

# Find an Optimal Location for a Hotel in Beijing

Applied Data Science Capstone by IBM/Coursera

# Introduction: Business Problem

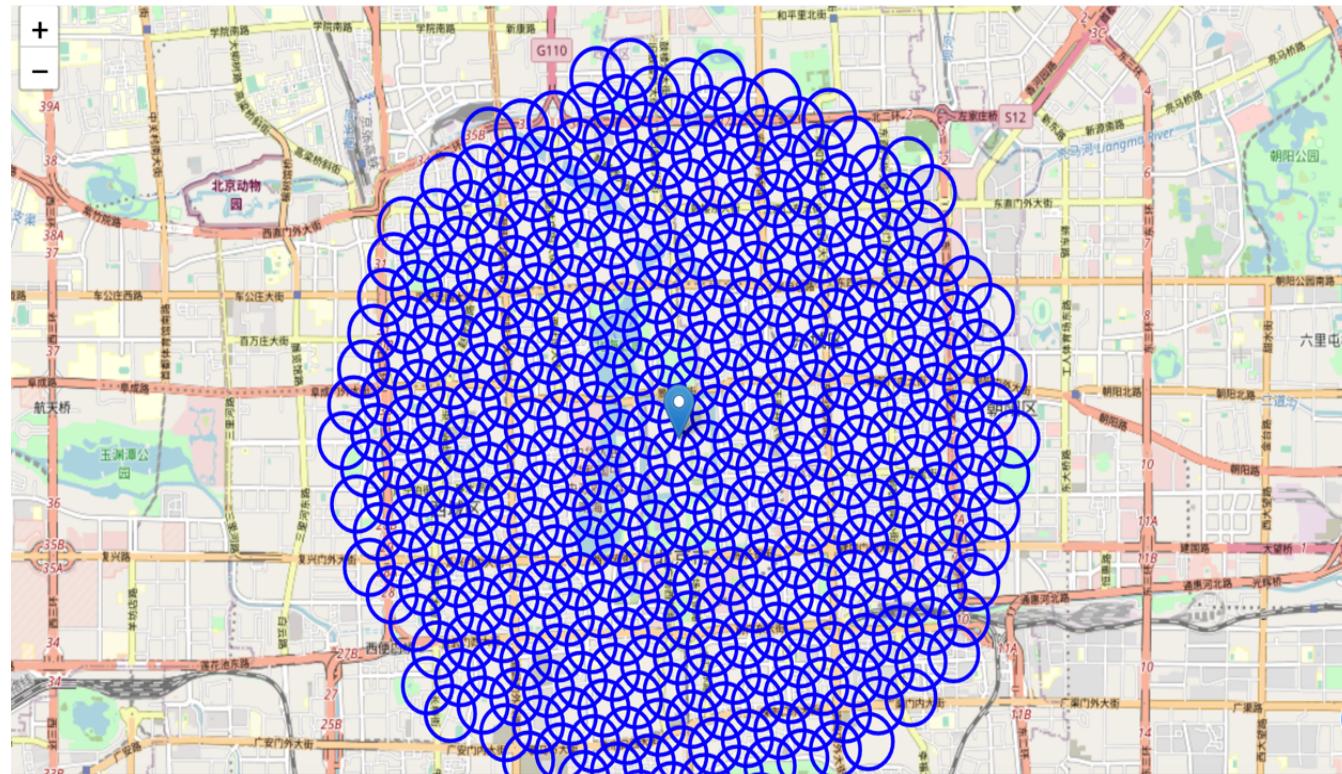
- In this project we will try to find an optimal location for a hotel. Specifically, this report will be targeted to stakeholders interested in opening a **hotel** in **Beijing**, China.
- Since there are lots of hotels in Beijing we will try to detect **locations that are not already crowded with hotels**. 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 promising neighborhoods based on this criteria.

# Data

- Based on definition of our problem, factors that will influence our decision are:
  - number of existing hotels in the neighborhood
  - number of and distance to hotels 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 **Google Maps API reverse geocoding**
  - number of hotels and their location in every neighborhood will be obtained using **Foursquare API**
  - coordinate of Beijing center will be obtained using **Google Maps API geocoding** of well known Beijing location (the Forbidden City)

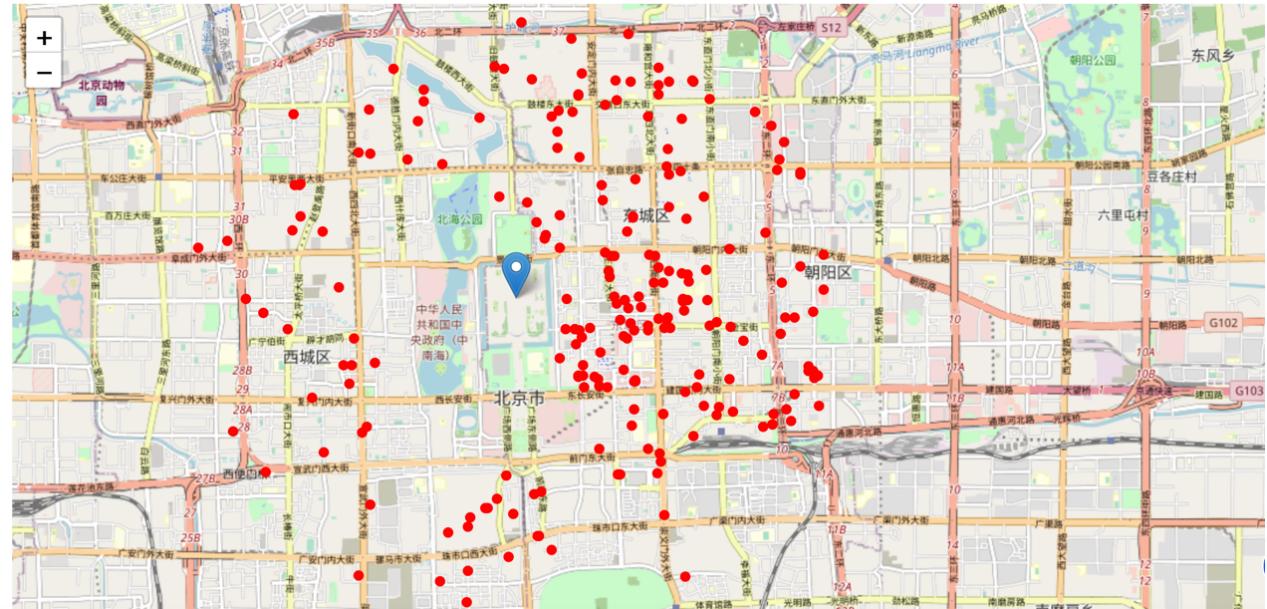
# Neighborhood Candidates

- Let's visualize the data we have so far: city center location and candidate neighborhood centers:



# Foursquare

- Now that we have our location candidates, let's use Foursquare API to get info on hotels in each neighborhood.
- We're interested in venues in 'hotel' category. Total: 224 Hotels data loaded.



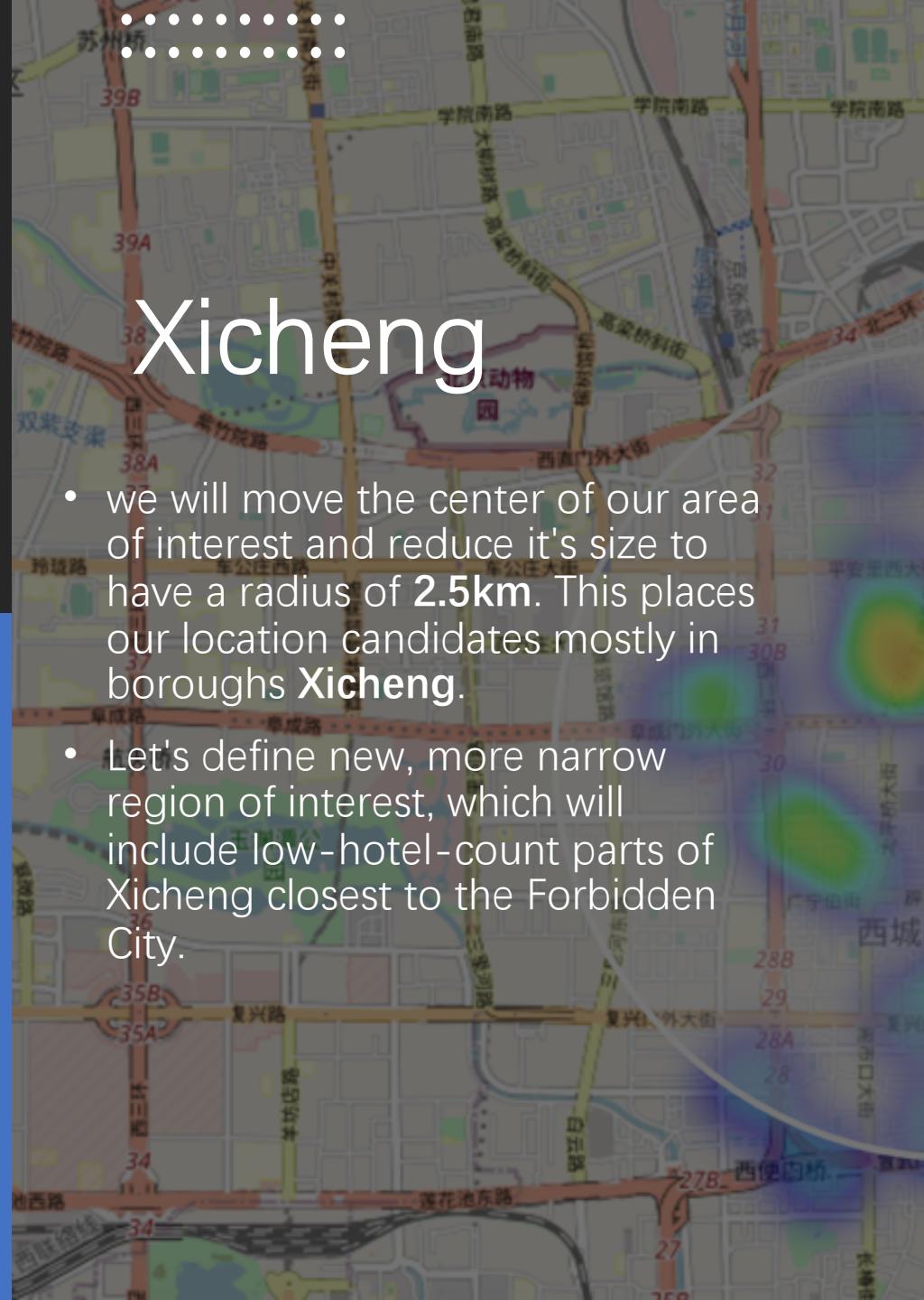
# Methodology

- In this project we will direct our efforts on detecting areas of Beijing that have low hotel density. 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 hotel within 6km from Beijing center**(the Forbidden City).
- Second step in our analysis will be calculation and exploration of '**hotel density**' across different areas of Beijing - we will use **heatmaps** to identify a few promising areas close to center with low number of hotels in general 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 hotels in radius of 250 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 hotels in every area candidate
- Looks like a few pockets of low hotels density closest to city center can be found **west** from the forbidden city.
- This map is not so 'hot' but it also indicates higher density of existing hotels directly **east** from the Forbidden City, with closest pockets of low hotel density positioned **south-west** from city center.





# Xicheng

- 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 **Xicheng**.
- Let's define new, more narrow region of interest, which will include low-hotel-count parts of Xicheng closest to the Forbidden City.



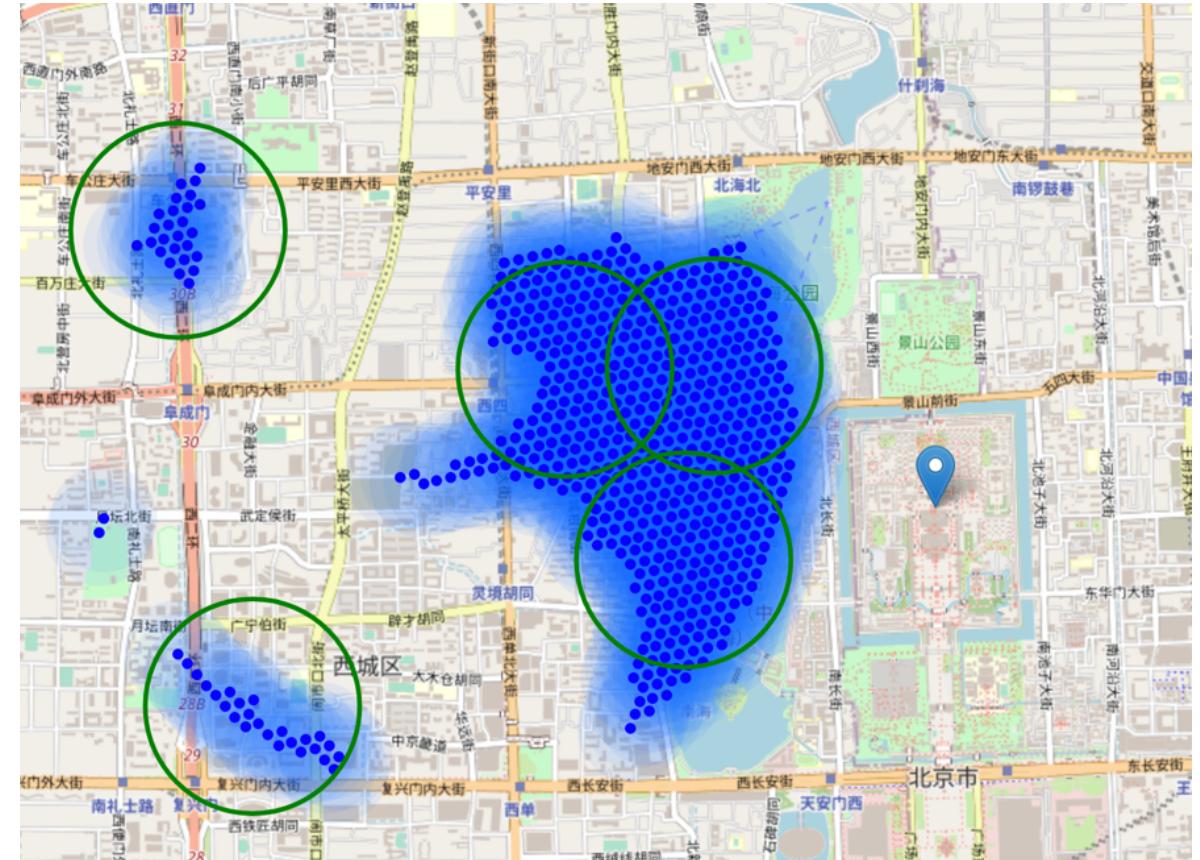
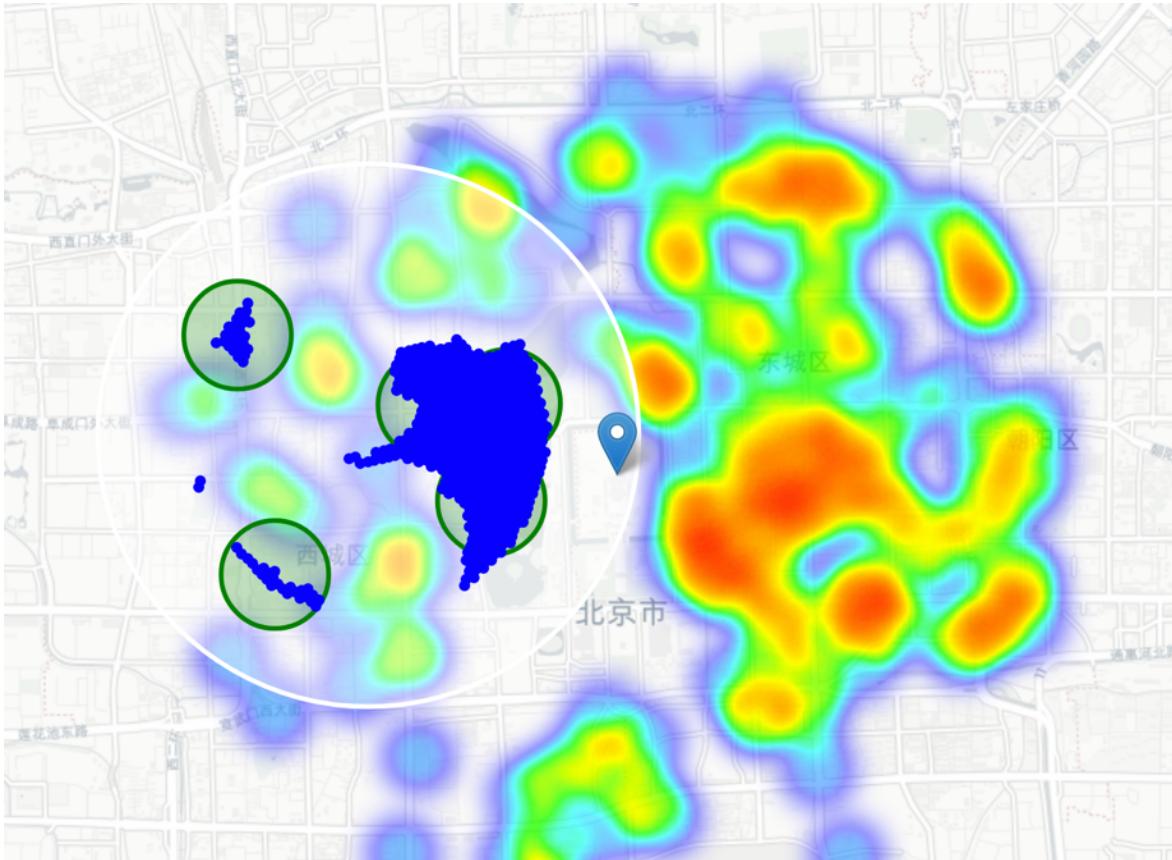
# Discover locations in Xicheng

- Let us now **filter** those locations: we're interested only in **locations with no more than one hotel in radius of 250 meters**.



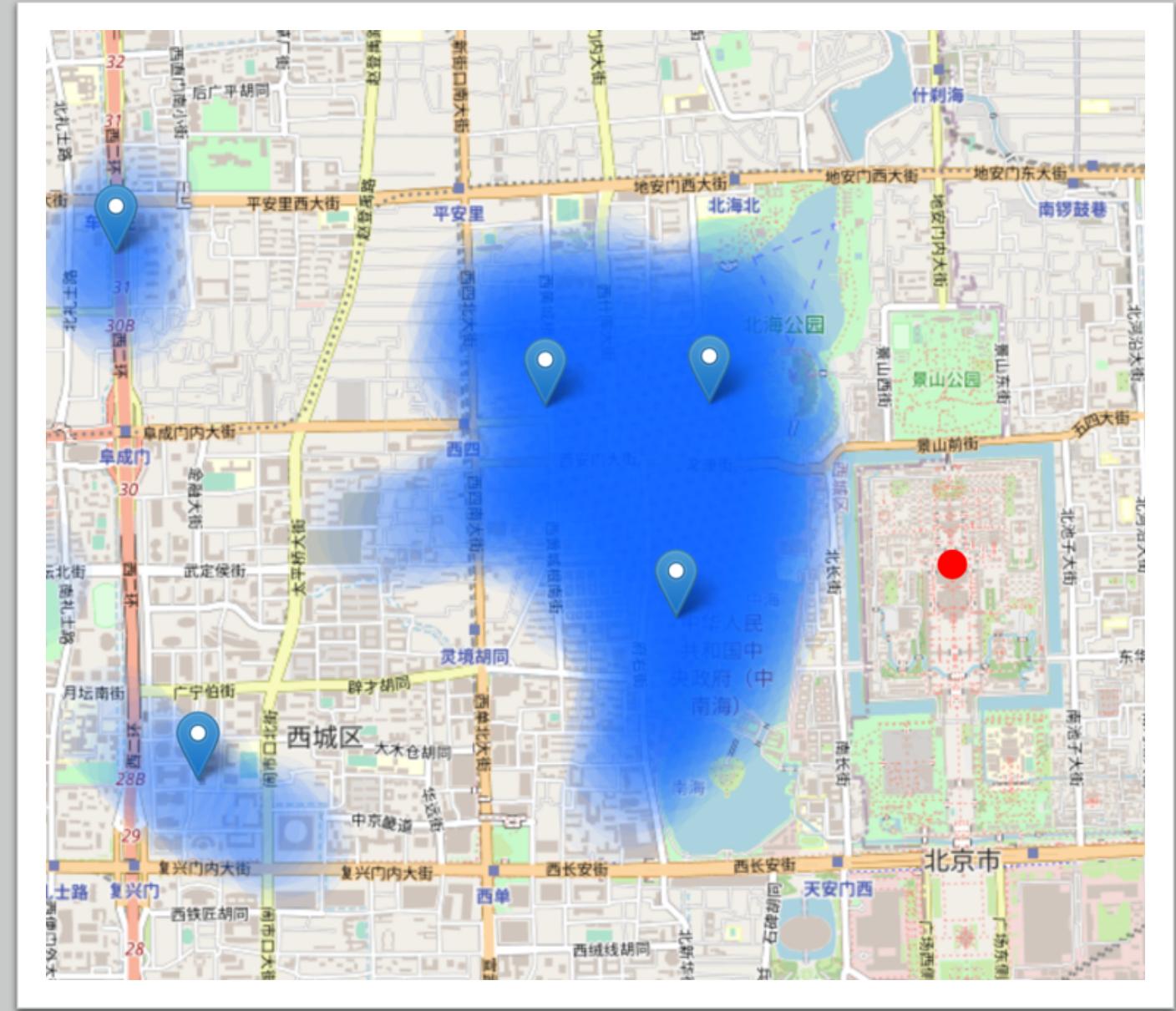
# Cluster

Let us now **cluster** those locations to create **centers of zones** containing good locations



# Final Results

- Finally, let's reverse geocode those candidate area centers to get the addresses
  - 北海公园
  - 西黄城根北街
  - 中华人民共和国中央政府（中南海）
  - 北京市公安局公共交通管理局，
  - 都城隍庙后殿



# Conclusion

Purpose of this project was to identify Beijing areas close to center with low number of hotels in order to aid stakeholders in narrowing down the search for optimal location for a new hotel. By calculating hotel density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Xicheng), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby hotels. 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.