# The Battle of the Neighborhoods

In this project we will try to find an optimal location for an Outfit store. Specifically, this report will be targeted to stakeholders interested in opening an Clothing Shop in Downtown Toronto, Canada.

Since there are lots of clothing stores in Toronto we will try to detect locations that are not already crowded with such outlets. We are also particularly interested in areas with no outfit store in vicinity. We would also prefer locations as close to city center as possible, assuming that first two conditions are met.

We will use our data science powers to generate a few most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

## Data

Based on definition of our problem, factors that will influence our decission are:

* number of existing stores in the neighborhood (any type of store)
* number of and distance to Clothing stores in the neighborhood, if any
* distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

* centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Geopy reverse geocoding
* number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
* coordinate of Downtown Toronto will be obtained using Geopy geocoding of well known toronto location (Richmond street)

### Neighborhood Candidates

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Toronto city center.

Let's first find the latitude & longitude of Toronto city center, using specific, well known address and Geopy geocoding API.

Now let's create a grid of area candidates, equaly spaced, centered around city center and within ~6km from Richmond street. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

### Foursquare

Now that we have our location candidates, let's use Foursquare API to get info on stores in each neighborhood.

We're interested in venues in 'shop' category, but only those that are proper shops - antique shop, automative shop etc. are not direct competitors so we don't care about those. So we will include in out list only venues that have 'clothing shop' in category name, and we'll make sure to detect and include all the subcategories of specific 'clothing stores' category, as we need info on clothing stores in the neighborhood.

So now we have all the stores in area within few kilometers from Richmond Street, and we know which ones are clothing stores! We also know which restaurants exactly are in vicinity of every neighborhood candidate center.

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new clothing store!

## Methodology

In this project we will direct our efforts on detecting areas of Toronto that have low store density, particularly those with low number of clothing. We will limit our analysis to area ~6km around city center.

In first step we have collected the required data: location and type (category) of every store within 6km from center (Richmond Street). We have also identified clothing stores (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'store density' across different areas of Toronto - we will use heatmaps to identify a few promising areas close to center with low number of stores in general (and no clthing stores in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two stores in radius of 250 meters, and we want locations without clothing stores in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

## Analysis

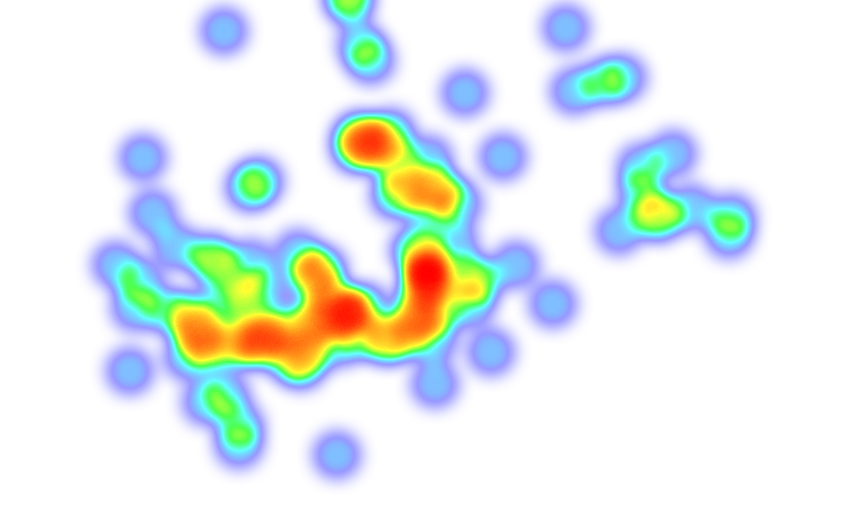
Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the number of stores in every area candidate:

OK, so on average clothing store can be found within ~500m from every area center candidate. That's fairly close, so we need to filter our areas carefully!

Let's crete a map showing heatmap / density of stores and try to extract some meaningfull info from that. Also, let's show borders of Toronto boroughs on our map and a few circles indicating distance of 1km, 2km and 3km from Richmond street.

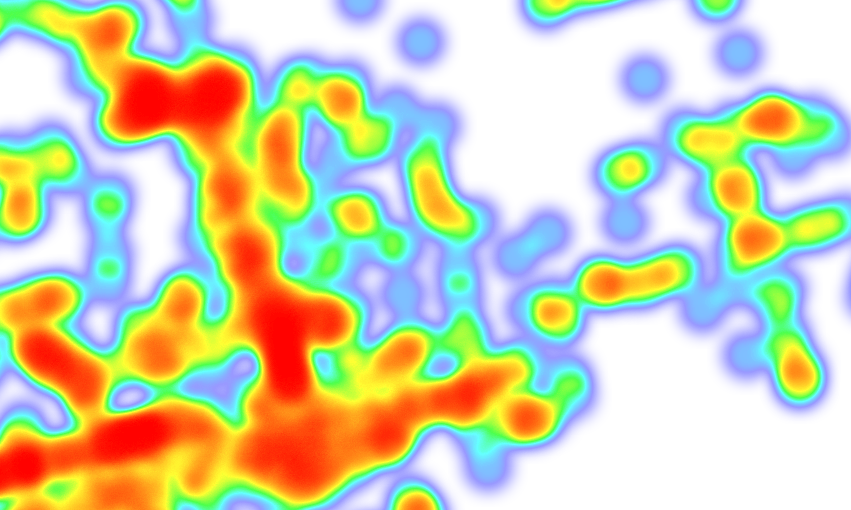
Looks like a few pockets of low res density closest to city center can be found north, north-east and south-east from Richmond Street.

Let's create another heatmap map showing heatmap/density of clothing\_stores only.



This map is not so 'hot' (clothing store represent a subset of ~20% of all stores in Toronto) but it also indicates higher density of existing clothing stores directly west and east from Richmond Street, with closest pockets of low clothing store density positioned east, south-east and south from city center.

Based on this we will now focus our analysis on areas north, south, south-east and north-east from center - we will move the center of our area of interest and reduce it's size to have a radius of 2.5km. This places our location candidates mostly in boroughs Kreuzberg and Friedrichshain (another potentially interesting borough is Prenzlauer Berg with large low store density north-east from city center, however this borough is less interesting to stakeholders as it's mostly residental and less popular with tourists).



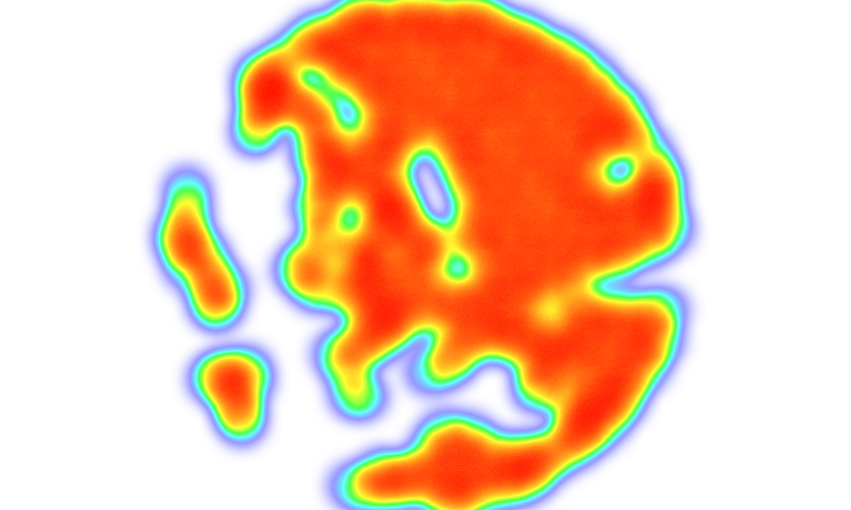
Not bad - this nicely covers all the pockets of low store density in Kreuzberg and Friedrichshain closest to center.

Let's also create new, more dense grid of location candidates restricted to our new region of interest (let's make our location candidates 100m appart).

Let us now filter those locations: we're interested only in locations with no more than two stores in radius of 250 meters, and no clothing stores in radius of 400 meters.

Looking good. We now have a bunch of locations fairly close to Alexanderplatz (mostly in Kreuzberg, Friedrichshain and south-east corner of Mitte boroughs), and we know that each of those locations has no more than two restaurants in radius of 250m, and no Italian restaurant closer than 400m. Any of those locations is a potential candidate for a new Italian restaurant, at least based on nearby competition.

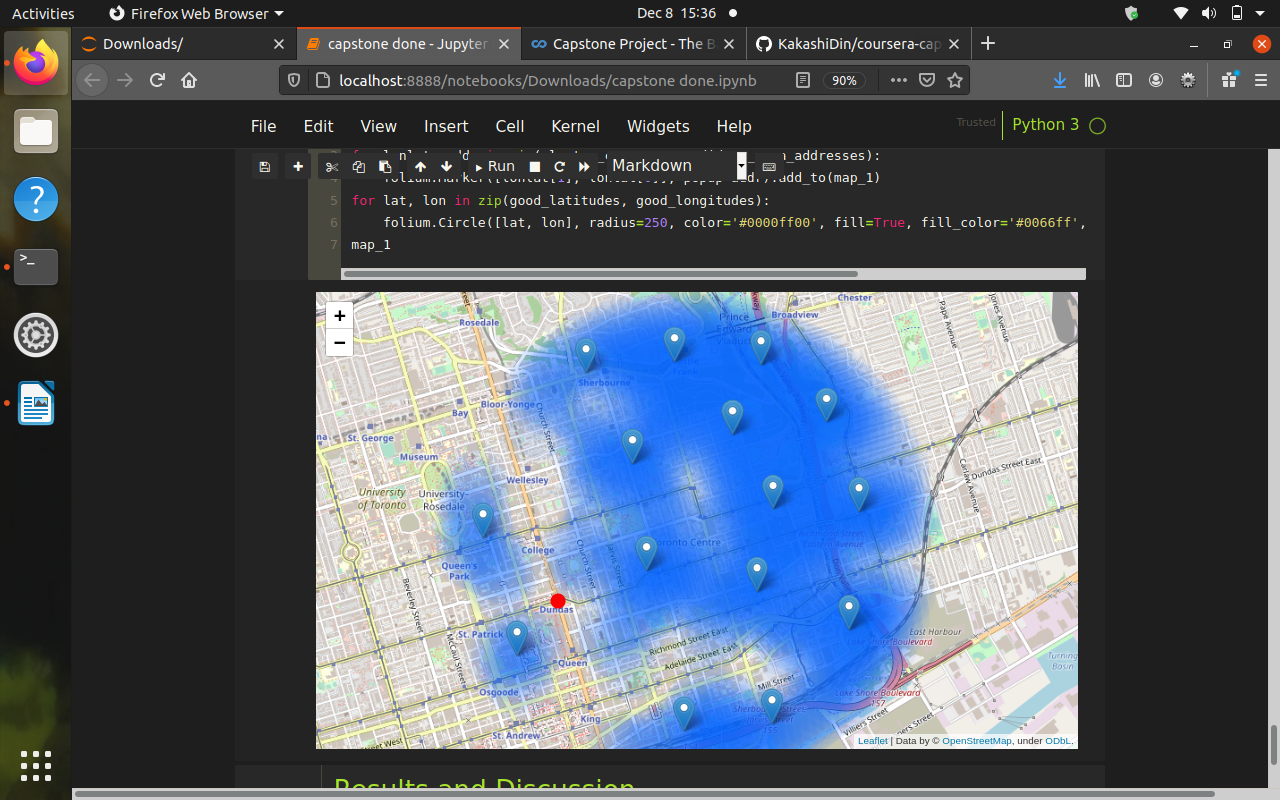
Let's now show those good locations in a form of heatmap:



Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

Let's see those zones on a city map without heatmap, using shaded areas to indicate our clusters:



## **Results and Discussion**

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of stores and no clothing stores nearby, all zones being fairly close to city center (all less than 4km from Richmond street, and about half of those less than 2km from Richmond street). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential store locations. Most of the zones are located in Kreuzberg and Friedrichshain boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

Our analysis shows that although there is a great number of restaurants in Toronto (~2000 in our initial area of interest which was 12x12km around Richmond street), there are pockets of low store density fairly close to city center. Highest concentration of restaurants was detected north and west from Richmond Street, so we focused our attention to areas south, south-east and east, corresponding to boroughs Kreuzberg, Friedrichshain and south-east corner of central Mitte borough. Another borough was identified as potentially interesting (Prenzlauer Berg, north-east from Richmond street), but our attention was focused on Kreuzberg and Friedrichshain which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics and a number of pockets of low store density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Richmond street) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two stores in radius of 250m and those with an clothing stores closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new store locations based on number of and distance to existing venues - both stores in general and clothing stores particularly. This, of course, does not imply that those zones are actually optimal locations for a new store! Purpose of this analysis was to only provide info on areas close to center but not crowded with existing stores (particularly clothing) - it is entirely possible that there is a very good reason for small number of stores in any of those areas, reasons which would make them unsuitable for a new store regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

## Conclusion

Purpose of this project was to identify Toronto areas close to center with low number of stores (particularly clothing stores) in order to aid stakeholders in narrowing down the search for optimal location for a new clothing stores. By calculating store density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Kreuzberg and Friedrichshain), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby stores. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission on optimal store location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.