Hipsterhood

Applied Data Science Capstone Project

by Helder Reis, March 2019

Motivation

- Nowadays there is no shortage of resources (websites, apps, magazines) for someone to find the "coolest", most "trendy" and modern places in any city in the world
- What seems to have become increasingly difficult is to find "normal" places where food is not served on bits of wood and roof slates, chips in mugs and drinks in jam jars

Project description

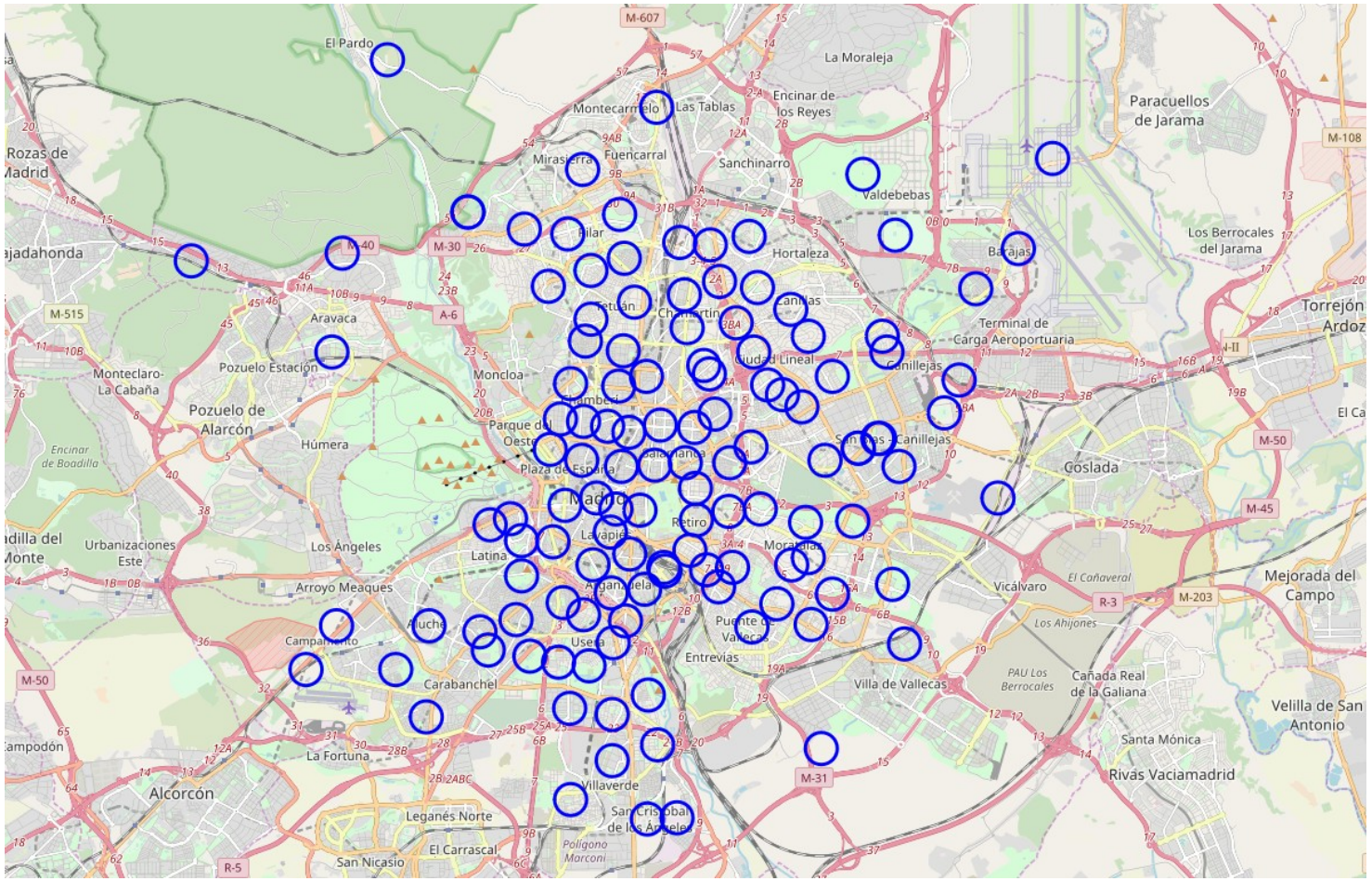
- **Hence the goal of this project is to find a "Neighborhood Hipster Rating" ("Hipsterhood" for short) to help all users find places they are comfortable with**
- We will find the rating of each neighborhood, then cluster them into 3 clusters, with 1 being "Mostly Traditional", 2 "Something for everyone" and 3 "Hipster Paradise"
- To better evaluate the results we will use the city of Madrid, where I've been living for a few years hence am familiar with

Data acquisition and cleaning

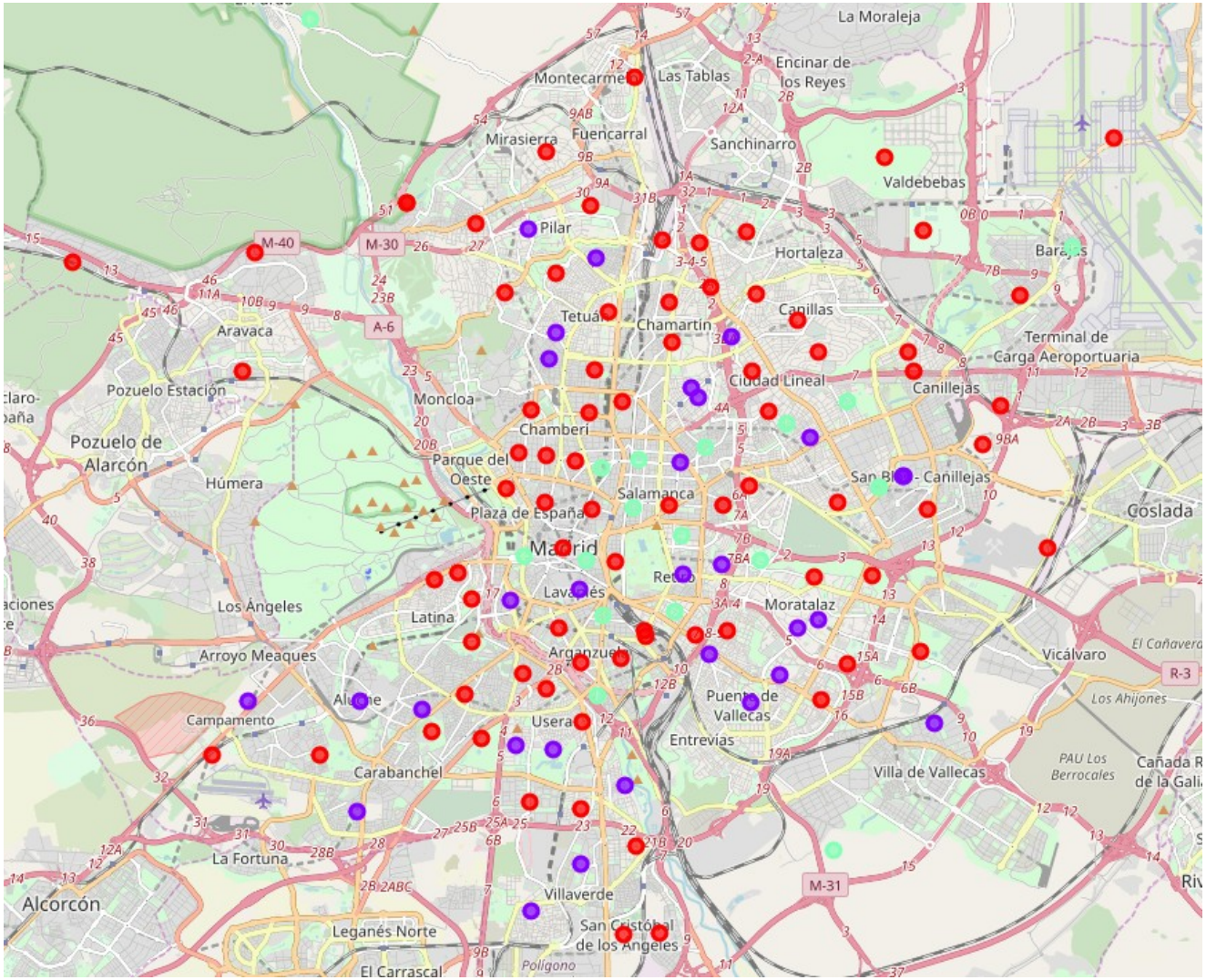
- We got the district, wards (neighborhoods) and coordinates for each of the ward from Wikipedia
- Foursquare has available a list of categories, to which we added a "hipster rating" (from 1 to 3, least to most “hipster”)
- From Foursquare we got a list of all venues in a 400 meters radius of each neighborhood, getting over 18k venues
- We multiplied each category by its “rating”
- We used k-Means to cluster the neighborhood into 3 clusters

Map of neighborhoods

- We used a radius of 400 meters around each neighborhood center



Clustered neighborhoods



"Mostly traditional"

"Something for everyone"

"Hipster paradise"

Conclusion and future directions

- While far from anything production ready or scientifically accurate, the classifier seems to be working within the expectations and could be the basis for an app or website
- Some improvements:
  - Filter by parent categories, keeping categories under Arts & Entertainment, Food and Nightlife Spot but ignoring the ones under Residence, Professional & Other Places.
  - Use venue labels to better classify the venue, since categories are quite generic.
  - Get venues by neighborhood used by Foursquare (even if not an official ward) instead of radius