Date Night!

A play in three acts… perhaps four
by Chris Davis for the IBM Data Science Capstone
February 14, 2021

Introduction

Planning a date night is a tricky affair. What to do? Where to go? What to eat? Is there entertainment? Do we get drinks afterwards? How should we get from one place to another? All these questions must be painstakingly researched and answered while making your boss believe that you're working on that overdue report. If only your report was on city traffic planning, then you wouldn't have to worry about your boss seeing Foursquare maps on your screen. Alas, your report is on the rate of sewage pump failure at political conventions, a topic that doesn't typically lend itself to romantic engagements.

So you must plan your date quickly, and it must be a REAL date. A "date" is defined by the International Society of Don't Screw This Up as

> *date* (n. fr. medieval Latin)
> 1) *the time at which an event occurs,*
> *2) the brown, oblong edible fruit of a palm (Phoenix Dactylifera),*
> *3) an outing comprising food, film, and fine beverages (necessarily in that order) with a person in whom you are interested romantically but have absolutely no chance if you screw this up.*

The stakes couldn't be higher! If only there were a service to plan the date automatically. 🤔

Business Problem

As a startup seeking an easy means to help the helpless plan "real" dates, we look to leverage Foursquare to quickly identify "date triples" in a given city. A "date triple" is a group of venues within close proximity of each other where one is a restaurant, one is an entertainment venue, and the last is a bar. The user of this kind of service could further refine the results based on dietary preference or neighborhood, but that is outside the scope of this initial prototype.

Data

Our algorithm requires three primary types of data: user location and range, venue categories, and Foursquare Places API.

Region, Range, and Proximity

A user is interested in planning a date in a particular region (e.g. city, borough, town, etc.), is willing to travel an approximate distance from their starting point (e.g. home, workplace, etc.), and wishes the venues comprising the date triple to be within a general proximity of each other.

Venue Categories

Foursquare provides a hierarchical list of categories[1] that allow us to query different types of venues. Reviewing this list, we select the following venues of interest:

Date Triple Group	Foursquare Venue Category	Foursquare Venue Identifier
Restaurant	Food	`4d4b7105d754a06374d81259`
Entertainment	Arts & Entertainment	`4d4b7104d754a06370d81259`
Bar	Nightlife Spot	`4d4b7105d754a06376d81259`

While Foursquare's categories aren't perfect for our purposes, we'll start with the most coarse categories and differentiate between options using clustering.

Foursquare Places API

We will use the Foursquare Places API to gather venue data, specifically version 2 of the venues/explore endpoint. Given the near-ubiquity of restaurants in any given city, we center our search around entertainment venues, then fill in the gaps for restaurants and bars based on where the entertainment venues are.

The venues/explore endpoint returns JSON text, of which we'll use the following fields:

JSON Path	Description
`response.groups[].items[]`	An array of venues matching our query
`item.venue.id`	The unique Foursquare ID of the venue
`item.venue.name`	The venue's name
`item.venue.location`	The venue's latitude and longitude
`item.venue.formattedAddress`	The human-readable address of the venue
`item.venue.categories[]`	The categories in which the venue is placed

Methodology

Using the user's region and range, we query the Foursquare API to identify entertainment venues within the range of the the center of the region. The region and range indicate the distance the user is willing to travel to reach the date venues.

With the list of entertainment venues in range of the selected region, we then query the Foursquare API for the restaurants and bars within proximity to each entertainment venue, and transform this data into 3-tuples consisting of a restaurant, an entertainment venue, and a bar (what we call a "date triple").

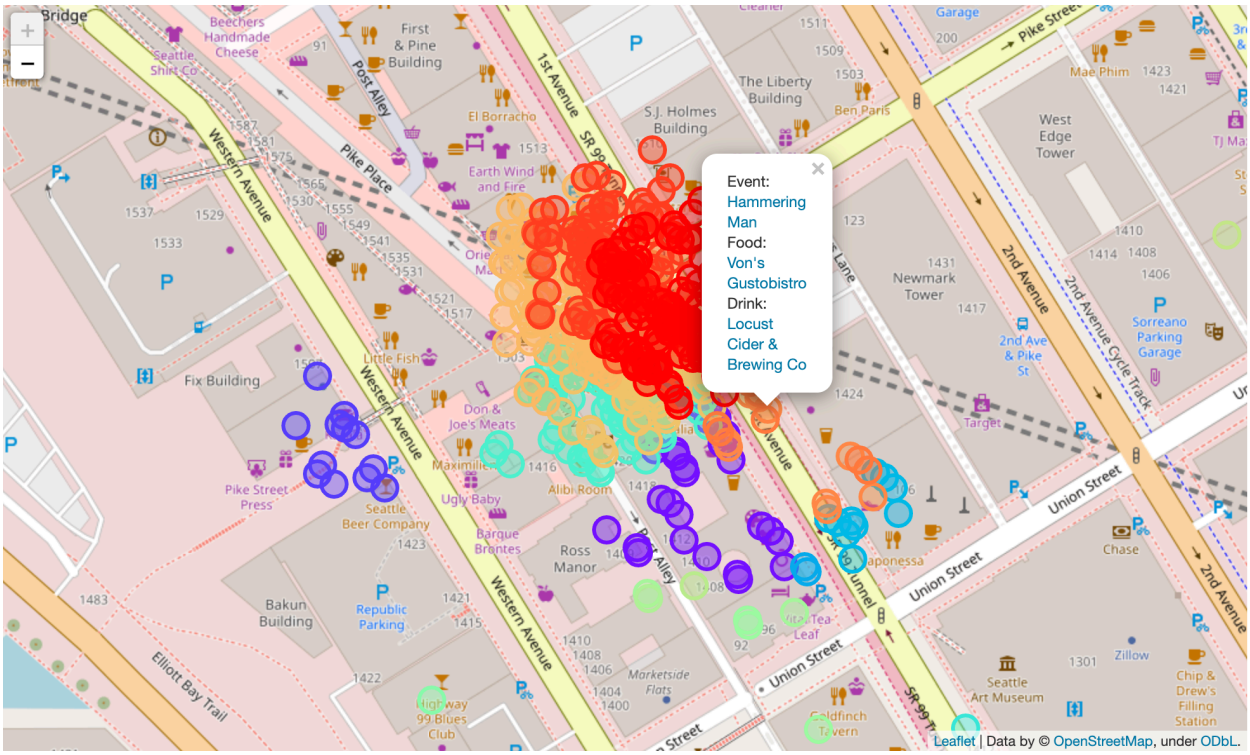
[1] https://developer.foursquare.com/docs/build-with-foursquare/categories/

Upon exploring these 3-tuples, we discovered that Foursquare will sometimes categorize the same venue as both a restaurant and a bar. Therefore, that venue would appear in both the proximate restaurant and proximate bar queries. We removed the 3-tuples that contained the same venue for both restaurant and bar.

Placing these 3-tuples in a data frame, we then calculated the 3-tuple's centroid through the individual averages of each venue's latitude and longitude. We also assigned a color value to each entertainment venue to present groupings of date possibilities based on color.

Finally, we plotted these centroids on a map, creating a label for each 3-tuple wherein links to the Foursquare pages for each venue are presented.

Results



The above example shows a variety of dates just West of Seattle's Central Business District. As you can see in the label, the event of this evening would be "Hammering Man," a public sculpture installation from the Seattle Art Museum[2]. The Foursquare API category of Arts & Entertainment is a very coarse category that includes public sculptures, hence the Hammering Man's inclusion in the results.

Discussion

Our initial analysis shows a wide variety of date options within our selected region and range, and presents the viewer with a high

[2] https://en.wikipedia.org/wiki/Hammering_Man

level understanding of the variety of dates available within that region and range.

However, our initial analysis is the beginning of a user experience designed to assist with date planning and exposes significant limitations with our current static analytical approach.

The most significant limitation in our approach was the Foursquare API. By using a developer account, we were limited to 500 daily queries of the Places API, of which only 50 could be "premium"—queries that provide venue-specific information such as price tier and recommendations. This extra data could be used to provide better recommendations to users, such as expensive versus cheap dates and venues to avoid.

Another area for improvement is user preferences for venue filtering. For example, a user may find it preferable to show only dates with nearby restaurants serving vegetarian or vegan options; or, they may wish to avoid movie theaters. Categorical filtering requires curation and foreknowledge of desirable categories, an activity that requires human intervention.

Conclusion

We attempted to use the Foursquare API and some basic data science to identify potential "date" combinations of food, fun, and drinks. Following our analysis, we presented findings to the user by way of a colorful and interactive map focused on their preferred region. While the map showed a wide variety of options, it still required the user to inspect the results to develop a sense of what would be a "good" date versus a "bad" one. Through the use of a wider selection of user preferences and venue-specific data of price tiers and recommendations, it may be possible to reduce the number of date options to a manageable number.