JAT's Capstone Project "Dinner and a Date"

Jerry Ackaret

Introduction to JAT's Capstone Project

▶ I decided to create something that was useful, and solved a real world problem.

▶ My wife volunteers for two non-profit organizations that use silent auctions to make money. Her specialty is organizing and decorating the baskets. One of the most popular we call "Dinner and a Date" baskets. They consist of tickets to a Venue and gift cards to a restaurant, as well as other 'trinkets'.

▶ Prior to my wife's involvement, the Fun Venue and Food Venues were randomly put together, resulting in the possibility of miles between them, not too conducive to a 'date'. She started pairing the two items for being in close proximity, thereby making the basket more appealing.

▶ **The problem was: Since no thought for this arrangement was put into the 'solicitation' of the items, there was no guarantee of arriving at a suitable combination.**

▶ So, my solution is to provide the 'Donation Solicitor' a list of Fun Venues, and a list of distance and regional associated Food Venues to facilitate the donation requests for what will make suitable combinations.

    ▶ Thereby increasing the potential value to the bidders, which in turn **increases the money for the non-profit**.

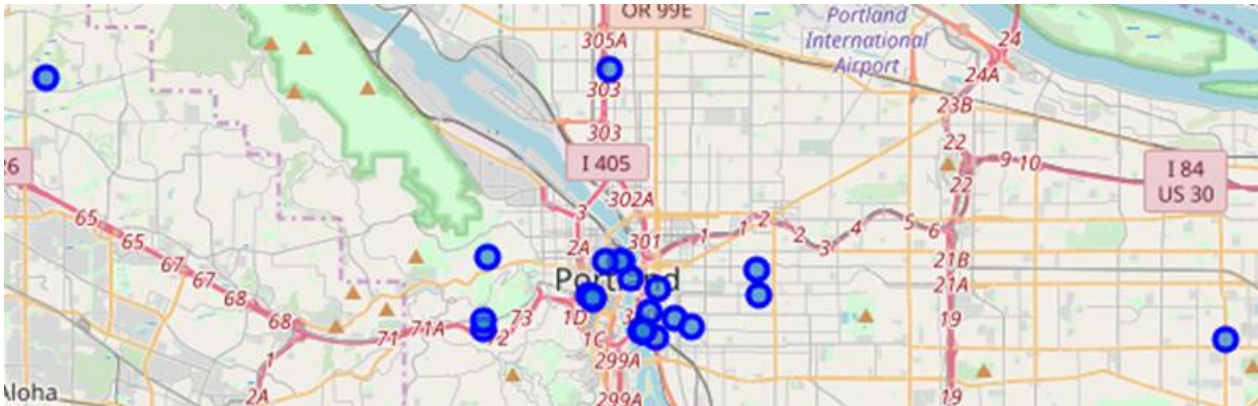# Data Used in JAT's Capstone Project

▶ Primarily, only Foursquare data is necessary to accomplish this project, with the exception of determining Portland, OR latitude and longitude, using geopy.geocoders.

▶ The Foursquare data to be used is in the Explore and Details Endpoints:

    ▶ Explore Endpoint

        ▶ Limit the Fun Venue results to Museums, Art Galleries, Concert Halls, Historic Sites, Stadiums, Zoos

        ▶ Near Portland, OR (there is a 'near' function inside of Foursquare)

    ▶ Example Output:

```
reasonName : giobaiinteractionkeason }]},
'venue': {'id': '43ee1bd9f964a5205f2f1fe3',
'name': 'Portland Art Museum',
'location': {'address': '1219 SW Park Ave',
 'lat': 45.51622988093649,
 'lng': -122.68359661102295,
 'labeledLatLngs': [{'label': 'display',
   'lat': 45.51622988093649,
   'lng': -122.68359661102295}],
 'postalCode': '97205',
 'cc': 'US',
 'city': 'Portland',
 'state': 'OR',
 'country': 'United States',
 'formattedAddress': ['1219 SW Park Ave',
  'Portland, OR 97205',
  'United States']},
'categories': [{'id': '4bf58dd8d48988d18f941735',
```

Data Used in JAT's Capstone Project

- Explore continued
  - Obviously, Foursquare returns far more information than is needed, so I scrubbed the data down to Name, Category, Latitude, Longitude, Address, and Venue ID.
    - The Latitude and Longitude was needed for mapping and the Venue ID to obtain details of the Fun Venue than the explore endpoint provides.

- Data Contraction: Two Fun Venues were dropped that fell outside of the core part of Portland and therefore were outside of the scope of the project.

- Two venues were dropped that fell outside of the core part of Portland, selecting their Circles gave me their names; 'Washington County Museum', and 'Madrugada Pottery'.

Data Used in JAT's Capstone Project

▶ Data Expansion: Continuing to use the Explore Endpoint, found the Food Venues that are around each of the Fun Venues, within the 750 meter radius that was chosen by the user (set early in the program).

  ▶ The Food Venue Name, ID, Latitude, Longitude, Distance to Fun Venue, Category, and Address were retrieved, and combined with the appropriate Fun Venue.

  ▶ Example of output:

Fun_Venue	Fun_Address	Fun_Latitude	Fun_Longitude	Fun_ID	Venue	Venue_ID	Venue_Latitude	Venue_Longitude
Portland Art Museum	1219 SW Park Ave	45.51623	-122.683597	43ee1bd9f964a5205f2f1fe3	Higgins Restaurant & Bar	41ddd100f964a520c51e1fe3	45.515464	-122.682074
Portland Art Museum	1219 SW Park Ave	45.51623	-122.683597	43ee1bd9f964a5205f2f1fe3	Behind The Museum Café	4ee11c6abe7b7e4d131503ff	45.516492	-122.684151
Portland Art Museum	1219 SW Park Ave	45.51623	-122.683597	43ee1bd9f964a5205f2f1fe3	Raven & Rose	50960315e4b0e34d634b48f5	45.514794	-122.682321
Portland Art Museum	1219 SW Park Ave	45.51623	-122.683597	43ee1bd9f964a5205f2f1fe3	Southpark Seafood & Oyster Bar	40b13b00f964a520fef51ee3	45.517909	-122.681596
Portland Art Museum	1219 SW Park Ave	45.51623	-122.683597	43ee1bd9f964a5205f2f1fe3	Addy's Sandwich Bar	4a8eec36f964a5202b1320e3	45.518534	-122.682720

Data to Use in JAT's Capstone Project

- Data Contraction: I did not want food venues that might not meet the level of a Dinner and a Date, and so dropped some categories:
  - 'Food Truck','Donut Shop','Bakery','Bagel Shop','Coffee Shop','Dessert Shop','Fast Food Restaurant', 'Food Court','Food Stand','Juice Bar','Snack Place'

- Data Contraction for Spreadsheet output: Further reduced the Food Venues to the closet ones. The number was a variable set by the user at the beginning of the program (10 in this instance).
  - Note, the entire set of over 1,500 Food Venues were kept separately to use for clustering.

- Detail Endpoint:

  - The explore endpoint does not have phone numbers or ratings, and so I pulled just the Venue IDs from Fun and Food, retrieved the phone numbers and ratings where they exist, and combined this data back into the master list using merges.

Methodology Used

- To determine if the radius was appropriately set, I used the groupby/count function to see how many Food Venue returns I got for each Fun Venue.

  - Some had too few, and I increased the radius from 500 to 750 meters.

- Once I had the Fun Venues down to the ones I desired, and the associated Food Venues down to the types and number that I desired, I exported the output to a spreadsheet.

  - Dropped all latitude, longitudes, and IDs as not needed for the Solicitor to make contacts.

  - The spreadsheet is created automatically when the code is run.

- To prepare for the k-means clustering, I had to insure there were Food Venues for each Fun Venue.

  - One did not (Pittock Mansion) which if allowed to stay would cause the Cluster value to be a floating point, disrupting the mapping feature.
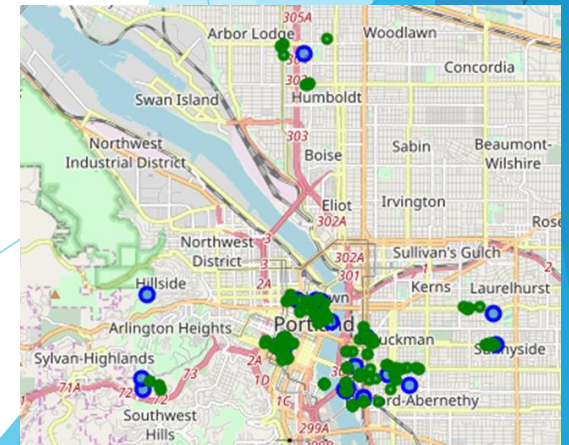
Methodology Used – Clustering continued

- I desired that the Solicitor have more data than just distance and rating, primarily answering:

  - Under the theory that the venues that have the best ratings of the categories that have the most options will tend to be the actual best venues.

- Since machine learning algorithms cannot usually use categorical variables, I used the one hot encoding method to turn them into binary vectors.

- Using k-means clustering, I determined each region of Portland's most common eating establishments.

  - Cluster 0 (Purple) on the east side of the river:

    - Primarily a Seafood and Mexican Restaurant area

  - Cluster 1 (Blue) in the West Hills:

    - is more Café and general Restaurant

  - Cluster 2 (Green) on the east side of Portland, but reasonably close to the river:

    - Mexican and Pizza shops

  - Cluster 3 (tan) on the west side of the river:

    - Primarily American Restaurants

  - Cluster 4 (red) further east away from the river is more Thai.
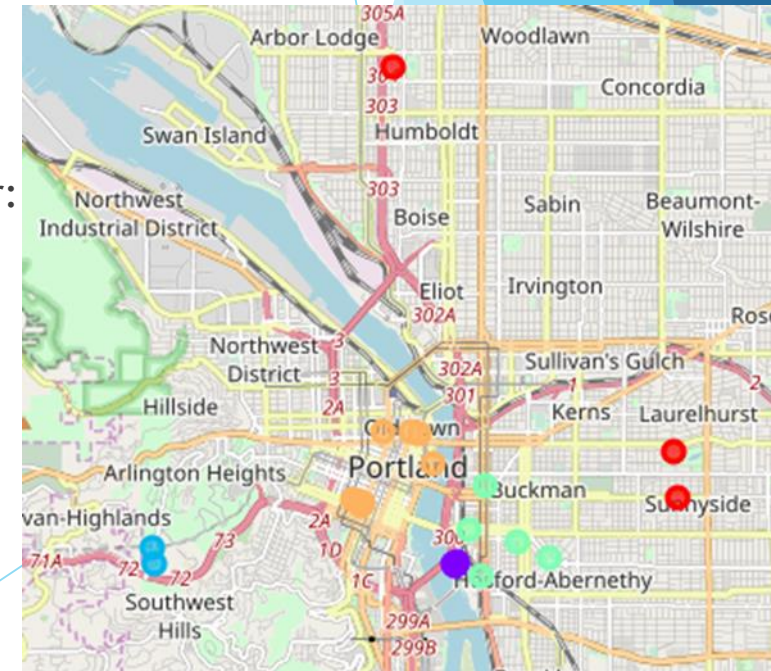
Discussion

- I believe that the results that I obtained will be beneficial in the coming year donation solicitation.

  - I have spoken to the chairperson of one of the non-profits that my wife volunteers for, and she is quite interested in getting a list to start working with.

- The underlying technique used can be also used in other ways.

  - First, changing the parameters of the Fun Venues would be of use to determining Fun Venues that are family oriented, such as water parks, bowling alleys, etc.

  - In addition, either of these can be used personally, by people that just want to explore finding new venues that would make for a good outing with one's significant other, or family.

Conclusion

► While the nature of my job as a Reliability Engineer in the design of servers means that this is probably my only exposure to Foursquare, I was glad to find a project to use for this Capstone that was useful for someone, and not just academic. I was pleased that one of my intended audience is already excited to see the results.