Capstone Project:
Identify what venues or shops could be opened at different locations
(Wolf-Dietrich Zabka)

Introduction and business problem

Small businesses are the backbone of the economy: They provide most of the services we need in our everyday life. For possible founders of new restaurants or other venues one of the first and hardest choices is to pick the right place. You need to have a solid customer base in the neighborhood you locate your business in, so you should locate your business in an area where you know people could use your venue and would appreciate its availability. On the other hand, you want that the service you provide will be in sufficient demand. Sure Chinatown is a great place for a Chinese Restaurant, but is it a great place to open another one? In this project, we aim to identify promising neighborhoods for venues.

Data to use

We will use the clustered neighborhoods of Manhattan as an example. Previously we clustered the neighborhoods of Manhattan to identify similar neighborhoods. The customer base within one cluster might be very similar. For each cluster, we identify how common certain venues are. For each venue in a cluster we calculate the average rank how common it is. In a next step, we look at the ranks of the same venues in each neighborhood within the cluster. Based on the current rank of a venue in a neighborhood and the average rank of the same venue in the cluster we will define a score that expresses the need to have a new venue in this neighborhood. With this, we can identify new venue options for a neighborhood which will be in demand.

Cluster 1

```
anhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue
1	Chinatown	Bakery	Chinese Restaurant	Cocktail Bar	Spa	Salon / Barbershop	Optical Shop	Bubble Tea Shop	Dessert Shop	Ice Cream Shop	American Restaurant	Hotpot Restaurant	Vietnamese Restaurant	Bar
5	Manhattanville	Seafood Restaurant	Coffee Shop	Italian Restaurant	Mexican Restaurant	Food & Drink Shop	Lounge	Bike Trail	Sushi Restaurant	Supermarket	Boutique	Gastropub	Bus Station	Café
6	Central Harlem	African Restaurant	Chinese Restaurant	Bar	Seafood Restaurant	French Restaurant	American Restaurant	Cosmetics Shop	Café	Cafeteria	Fried Chicken Joint	Spa	Caribbean Restaurant	Boutique
8	Upper East Side	Italian Restaurant	Exhibit	Coffee Shop	Bakery	Gym / Fitness	Spa	French Restaurant	Juice Bar	Yoga Studio	Hotel	American Restaurant	Salad Place	Grocery Store

Fig 1: Data of neighborhood cluster 1 used for the start of the analysis

Methodology section

We start with clustered neighborhoods of Manhattan as previously in the course described. To describe the methodology, we will focus here on cluster 1.

1. We add the 300 most common venues to the existing clustering algorithm.
2. We make a list of all existing venues in the cluster (a list of all venues existing for all neighborhoods in the cluster)
3. Each venue is assigned a rank in each neighborhood. The most common venue gets the rank 1. The least common venue gets the rank 300. If a venue does not exist, it gets the rank 300 assigned.
4. For each venue, we calculate its average rank in the cluster
5. The “need for a new venue”-score in a neighborhood is defined as “average rank in cluster” - “Current rank in neighborhood”
6. The venues with the highest “need for a new venue” will be listed

Results

We make a histogram of the “need for a new venue”-score for all neighborhoods. The score is normal-distributed. A low score for a venue below 0 indicates that this venue exists in a sufficient manner in a neighborhood. If the score for a venue in a neighborhood is positive, this type of venue is underrepresented in a neighborhood. If the score is above 150, it would be great to open a new venue in the respective neighborhood. Here I give for four neighborhoods the most-urgently needed venues:

- In the neighborhood: Manhattanville, we recommend to open:
  Pizza Place, Thai Restaurant
- In the neighborhood: Central Harlem, we recommend to open:
  Italian Restaurant, Mexican Restaurant, Sushi Restaurant, Thai Restaurant
- In the neighborhood: Upper East Side, we recommend to open:
  Ice Cream Shop, Park
- In the neighborhood: Yorkville, we recommend to open:
  Hotel

For all other neighborhoods, the results can be found in the python notebook.

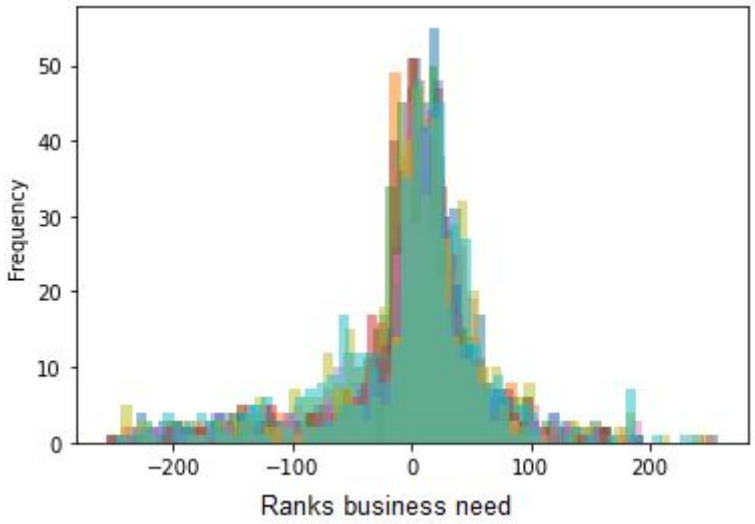

Fig 2: Distributions of occurring ranks of venues for all neighborhoods

Discussion

This method can generally be used to identify venues which are underrepresented in certain areas. If a person wants to open a new venue, the provided score gives solid evidence if there is a business need for the venue in the intended location. Further, it enables to proactively give business ideas based on what venues exist in similar neighborhoods, but are underrepresented in the neighborhood in scope.

Conclusion

A method was shown to identify similar neighborhoods based on existing venues. Due to comparison of the prevalence of existing venues, it is possible to point out underrepresented venues of neighborhoods.