# APPLIED DATA SCIENCE CAPSTONE PROJECT

## FINDING THE BEST NEIGHBORHOOD TO LIVE

**Loïc BRISSOT**

# CONTENTS


- INTRODUCTION
- BUSINESS PROBLEM
- DATA
- METHODOLOGY
- DETERMINE THE OPTIMUM NEIGHBORHOOD
- RESULTS
- CONCLUSION

# INTRODUCTION

- WHEN LIVING IN A METROPOLIS, THERE ARE OFTEN A LOT OF JOB OPPORTUNITIES AVAILABLE BUT CHANGING EMPLOYER CAN ALSO REQUIRE YOU TO CHANGE OF HOME.
- SEVERAL FRIENDS OF MINE WILL CHANGE OF EMPLOYER AND APARTMENT IN THE NEXT FEW YEARS.
- FINDING THE BEST PLACE TO LIVE REQUIRES TO ANALYZE MULTIPLE FACTORS AND THOUSANDS OF DATA POINTS.

# BUSINESS PROBLEM

- THE PURPOSE OF THIS WORK IS TO FIND THE OPTIMUM NEIGHBORHOOD TO LIVE IN THE CITY OF PARIS.
- EVERY PERSON HAS DIFFERENT TASTES AND DIFFERENT NEEDS THAT HAVE TO BE TAKEN INTO ACCOUNT TO DETERMINE THE PLACE THAT WOULD FIT THEM THE BEST.
- A PERSON HAS TO TRAVEL A LOT BETWEEN THEIR HOME AND THEIR WORKPLACE. TRAVEL TIME IMPACTS A LOT ON THE QUALITY OF LIFE AND HAPPINESS.

# DATA

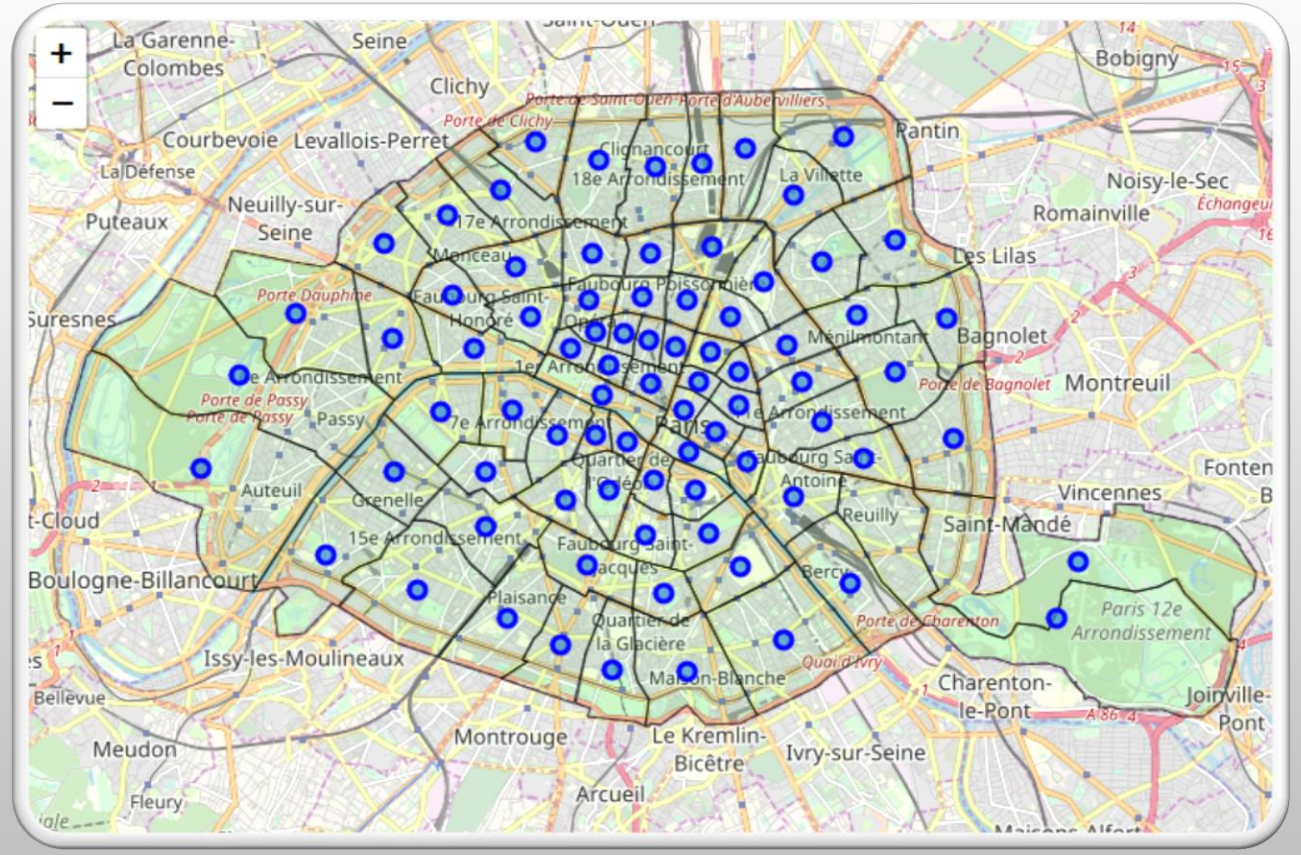
- ADMINISTRATIVE DIVISIONS OF THE NEIGHBORHOODS IN PARIS: HTTPS://OPENDATA.PARIS.FR/EXPLORE/DATASET/QUARTIER_PARIS/INFORMATION/
- VENUES IN PARIS: HTTPS://DEVELOPER.FOURSQUARE.COM/
- TRAVEL TIMES FROM A TO B IN PARIS: HTTPS://CITYMAPPER.COM/TOOLS/905/TRAVEL-TIME

# METHODOLOGY

OUR SUBJECT IN THIS PROJECT IS JEAN-PHILIPPE. HE WORKS DOWNTOWN PARIS, LIKES PLAYING TENNIS, EATING CHINESE FOOD AND GOING TO THE MOVIE THEATRE.

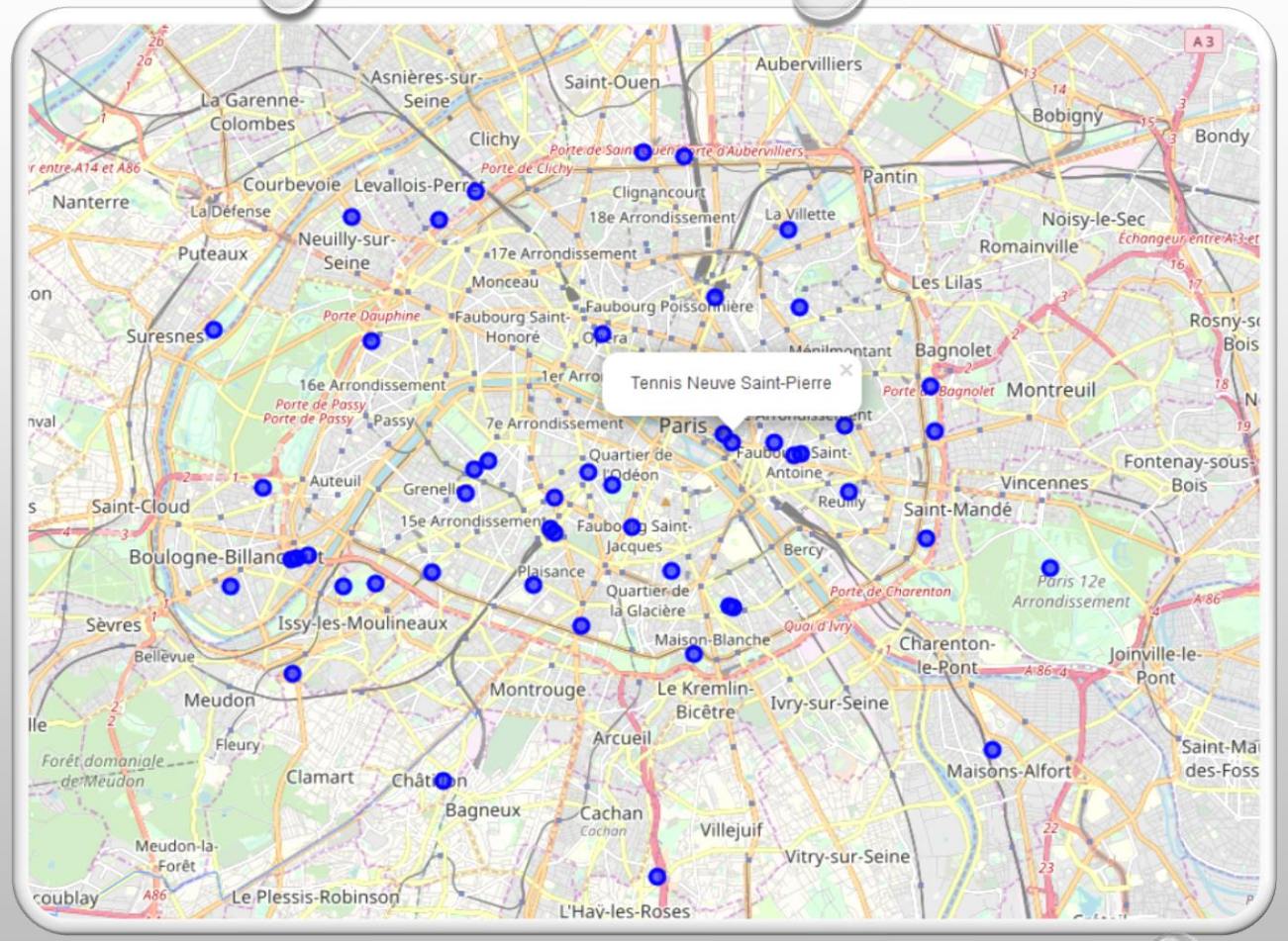
CONSIDERING THE CONTRIBUTION OF DIFFERENT FACTORS TO THE QUALITY OF LIFE, THE FOLLOWING SCORE HAS BEEN DEFINED:

$$FinalScore = 0.4 * TravelScore + 0.3 * SportScore + 0.1 * RestaurantScore + 0.2 * CultureScore$$




# DATA EXPLORATION

THERE ARE 80 ADMINISTRATIVE NEIGHBORHOODS IN PARIS.



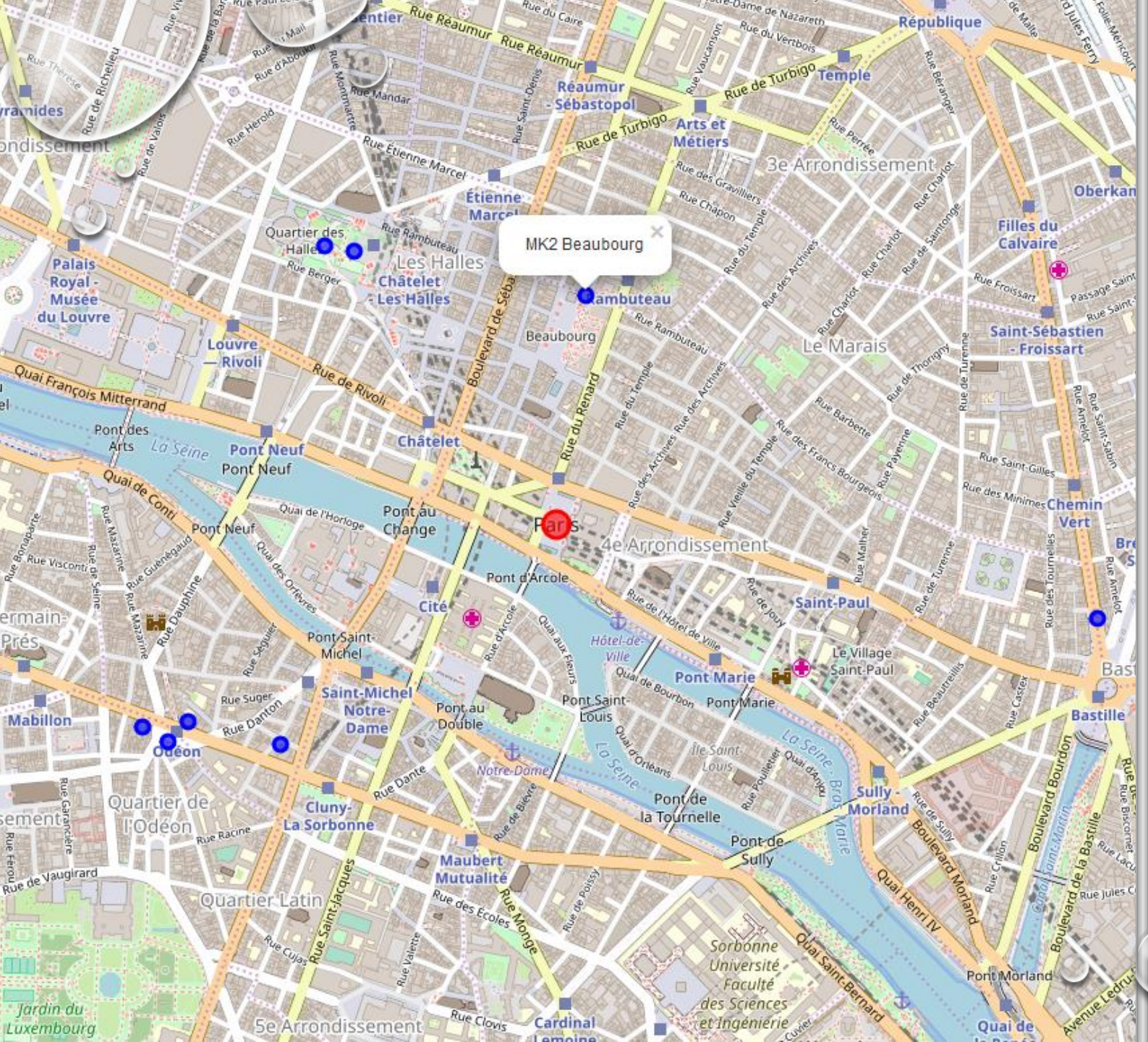

# DATA EXPLORATION

FOURSQUARE CAN HELP US FIND THE TENNIS COURTS INSIDE THE CITY

| | name | categories | address | cc | city | country | crossStreet | distance | formattedAddress | labeledLatLngs | lat | lng | neighborhood | postalCode | state | id | rating |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Trois Fois plus de Piment | Chinese Restaurant | Rue Saint-Martin 184 | FR | Paris | France | Rue aux Ours | 691 | [Rue Saint-Martin 184 (Rue aux Ours), 75003 Pa... | [{'label': 'display', 'lat': 48.8628, 'lng': 2... | 48.862800 | 2.352370 | NaN | 75003 | Île-de-France | 560d03b5498eaea729b6bf10 | 9.1 |
| 1 | Raviolis Chinois Nord-Est | Chinese Restaurant | 115 rue Saint-Denis | FR | Paris | France | NaN | 707 | [115 rue Saint-Denis, 75001 Paris, France] | [{'label': 'display', 'lat': 48.86284447210104... | 48.862844 | 2.349621 | NaN | 75001 | Île-de-France | 557f14fc498e61a8a94c720f | 8.8 |
| 4 | Chez Shen | Chinese Restaurant | 39 rue au Maire | FR | Paris | France | NaN | 964 | [39 rue au Maire, 75003 Paris, France] | [{'label': 'display', 'lat': 48.86485050214374... | 48.864851 | 2.355559 | NaN | 75003 | Île-de-France | 4b3f52dbf964a52094a525e3 | 8.7 |
| 3 | Boutique yam'Tcha | Chinese Restaurant | 4 rue Sauval | FR | Paris | France | NaN | 876 | [4 rue Sauval, 75001 Paris, France] | [{'label': 'display', 'lat': 48.8617098431057,... | 48.861710 | 2.342380 | NaN | 75001 | Île-de-France | 548884fc498e7a7ca55edf84 | 8.6 |
| 5 | Chez Vong | Chinese Restaurant | 10 rue de la Grande-Truanderie | FR | Paris | France | NaN | 707 | [10 rue de la Grande-Truanderie, 75001 Paris, ... | [{'label': 'display', 'lat': 48.86281389498094... | 48.862814 | 2.349376 | NaN | 75001 | Île-de-France | 4adcda11f964a520e73521e3 | 7.8 |

# DATA EXPLORATION

FOURSQUARE CAN HELP US FIND THE NEARBY CHINESE RESTAURANTS AND THEIR RATINGS




# DATA EXPLORATION

FOURSQUARE CAN HELP US FIND WHERE ARE THE NEARBY MOVIE THEATRES

| | numero_quartier | arrondissement | quartier | latitude | longitude | coordonees_centre | travel_time |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | St-Germain-l'Auxerrois | 48.860650 | 2.334910 | [48.8606501352, 2.3349103293] | 12 |
| 1 | 2 | 1 | Halles | 48.862289 | 2.344899 | [48.8622891081, 2.3448988583] | 12 |
| 10 | 11 | 3 | Archives | 48.859192 | 2.363205 | [48.8591924127, 2.3632050573] | 21 |
| 11 | 12 | 3 | Sainte-Avoie | 48.862557 | 2.354852 | [48.862557245, 2.3548515182] | 11 |
| 12 | 13 | 4 | Saint-Merri | 48.858521 | 2.351667 | [48.8585213723, 2.3516669671] | 13 |
| 13 | 14 | 4 | Saint-Gervais | 48.855719 | 2.358162 | [48.8557186509, 2.3581623339] | 17 |
| 14 | 15 | 4 | Arsenal | 48.851585 | 2.364768 | [48.851585175, 2.3647679539] | 21 |
| 15 | 16 | 4 | Notre-Dame | 48.852896 | 2.352775 | [48.8528955862, 2.3527750121] | 19 |
| 16 | 17 | 5 | Saint-Victor | 48.847664 | 2.354093 | [48.8476636355, 2.3540931534] | 25 |
| 17 | 18 | 5 | Jardin-des-Plantes | 48.841940 | 2.356894 | [48.8419401934, 2.3568938896000002] | 29 |
| 18 | 19 | 5 | Val-de-Grace | 48.841684 | 2.343861 | [48.841684288, 2.3438609263] | 25 |
| 19 | 20 | 5 | Sorbonne | 48.849045 | 2.345747 | [48.8490447659, 2.3457466002] | 23 |
| 2 | 3 | 1 | Palais-Royal | 48.864660 | 2.336309 | [48.8646599781, 2.336308919] | 9 |
| 20 | 21 | 6 | Monnaie | 48.854384 | 2.340035 | [48.8543844036, 2.3400353711] | 21 |
| 21 | 22 | 6 | Odeon | 48.847801 | 2.336339 | [48.847800629299996, 2.3363388276] | 26 |
| 22 | 23 | 6 | Notre-Dame-des-Champs | 48.846428 | 2.327357 | [48.846427594, 2.3273568782] | 19 |
| 23 | 24 | 6 | Saint-Germain-des-Prés | 48.855289 | 2.333657 | [48.85528872, 2.3336568681] | 18 |
| 24 | 25 | 7 | Saint-Thomas-d'Aquin | 48.855263 | 2.325588 | [48.8552632694, 2.3255876526] | 17 |
| 25 | 26 | 7 | Invalides | 48.858515 | 2.316445 | [48.8585149661, 2.3164447403] | 20 |
| 26 | 27 | 7 | Ecole-Militaire | 48.850359 | 2.311031 | [48.8503592615, 2.3110309363] | 27 |

# DATA EXPLORATION

CITYMAPPER WILL CALCULATE TRAVEL TIMES FOR US

# DETERMINE THE OPTIMUM NEIGHBORHOOD

## CALCULATING SCORES

$$cultureScore = \begin{cases} 1 & \text{if there is a movie theatre with distance} \leq 1000 \\ 0 & \text{else} \end{cases}$$

| | quartier | numero_quartier | movie_score |
|---|---|---|---|
| 0 | St-Germain-l'Auxerrois | 1 | 1 |
| 1 | Halles | 2 | 1 |
| 2 | Palais-Royal | 3 | 1 |
| 3 | Place-Vendôme | 4 | 1 |
| 4 | Gaillon | 5 | 1 |

$$restaurantScore = \frac{1}{3}\left(\max \sum_{i=1}^{N} restaurantRating\right)$$

| | quartier | numero_quartier | food_score |
|---|---|---|---|
| 0 | St-Germain-l'Auxerrois | 1 | 0.83 |
| 1 | Halles | 2 | 0.89 |
| 2 | Palais-Royal | 3 | 0.85 |
| 3 | Place-Vendôme | 4 | 0.78 |
| 4 | Gaillon | 5 | 0.80 |

$$sportScore = \begin{cases} 1 & \text{if distance} \leq 1000 \\ 1 - \frac{distance - 1000}{1500} & \text{if } 1000 \leq \text{distance} \leq 2000 \\ 0 & \text{if distance} \geq 2500 \end{cases}$$

| | quartier | numero_quartier | sport_score |
|---|---|---|---|
| 0 | St-Germain-l'Auxerrois | 1 | 0.66 |
| 1 | Halles | 2 | 0.42 |
| 2 | Palais-Royal | 3 | 0.36 |
| 3 | Place-Vendôme | 4 | 0.16 |
| 4 | Gaillon | 5 | 0.02 |

$$travelScore = \begin{cases} 1 & \text{if travelTime} \leq 20\text{mn} \\ 1,5 - travelTime/40 & \text{if travelTime} > 20\text{mn} \end{cases}$$

$$\begin{cases} \alpha_1 = 0.4 & \text{the coefficient for the travel score} \\ \alpha_2 = 0.3 & \text{the coefficient for the sport score} \\ \alpha_3 = 0.1 & \text{the coefficient for the restaurant score} \\ \alpha_5 = 0.2 & \text{the coefficient for the culture score} \end{cases}$$

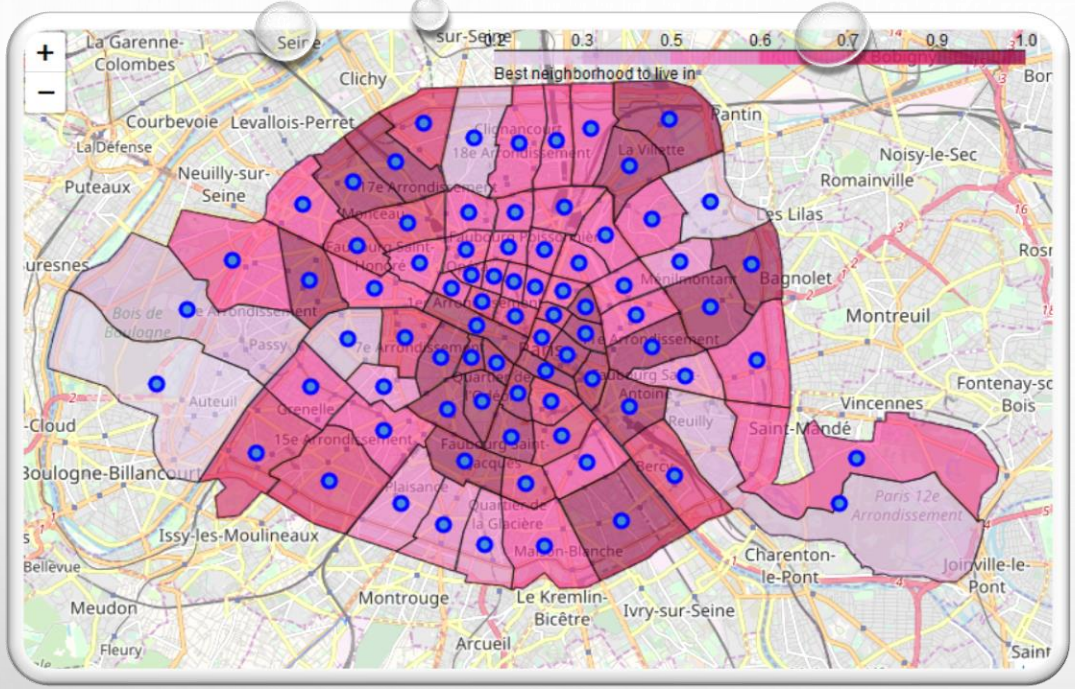
| | quartier | numero_quartier | travel_score |
|---|---|---|---|
| 75 | Combat | 76 | 0.8 |
| 76 | Belleville | 77 | 0.875 |
| 77 | Saint-Fargeau | 78 | 0.9 |
| 78 | Père-Lachaise | 79 | 1 |
| 79 | Charonne | 80 | 0.775 |

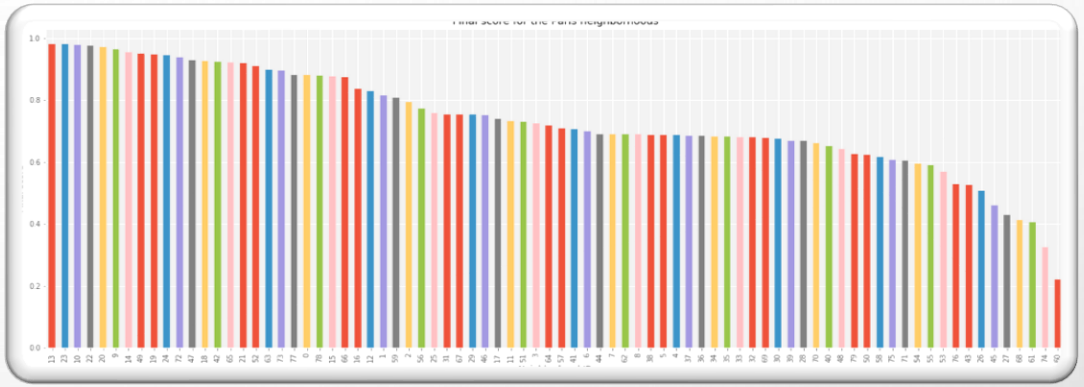
$$neighborhoodScore = \alpha_1 * travelScore + \alpha_2 * sportScore + \alpha_3 * restaurantScore + \alpha_4 * cultureScore$$

CALCULATING SCORES

# DETERMINE THE OPTIMUM NEIGHBORHOOD

| | quartier | numero_quartier | arrondissement | travel_score | sport_score | food_score | movie_score | final_score |
|---|---|---|---|---|---|---|---|---|
| 23 | Saint-Germain-des-Prés | 24 | 6 | 1 | 1 | 0.81 | 1 | 0.981 |
| 13 | Saint-Gervais | 14 | 4 | 1 | 1 | 0.81 | 1 | 0.981 |
| 10 | Archives | 11 | 3 | 0.975 | 1 | 0.88 | 1 | 0.978 |
| 22 | Notre-Dame-des-Champs | 23 | 6 | 1 | 1 | 0.76 | 1 | 0.976 |
| 20 | Monnaie | 21 | 6 | 0.975 | 1 | 0.82 | 1 | 0.972 |
| 9 | Enfants-Rouges | 10 | 3 | 1 | 0.92 | 0.89 | 1 | 0.965 |
| 14 | Arsenal | 15 | 4 | 0.975 | 1 | 0.64 | 1 | 0.954 |
| 49 | Gare | 50 | 13 | 0.925 | 1 | 0.80 | 1 | 0.95 |
| 19 | Sorbonne | 20 | 5 | 0.925 | 1 | 0.77 | 1 | 0.947 |
| 24 | Saint-Thomas-d'Aquin | 25 | 7 | 1 | 0.92 | 0.70 | 1 | 0.946 |



# DETERMINE THE OPTIMUM NEIGHBORHOOD

## CALCULATING SCORES

# RESULTS

- 30 NEIGHBORHOODS WITH A RATING ABOVE 0.8 AND 8 WITH A RATING ABOVE 0.95
- THE TOP SCORES ARE PRETTY CLOSE AND THE RESTAURANT SCORE IS THE ONLY DIFFERENTIATING FACTOR BETWEEN THOSE WELL-RANKED NEIGHBORHOODS
- THE CULTURE AND SPORT SCORE ARE QUITE EFFECTIVE AT DISCRIMINATING GOOD AND BAD NEIGHBORHOODS.
- THE TRAVEL SCORE IS NOT EFFECTIVE AT DISCRIMINATING BECAUSE JEAN-PHILIPPE'S WORKPLACE IS IN A EASILY ACCESSIBLE LOCATION
- THE RESTAURANT SCORE IS NOT SIGNIFICANT AND SKEWED BY THE LACK OF DATA IN CERTAIN NEIGHBORHOODS. MOST VENUES ARE ALSO OVERRATED AND THE VARIATIONS IN RATING ARE NOT MEANINGFUL.

# DISCUSSIONS

- THE RESTAURANT SCORE FORMULA NEEDS TO BE OVERHAULED.
- SOME EXTRA DATA WOULD BE GREAT TO BRING A MEANINGFUL DIFFERENTIATION. IDEALLY WE COULD USE RENT PRICE DATA AND PONDER IT AS MUCH AS TRAVEL TIME.
- USING LOCATION OF APARTMENTS THAT CAN BE RENTED WOULD BE MORE INTERESTING AND MORE ACTIONABLE.


THANKS TO CITYMAPPER,
FOURSQUARE AND
OPENDATAPARIS FOR THEIR DATA