business
0	London	0.856
1	Paris	0.704
2	Madrid	0.224
3	Milano	0.216
4	München	0.000

Highest position in our list gets London followed by Paris.

## Getting the overall rating.

	City	Starting a Business	gdp_25_35	Tourists	Rank
0	London	0.856	1.000000	0.729684	1.000000
1	Paris	0.704	0.902020	0.658051	0.884600
2	Madrid	0.224	0.209686	1.000000	0.437167
3	Milano	0.216	0.168930	0.553120	0.296832
4	München	0.000	0.257116	0.707566	0.333630

The overall ranking leader is London.

## 4.2. Finding the best location in London.

Our further research is focused on finding the most suitable location in London. For this purpose, we made a list of neighborhoods ordered by number of shoe shops located there and another 5 lists of neighborhoods ordered by number of venues preferred by young professional: trending café, fashion restaurants, art-galleries, university campuses and tourist attractions.

We used Foursquare API for that purpose.



	Location	Borough	Lat	Lng	Shoe store	Art gallery	College and University	Cafe	Vegetarian Vegan Restaurant	Travel and Transport
0	Fitzrovia	Camden	51.518529819132	-0.13784759302216	50.0	50.0	49.0	50.0	17.0	42.0
1	Soho	Westminster	51.517585379522	-0.13500343088978	50.0	50.0	49.0	50.0	21.0	42.0
2	Mayfair	Westminster	51.508826777815	-0.14977175677031	48.0	49.0	23.0	50.0	2.0	43.0
3	St James's	Westminster	51.508598701761	-0.1353704316591	48.0	47.0	33.0	50.0	15.0	39.0
4	Covent Garden	Westminster	51.512009617972	-0.12370182518566	47.0	46.0	43.0	50.0	8.0	47.0
5	Marylebone (also St Marylebone)	Westminster	51.517813528559	-0.1494075869483	47.0	49.0	45.0	50.0	6.0	34.0
6	Holborn	Camden	51.517355462669	-0.1205994650938	41.0	38.0	50.0	50.0	10.0	40.0
7	St Giles	Camden	51.517355462669	-0.1205994650938	41.0	38.0	50.0	50.0	10.0	40.0
8	Aldwych	Westminster	51.512815974418	-0.11790391775668	40.0	33.0	50.0	49.0	4.0	42.0
9	Shepherd's Bush	Hammersmith and Fulham	51.503649663134	-0.22202608436881	32.0	0.0	17.0	29.0	3.0	39.0
10	King's Cross	Westminster	51.508826777815	-0.14977175677031	48.0	49.0	23.0	50.0	2.0	43.0

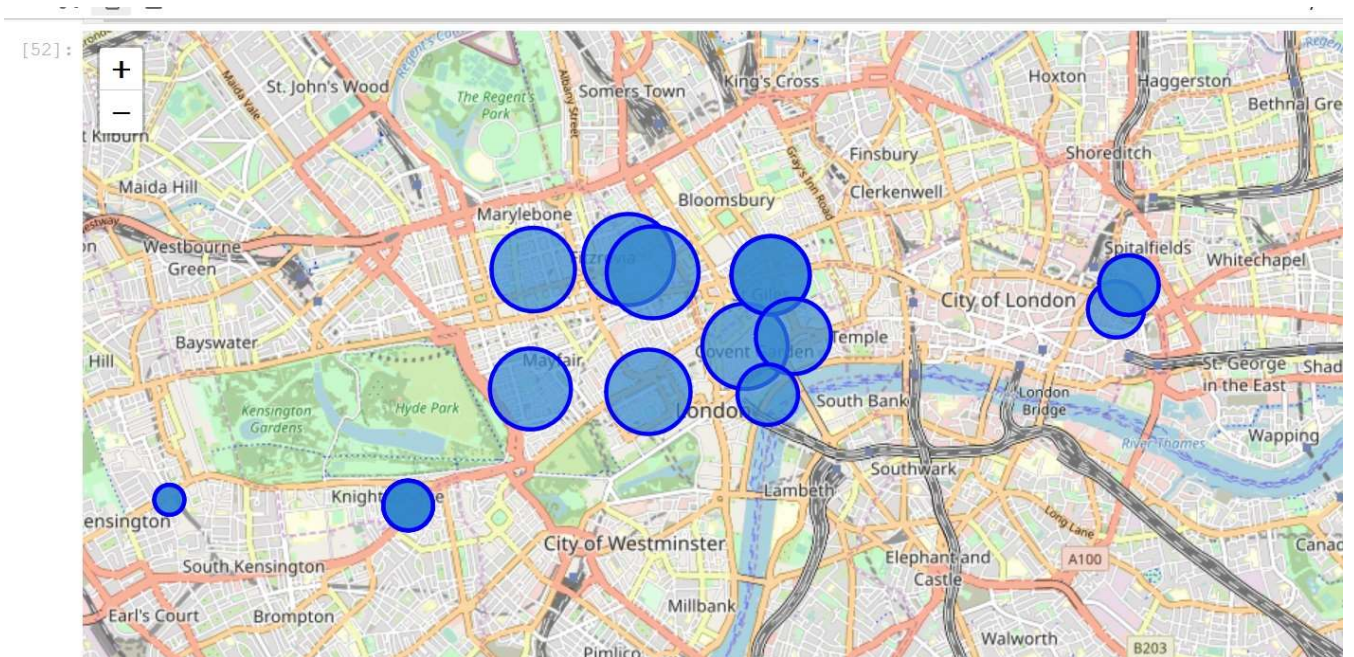
We can see that our main criteria with weight 0.5 “shoe stores” has its maximum for Fitzrovia and Soho, followed by Mayfair and St. James’s.

The others criteria with total weight of 0.5, related to venues preferred by young people 25-35, are shown in the table too, have their maximums at the same locations and probably will not change the order. Let’s do a simple mathematical calculation and find the overall rank.

	Location	Borough	Lat	Lng	Shoe store	Art gallery	College and University	Cafe	Vegetarian Vegan Restaurant	Travel and Transport	Rating
	Soho	Westminster	51.517585379522	-0.13500343088978	50.0	50.0	49.0	50.0	21.0	42.0	46.2
	Fitzrovia	Camden	51.518529819132	-0.13784759302216	50.0	50.0	49.0	50.0	17.0	42.0	45.8
	Covent Garden	Westminster	51.512009617972	-0.12370182518566	47.0	46.0	43.0	50.0	8.0	47.0	42.9
	St James's	Westminster	51.508598701761	-0.1353704316591	48.0	47.0	33.0	50.0	15.0	39.0	42.4
	Marylebone (also St Marylebone)	Westminster	51.517813528559	-0.1494075869483	47.0	49.0	45.0	50.0	6.0	34.0	41.9
	Mayfair	Westminster	51.508826777815	-0.14977175677031	48.0	49.0	23.0	50.0	2.0	43.0	40.7
	Holborn	Camden	51.517355462669	-0.1205994650938	41.0	38.0	50.0	50.0	10.0	40.0	39.3
	St Giles	Camden	51.517355462669	-0.1205994650938	41.0	38.0	50.0	50.0	10.0	40.0	39.3
	Aldwych	Westminster	51.512815974418	-0.11790391775668	40.0	33.0	50.0	49.0	4.0	42.0	37.8
	Charing Cross	Westminster	51.508368858456	-0.12096929668046	25.0	43.0	39.0	48.0	3.0	46.0	30.4
	Spitalfields	Tower Hamlets	51.516655105759	-0.077388723932787	20.0	36.0	47.0	50.0	18.0	49.0	30.0
	Whitechapel	Tower Hamlets	51.516655105759	-0.077388723932787	20.0	36.0	47.0	50.0	18.0	49.0	30.0

So, the bests locations are Soho, Fitzrovia and Covent Garden.

Let's visualize our results.



The blue circles represent the top locations for opening "TP collection" shoe store in London, the size of the circle is proportional to location's rank.

## 5. Discussion

During this study I noted some interesting observations:

### 5.1. Quality of the data.

The main data used was from Eurostat, but even it must be carefully checked. For example, report with data for metropolitan regions includes data not only for cities and greater cities, but for countries which must be dropped manually. There were many cases with missing values too.

Additionally, ArcGIS Online Geocoding Service works well for central London, but for all neighbors located outside without any indication returns the same coordinates.

### 5.2. The data itself.

According Eurostat, Paris has more total population than London (9.80mil vs 8.87mil), but London has more young population aged 25-35 (1.64mil vs 1.5mil) and bigger GDP. Other cities in Top 5: Madrid, Milano, Munich have average 30% lower values.

---

As we can see from the map of London with superimposed top locations, there is very clear cluster located on North Bank between Marylebone and Mayfair from the West, and Bloomsbury and Temple from the East. Probably this is the area with highest rent rates too.

## 6. Conclusion

Finding the best location to open the shoe shop is a complex task and generally it consists of two parts: a matter of fashion i.e. will the people in that location would like this kind of shoes.  
does the population have enough purchasing power to buy them?

In our study we found the best city for “TP Collection” based on purchasing power of generation 25-35, which as agreed with the client, is our target group.

We tried to take in account another minor factors such as visiting city tourists and ease of doing business. During selection of location the main criteria was available shoe shops and minor factor was presence of places, often visited by our target group.

This is the rational part of process of finding the best location, based on easily countable figures.

Solving the first part of the task - will the people in that location would like this kind of shoes, is much more interesting. It is more complex, and there is no publicly available data for it.

Possible way to estimate demand for design shoes would be to analyze customer behavior related to other fashion products like clothes, bags etc., which is out boundaries of this course.