

Data Description and Methodology

The data sources used in implementing this project are:

1. Neighborhood data for New York scraped from Wikipedia
2. Neighborhood data for Paris scraped from Wikipedia

The data scraped from Wikipedia on the New York neighborhoods contain columns describing the following.

1. The name of the neighborhood
2. The population of the neighborhood
3. The area in Km² of the neighborhood
4. The community board to which the neighborhood belongs

The data scraped from Wikipedia on the Paris neighborhoods contain columns describing the following.

1. The name of the neighborhood
2. The population of the neighborhood
3. The area in Km² of the neighborhood
4. The Arrondissement to which the neighborhood belongs

Other columns which are then added to the collected data for each of the neighborhoods include

1. The latitude and longitude of each neighborhood
2. The number of coffee shops open in each neighborhood
3. The number of people served per coffee shop in each neighborhood assuming each coffee shop has an equal chance of serving the same number of people.

A visualization of the number of people potentially served by each coffeeshop shows the potential for profit assuming the coffee shops all sell coffee at the same price.

A higher number of people served per coffeeshop means higher earning potential if we cite our coffeeshop in that neighborhood, hence encouraging the investor to open the coffeeshop in that neighborhood.