

# Placement Pizzeria Paris

PREDICTION OF THE BEST LOCATION TO OPEN A PIZZERIA IN PARIS

# **1. INTRODUCTION**

## **1.1 Context**

## **1.2 Problem**

# 1.1 Context

**Paris**

**44.896 restaurant**



**people eat one meal on 5 outside**

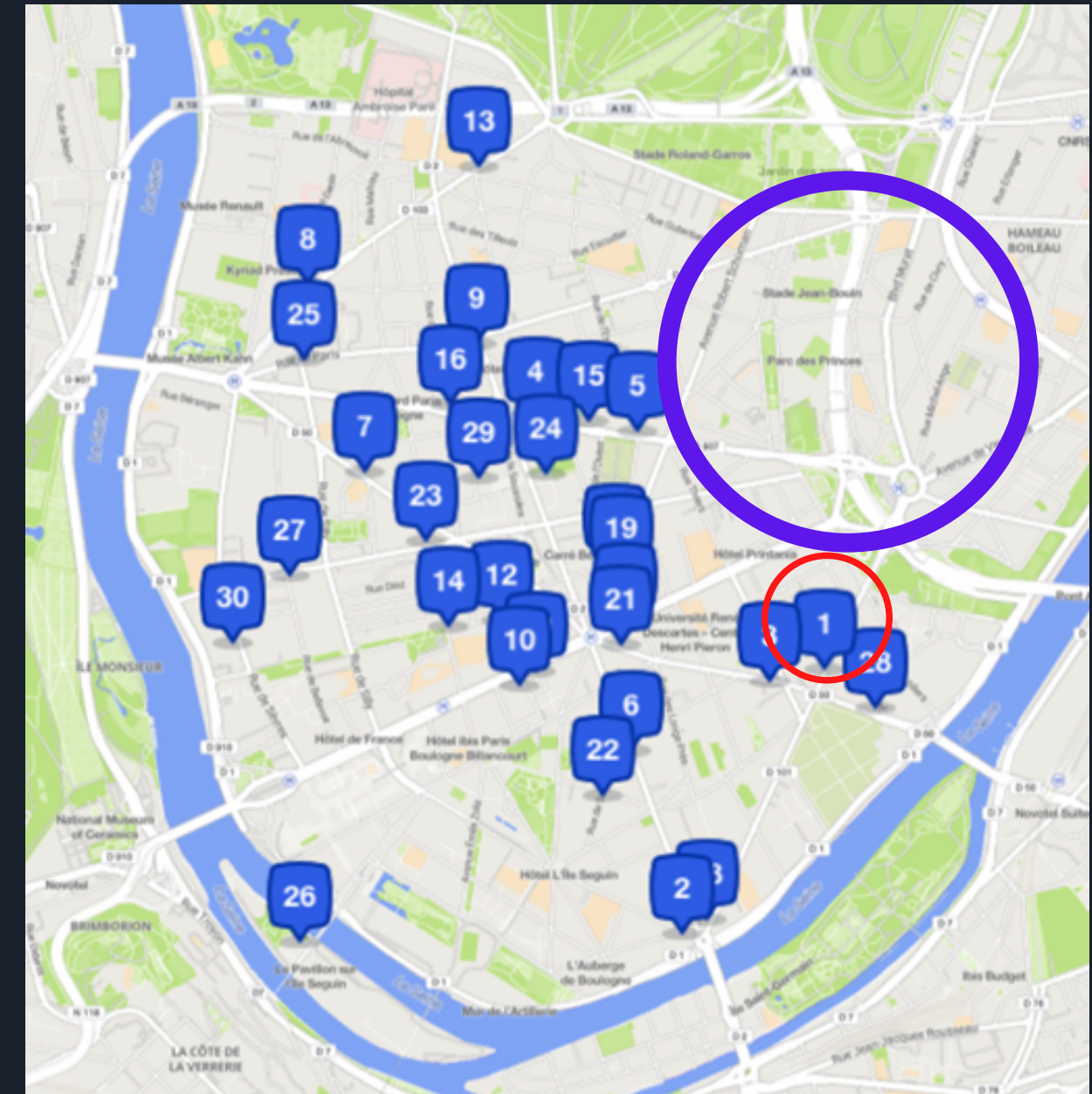


**every restaurant should has 9.6 clients for each meal on the global scope of Paris, but certain neighborhood are more interessting than others to open a restaurant.**

# 1.2 Problem

**It is almost impossible to have no concurrence in a big city when we open a restaurant, especially a pizzeria**

**each blue point represents a pizzeria, the red circle is a pizzeria that open just a few years ago, and which got a phenomenal success, maybe we can explain this success by the fact that there is a large portion of the city with no pizzeria and a large population in this area.**



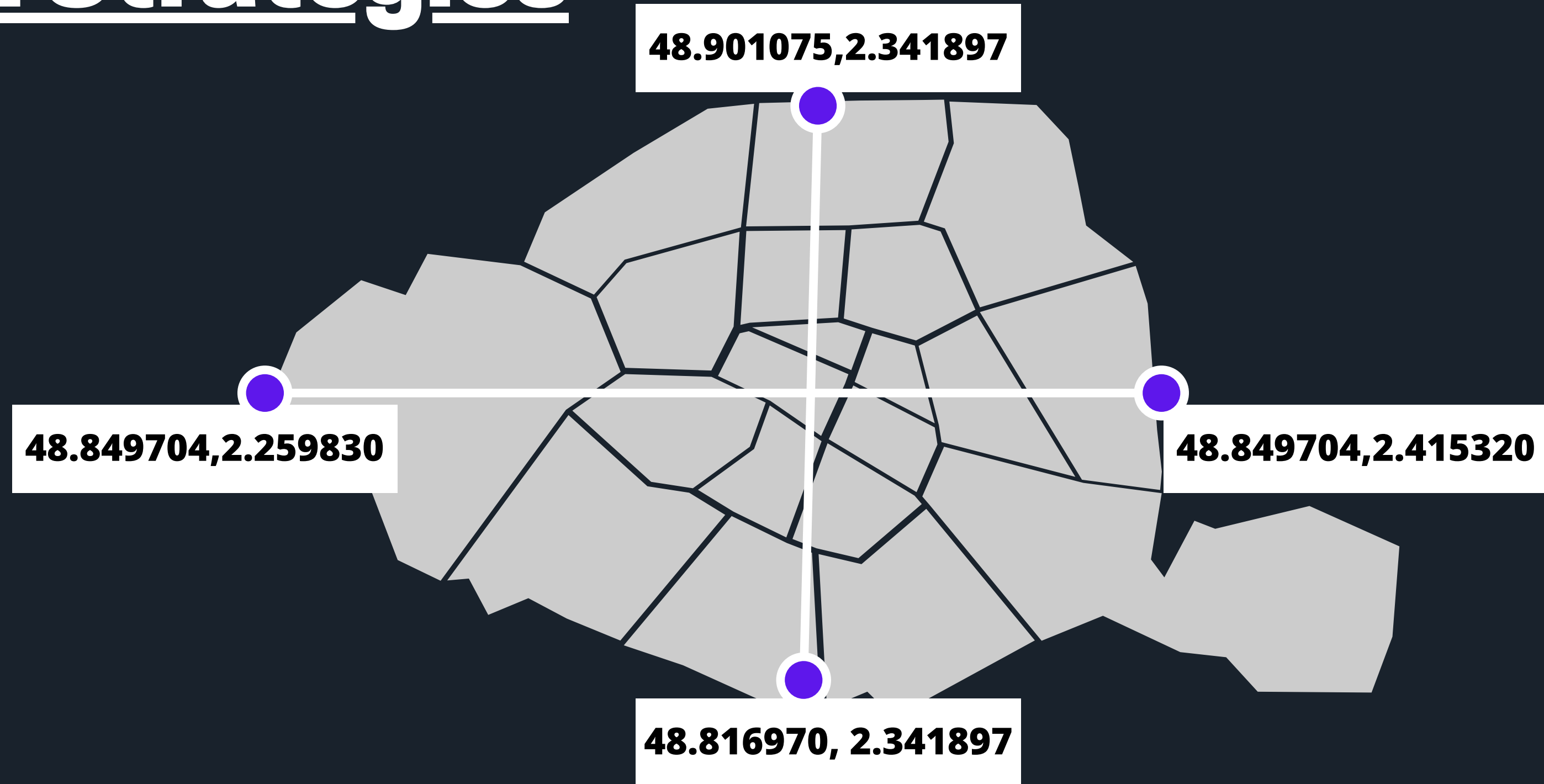
**Starting in the best zone can just speed up the success process and touch more people quicker even if the quality matter.**

# **1. DATA**

## **2.1 Strategies**

## **2.2 Data structure**

# 2.1 Strategies



We will take a 200m range for each location, we have **2622 possible areas**

## **2.2 Data structure**

**Here are the features that we will use**

**The density of population in the area**

**The number of pizzerias in a 500m range**

**The number of other restaurants in a 300m range**

**The presence of tourist places in a 1km range**

**The number of stores in a 300m range**

**Price per square meter of the district**

# **3. Methodology.**

## **3.1 Data collection**

## **3.2 Method**

## **3.3 Algorithm**



## 3.1 Data collection



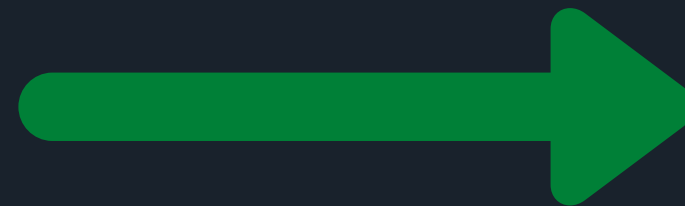
Foursquare API



- Number of restaurants
- Number of shops
- Number of pizzerias
- Number of tourist places



Excel Tabular

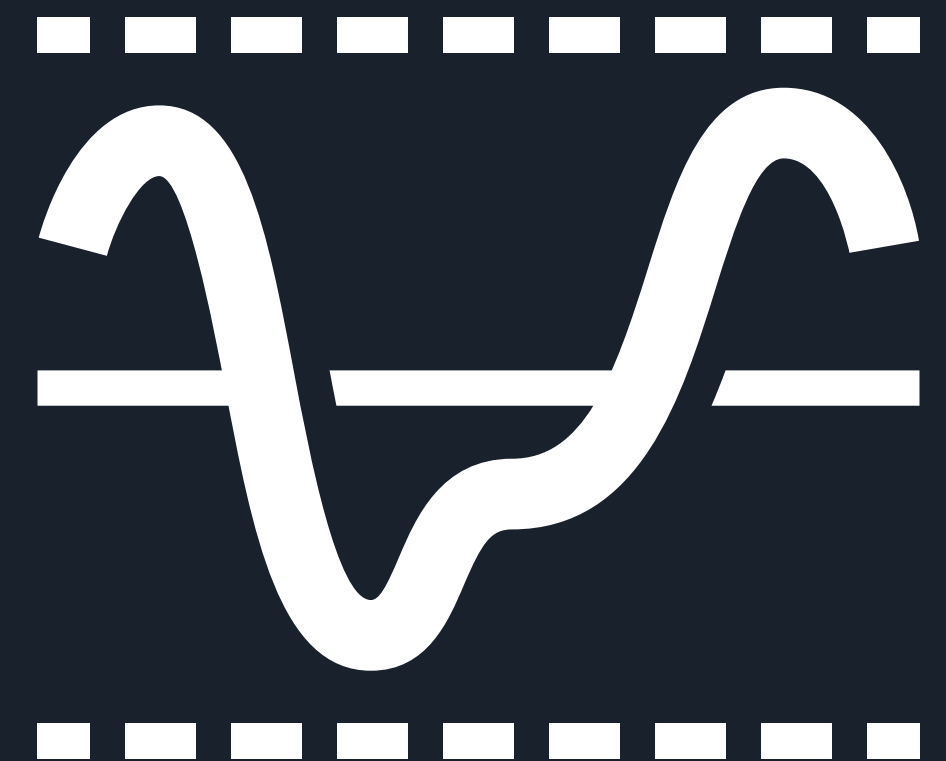


- Population density and
- price per square meter.

## 3.2 Method

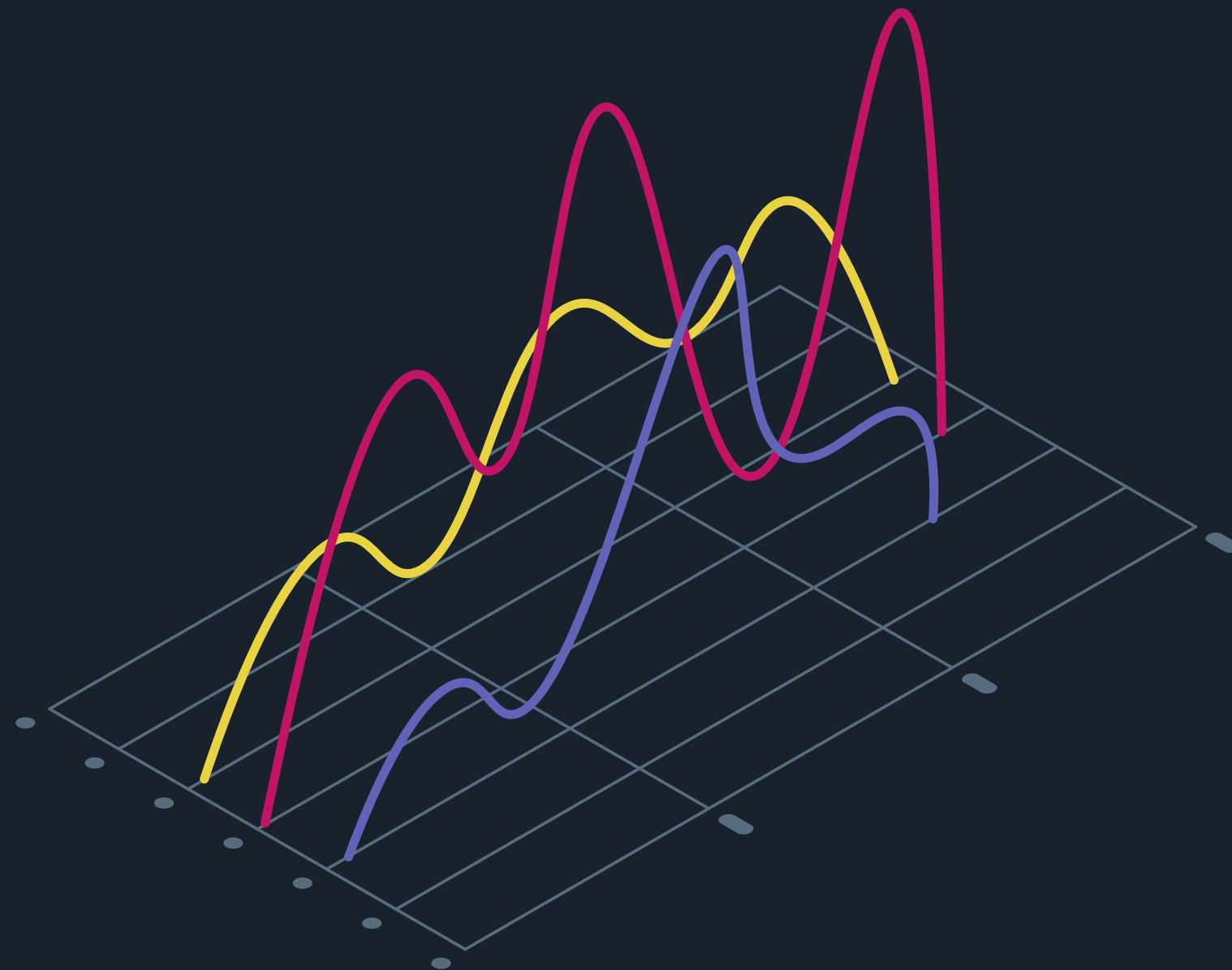
For target variable to estimate the number of clients of a restaurants, we will use the **number of comments**, it is not a very good metrics but we don't have access to the premium foursquare API, which have a metrics to estimate the number of clients per day.

The best 30% of the score will define the **threshold**, we will then calculate the score for each neighborhood, if it is above the threshold then the neighborhood will be defined as being interesting.



## 3.3 Algorithm

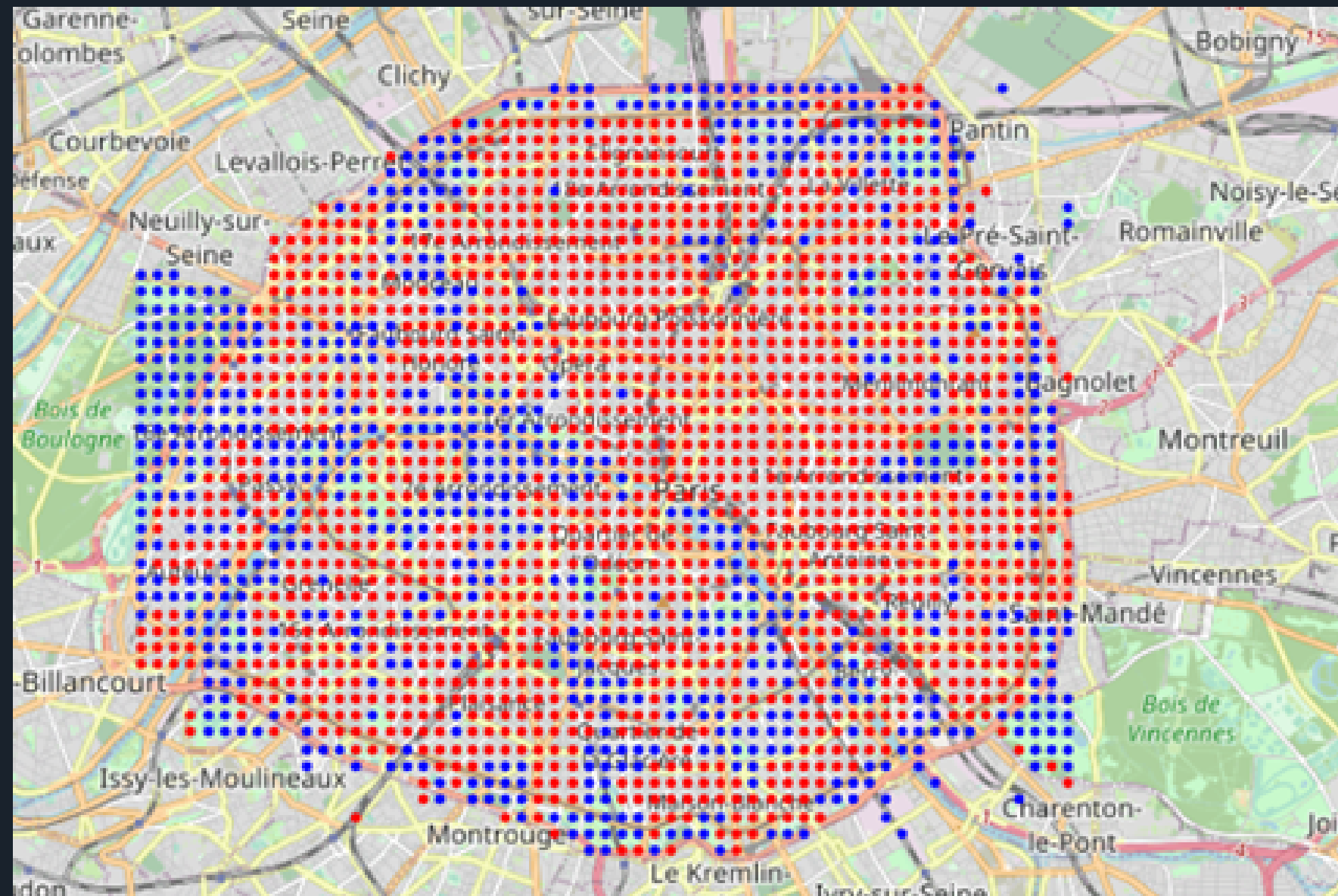
The variable that we want to predict is a continuous value so we will use a **regression technique**



# 3.3 Algorithm

**Red point : Training Set**

**Blue points : Prediction**



# **4. Results**

## **4.1 Data cleaning (correlation analysis)**

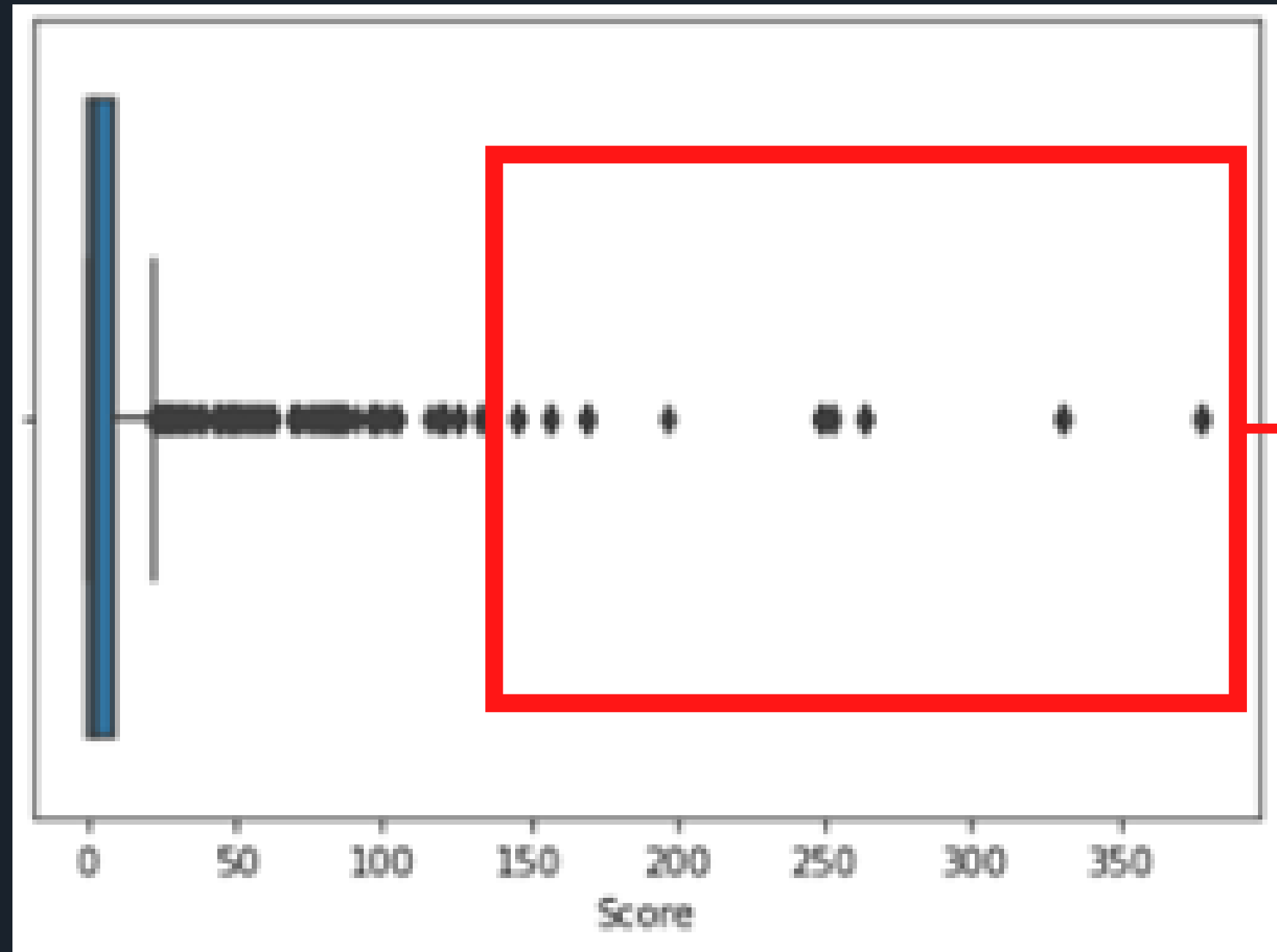
## **4.2 Model results**

# 4.1 Data cleaning (correlation analysis)

|       |        |       |       |       |        |        |
|-------|--------|-------|-------|-------|--------|--------|
| 1     | 0.29   | 0.46  | 0.23  | 0.26  | 0.098  | 0.26   |
| 0.29  | 1      | 0.48  | 0.19  | 0.19  | -0.033 | 0.085  |
| 0.46  | 0.48   | 1     | 0.39  | 0.39  | -0.11  | 0.23   |
| 0.23  | 0.19   | 0.39  | 1     | 0.71  | -0.57  | 0.17   |
| 0.26  | 0.19   | 0.39  | 0.71  | 1     | -0.55  | 0.22   |
| 0.098 | -0.033 | -0.11 | -0.57 | -0.55 | 1      | -0.015 |
| 0.26  | 0.085  | 0.23  | 0.17  | 0.22  | -0.015 | 1      |

→ **Really bad correlation  
(Population density)**

# 4.1 Data cleaning (correlation analysis)



## 4.2 Model results

**Here is the map with all of the good neighborhood to open a Pizzeria (score > 6 for the threshold)**

