

Data tells where to build new hotels in Berlin

I) Introduction

1. Background

Berlin, the capital and the largest city of Germany, has always been among the top popular cities. The city attracts millions of tourists every year to explore their cultural richness. Apart from that, Berlin is also a major center of politics, media, and science. Therefore, the need for accommodation in this city is always on high demand.

An entrepreneur wants to invest in building up a chain of hotels in Berlin, Germany. He is new to this city, so he wants to find the best area in Berlin where he should set up the hotel chain. The common sense for solving this problem is to have the hotel nearest to the famous tourist attractions. Other suggestion could be close to main public stations. However, every city has their own hidden gems, where can attract more travelers than it looks like, which can be an extraordinary sight or a coffee shop. Let's take an example in Berlin. Being so well-known for the Brandenburg Gate or the Berlin Wall, but it should not deny that a Hatch Sticker Museum could also attract many people to come and visit the art and history of the graffiti sticker. Or perhaps the best place to try currywurst in Berlin could be the small food stall on the street. The travelers who visited Berlin previously, are those who can tell us what the best places to be in Berlin are.

On the other hand, social media has become an integral part of all of our lives. People use social media to connect with friends, to share experience, or to catch up with trends. As such, tourists have also shared their experiences by providing likes, comments or ratings for the places they have visited. Thanks to that, business has an extremely important source of data to explore. Understanding that popular trend, we decide to use the data from people visiting Berlin on Foursquare to find where this entrepreneur should set up his hotel chain.

2. Problem

How to find the best possible places in Berlin to build up a new hotel chain in Berlin using the data retrieved from Foursquare?

3. Solution

So the proposed idea in order to solve this challenge for the entrepreneur is to build the **hotel chain** in the location that in the **centroids** of interesting places which receive **high ratings** from travelers on Foursquare. Since Berlin is a big city with the area of approximately 900 km², we initially plan to have at least 5 hotels in the city. As such, we will use the *k-Clustering* to identify the 5 main clusters, and use the centroids of these clusters to suggest as the best place to build hotels

4. Assumption

To do that, several assumptions have been made for this project:

- A place is considered as interesting if the rating is 7.5 (out of 10) and above
- The entrepreneur has sufficient capital, so we do not consider the impact of capital in decision making process
- Considering the area of Berlin is about 891.8km², we make the assumption that the most interesting places should be within the radius of 10 km from Berlin city center
- With the sandbox account, we have the limit of the premium calls, so we are unfortunately unable to explore the full set of data.