Determining the best place in New York parks to open a mobile barista

1 Introduction

1.1 Background

New York is a big modern city with huge sky scrapers and a lot of going around. However, there are a lot of green spaces in the city as well, where people like to spend their free time and enjoy the time outside. It is common for people to enjoy some snacks, ice-cream or coffee whilst being outside, but what if a coffee-shop is not present nearby? Or the one that's very close has poor reputation or poor customer-service.

1.2 Problem

To tackle this problem, we want to start a mobile barista business. This would help to tackle the problems where there's no opportunities to establish a traditional coffee house, e.g. Central park, since you can't just build a building in the middle of it. In this project, we will try to identify the parks in New York, where you can't get decent coffee around and identify how many mobile baristas locations should we have if we want to cover the park fully without having to walk more than 500 meters.

1.3 Interest

The interested party in this case is us, since we want to start the business. Such a data science project will help to identify what are the best options for the start of the mobile barista business and how small or big we should start of.

2 Data

2.1 Data sources

The data for this project will be obtained from Foursquare as well as https://data.cityofnewyork.us where some open data about New York is available. The data from Foursquare will be used to identify whether a certain park has good coffee-shops nearby based on their rating on Foursquare. Unrated coffee-shops will be considered as poor quality coffee-shops and thus won't be considered as competition for the mobile barista. Also, coffee-shops not within walking distance of 500 metres from a specific park will also be ignored.

The data about park locations will come from the open data portal mentioned above. The coordinates of parks in NYC are given a geojson file format, which fits well for the analysis we intend to perform.

The issues faced here is that Foursquare only allows 500 "premium" calls to their API to get places ratings and more details etc. For this project, the data was stored on computer after the quota is ran out. The main data obtained from Foursquare are venues category, rating and count of likes. The use of it is explained in the following section.

3 Methodology

The way we're tackling this problem is to identify the parks that have no coffee shops/cafes within 500 metres (or the ones present have no rating). This will be the potential list of parks that may be of interest to us. The next step is to identify the parks that have very good rated coffee shops/cafes within 500 meters and analyse what other well rated venues there are around to identify whether there's a common trend, e.g. there might be some well-known clothing stores, which attract a lot of clients. Such a place would be potential for a mobile barista, since there will be people around willing to spend money.

After we identify such trending places within the parks with good coffee shops, we can scan the remaining parks, to see whether there are similar venues there, which would indicate the potential of having our mobile barista positioned there.

4 Results

Once we got the data from New York city open data portal, we have identified that there are 103 parks listed in the dataset. For each park we have extracted the data from Foursquare about the venues within 500 meters of the central point of park. This gave us 2684 venues to work with. The next step is to find the best rated coffee shops/cafes near the parks. We select top 5 rated coffee shops like that:

Park_name	Venue	Rating
Battery Park	Starbucks	9.2
Morningside Park	Up Coffee Co.	8.9
Starlight Park	Prospect Coffee Shop	8.4
Eastern Parkway	Jenny Coffe Shop #2	8.2
Washington Square Park	The Uncommons	8.0

As one can see, the best rated venue is Starbucks close to the Battery Park. Let's look at these 5 parks to see which venues excluding the coffee shops are well rated there:

```
top_rated_venues_not_caffe[top_rated_venues_not_caffe['Park_name'].apply(lambda x: x in top5_parks)]['Venue Category'].
```

```
array([list(["Women's Store", 'Wine Shop', 'Wine Bar', 'Supplement Shop', 'Steakhouse']),
       list(['Yoga Studio', 'Wings Joint', 'Vegetarian / Vegan Restaurant', 'Thai Restaurant', 'Sushi Restaurant']),
       list(['Wine Shop', 'Wine Shop', 'Wine Bar', 'Wine Bar', 'Sports Bar']),
       list(['Video Game Store', 'Supplement Shop', 'Supermarket', 'Shopping Mall', 'Shoe Store']),
       list(['Yoga Studio', 'Wine Shop', 'Wine Bar', 'Wine Bar', 'Wine Bar'])],
      dtype=object)
```

Again, we select top 5 venues from each park and determine their categories. This will give us some sense what is around the coffee shops, to make our sweet spot choice for mobile barista easier. As one can see women's store, yoga studio, wine bar/wine shop is quite

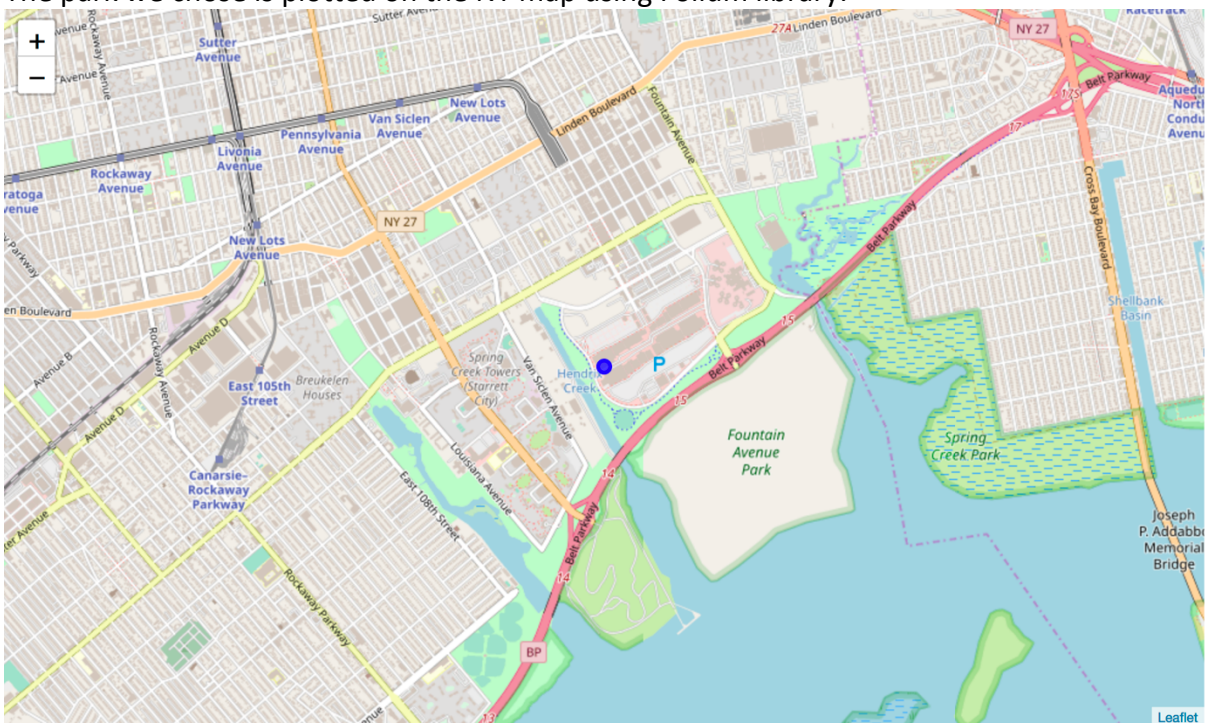
frequent among the best rated venue. Let's have a look at the potential list of parks, that have any of these venues:

```
parks_venues.apply(lambda x: lookup_venues(x)).sort_values(ascending=False)[:3]
```

```
Park_name
Spring Creek Park     3
Franz Sigel Park      2
Highbridge Park       1
Name: Venue Category, dtype: int64
```

We can see that several parks, which don't have coffee shops, have venues like the ones we are interested in, but the most of them are situated close to the Spring Creek Park. Let's get the data for that park and see where the park is located, since I'm not a New Yorker, and have no idea where this is. Let's remember, that data helps us to make decisions, but you don't have to follow them 100% if the result doesn't look logical from the business perspective.

The park we chose is plotted on the NY map using Folium library:



It looks like this could be a good spot for our mobile barista, as there's quite a bit of green spaces, which always a lot of people, which is good for us.

5 Discussion

We have noted that there are a lot of parks within New York and thus a lot of potential places for our future mobile baristas business. We can narrow down the list of potential areas based on the criteria that we choose for our business and this is where the data science skills help. As long as we can obtain the data necessary to support our hypotheses, we can run an experiment to see whether it's true or not, without actually having to try it out in the real world. The interesting observation made here is that there are lot of parks, where there are plenty of coffee shops that are already well rated. Maybe the next step could be to reduce

the radius to only 100 or 200 meters, so we definitely get that competitive edge of people not having to leave the park to enjoy the coffee.

6 Conclusion

In this small study we have seen how effectively the data can help us to make decisions. We have also seen that there may be data sources that you have never thought of us and they're openly available to you to do the analysis that you need. Joining data analysis with business perspective, can give you the edge while competing with other businesses and thus enable you to have better opportunities to win. In our case we were able to drill down the list of 103 potential parks to just 2-3 parks, where we could potentially have a better opportunity to open our mobile barista businesses. And this came from the understanding that we need to look into why other venues are successful, but we barely scratched the surface, there are plenty more options to explore. For example, a very powerful insight might be gained from doing text analytics on customer reviews and check in data in "best performing" and "worst performing" coffee shops, to understand what people would expect from your business and where the bar is set.