Location Evaluation for Gastronomy Businesses based on Geospatial Data

By

Christopher Husmann

Introduction

- The location is essential for the long-term success of a gastronomy business
- The drastic emergency acts, which were ratified to enforce social distancing, hitting gastronomy especially hard
  - as many as 75% of the independent restaurants of New York may not survive the COVID-19 pandemic
- At the end of the COVID-19 crisis the expected shortage of gastronomy businesses of all kinds will present tremendous business opportunities
- Reliable Location evaluation for Gastronomy Businesses will be more valuable than ever

Dataset

- County Business Patterns: 2017
  - Localized economic data gathered by the US government
    - Number of establishments
    - Employment during the week of March
    - Total annual payroll for all employees in a given postal code
- Foursquare database
  - Registered venues in the US grouped by postal code
  - Detailed information on all US venues of a certain kind (e.g., ice cream shop)

Methodology

1. Gather localized economic data for all post codes of the US
2. Build data set with all on Foursquare registered venues in the US and group them by postal code
3. Query detailed information of all venues of interest (e.g., all ice cream shops)
4. Label location of the most successful business among the venues of interest as promising location
5. Identify key venues which characterize promising location
6. Train logistic regression model

Feature reduction

In Foursquare hundreds of different business types are recorded

- Using all of these as features leaves ML method prone to overfitting especially for small and moderate feature sets.

Venues that characterize promising locations should:

1. be found frequently near successful businesses of interest
2. have a significantly above average density at promising location

Base on the above describing we introduce the two thresholds α and β to identify key venues. In particular a venue x is used as feature if :

$$\alpha < f_+(x) \text{ and } \beta < \left| \frac{\overline{f_+}(x)}{\overline{f_-}(x)} - 1 \right|$$

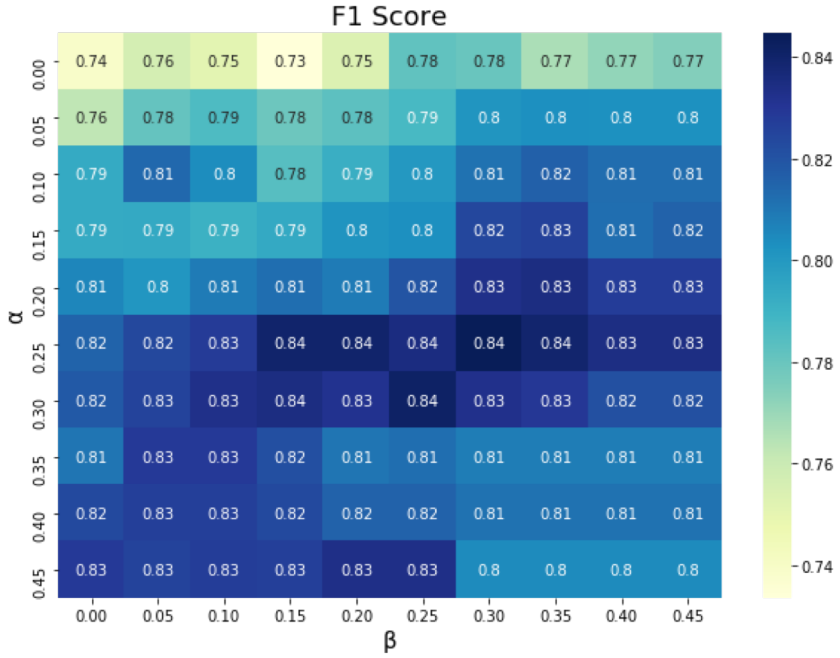
Where α and β are the two threshold values, $f_+(x)$ is the frequency of occurrence for venue x near promising locations ($f_+(x = coffe\ shop) = 1.2$ means in average 1.2 coffee shops are found near promising locations) and $\overline{f_+}(x)$ and $\overline{f_-}(x)$ nominate the normalized (by average) frequency of occurrence for promising and non-promising locations, respectively.

Results

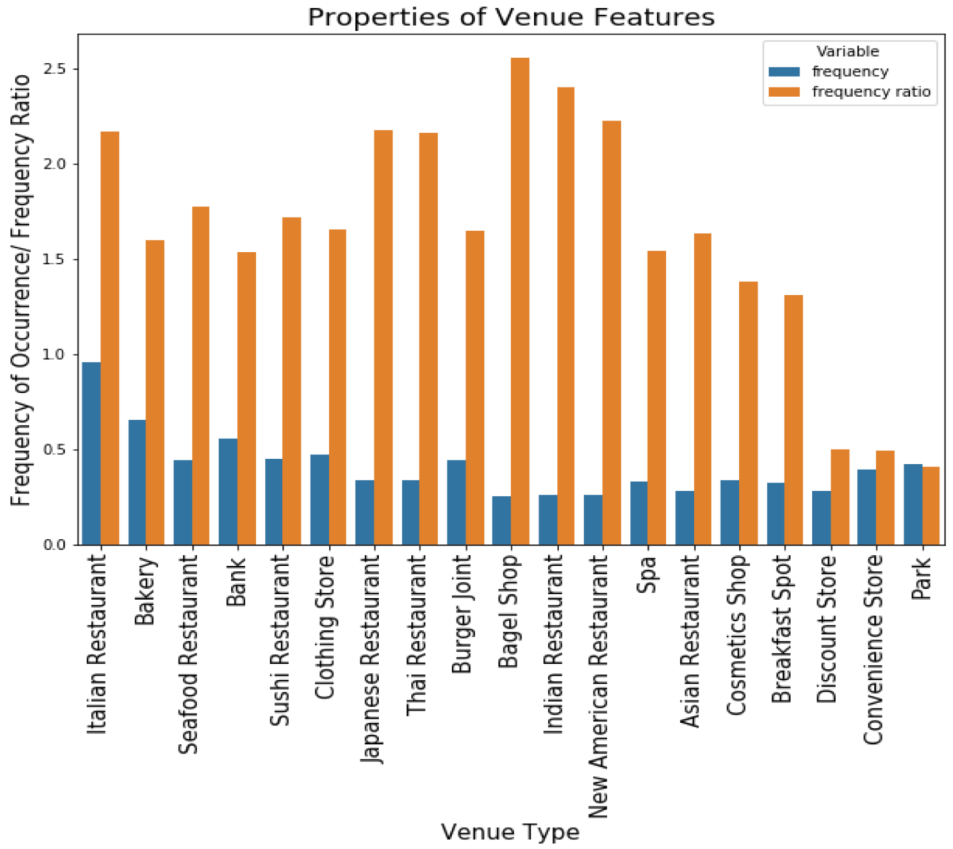
- In our evaluation the business of interests are ice cream stores
- We used likes as in directed measure for popularity and successes of a business
- In total 1510 ten US-based ice creams store are registered at Foursquare
- The location of the 450 shops with the most likes are labelled as promising
- 639 different venues have been in the vicinity of these ice cream stores

Determine threshold values and key venues

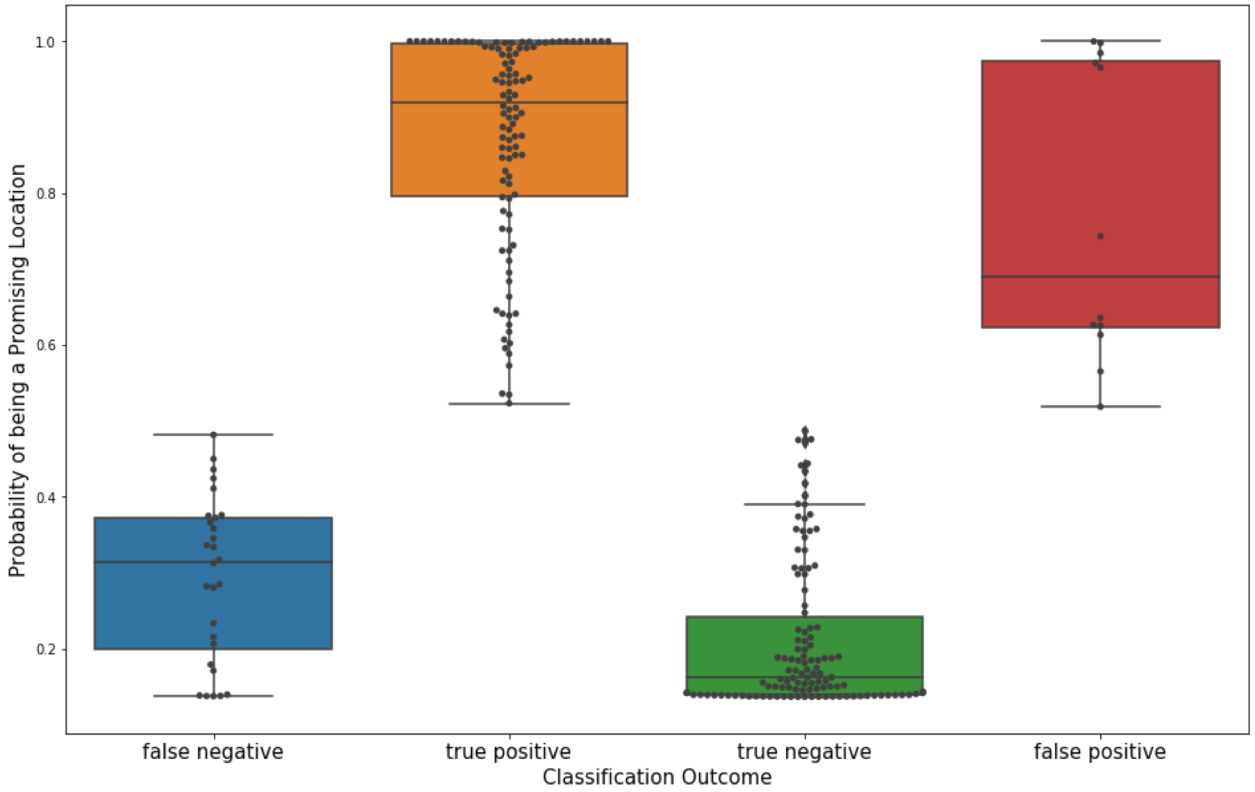
Thresholds are determined via gird search



Key venues for ice cream stores:



Classification Results



	Labelled non-promising	Labelled promising
Non-promising	45% (True negative)	4.5% (False negative)
Promising	10.5% (False negative)	40% (False negative)

Conclusion

- Evaluated the suitability of location for a specific given gastronomy business based on geospatial data
  - Localized economical data
  - Geospatial data for Foursquare
- We presented a method to reduce the location data of Foursquare to few key features
- We showed, at the example of ice cream stores, that locations can be reliably classified into the categories **promising** and **not promising** based on geospatial data