Hitting the Bullseye with NLP



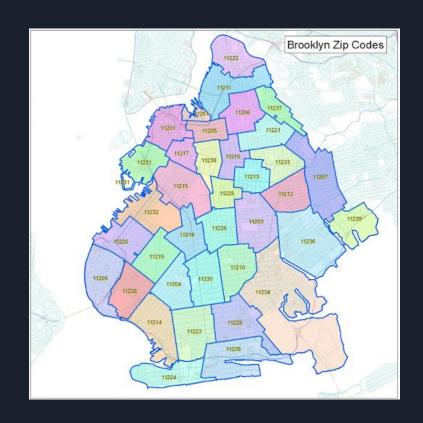
Problem Statement:

Can we use Natural Language Processing and data from Yelp to locate businesses/enterprises with a specific feature?

Other Questions needing Answers

- Why is this a problem?
- Why did I pick this?
- How can we attempt to solve this problem?

- Step 1 Picking a target
 - Brooklyn sets the stage
 - Proof of concept "city"
 - Look at all the bars
 - Yelp delivers the info



- Step 2 Gathering the data
 - Collecting via Yelp APIs
 - Yelp Fusion
 - Yelp GraphQL
 - Web Scraping
 - requests
 - BeautifulSoup



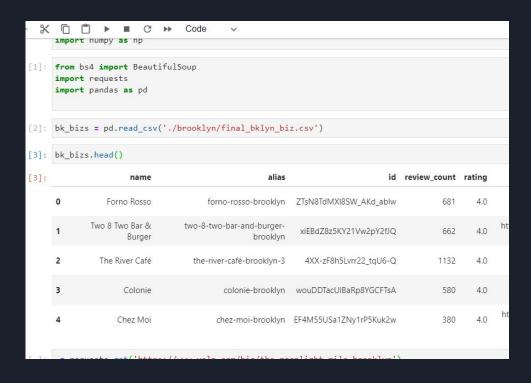
- Step 2 Gathering the data
 - Yelp APIs
 - Yelp Fusion

```
{'businesses': [{'id': '1Q3oaJahyGRogDWgpo7PIw',
   'alias': 'five-leaves-brooklyn-2',
   'name': 'Five Leaves',
   'image url': 'https://s3-media1.fl.yelpcdn.com/bphoto/vSBFMjqIMtD2vpOV2s6Wfw/o.jpg',
   'is closed': False.
   'url': 'https://www.yelp.com/biz/five-leaves-brooklyn-2?adjust creative=bMPz0NoFK6qQZA
OZAimOncxVw'.
   'review count': 1609,
   'categories': [{'alias': 'newamerican', 'title': 'American (New)'},
   {'alias': 'coffee', 'title': 'Coffee & Tea'},
   {'alias': 'bars', 'title': 'Bars'}],
   'rating': 4.0,
   'coordinates': {'latitude': 40.72371830945, 'longitude': -73.95159863632},
   'transactions': [].
   'price': '$$'.
   'location': {'address1': '18 Bedford Ave',
    'address2': ''
    'address3': '',
    'city': 'Brooklyn'.
    'zip code': '11222',
    'country': 'US',
    'state': 'NY',
    'display_address': ['18 Bedford Ave', 'Brooklyn, NY 11222']},
   'phone': '+17183835345'.
   'display_phone': '(718) 383-5345',
   'distance': 779.9014324141493}.
```

- Step 2 Gathering the data
 - Businesses
 - Yelp APIs
 - GraphQL
 - gql python library

```
while remaining > 0:
    query = gql('''
                query($zipcode: String $offset: Int)
                    search(location: $zipcode categories: "bars, beergarden, beergardens, beerhall, gastropubs,
                        total
                            business {
                                name
                                alias
                                review count
                                rating
                                url
                                location {
                                    city
                                    postal code
                                coordinates{
                                  latitude
                                  longitude
    param = {"zipcode": zipcode,
            "offset": offset}
    response query = client.execute(query, variable values = param)
```

- Step 2 Gathering the data
 - Reviews
 - Web Scraping
 - requests
 - BeautifulSoup



tarter-can X Untitled1.ipynt X Untitled2.ipynt X Untitled3.ipynt X Zip_codes.ipynl X Untitled3.ipynt X Zip_codes.ipynl X

!function(t){function e(n){if(i[n])return i[n].exports:var r=i[n]={ ction(t,i,n){e.o(t,i)||Object.defineProperty(t,i,{configurable:!1,enumerable:!0,get (i, "a",i),i},e.o=function(t,e){return Object.prototype.hasOwnProperty.call(t,e)},e. s}),i.d(e, "a", function(){return o});var n=i(5),r="global",s="biz",o="biz gtag",a=[r n("."):throw new Error("google analytics attempted to set "+t+" to an unrecognized ar i=null!=arguments[e]?arguments[e]:{},n=Object.keys(i);"function"==typeof Object. opertyDescriptor(i,t).enumerable}))),n.forEach(function(e){r(t,e,i[e])})}return t}f 0}):t[e]=i,t}var s=i(2),o=i,n(s),a=i(3),c=i(4),u=i(0),f=i(6);0bject.kevs||o,a,shim(n+=1)i[n]=[e[n],t[e[n]]];return i});var l=function(){function t(t){r(this,"clientIQ ons", void 0), r(this, "metrics", void 0), r(this, "temporaryDimensions", void 0), r(this, " void 0), window.ga | (window.ga=function(){}), window.gtag | (window.gtag=function(){}) oraryDimensions={},this.temporaryMetrics={},Object.entries(this.trackers).forEach(f r=n({cookieDomain:this.domain,name:t,clientId:this.clientID},this.userID?{userId:th etrics[t]={},this.addDimensionsToFriendlyMap(t,this.dimensions),this.addDimensionsT setDimension(String(e),i.dimensions[t][e][1])}),this.metrics[t]&&Object.kevs(this.m ("set".t). "anonymizeIp". !0) }.e. addDimensionsToFriendlyMap=function(t.e) {var i=this: {var e=this;Object.keys(this.trackers).forEach(function(i){e.firePageview(String(i) ction(i){(i!==u.c||e.enableHighVolumeEvents)&&e.firePageview(String(i),t)})},e.fire Object.kevs(r).map(function(t){return r[t]}).reduce(function(t,e){var i,r=e[0],s=e[(o).map(function(t){return o[t]}).reduce(function(t,e){var i.r=e[0].s=e[1]:return r ain:this.domain,client id:this.clientID},s,a,this.userID?{user id:this.userID}:{},e ew")},e.setDimensionOneEvent=function(t,e){var i=this;Object.keys(this.friendlyMap) rl=e.toString())})},e.setDimension=function(t.e){var i=this:Object.kevs(this.friend imension"+r,e.toString())})},e.setDimensions=function(t){var e=this;Object.keys(t). tric"+t,e)},e.setMetricOneEvent=function(t,e,i){this.temporaryMetrics[i]["metric"+t {var c=o.trackers[a];window.gtag("event",e,n({send to:c,event category:t,non intera s}))}else if(a!==u.c){var l=n({hitTvpe:"event".eventCategorv:t.eventAction:e.eventU Object(f.a)(String(a),1)}o.temporaryDimensions[a]={},o.temporaryMetrics[a]={}})},e. 1:n};Object.keys(this.trackers).forEach(function(t){Object(f.a)(String(t),r)})},e.t i,n,r)}},e.trackEventWithTime=function(t,e,i){this.trackEvent(t,e,i,Object(c,a)())} (c.a)(),i=Object(a.a)(e);t.trackEventHighVolume("time on page","unload",i,e)})},e.i s")},3e4)},e.applyConfig=function(t){this.trackers=t.trackers,this.domain=t.domain, mensions=t.js dimensions,this.enableHighVolumeEvents=t.enable high volume events},e trackers.domain:this.domain.clientID:this.clientID.user.id:this.userID.dimensions:t



"Outstanding atmosphere & drinks. The best old fashions I've ever had!! Every ti "So admittedly I had bookmarked this spot ages ago when I saw they were a whiske friend and he wanted a nightcap and I'm like 'well, I yelped the hell out of the Where do people drink things? I love the atmosphere of the place. It isn't remar tles of bourbon are grinning up at you as you enter. For single pours of whiskey hey have a pretty extensive craft beer menu even surprisingly at affordable-ish ited - The Game of Thrones cocktails were still on the menu. Be still my heart. he Stark family and I also had the Ned Stark and then later banged back a White yrup, Blueberry Port & Blueberries. I liked the sweet and spiciness going on. Re & Three Chili Syrup. Definitely more my cup of tea. This was my favorite drink lker: Stillhouse Moonshine, Lemon Juice, Simple Syrup, Egg White & Creme de Pect taste to it. Maybe I shouldn't be mixing moonshine with bourbon. Bottom line, the skey to explore. Definitely worth checking out."

'The , Moonlight is back baby. \xa0We missed you guys sooooo much, and the Old to vibe. love it. \xa0Great to see you all...'

'Tonight I came on a online date here. The date turned out to be a low class id: mbiance to the service. They close around 10pm in order to shut down by the mand wasnt interested he asked me to pay him the money back.'

'Just wanted to shoutout what a wonderful spot this beautiful bar is! I feel ver eople on each side of the bar makes me smile :) Cheers to the Mile!'

Best bar in the neighborhood. Super friendly service, great selection of whiske "Went here to grab drinks with a friend and found a spot at the bar. This place or reen. Vibe was fun and drinks were decent—and very reasonable. Not a bad spot to "Loved Moonlight Mile. I was in the area to check out the barge but when we four tion of draught beers. The outdoor seating was the cherry on top, which is perfe "I had the White walker drInk. It was the first drink I tried and had 3 over all it usually plays what it wants. It was a Friday evening. Place was not crowded a "Best old fashioned there is! Love the outdoor seating - great for the summer. ("Really nice and comfortable setup. Me and the Mrs. hung out at this place on a ou enjoy beer as well and they have a nice selection of beers on draft from Pils tenjoy Ned Stark so much and kept complaining that it was sweet (or she was jus Extremely friendly and welcoming people too! With 2 whiskeys and 3 beers, our ta "Moonlight Mile is like some kind of magical Goldilocks bar - everything here is bration and, while patrons came and went all night, somehow the crowd size and respective to the summer of the patrons came and went all night, somehow the crowd size and respectively.

- Step 3 "Featurizing" the reviews
 - Business Corpus

	name	alias	id	review_count	rating	url	location	coordinates	review_text
0	Forno Rosso	forno-rosso-brooklyn	ZTsN8TdMXI8SW_AKd_ablw	681	4.0	https://www.yelp.com/biz/forno- rosso-brooklyn?	{'city': 'Brooklyn', 'postal_code': '11201'}	{'latitude': 40.694467, 'longitude': -73.98293}	SQUID INK TAGLIOLINI WITH SHRIMP IN BRANDY SAU
1	Two 8 Two Bar & Burger	two-8-two-bar-and-burger- brooklyn	xiEBdZ8z5KY21Vw2pY2fJQ	662	4.0	https://www.yelp.com/biz/two-8-two- bar-and-bur	{'city': 'Brooklyn', 'postal_code': '11201'}	{'latitude': 40.688536, 'longitude': -73.989728}	Walking around Brookly around lunchtime is a g
2	The River Café	the-river-café-brooklyn-3	4XX-zF8h5Lvrr22_tqU6-Q	1132	4.0	https://www.yelp.com/biz/the-river- caf%C3%A9-b	{'city': 'Brooklyn', 'postal_code': '11201'}	{'latitude': 40.70318, 'longitude': -73.9946499}	The perfect ambience with beautiful classical
3	Colonie	colonie-brooklyn	wouDDTacUIBaRp8YGCFTsA	580	4.0	https://www.yelp.com/biz/colonie- brooklyn?adju	{'city': 'Brooklyn', 'postal_code': '11201'}	{'latitude': 40.6909, 'longitude': -73.9958}	have passed by this restaurant at least 100t
4	Chez Moi	chez-moi-brooklyn	EF4M55USa1ZNy1rP5Kuk2w	380	4.0	https://www.yelp.com/biz/chez-moi- brooklyn?adj	{'city': 'Brooklyn', 'postal_code': '11201'}	('latitude': 40.6907305802403, 'longitude': -7	Wow. I ordered a few things for pickup and was
1444							S***		***

- Step 4 Scoring the reviews
 - spaCy for NLP
 - Identify the keyword match patterns
 - Count matches for points
 - Supervise the outcomes and assign labels
 - Tune match list for optimal accuracy

A little Insurance

 Force fed known positive results to confirm whether language detection and scoring concept were working darts

Restaurants
Home Services
Auto Services
More

1. Zablozki's

1. Zablozki's

1. Yablozki's

1. Yabloz



2. The Diamond

******* 110

(718) 383-5030 43 Franklin St Greenpoint

(718) 383-9699

Greenpoint

1041 Manhattan Ave

\$\$ · Bars

"shuffleboard, darts, a nice outdoor area. Overall, great casual atmosphere. Its a little off the beaten path so it's never too crowded to comfortably order at the bar. Great service, go out of your way to go here." more



3. Tommy's Tavern

33

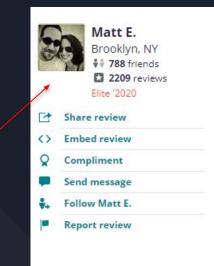
\$ · Bars

"almost seemed that he knew what I wanted to order. Bored? Not in this bar. They have a lovely **dart** board and a pool table that is placed strategically in the center of the bar. ****Please DO NOT place your beverage on the table***** the condensation from the glass will damage the fine solid oak...." more

Results

	name	alias	id	review_count	rating	url	location	coordinates	reviews	score
27	Floyd, NY	floyd-ny- brooklyn- 2	YL1LxYufuwX5sq4RwW-2Gw	210	4.0	https://www.yelp.com/biz/floyd-ny- brooklyn-2?a	{'city': 'Brooklyn', 'postal_code': '11201'}	{'latitude': 40.6907799, 'longitude': -73.9957}	I'm basically just rating this place as a reac	1
69	The Brazen Head	the- brazen- head- brooklyn	QiKtEJ6CFA3nSPRIRsq5Ew	126	3.5	https://www.yelp.com/biz/the- brazen-head-brook	{'city': 'Brooklyn', 'postal_code': '11201'}	{'latitude': 40.689235, 'longitude': -73.991653}	\nShare review\n\nEmbed review\n\nCompliment\n	5
76	White Horse Tavern - Financial District	white- horse- tavern- financial- district- new-york	IHGQM_w1x5Dxiq3aRPCTQ	372	4.0	https://www.yelp.com/biz/white- horse-tavern-fi	{'city': 'New York', 'postal_code': '10004'}	{'latitude': 40.70372, 'longitude': -74.01237}	3 floors. Good crowd. All ages. Fantastic happ	1
77	Zablozki's	zablozkis- brooklyn- 2	yaD4coYkCLtq6sn3692CRw	149	4.0	https://www.yelp.com/biz/zablozkis- brooklyn-2?	{'city': 'Brooklyn', 'postal_code': '11211'}	{'latitude': 40.7183462871818, 'longitude': -7	\nShare review\n\nEmbed review\n\nCompliment\n	2
79	Tommy's Tavern	tommys- tavern- brooklyn	BXQi7hBc2iUK60iq_eHnKg	33	2.0	https://www.yelp.com/biz/tommys- tavern-brookly	{'city': 'Brooklyn', 'postal_code': '11222'}	{'latitude': 40.73431, 'longitude': -73.955215}	My fingers can't move quick enough to express	1

False Positive - Floyd





Very cool place. Comfy vintage sofas like a nice coffee shop. The indoor bocce court alone is worth 4 stars but when you throw in a Pac Man machine and a place to get fried Twinkies next door (Chip Shop) you've got yourself a pretty cool place. The beers are a little pricer than I'd ideally like them if you're not getting their crappy canned beer (\$3). Bottles (even Bud Light) are \$5, and drafts are \$6.

Word of the wise, if you want to play Bocce get there early or sign up on the waiting list and expect to wait awhile (if you're coming on a Friday/Saturday night) because its a hit with the crowd. They used to be a very dog friendly bar, and still might be but now that they sell their own beer cheese (\$7) I'm not sure if they're still allowed to be.

A dart board, patio area, or cheaper beers might have been the only way to squeeze the 5th star from me, otherwise a nice place to throw a couple down.





THIS GUY!

Next Steps

- Apply for Yelp Fusion VIP
- Train the model
 - Prodigy (spaCy sibling)
- Scale up to take additional features
- Include other review/data sets
- Improve Speed for near real time
 - Amazon Web Services EMR for compute and parallelization
 - MRJobs Mapper Reducer developed by Yelp
 - Create an app