

Restaurant Recommender System

Applied Data Science Capstone Project

Introduction

- Current gastronomical scenario:
 - Highly appreciated as one of main kinds of leisure.
 - Global cultural interchange provides massive variety.
- Problem: Massive offer makes difficult for costumer to find places of his taste.
- Solution: A recommendation system can be developed to help people find restaurants nearby that attend to their preferences, and be available as a mobile app.
- Interest:
 - Users can find more easily restaurants by their preference.
 - Restaurants will get publicity.

Data acquisition and cleaning

- Data sources:
 - Foursquare API will be used to locate and extract information on restaurants near the user.
 - User coordinates will be assume to be acquired by device GPS.
 - For this instance, the user volunteer lives in Rio de Janeiro, Brazil, whose data were manually gathered. (It's important to consider that, in the case of an actual business, the same data could be automatically gathered by Foursquare, under a premium account.)

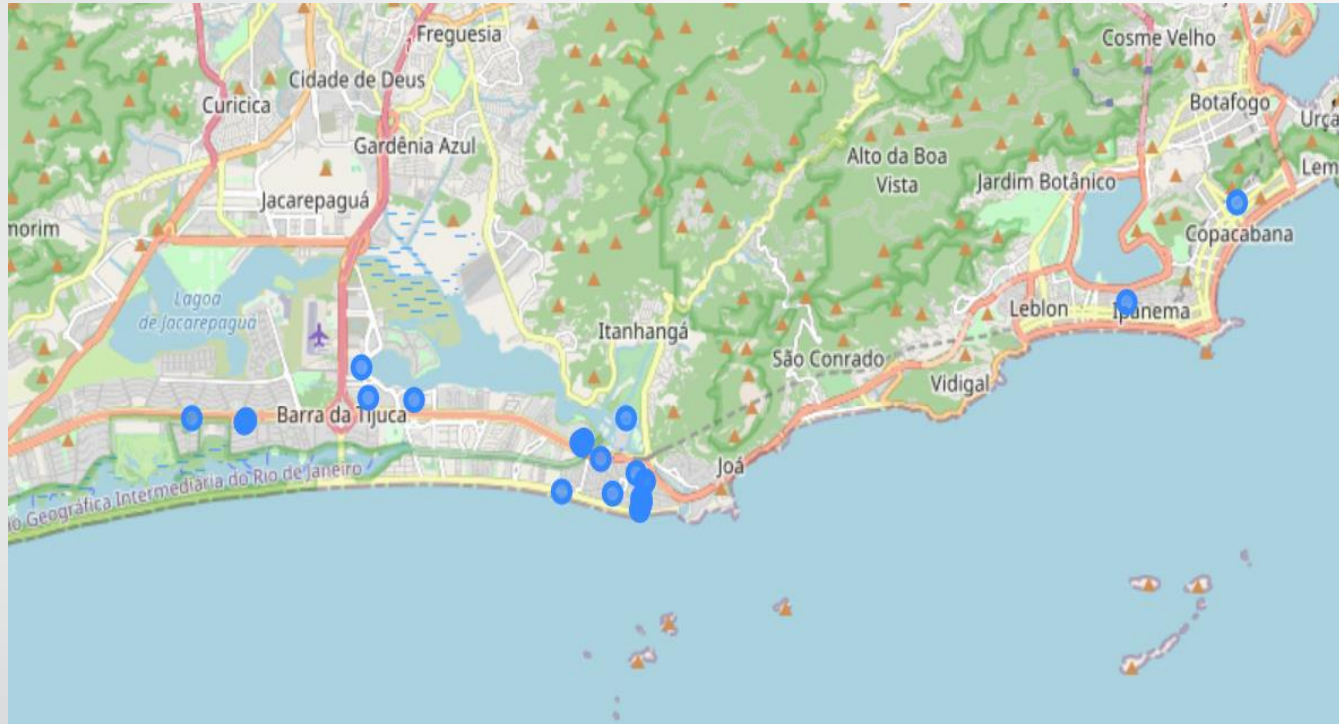
Data acquisition and cleaning

- User experience features and variable type:

Venue id	Name	Rate	Category	Category id	Latitude	Longitude
String	String	Int (0 to 5)	String	String	Float	Float

- The venues' features returned by the explore query were filtered and modelled the same way as the user experience's.

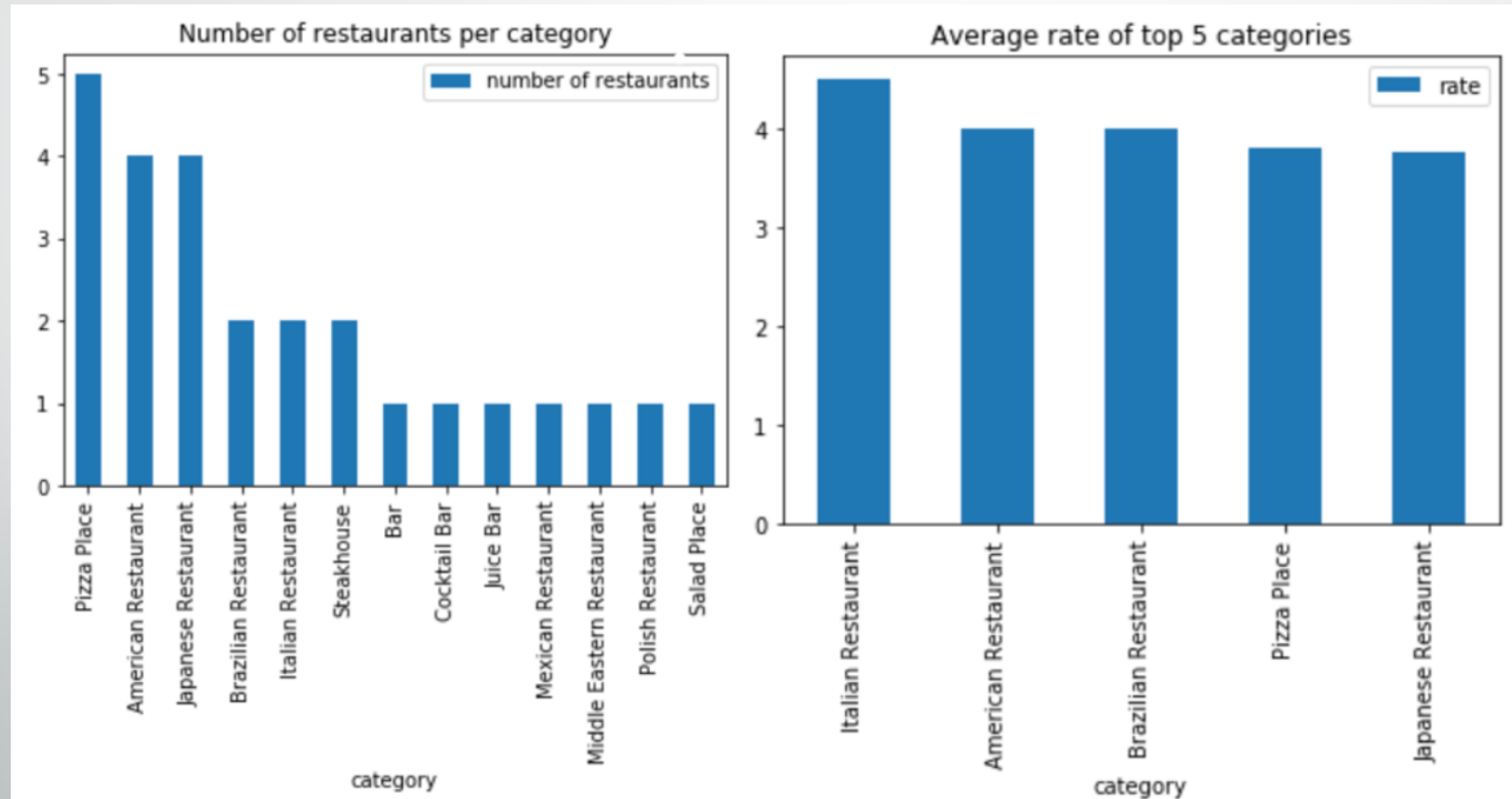
Exploratory Data Analysis



- That higher density is no coincidence, as being the area where the user lives in.

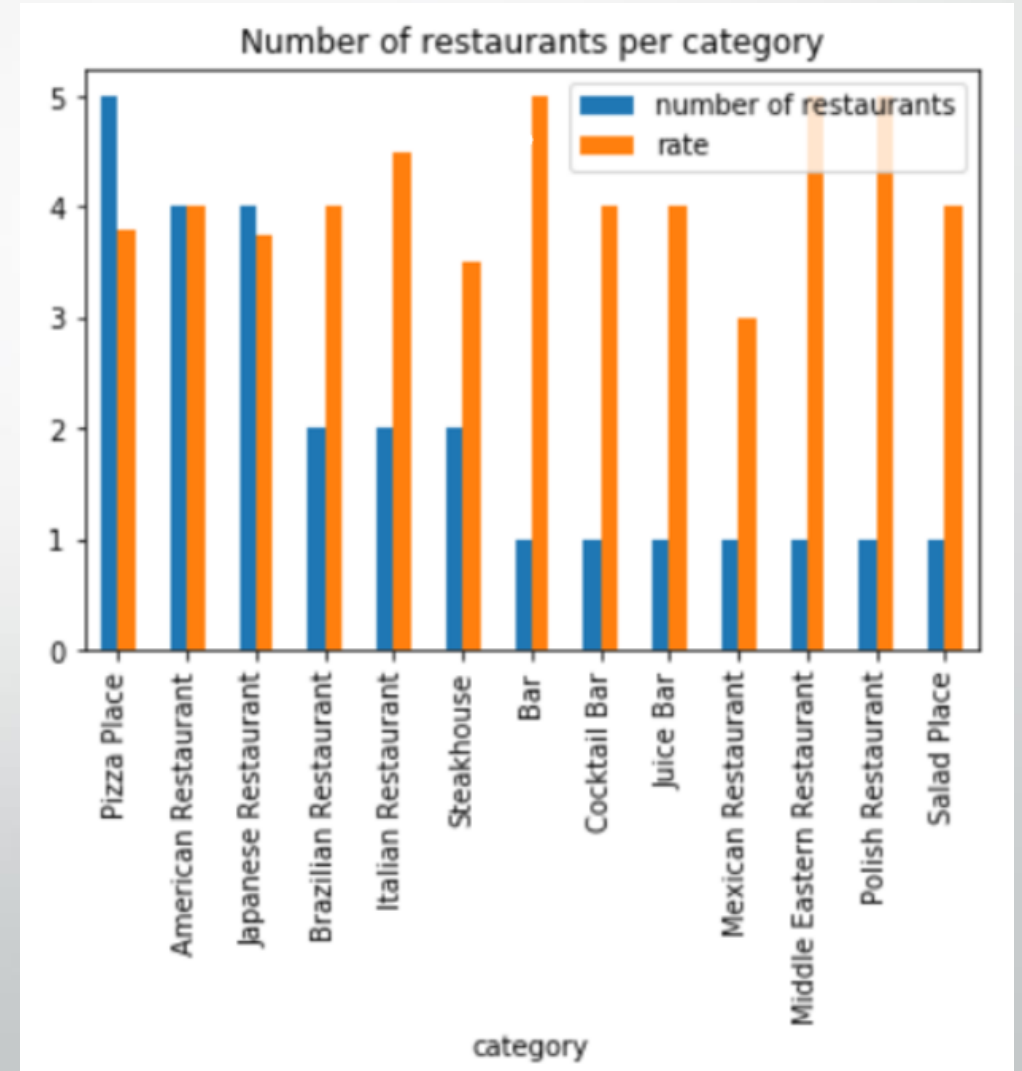
Exploratory Data Analysis

- User experience:



Exploratory Data Analysis

- We can see that the rate deviation is higher between the less frequent categories. But it is no surprise that, with the lack of samples, the data turns biased.



Recommendation System Modelling

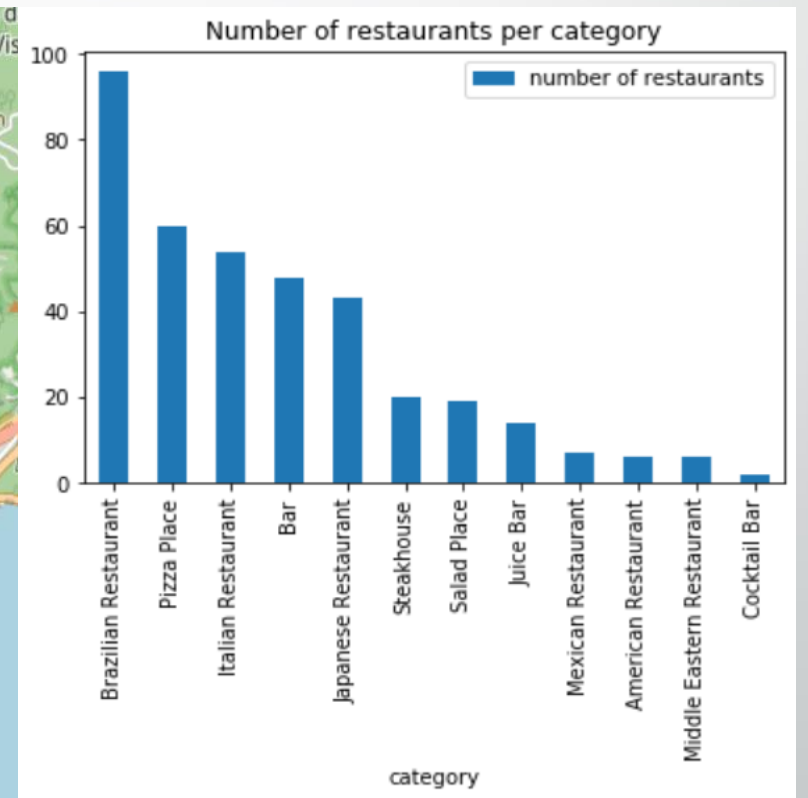
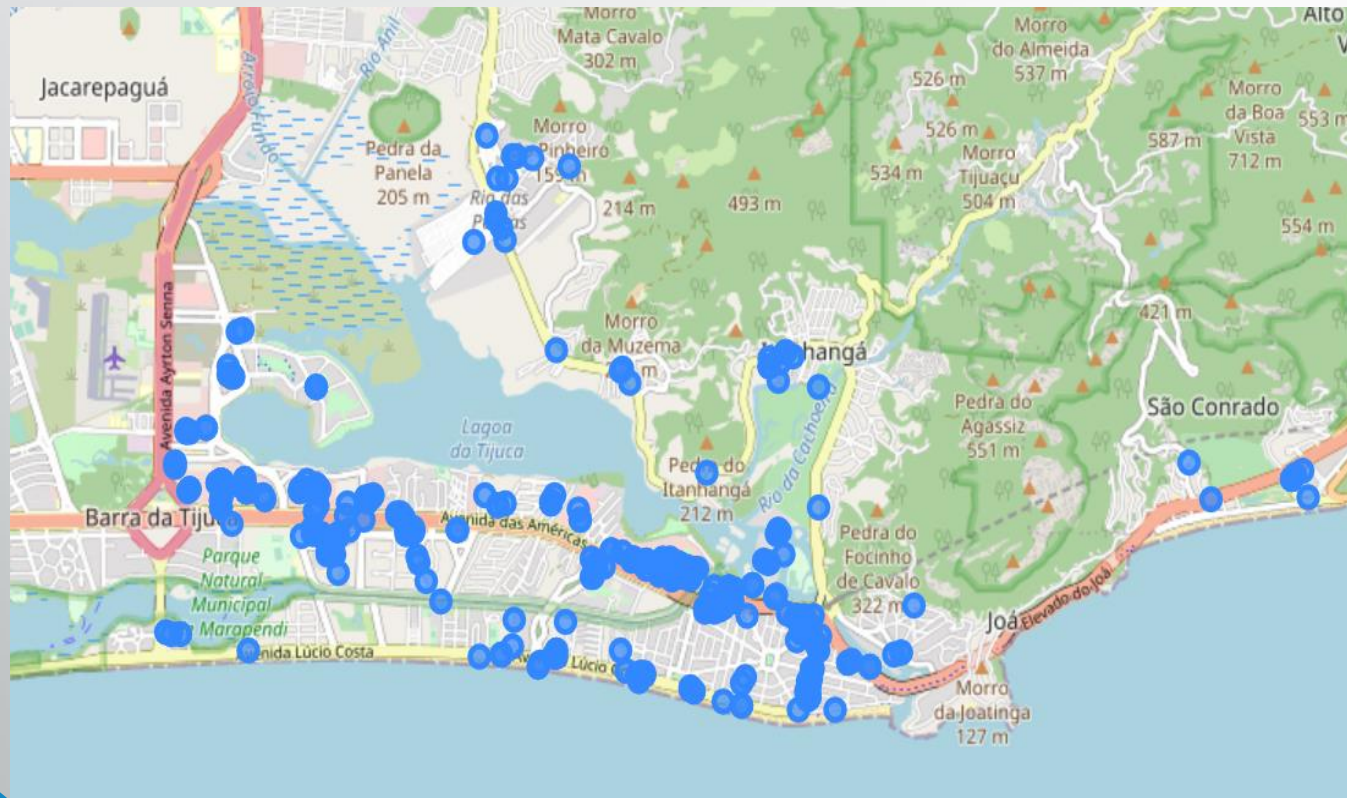
- New Venues Gathering: API URL parameters:

Client id	Client Secret	Version	Radius	Limit	Category	Latitude	Longitude
-	-	20180605	6000	100	Looped over each	-23.013457	-43,309648

- Result's Features: venue id, name, category, latitude and longitude, distance.

Recommendation System Modelling

- Gathered venues:



Recommendation System Modelling

- Data Preparation: Venue features and user experience set as numerical.

index	American Rest.	Pizza Place	Japanese Rest.	...	Bar	Juice Bar	Mexican Rest.	Italian Rest.	Distance
1	0	0	1	...	0	0	0	0	0.00828
2	0	0	0	...	0	0	0	1	0.01223
3	0	0	0	...	1	0	0	0	0.01194

index	American Rest.	Pizza Place	Japanese Rest.	...	Bar	Juice Bar	Mexican Rest.	Italian Rest.
1	0	0	5	...	0	0	0	0
2	0	0	0	...	0	0	0	4
3	0	0	0	...	4	0	0	0

Recommendation System Modelling

- Data Preparation: User profile calculated. Bias value was included as a weight for the distance feature.

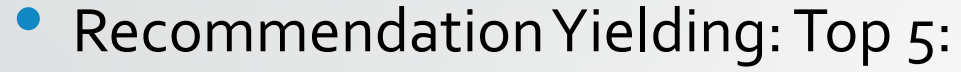
American Rest.	Pizza Place	Japanese Rest.	...	Bar	Juice Bar	Mexican Rest.	Italian Rest.	Distance Bias
0,8421	0,6573	0,5864		0,2685	0,2648	0,2567	0,3489	-120

Recommendation System Modelling

- Recommendation Yielding: Top 5:

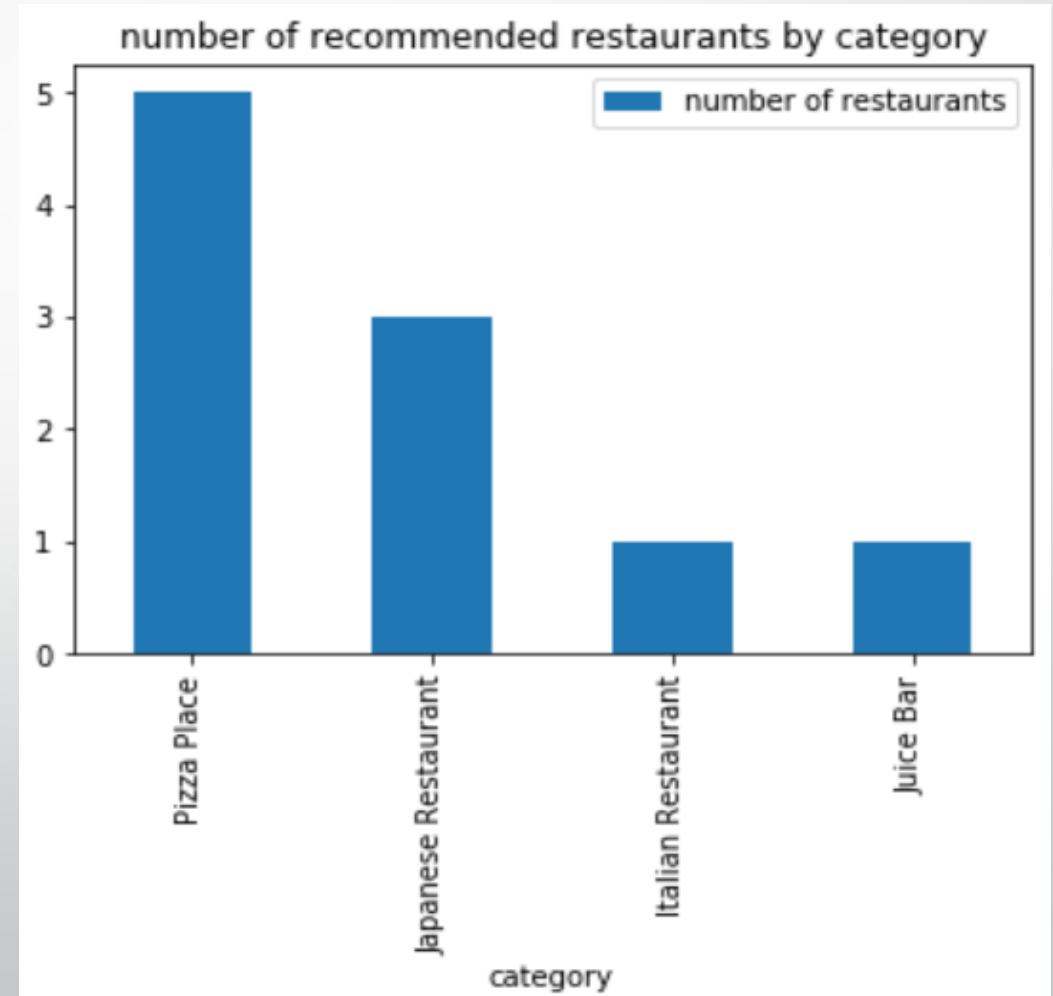
id	name	category	lat	lng	distance	Recommendation
4c81acfed4e2370464956088	Pizzaria Zona Sul	Pizza Place	-23.012214	-43.311411	0.002160	0.568043
587c01a8c7ec6b7a1e21b1ad	Skipper 1992 Pizza Bar	Pizza Place	-23.013088	-43.305379	0.004311	0.137705
4cobd1fda1b32d7f2c899bfo	Capricciosa	Pizza Place	-23.013811	-43.305366	0.004317	0.136652
4bcbaaa368f976boaf9d6183	Pe'ahi	Japanese Restaurant	-23.014807	-43.306366	0.003556	0.078222
4e9d880010814f7188b647d2	Gero	Italian Restaurant	-23.012820	-43.311783	0.002217	0.030332

- Recommendation Yielding: Top 5:



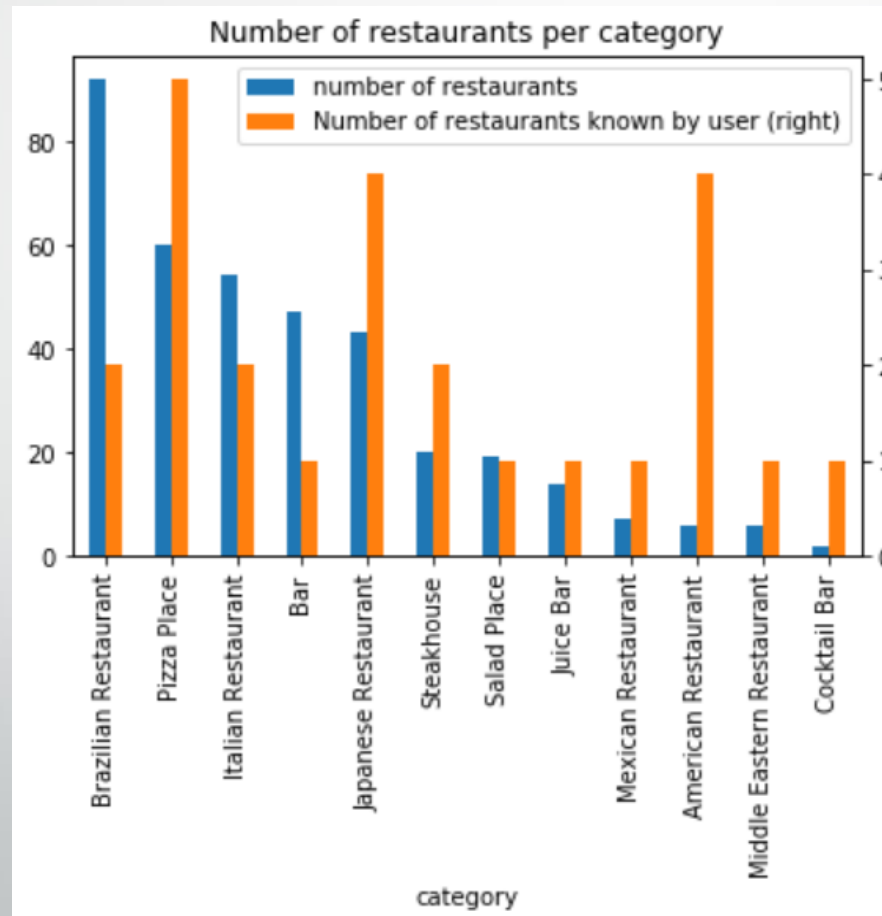
Result Analysis

- It can be seen on the map that the algorithm successfully yielded venues close to the user, the farthest one being mere 8 minutes away by feet.
- Most of the venues also belong to the categories the user frequents the most and rates highly, as it should be



Result Analysis

- Comparison between the user's experience and the local offer:



Result Analysis

- As expected, the recommended system yielded venues:
 - close to the user;
 - from categories with high availability, which makes them easier to be found nearby;
 - from categories the user rates highly.
- Although, if a restaurant belongs to a category that the user has experience and high satisfaction with, but low availability, it still might be recommended as well, as long as the user finds herself close to it.

Conclusion and Future Directions

- The model implemented showed itself successful by yielding the expected results given the user profile
- Possible improvements:
 - With an Foursquare premium account, more information could be gathered about each venue, adding a new layer to the recommendation analysis, which could make its results even more fitting.
 - Besides, if there were a record of not only of which venues the user knows, but also how often she goes to each of them, her profile could be more accurate as well.