# Using Yelp Data for Restaurant Mapping

Understanding Yelp restaurants in Pittsburgh and how wifi impact rankings.

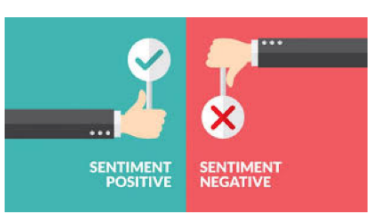
# Original Plan

## Twitter Data

Download twitter data to understand people's perceptions of ride-hailing services such as Uber and Lyft.

Map perceptions based on sentiment, positive vs. negative sentiments to ride-hailing services.

Unfortunately the dataset I compiled did not have enough LAT and LONG values. Many tweets had no location identifier.




Write a Review
Events
Talk
Log In
Sign Up
yelp
Find burgers, barbers, spas, handymen…
Near Downtown, Pittsburgh, PA
Restaurants
Nightlife
Home Services
Delivery
Bake Shack
Photo by Hope P.

# What is Yelp?

Yelp is a local-search service powered by a crowd-sourced review forum.

## Stats

Yelp was started in 2004, and now has 26 million monthly active users. Yelp is present in over 200 cities and 32 countries.

## Uses

Yelp users can find great local businesses like restaurants, dentists, hair stylists, and mechanics.

## Yelp Business Profile

Business Name

Categories (attributes)

Review Count

Stars

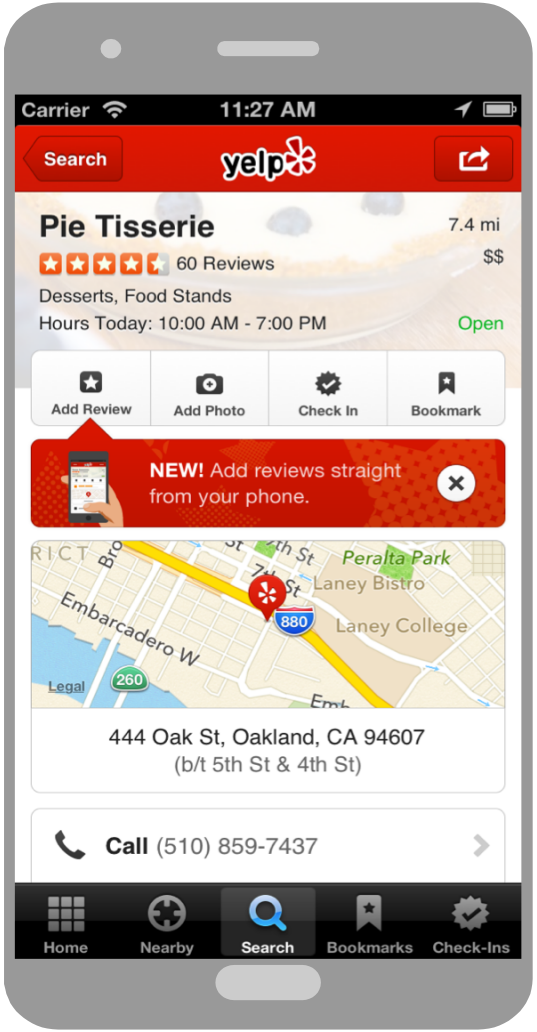
Location/ Map

Photos

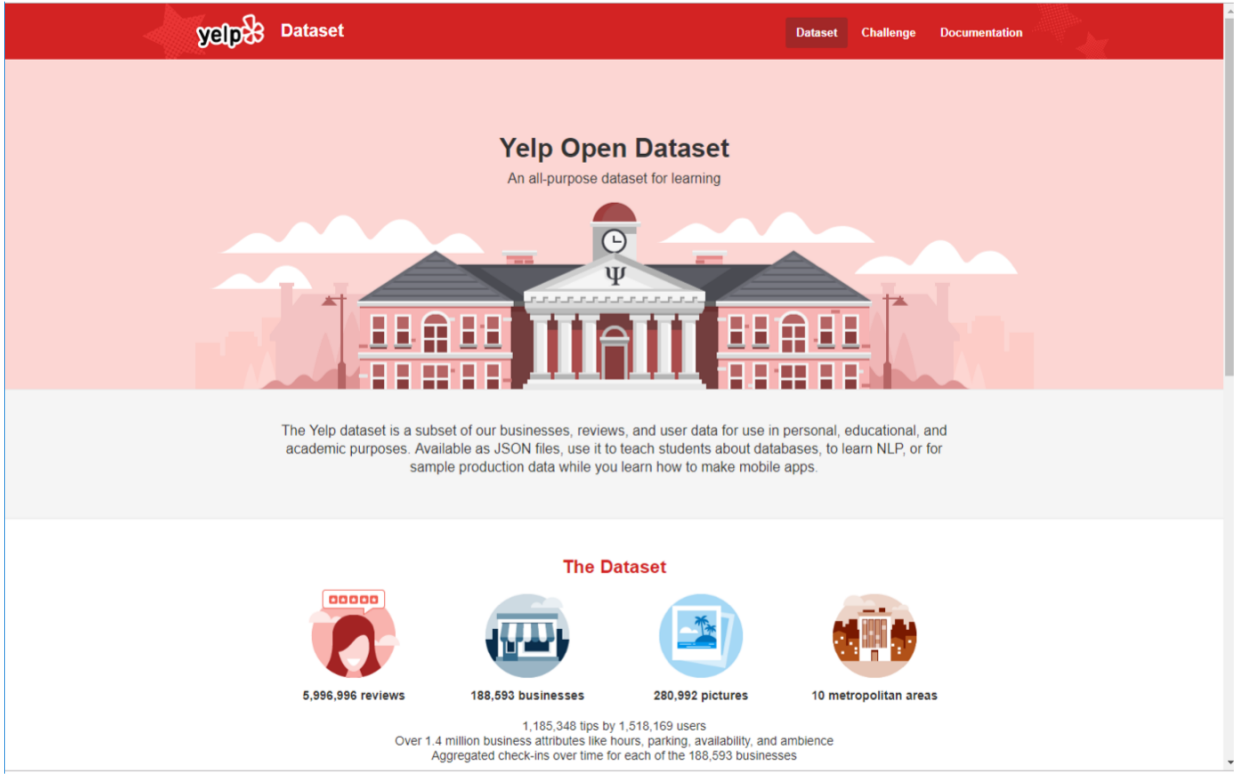
Hours of Operation

Price Range

& More



# Data Collection: Yelp




Yelp hosts an open data research contest each year. They provide JSON files for 12 major cities, Pittsburgh was one of the cities with Yelp Data.

# Data Collection: Yelp

## Data Files:

- Reviews
- Users
- ✓ Businesses
- Photos
- Checkin
- Tip
- Photo

| | NAME | ADDRESS | CITY | STATE | ZIPCODE | LAT | LONG | STARS | REVIEW_COUNT | IS_OPEN |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Dunkin Donuts | 145 Mcmurray Rd | Upper St Clair | PA | 15241 | 40.32602 | -80.06520 | 2.0 | 5 | 1 |
| 2 | China Palace | 5440 Walnut St | Pittsburgh | PA | 15232 | 40.45087 | -79.93392 | 3.0 | 110 | 1 |
| 3 | Pizza Bellagio | 4635 Centre Ave | Pittsburgh | PA | 15213 | 40.45241 | -79.95067 | 1.5 | 18 | 1 |
| 4 | Pittsburgh Poke | 500 Liberty Ave | Pittsburgh | PA | 15222 | 40.44132 | -80.00351 | 4.5 | 94 | 1 |
| 5 | Schenley Park Visitors Center | 101 Panther Hollow Rd | Pittsburgh | PA | 15213 | 40.43754 | -79.94888 | 5.0 | 12 | 1 |
| 6 | Vocelli Pizza | 2011 Waverly Ave | Swissvale | PA | 15218 | 40.42191 | -79.88635 | 2.5 | 10 | 1 |
| 7 | Crazy Mocha | 1836 Centre Ave | Pittsburgh | PA | 15219 | 40.44260 | -79.98229 | 5.0 | 3 | 1 |
| 8 | The Commoner Corner | 458 Strawberry Way | Pittsburgh | PA | 15219 | 40.44154 | -79.99614 | 3.5 | 13 | 0 |
| 9 | Arby's | 3200 Saw Mill Run Blvd | Pittsburgh | PA | 15227 | 40.37644 | -79.98411 | 3.0 | 3 | 0 |
| 10 | Cole's Pub | 527 US 30 | Imperial | PA | 15126 | 40.45143 | -80.25184 | 4.0 | 35 | 1 |
| 11 | Pizza Marsala | 10729 State Rt 30 | Irwin | PA | 15642 | 40.32014 | -79.68454 | 4.0 | 12 | 1 |
| 12 | Timeout Pizza Pub & Grub | 775 Freeport Rd | Cheswick | PA | 15024 | 40.54128 | -79.81897 | 2.0 | 5 | 0 |
| 13 | Go Pretzel | 807 Liberty Ave | Pittsburgh | PA | 15222 | 40.44267 | -79.99892 | 3.5 | 13 | 0 |
| 14 | Steel Cactus | 115 Federal St | Pittsburgh | PA | 15212 | 40.44758 | -80.00436 | 2.5 | 65 | 1 |
| 15 | Robert Wholey and Co Fish Market | 1711 Penn Ave | Pittsburgh | PA | 15222 | 40.45011 | -79.98587 | 4.0 | 210 | 1 |
| 16 | Bull River Taco | 1707 Murray Ave | Pittsburgh | PA | 15217 | 40.43773 | -79.92273 | 3.5 | 60 | 1 |
| 17 | DiBella's Subs | 6501 Steubenville Pike | Pittsburgh | PA | 15205 | 40.44800 | -80.15872 | 4.0 | 31 | 1 |
| 18 | Union Standard | 524 William Penn Pl | Pittsburgh | PA | 15219 | 40.44042 | -79.99720 | 4.0 | 117 | 1 |
| 19 | Pizzarita | 580 Burchfield Rd | Allison Park | PA | 15101 | 40.54453 | -79.96287 | 3.0 | 10 | 1 |
| 20 | Smoq Pitt | 600 Brookline Blvd | Pittsburgh | PA | 15226 | 40.39518 | -80.02250 | 3.5 | 85 | 0 |

# Data Collection: Shapefiles

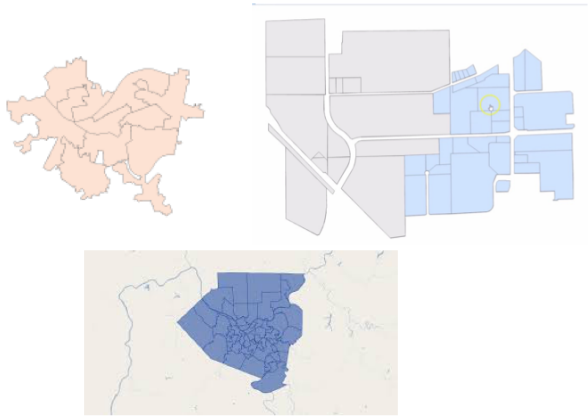
## Sources

Allegheny County, City of Pittsburgh

American Community Survey, United States Census Bureau

Western Pennsylvania Regional Data Center

Yelp, Pennsylvania Spatial Data Access



## Data Types

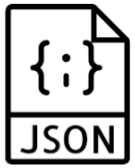
Pittsburgh Neighborhood (city)

Allegheny County Neighborhoods (region)

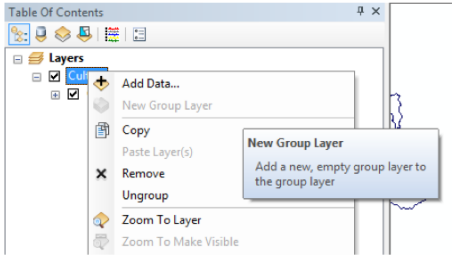
Streets

Parcels/ Building Footprints

# ArcMap Process


Download Yelp Dataset into R for cleaning.



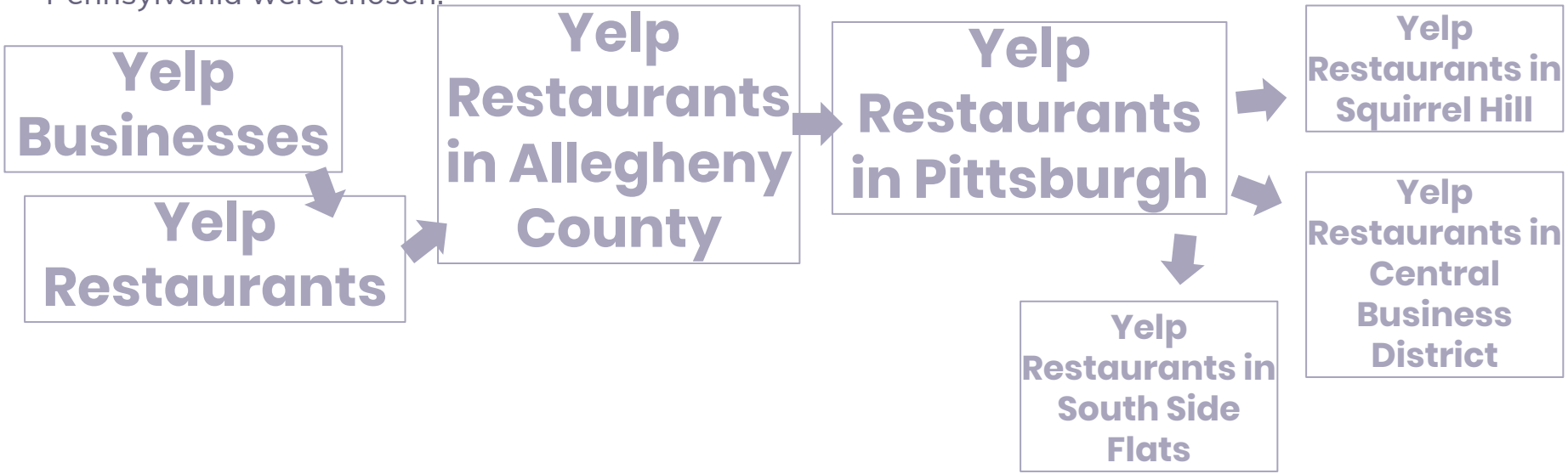

Import all data tables and polygon layers into ArcMap.

Determine hypothesis/ question and test results.

Conducting spatial joins to analyze restaurants for each neighborhood.

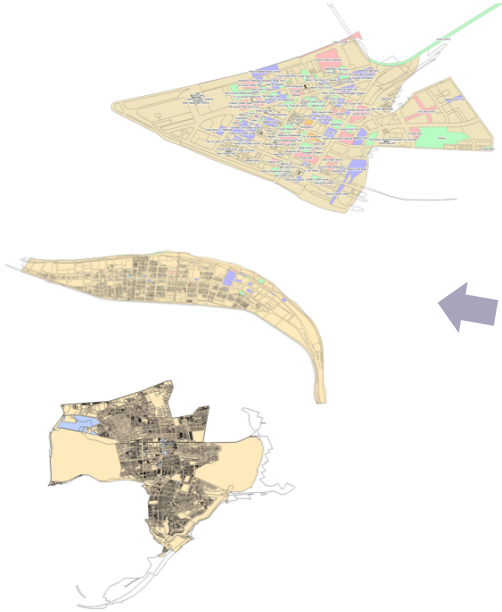
# Filtering Data

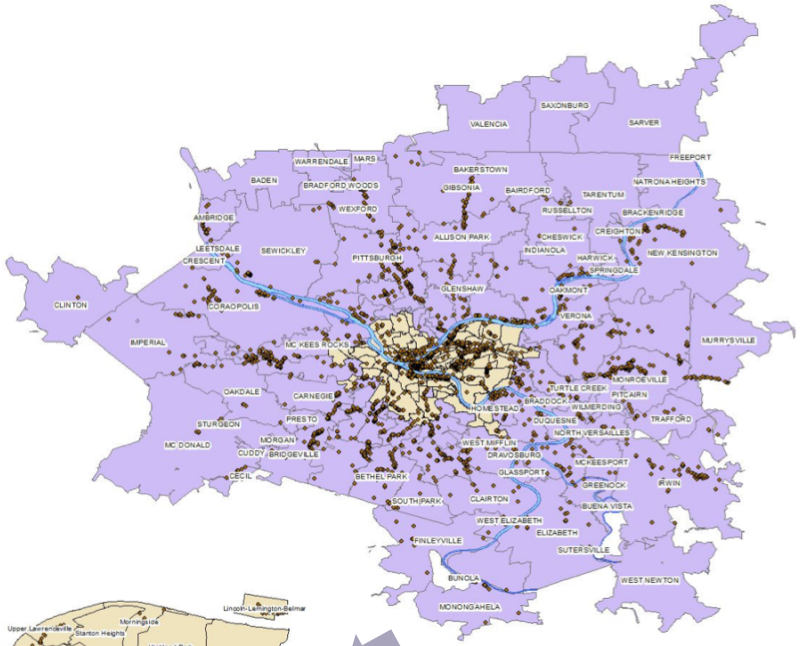
I used R to filter and clean data before uploading it to ArcGIS. The Yelp Dataset had to be converted from JSON format to CSV. Other filtering steps were needed to ensure that only Restaurants in Pennsylvania were chosen.



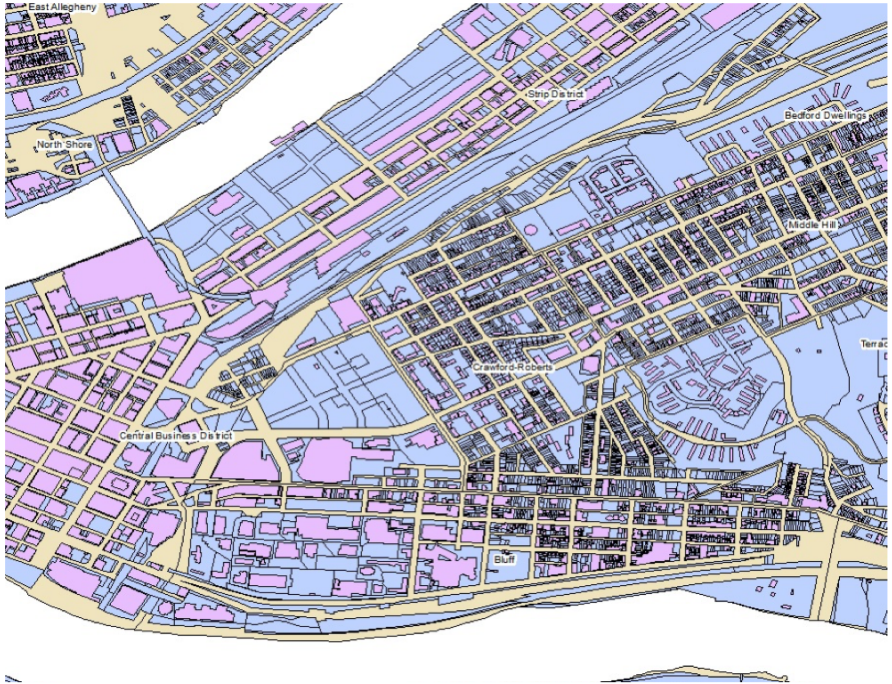
# Filtering Data

After filtering, the dataset became more manageable and I was able to determine which areas I wanted to examine and restaurant characteristics I wanted to analyze.





# ArcMap Process





Parcels are boundary outlines of individual properties (2018)


Footprints are  building outlines by roof line (2018)

*Source: Pennsylvania Spatial Data Access*

# Descriptive Stats

## 3,680 Restaurants on Yelp in Allegheny County

And 1,540 are located in the city of Pittsburgh

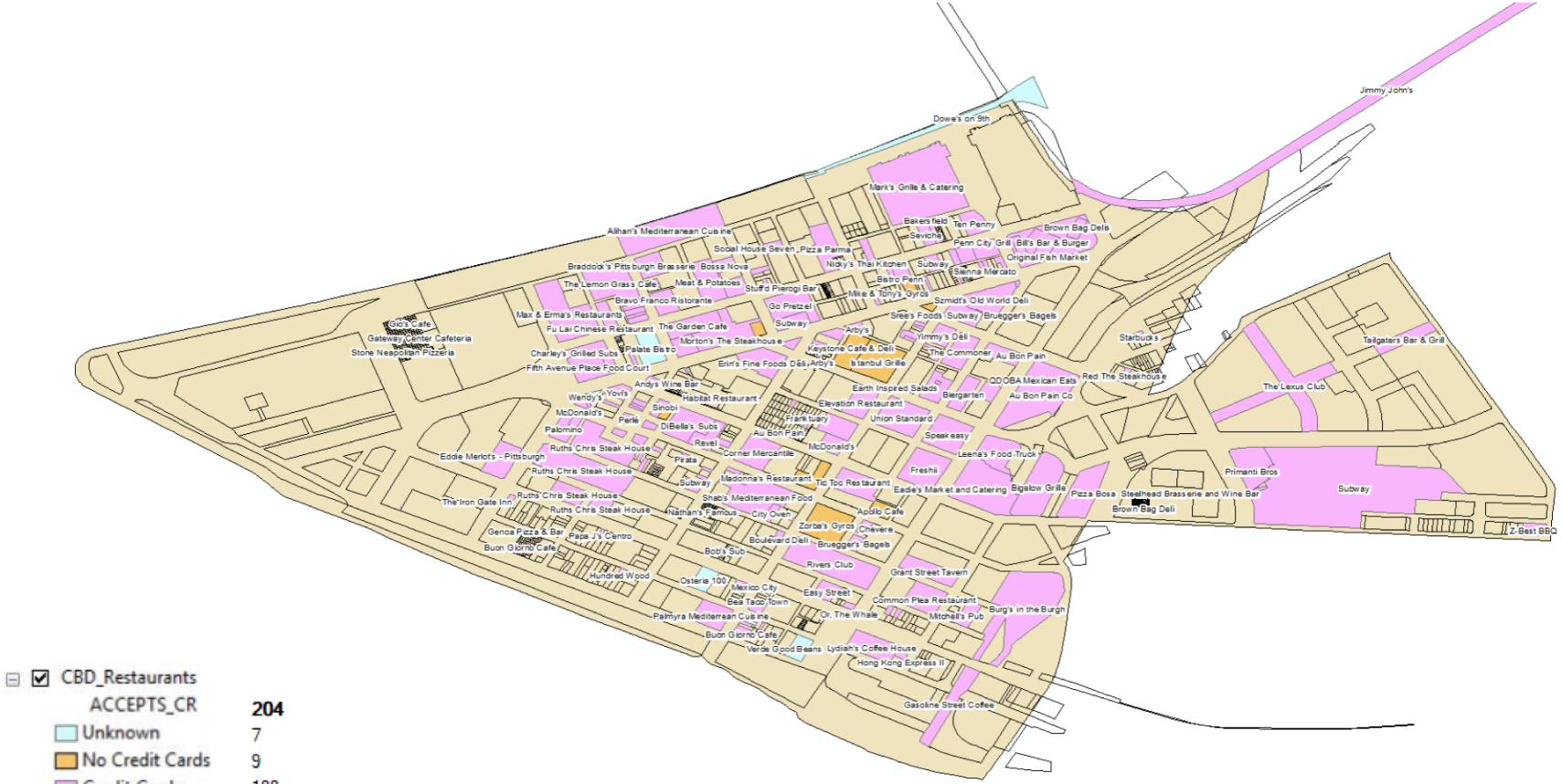
## 115 Restaurant Types/ Categories

Including Pizza, American Food, Bars, Fast Food, & Many More

## 38 Yelp Restaurant Attributes

Such as Noise Level, GoodforKids, Business Parking, & More

# Central Business District – 204

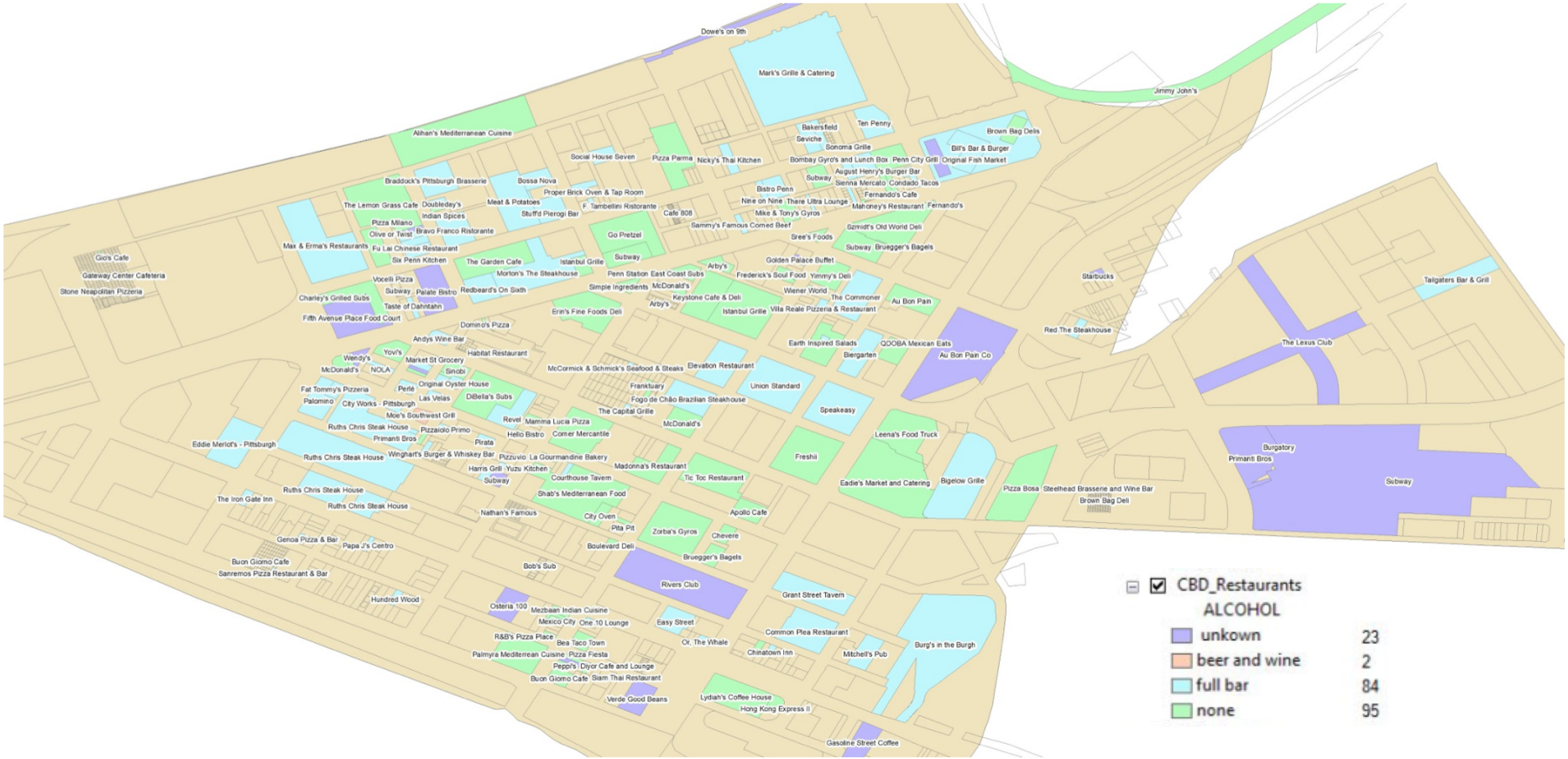



CBD_Restaurants

| ACCEPTS_CR | 204 |
|---|---|
| Unknown | 7 |
| No Credit Cards | 9 |
| Credit Cards | 188 |

# Central Business District – 204

# Central Business District – 204




CBD_Restaurants

ALCOHOL

| | |
|---|---|
| unkown | 23 |
| beer and wine | 2 |
| full bar | 84 |
| none | 95 |

# South Side Flats – 130

SouthSideFlats_Parcel_Join2

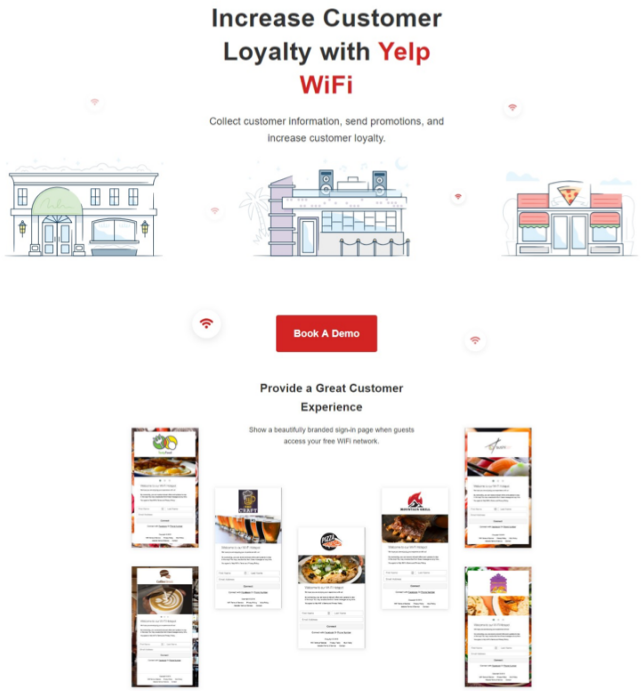
WHEELCHAIR

| | |
|---|---|
| Unknown | 60 |
| No Wheelchair Access | 11 |
| Wheelchair Access | 59 |

# Question 1:

Do restaurants WITH WIFI have more STARS than restaurants with NO WIFI?

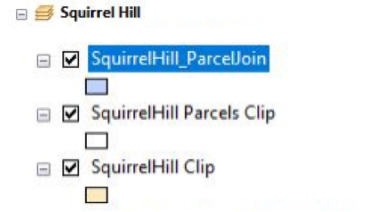
(Is there a relationship between WIFI and Rank?)

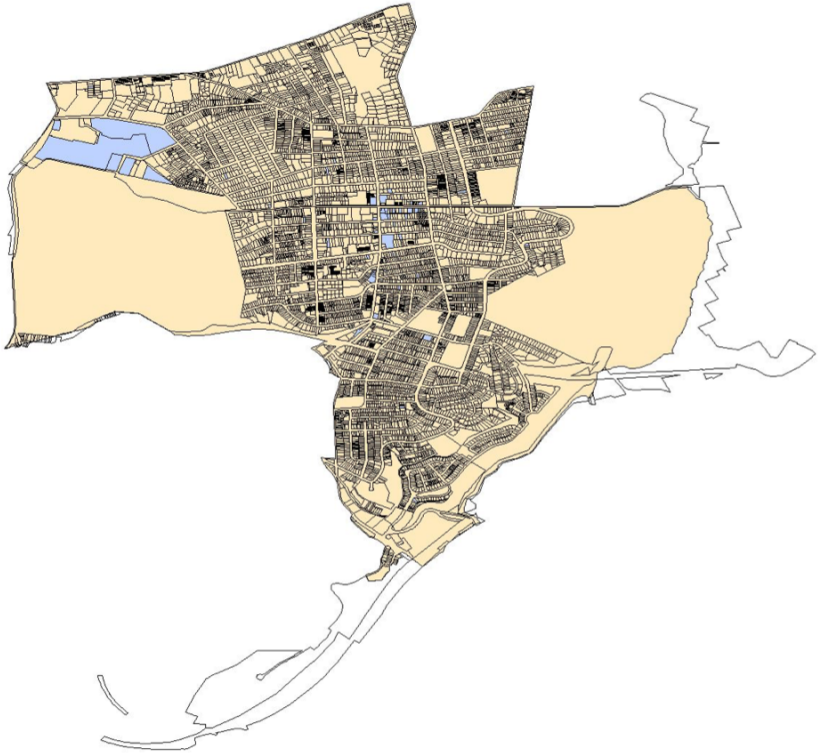


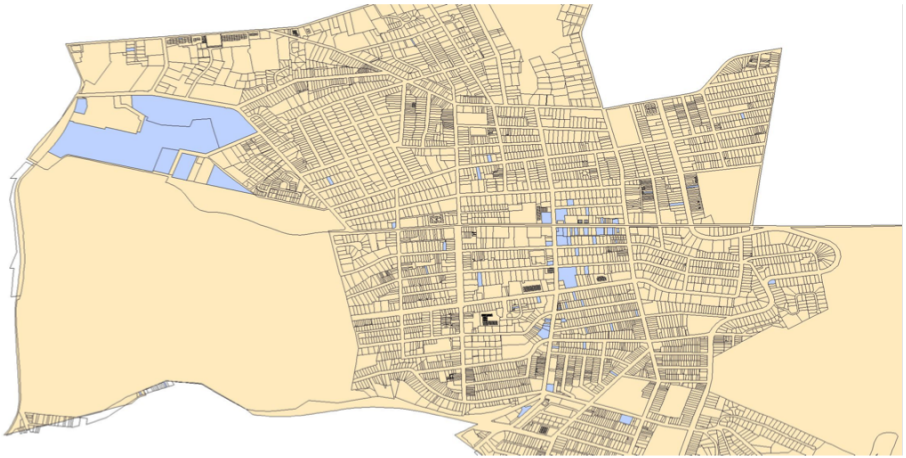

# Map

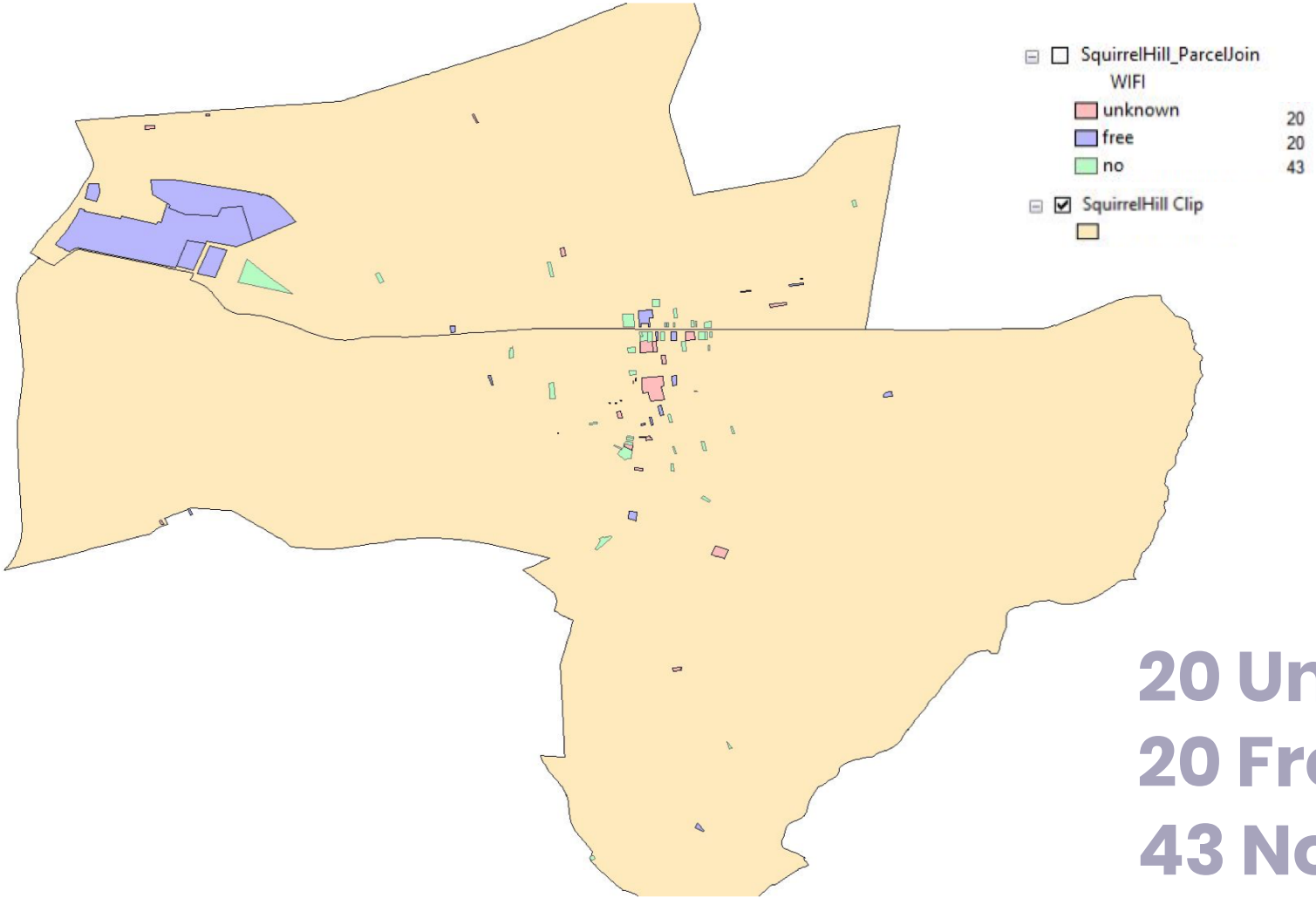
## Squirrel Hill Yelp Restaurants 83

(Actually 103)

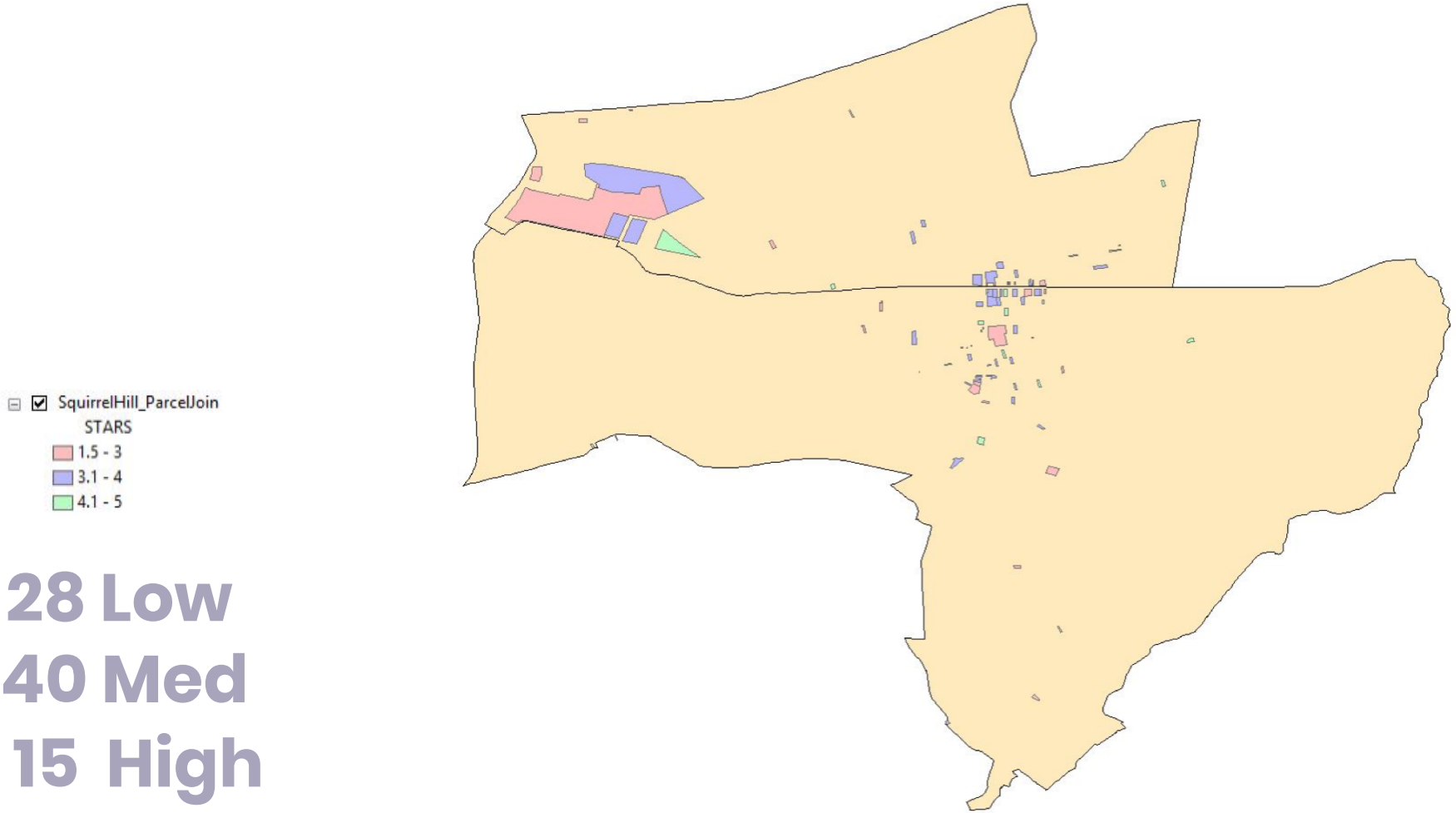


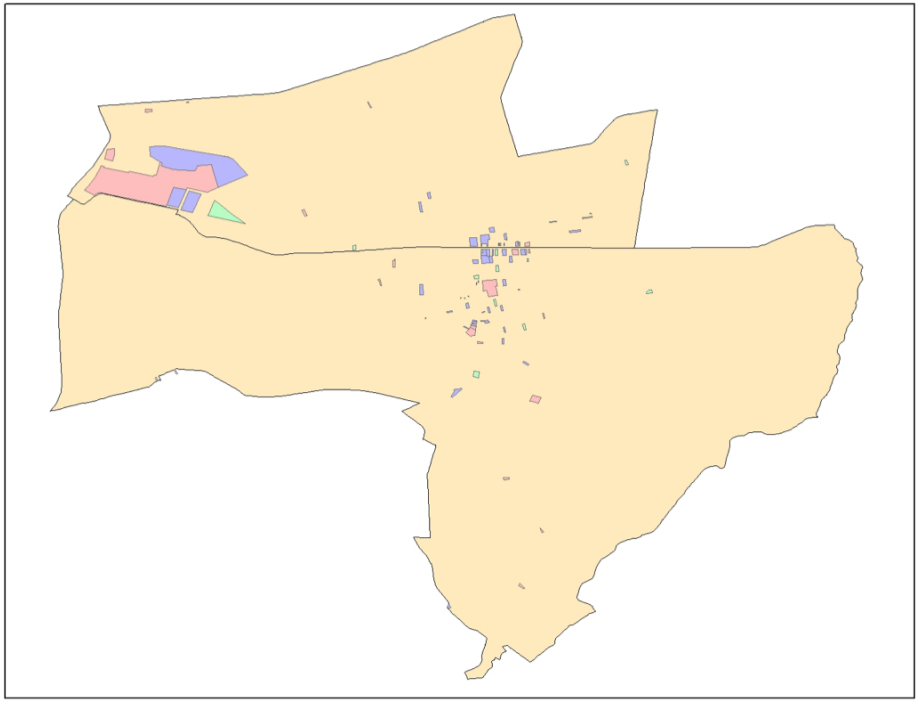


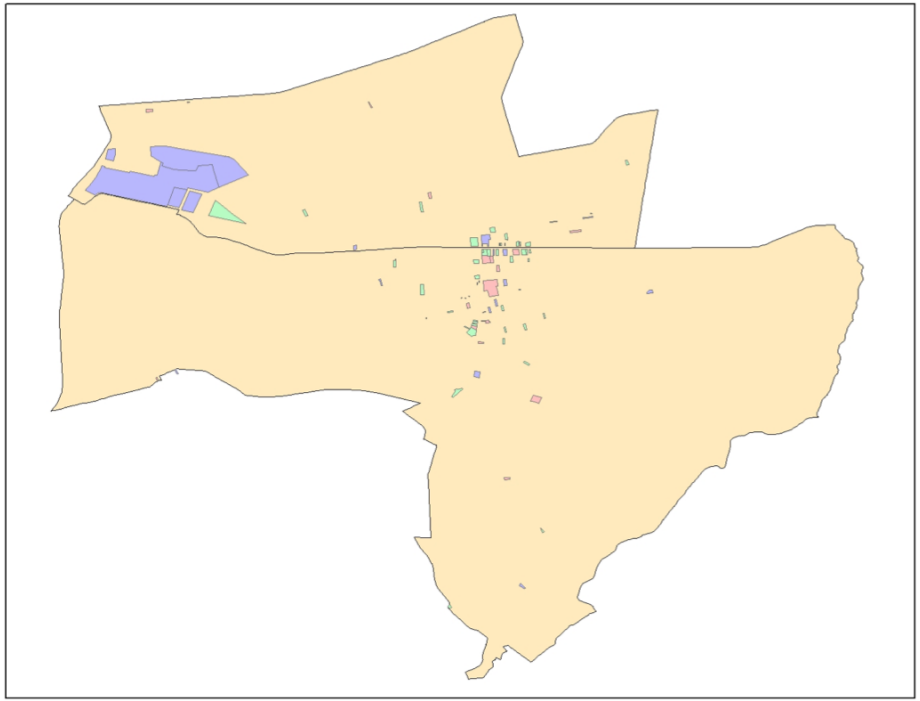



**20 Unknown**
**20 Free**
**43 No**




**28 Low**
**40 Med**
**15 High**





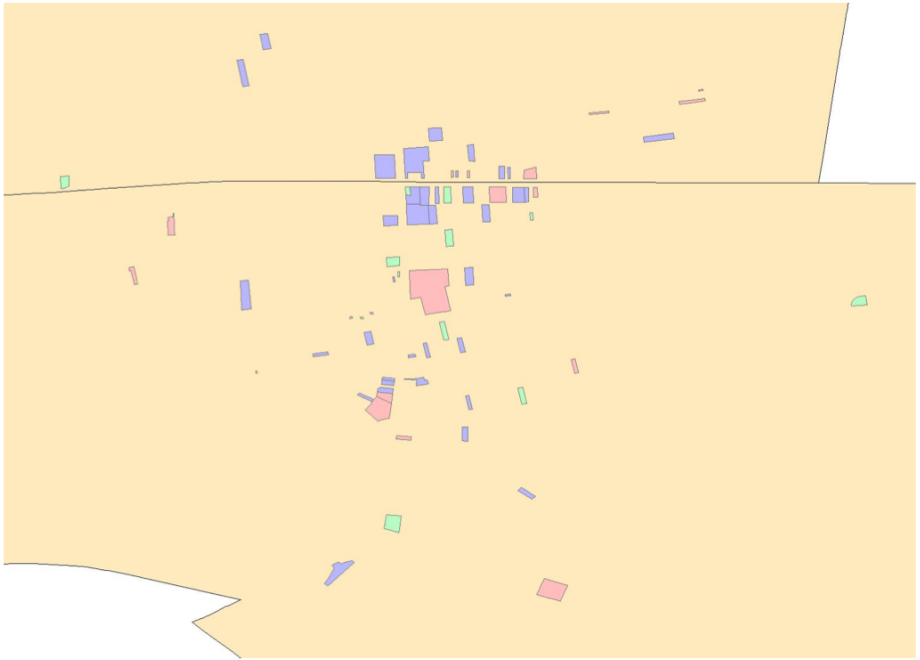
SquirrelHill_ParcelJoin
STARS
1.5 - 3
3.1 - 4
4.1 - 5

STARS

SquirrelHill_ParcelJoin
WIFI
unknown
free
no

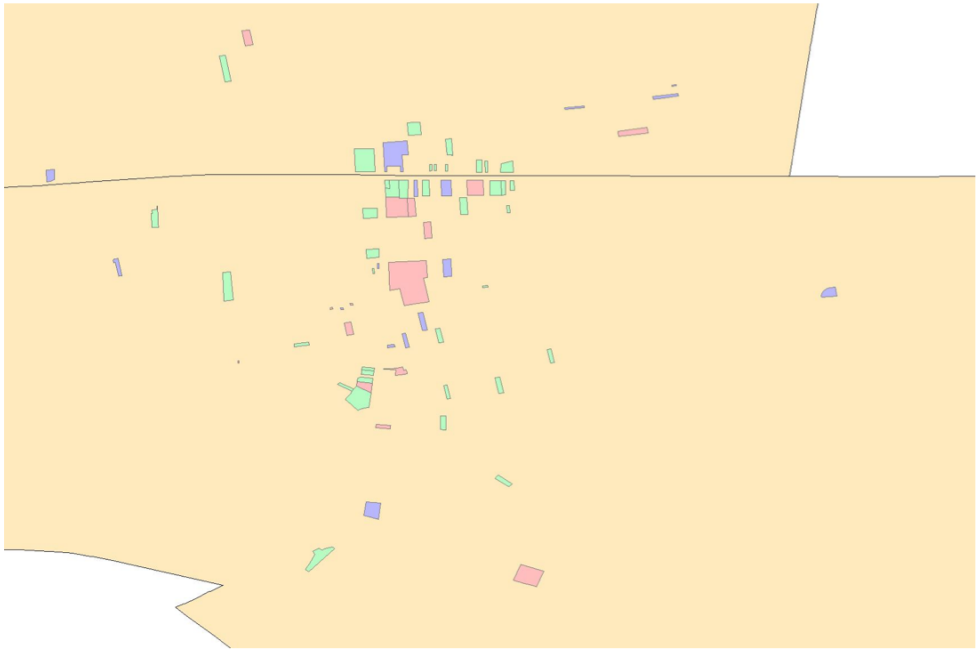
WIFI

# Map



SquirrelHill_ParcelJoin
STARS
1.5 - 3
3.1 - 4
4.1 - 5

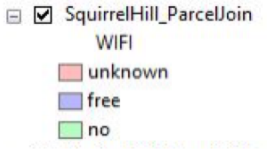
STARS



SquirrelHill_ParcelJoin
WIFI
unknown
free
no

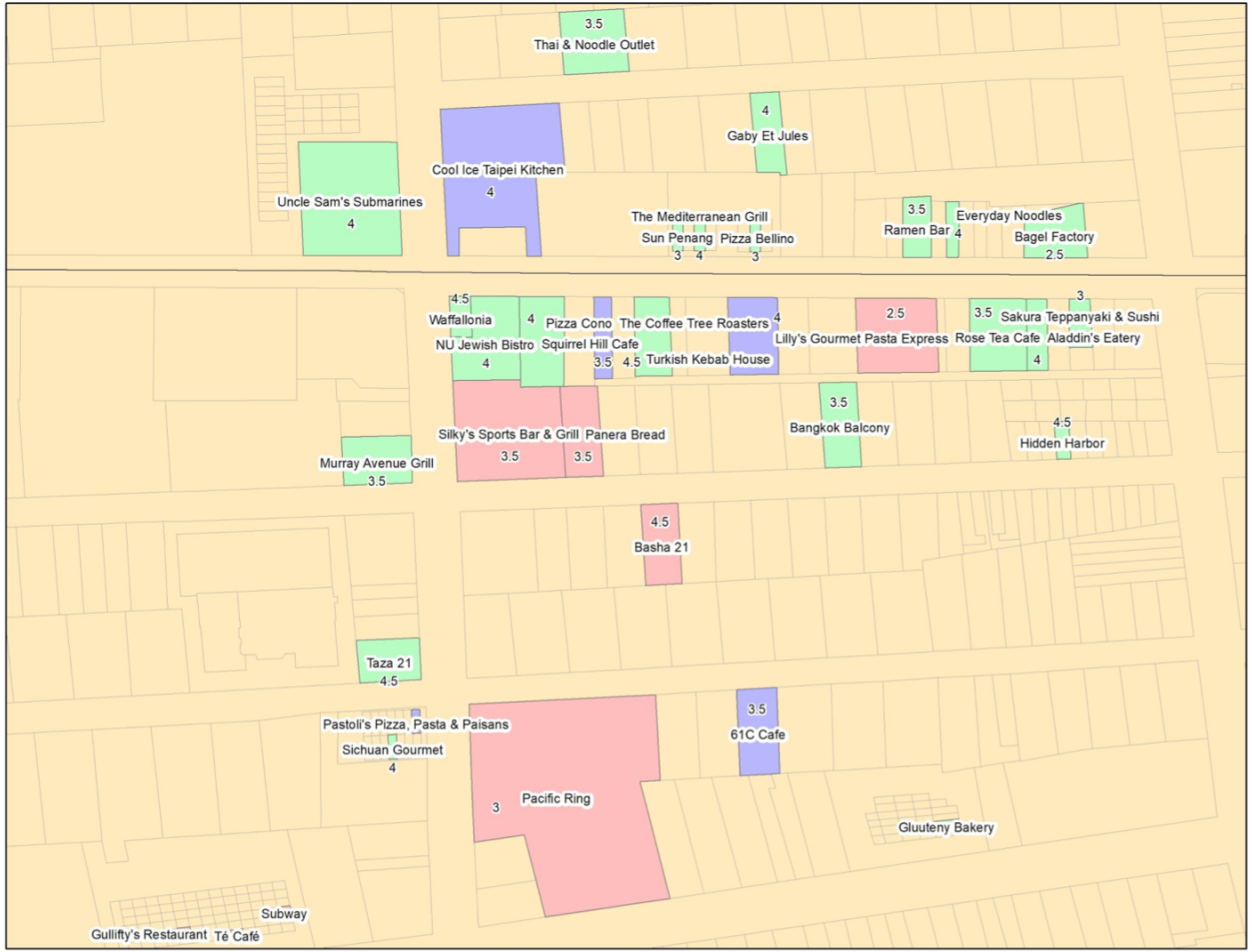
WIFI

# Findings




Both restaurants with and without WIFI have high ranks on Yelp. After determining the correlation coefficient, there was no linear relationship between WIFI and RANK.

```
x <- shill[24] #stars
y <- shill[30] #wifi
cor(x, y, use="complete.obs", method = "pearson")
#-0.04367569
```



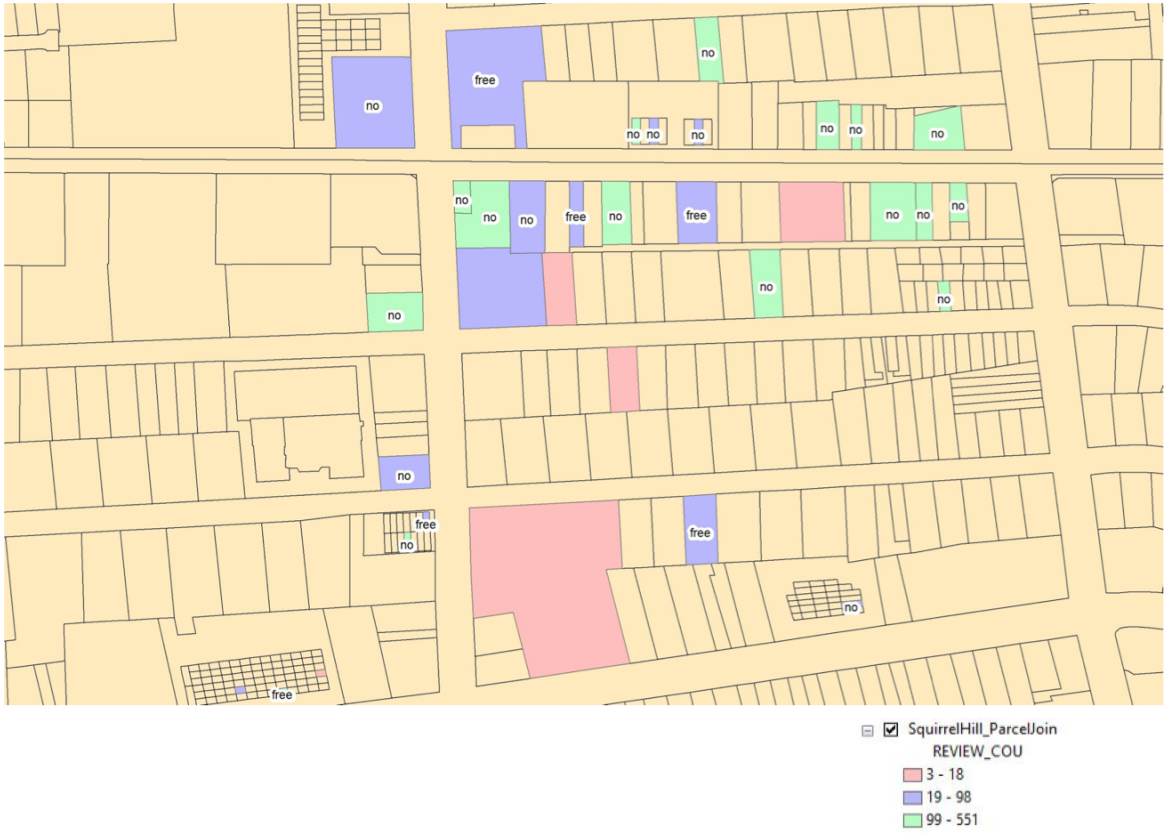
# Findings

There was no apparent correlation between WIFI and RANK.

```
x <- shill[24] #stars
y <- shill[30] #wifi
cor(x, y, use="complete.obs", method = "pearson")
#-0.04367569
```

Weak negative linear relationship between REVIEW_COUNT and WIFI.

```
x <- shill[25] #review_count
y <- shill[30] #wifi
cor(x, y, use="complete.obs", method = "pearson")
#-0.3693391
```

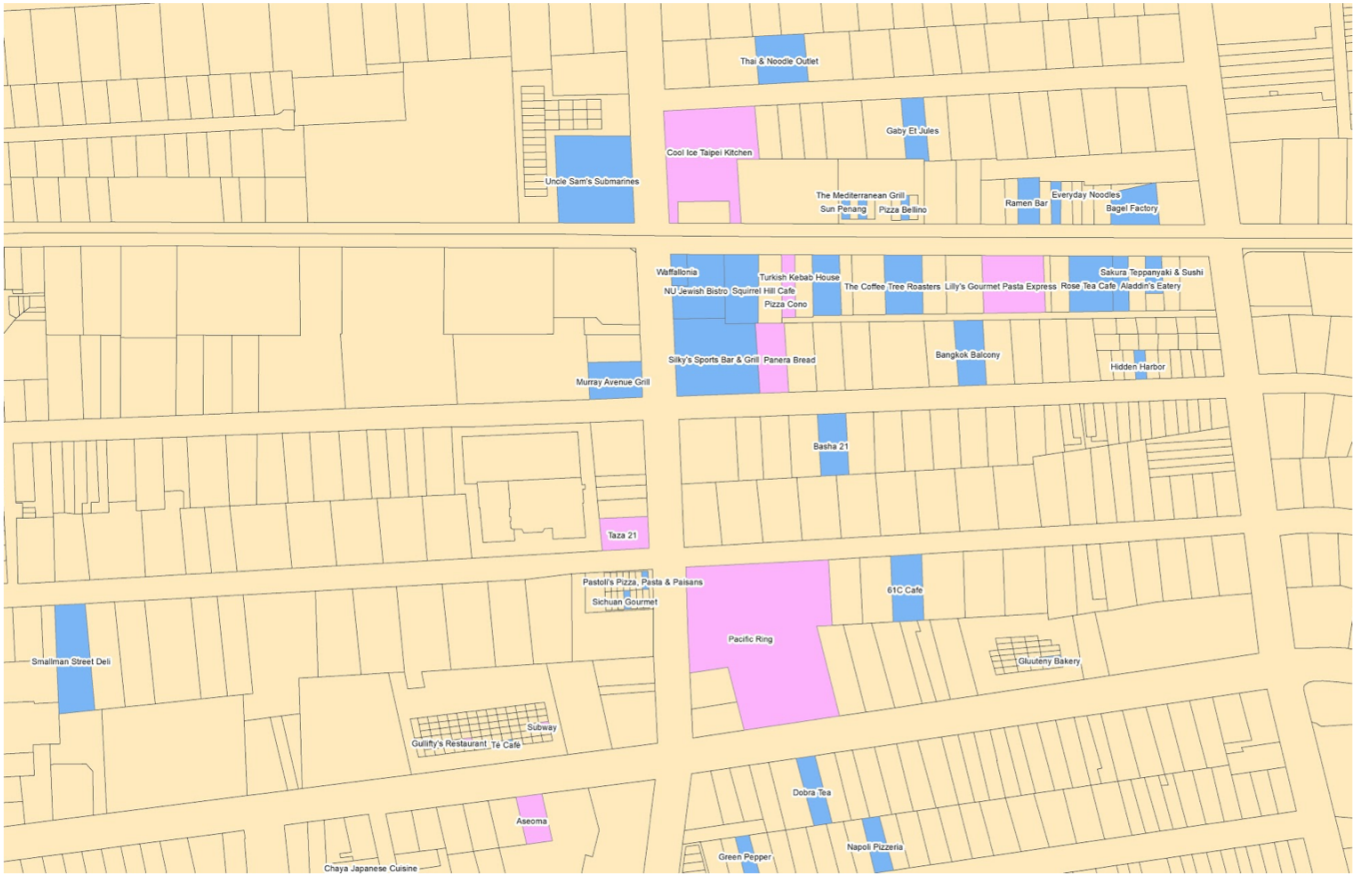



↑Review Count ↓WIFI Access

# Findings

Almost a 24% of the 83 restaurants in Squirrel Hill restaurants are closed.

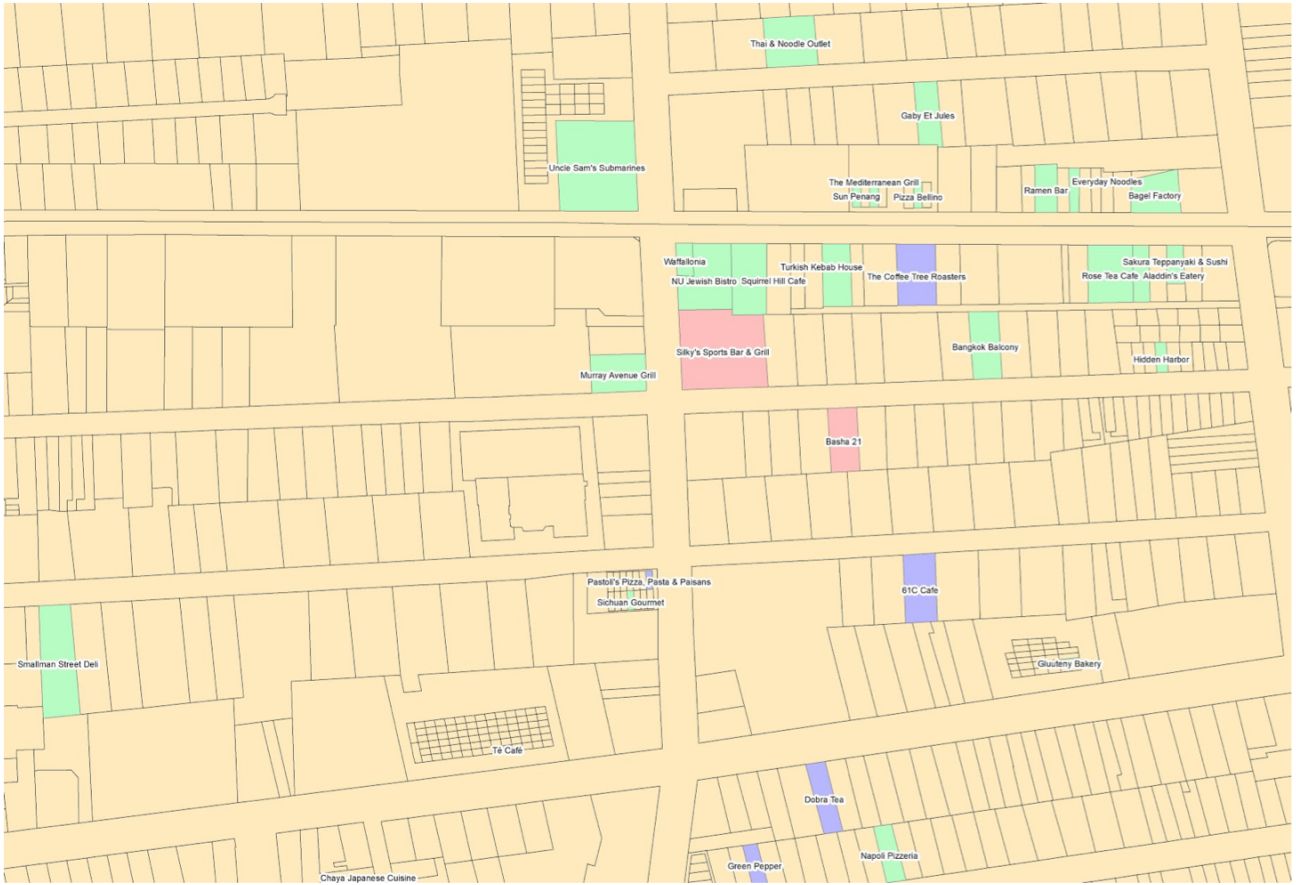

**21 Closed**
**62 Open**



# Findings

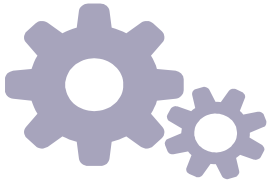
shill_IsOpen
WIFI
unknown
free
no

**7 Unknown**
**15 Free**
**40 No**



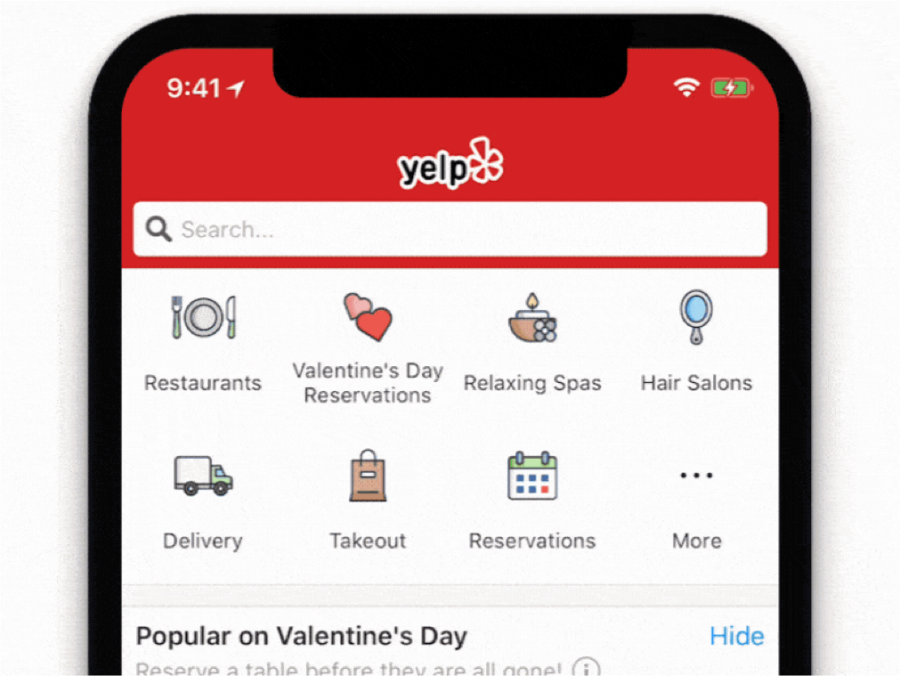
# Future Questions

Does picture quality or picture amount influence ranking/ star amount?
If an establishment has a significant amount of pictures does this help their ranking?



Positive vs. Negative reviews, how are the weighted?
Do negative reviews on a profile distract people from the positive reviews?

Lessons Learned?


9:41
yelp
Search...
Restaurants
Valentine's Day Reservations
Relaxing Spas
Hair Salons
Delivery
Takeout
Reservations
More
Popular on Valentine's Day
Hide

# Thank You

# Work cited for shapefile links

- city and county shapefiles https://data.wprdc.org/dataset/city-of-pittsburgh-gis-data
- buildings https://data.wprdc.org/dataset/allegheny-county-building-footprint-locations
- parcels http://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=1214 , http://openac-alcogis.opendata.arcgis.com/datasets/allegheny-county-parcel-boundaries
- find identical https://gis.stackexchange.com/questions/115672/remove-duplicated-items-in-arc-attribute-table
- select attributes cheat sheet https://www.esri.com/news/arcuser/0405/files/fieldcalc_1.pdf
- understanding the symbology tabs http://pro.arcgis.com/en/pro-app/help/mapping/layer-properties/symbolize-feature-layers.htm
- understanding summarize tools https://doc.arcgis.com/en/arcgis-online/analyze/perform-analysis.htm

# Work cited for academic articles


Tourist dining preferences based on reviews
https://journals.sagepub.com/doi/full/10.1177/0047287517744672?utm_source=summon&utm_medium=discovery-provider&
Promotional marketing, word of mouth
http://go.galegroup.com/ps/i.do?p=AONE&u=upitt_main&id=GALE|A344209646&v=2.1&it=r&sid=summon
Reviews and gentrification
https://journals.sagepub.com/doi/full/10.1177/1469540515611203?utm_source=summon&utm_medium=discovery-provider
Yelp wifi https://www.yelpwifi.com/
Yelp content guidelines https://www.yelp.com/guidelines
Yelp ranking factors http://www.localvisibilitysystem.com/2012/08/23/yelp-ranking-factors/
Yelp contest page https://www.yelp.com/dataset/documentation/main
Filtering with stringr https://blog.exploratory.io/filter-with-text-data-952df792c2ba
Correlations https://www.dummies.com/education/math/statistics/how-to-interpret-a-correlation-coefficient-r/