ALL FOOD IN THE HOOD

Yelp API comparison study of eateries and regions in the Bay Area

Phillip Choi · Natalie Odgerel St · Loba Quasem · Carly Russell

PRESENTATION MENU

A PROJECT IN FOUR COURSES

Click item for more details

Cocktails List

Api-tizer

Entrée

Dessert

Total Bill

INTRODUCTION

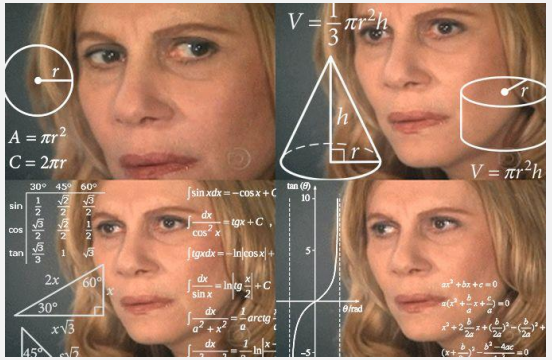
The Bay Area's diverse & innovative nature nourishes a hub of unique eats and is home to some of the top restaurants in the world

Problem Statement
Evaluating Bay regions for a foodie to visit or reside

CHALLENGES:

Thousands of Restaurants

Vast Area



Evaluating

HYPOTHESIS

Dense urban areas in the Bay Area,

eg. SF, Oakland

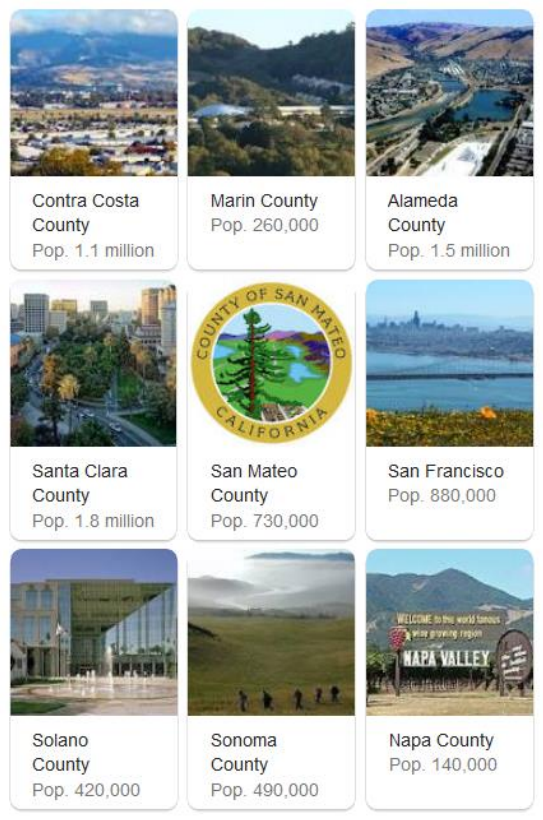
WILL HAVE:

More top-rated restaurants

More cuisine types

PROJECT SCOPE

Grade & **Compare** 9 Bay Counties



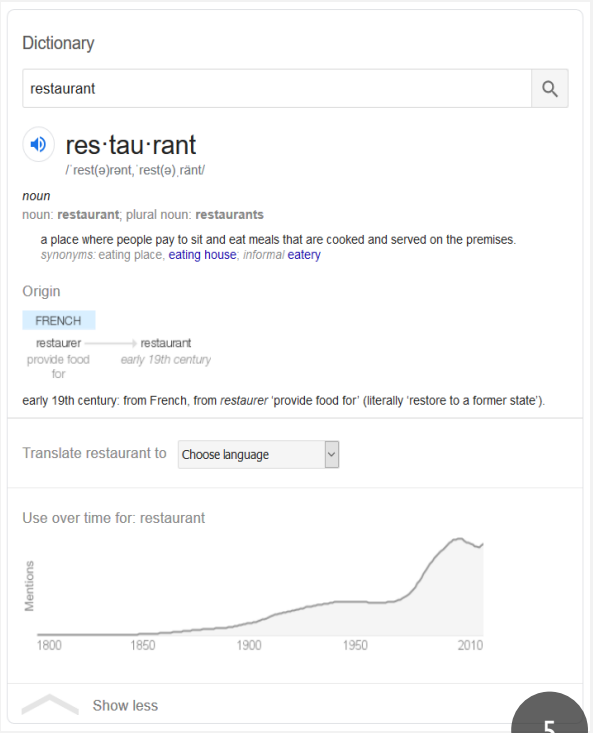
Yelp as sole resource

Defining:

-restaurant

-top-rated

-categories



YELP API

yelp fusion

Applicable Query Parameters:

- -Term search
- -Location search
- -Geocoordinates
- -Search radius
- -Sort by Yelp rating

```
def get_restaurants(lat, lng, api_key):
    url = "https://api.yelp.com/v3/businesses/search"
    headers = {"Authorization": "Bearer %s" % api_key}
    restaurant_data = []
    yelp_data = []
    count = 0

    for offset in range(0, 1000, 50):

        # Set parameters and pass into API calls, radius 8046 meters = 5 miles
        params = {"term":"restaurants", "latitude":lat, "longitude":lng, "radius":3412,
                  "limit":50, "offset":offset}
        req = requests.get(url, params=params, headers=headers).json()
        count += 1
        print(f'Now processing set {count} of max 20')

        if req["businesses"] == []:
            break
        else:
            for business in req['businesses']:

                business_dic = {}
                business_dic['Query ID'] = str(lat) + str(lng)
                business_dic['Query Lat'] = lat
                business_dic['Query Lng'] = lng
```

Each query returns max 1k results

YELP API


```
{
  "total": 8228,
  "businesses": [
    {
      "rating": 4,
      "price": "$",
      "phone": "+14152520800",
      "id": "E8RJkjfdcwgtyoPMjQ_Olg",
      "alias": "four-barrel-coffee-san-francisco",
      "is_closed": false,
      "categories": [
        {
          "alias": "coffee",
          "title": "Coffee & Tea"
        }
      ],
      "review_count": 1738,
      "name": "Four Barrel Coffee",
      "url": "https://www.yelp.com/biz/four-barrel-coffee-san-francisco",
      "coordinates": {
        "latitude": 37.7670169511878,
        "longitude": -122.42184275
      },
      "image_url": "http://s3-media2.fl.yelpcdn.com/bphoto/MmgtASP3l_t4tPCL1iAsCg/o.jp
      "location": {
        "city": "San Francisco",
        "country": "US",
        "address2": "",
        "address3": "",
        "state": "CA",
        "address1": "375 Valencia St",
        "zip_code": "94103"
      },
      "distance": 1604.23,
      "transactions": ["pickup", "delivery"]
    },
    // ...
  ],
  "region": {
    "center": {
      "latitude": 37.767413217936834,
      "longitude": -122.42820739746094
    }
  }
}
```


Sample Yelp API response

Datapoints in red fit for study

Value in “total” is the number of results in query

Max 7228 results omitted in sample query

ASSUMPTIONS

Foodie's preferences

Yelp is sole reference

Yelp data complete and error-free

All reviews unbiased and standardized

Ratings = Quality of Food

MODELING

For set of all restaurants in Bay:

Total Restaurants : T_x

Total Restaurants+ : T_{x+}

Avg Rating : $\mu_r = \frac{1}{T_x} * \sum_{i=1}^{T_x} r_i$

Rating SD : $\sigma_r = \sqrt{\frac{1}{T_x} * \sum_{i=1}^{T_x} (r_i - \mu_r)^2}$

Evaluate:

Avg Restaurants per County : $\mu_x = \frac{T_x}{9}$

Avg Restaurants+ per County : $\mu_{x+} = \frac{T_{x+}}{9}$

+ denotes restaurants with ratings ≥ 3.5

MODELING

For each county:

Total Restaurants : $T_{x \in c}$

Total Restaurants+ : $T_{x+ \in c}$

Avg Rating : $\mu_{r \in c} = \frac{1}{T_{x \in c}} * \sum_{i=1}^{T_{x \in c}} r_i$

Evaluate:

Restaurants per County SD : $\sigma_x = \sqrt{\frac{1}{9} * \sum_{i=1}^{9} (T_{x \in c_i} - \mu_x)^2}$

Restaurants+ per County SD : $\sigma_{x+} = \sqrt{\frac{1}{9} * \sum_{i=1}^{9} (T_{x+ \in c_i} - \mu_{x+})^2}$

MODELING

For each county:

Total Categories :	$T_{CAT \in c}$
Total Categories+ :	$T_{CAT+ \in c}$

Evaluate:

Avg Category per County :	μ_{CAT}	$= \frac{1}{9} * \sum_{i=1}^{9} T_{CAT \in c_i}$
Category per County SD :	σ_{CAT}	$= \sqrt{\frac{1}{9} * \sum_{i=1}^{9} (T_{CAT \in c_i} - \mu_{CAT})^2}$
Avg Category+ per County :	μ_{CAT+}	$= \frac{1}{9} * \sum_{i=1}^{9} T_{CAT+ \in c_i}$
Category+ per County SD :	σ_{CAT+}	$= \sqrt{\frac{1}{9} * \sum_{i=1}^{9} (T_{CAT+ \in c_i} - \mu_{CAT+})^2}$

+ denotes a category containing 1 or more restaurants with rating ≥ 3.5

MODELING

For each county, grade following characteristics using listed conditions:

Avg Rating Score: $\mathbf{s}(\mu_{r\in c}) = \begin{cases} 60, & \mu_{r\in c} < \mu_r - \sigma_r \\ 70, & \mu_r - \sigma_r \le \mu_{r\in c} < \mu_r \\ 80, & \mu_r \le \mu_{r\in c} < \mu_r + \sigma_r \\ 90, & \mu_r + \sigma_r \le \mu_{r\in c} < \mu_r + 2\sigma_r \\ 100, & \mu_{r\in c} \ge \mu_r + 2\sigma_r \end{cases}$

Total Restaurants Score: $\mathbf{s}(T_{x\in c}) = \begin{cases} 60, & T_{x\in c} < \mu_x - \sigma_x \\ 70, & \mu_x - \sigma_x \le T_{x\in c} < \mu_x \\ 80, & \mu_x \le T_{x\in c} < \mu_x + \sigma_x \\ 90, & \mu_x + \sigma_x \le T_{x\in c} < \mu_x + 2\sigma_x \\ 100, & T_{x\in c} \ge \mu_x + 2\sigma_x \end{cases}$

Total Restaurants+ Score: $\mathbf{s}(T_{x+\in c}) = \begin{cases} 60, & T_{x+\in c} < \mu_{x+} - \sigma_{x+} \\ 70, & \mu_{x+} - \sigma_{x+} \le T_{x+\in c} < \mu_{x+} \\ 80, & \mu_{x+} \le T_{x+\in c} < \mu_{x+} + \sigma_{x+} \\ 90, & \mu_{x+} + \sigma_{x+} \le T_{x+\in c} < \mu_{x+} + 2\sigma_{x+} \\ 100, & T_{x+\in c} \ge \mu_{x+} + 2\sigma_{x+} \end{cases}$

Total Categories Score: $\mathbf{s}(T_{CAT\in c})$ = *similar structure as above*

Total Categories+ Score: $\mathbf{s}(T_{CAT+\in c})$ = *similar structure as above*

OVERALL COUNTY SCORE: $\mathbf{0.3}\,[\mathrm{S}(T_{x+\in c}) + \mathrm{S}(T_{CAT+\in c})] \;+\; \mathbf{0.2}\,[\mathrm{S}(\mu_{r\in c})] \;+\; \mathbf{0.1}\,[\mathrm{S}(T_{x\in c}) + \mathrm{S}(T_{CAT\in c})]$

DATA COLLECTION

Results of Yelp API call attempts using search parameter:

City or Zip

- 101 cities, 394 zip codes
- Returns best match
- 57k results
- 50% of queries over limit
- Missing & duplicate entries
- Default search radius too broad
- Invalid datapoints

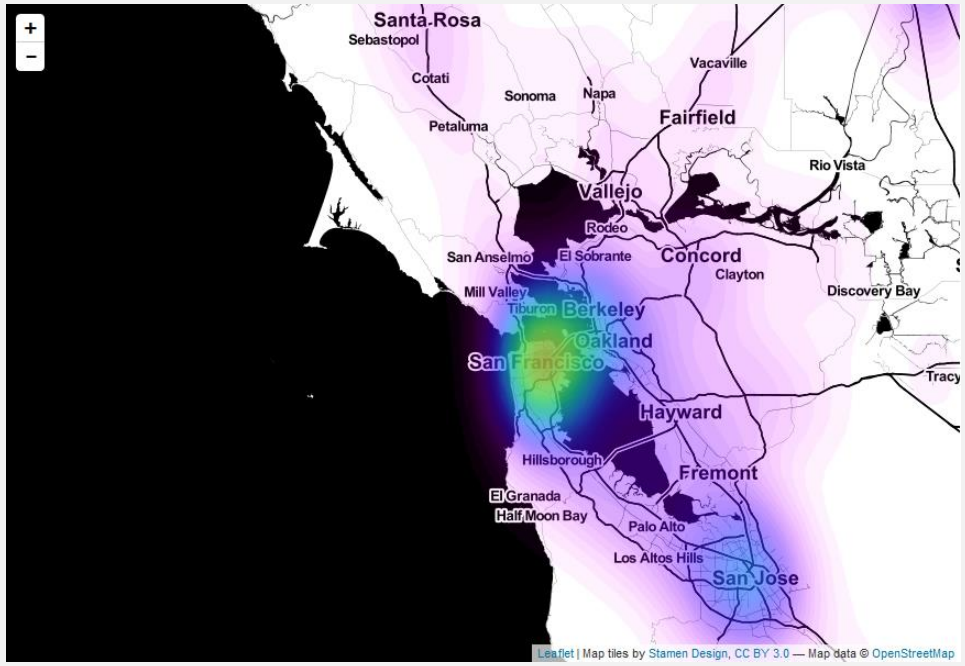
Geocoordinates with Search Radius Overlap

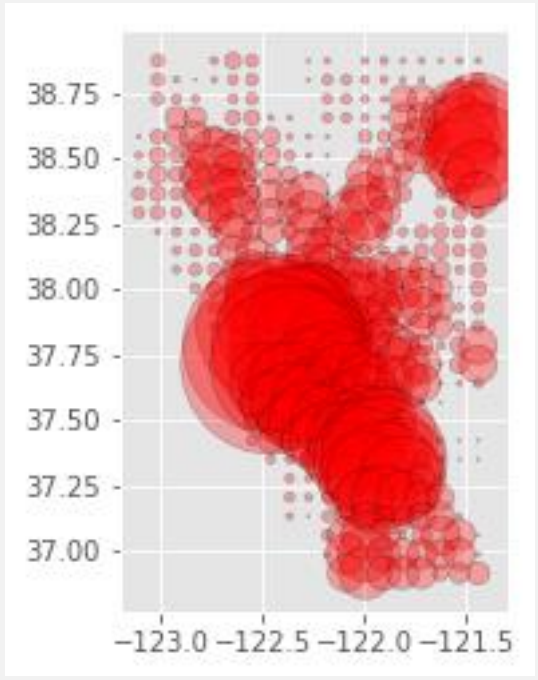
- Coordinates grid: 5 mi intervals
- Search radius: 5 mi
- 78k results
- 532 queries, 30 over limit
- More duplicates, fewer missing
- Minimal invalid datapoints

Geocoordinates with Hypotenuse Theory

- Coordinates grid: 3 mi intervals
- Search radius: 2.12 mi
- 54k results
- Full coverage and min overlap
- 1300+ queries, 2 over limit
- Even fewer duplicates and missing
- Minimal invalid datapoints

DATA COLLECTION



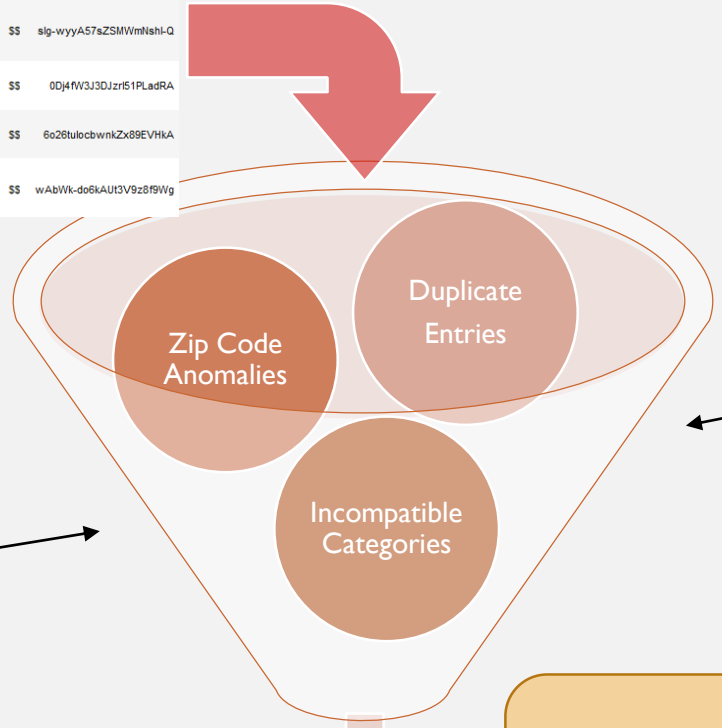


Heat maps from API calls on geocoordinates with radius overlap

DATA PREP

Query ID	Query Lat	Query Lng	Name	Category	Biz Address	Biz City	Biz Zip	Biz Lat	Biz Lng	Rating	Review Count	Price	Yelp ID
04-123.017222	38.842665	-123.017222	Railroad Station Bar and Grill	['American (New)', 'Southern', 'Pubs']	236 S Cloverdale Blvd	Cloverdale	95425	38.8030784358586	-123.01537800547199	4.0	223	$$	slg-wyyA57sZSMWmNshl-Q
04-123.017222	38.842665	-123.017222	Hamburger Ranch & Bar-B-Que	['Barbeque', 'Burgers']	31195 N Redwood Hwy	Cloverdale	95425	38.817817687988295	-123.023628234863	4.0	381	$$	0Dj4fW3J3DJzrl51PLadRA
04-123.017222	38.842665	-123.017222	Trading Post	['Bakeries', 'American (New)', 'Bars']	102 S Cloverdale Blvd	Cloverdale	95425	38.8051039	-123.0168659	4.0	130	$$	6o26tulocbwnkZx89EVHkA
04-123.017222	38.842665	-123.017222	Cloverdale Ale Company	['Pubs', 'Beer Bar', 'American (Traditional)']	131 E 1st St	Cloverdale	95425	38.8056374443952	-123.015944433181	4.0	22	$$	wAbWk-do6kAUt3V9z8f9Wg

Raw Data: 54k rows



94562	Napa
94567	Napa
94573	Napa
94574	Napa
94576	Napa
94581	Napa
94599	Napa
94102	SF
94103	SF
94104	SF
94105	SF
94107	SF
94108	SF
94109	SF

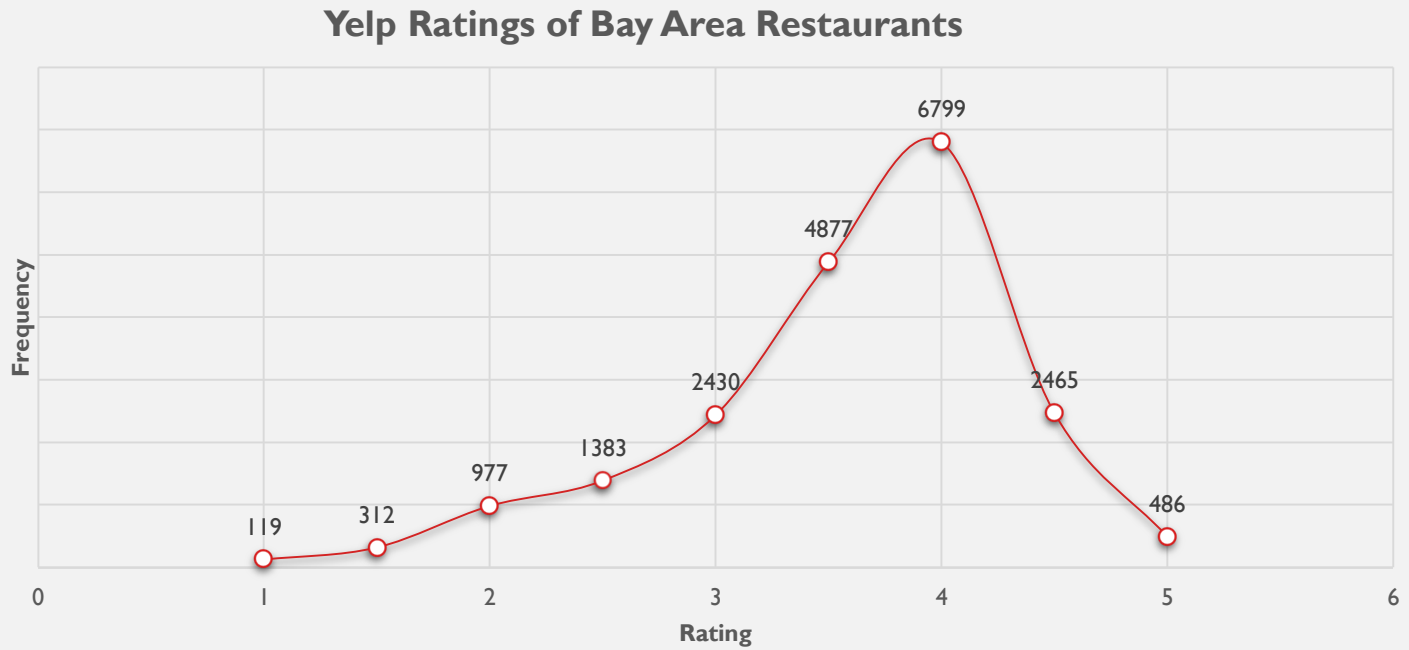
County Zips Data: 435 rows

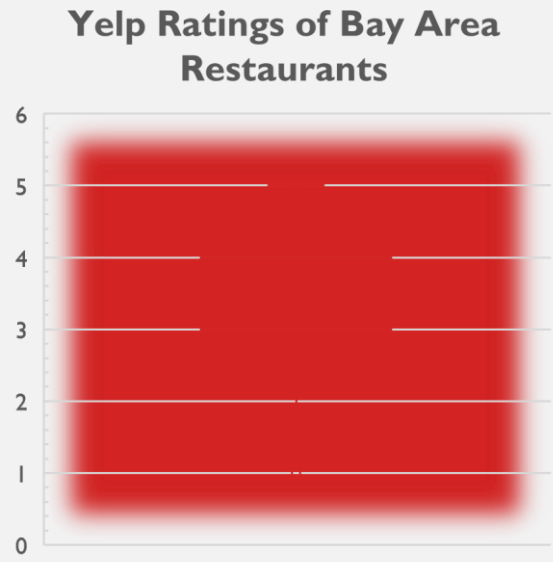
1	Couriers & Delivery Services
2	Community Service/Non-Profit
3	Bike Repair/Maintenance
4	Contractors
5	Convenience Stores
6	Department Stores
7	Cooking Schools
8	Caterers
9	Bowling
10	Bookstores

Rejected Categories: 50+ rows

Output:
20k rows
Unique, Bay-located, valid categories
Merge with county zips to add county col

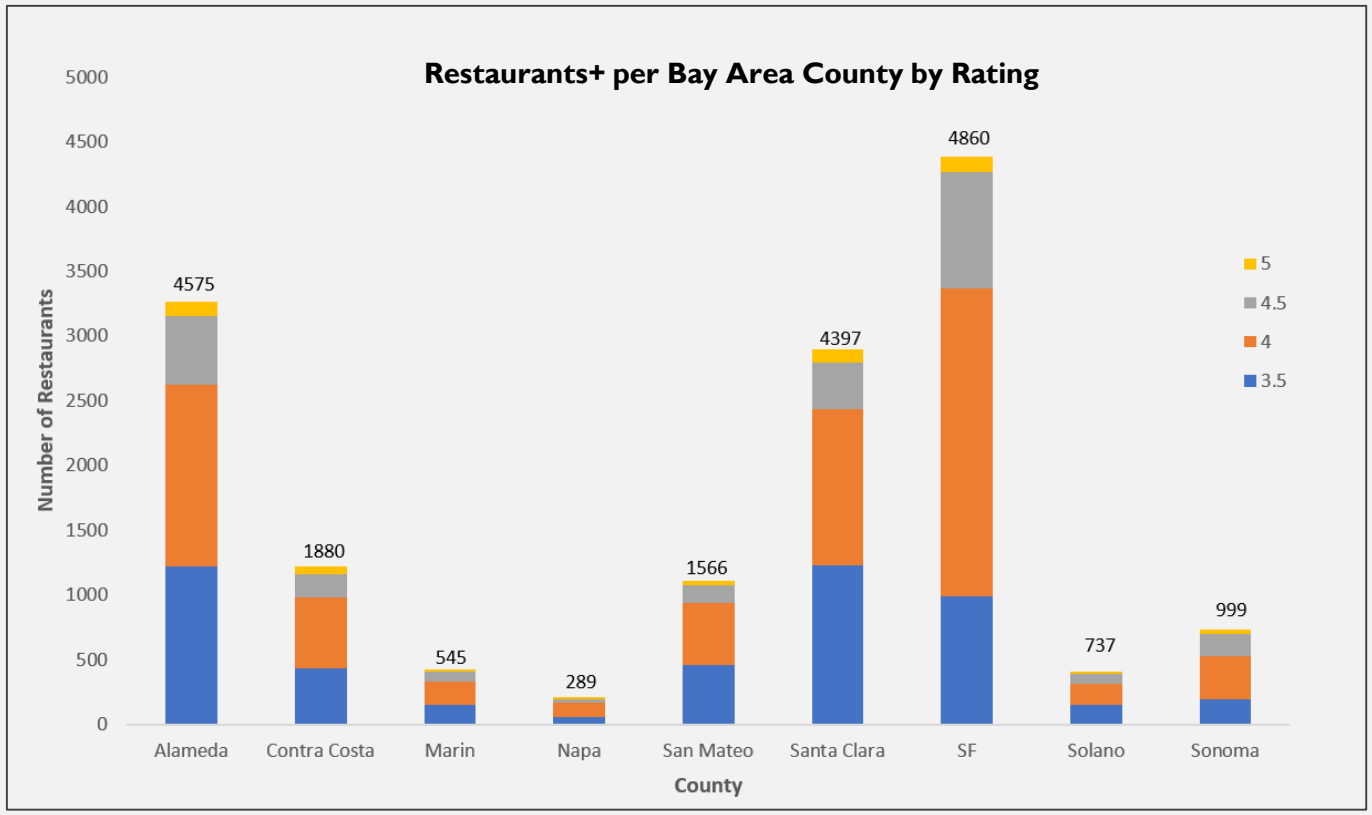
GRAPHS

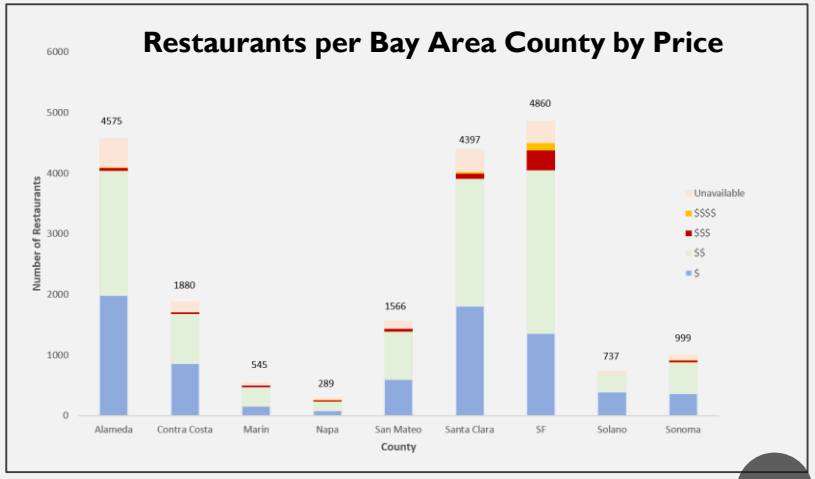

Yelp Ratings of Bay Area Restaurants
Frequency
Rating
0
1
2
3
4
5
6
119
312
977
1383
2430
4877
6799
2465
486


Yelp Ratings of Bay Area Restaurants
0
1
2
3
4
5
6

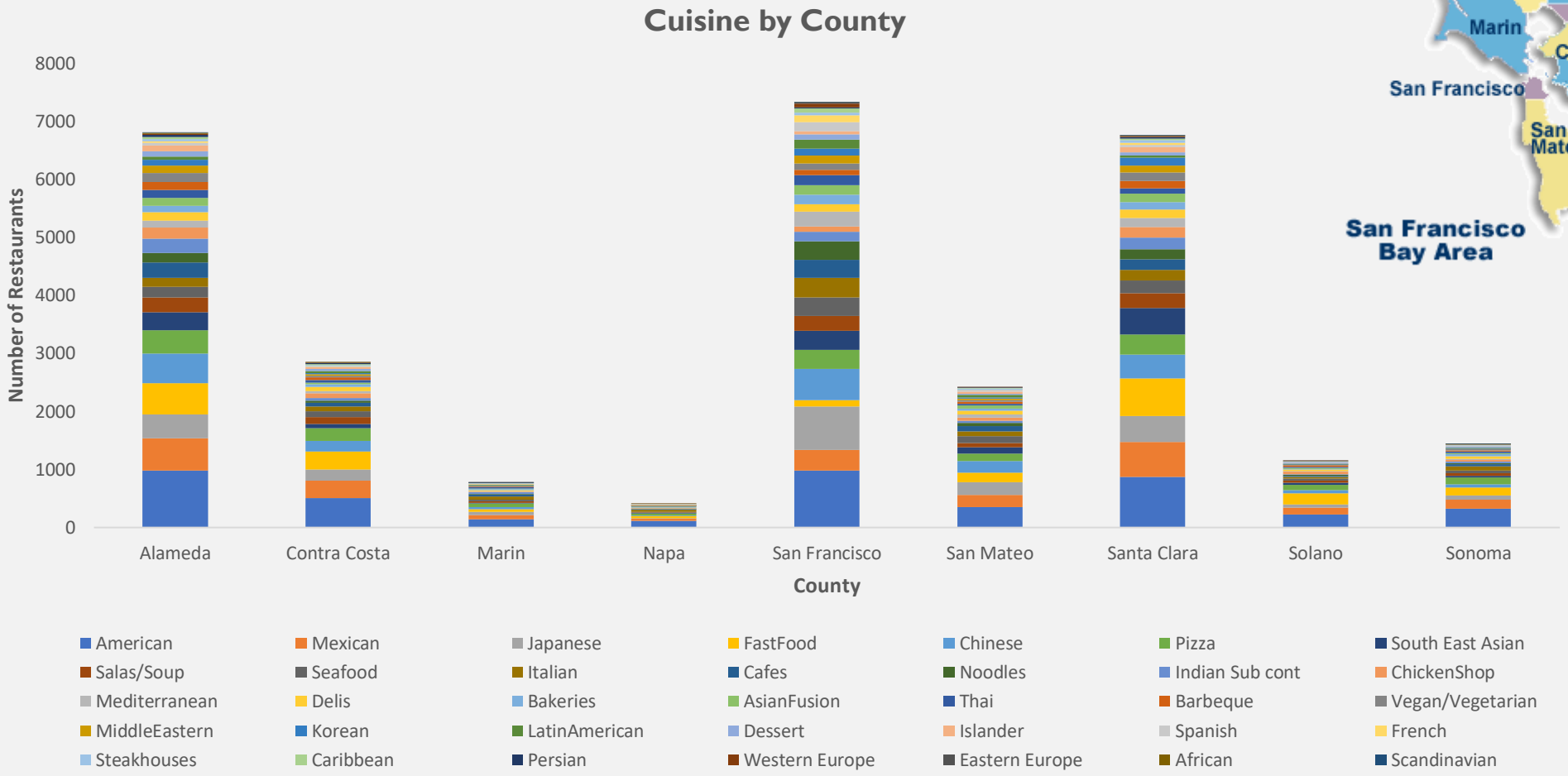
GRAPHS


Sonoma
Napa
Solano
Marin
Contra Costa
San Francisco
Alameda
San Mateo
Santa Clara
San Francisco Bay Area


Restaurants+ per Bay Area County by Rating
Number of Restaurants
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
4575
1880
545
289
1566
4397
4860
737
999
5
4.5
4
3.5
Alameda
Contra Costa
Marin
Napa
San Mateo
Santa Clara
SF
Solano
Sonoma
County


Restaurants per Bay Area County by Price
Number of Restaurants
6000
5000
4000
3000
2000
1000
0
4575
1880
545
289
1566
4397
4860
737
999
Unavailable
$$$$
$$$
$$
$
Alameda
Contra Costa
Marin
Napa
San Mateo
Santa Clara
SF
Solano
Sonoma
County

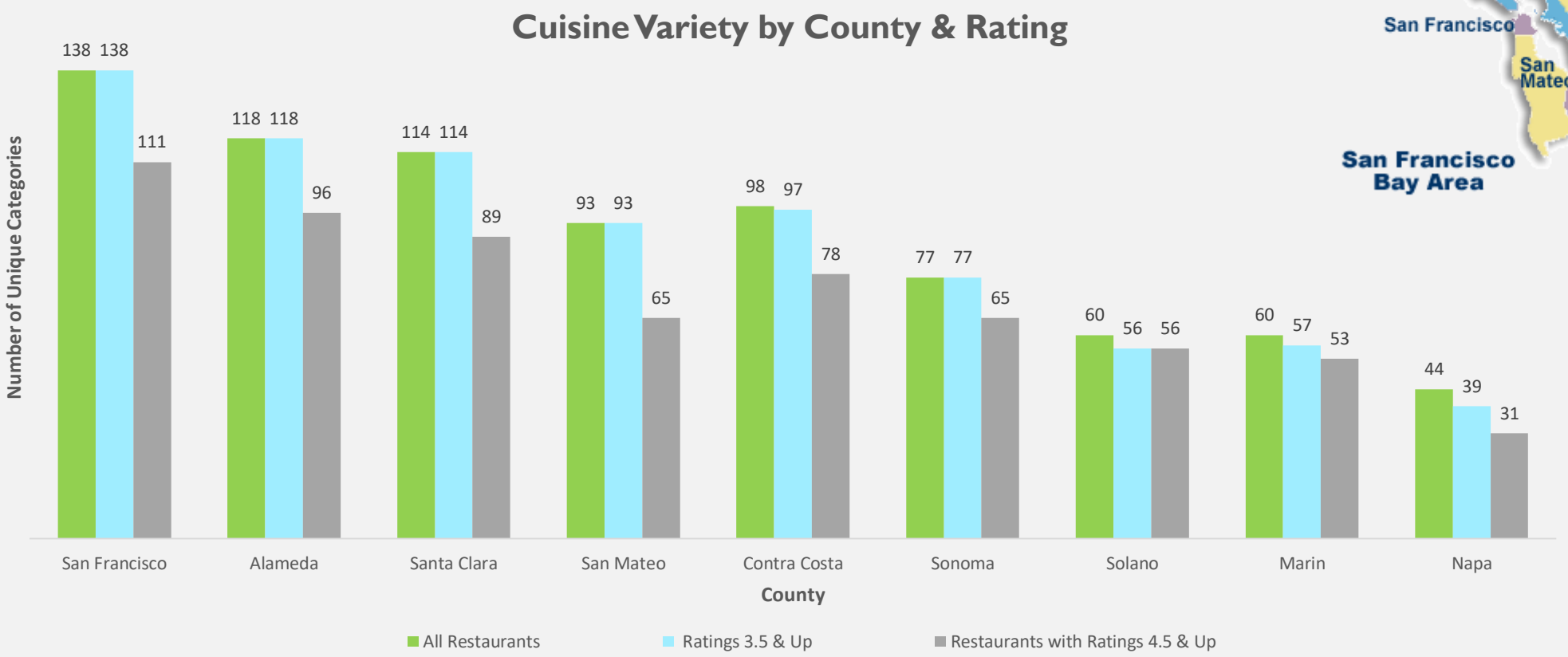
GRAPHS


Cuisine by County
Number of Restaurants
8000
7000
6000
5000
4000
3000
2000
1000
0
Alameda
Contra Costa
Marin
Napa
San Francisco
San Mateo
Santa Clara
Solano
Sonoma
County
American
Mexican
Japanese
FastFood
Chinese
Pizza
South East Asian
Salas/Soup
Seafood
Italian
Cafes
Noodles
Indian Sub cont
ChickenShop
Mediterranean
Delis
Bakeries
AsianFusion
Thai
Barbeque
Vegan/Vegetarian
MiddleEastern
Korean
LatinAmerican
Dessert
Islander
Spanish
French
Steakhouses
Caribbean
Persian
Western Europe
Eastern Europe
African
Scandinavian


Sonoma
Napa
Solano
Marin
Contra Costa
San Francisco
Alameda
San Mateo
Santa Clara
San Francisco Bay Area

GRAPHS


Sonoma
Napa
Solano
Marin
Contra Costa
San Francisco
Alameda
San Mateo
Santa Clara
San Francisco Bay Area


Cuisine Variety by County & Rating
Number of Unique Categories
138 138 111
118 118 96
114 114 89
93 93 65
98 97 78
77 77 65
60 56 56
60 57 53
44 39 31
San Francisco
Alameda
Santa Clara
San Mateo
Contra Costa
Sonoma
Solano
Marin
Napa
County
All Restaurants
Ratings 3.5 & Up
Restaurants with Ratings 4.5 & Up

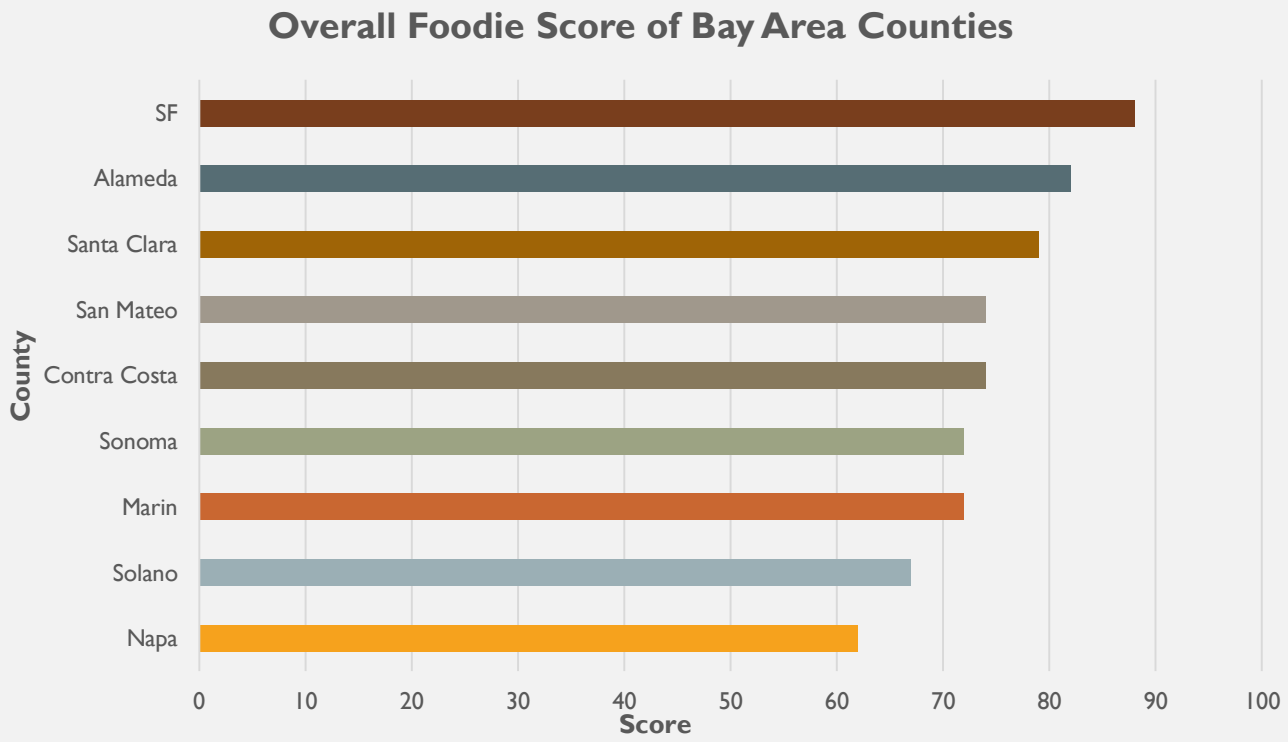
ANALYSIS

County	Avg Rating	Total Restaurants	Total Restaurants+	Unique Categories	Unique Categories+
Alameda	3.52	4,575	3,266	118	118
Contra Costa	3.44	1,880	1,216	98	97
Marin	3.63	545	421	60	57
Napa	3.55	289	206	44	39
San Mateo	3.50	1,566	1,106	93	93
Santa Clara	3.44	4,397	2,893	114	114
SF	3.89	4,860	4,384	138	138
Solano	3.26	737	406	60	56
Sonoma	3.61	999	729	77	77
TOTAL BAY AREA	**3.58 ± 0.75**	**19,848**	**14,627**	**154**	**149**
	PER COUNTY AVG ± SD	**2,205 ± 1,764.49**	**1,625 ± 1418.19**	**89 ± 29.47**	**87.67 ± 31.07**

RESULTS

County	Avg Rating Score	Total Restaurants Score	Total Restaurants+ Score	Unique Categories Score	Unique Categories+ Score	OVERALL GRADE
Alameda	70	90	90	80	80	82
Contra Costa	70	70	70	80	80	74
Marin	80	70	70	70	70	72
Napa	70	60	60	60	60	62
San Mateo	70	70	70	80	80	74
Santa Clara	70	90	80	80	80	79
SF	80	90	90	90	90	88
Solano	70	70	70	70	60	67
Sonoma	80	70	70	70	70	72

RESULTS


Overall Foodie Score of Bay Area Counties
SF
Alameda
Santa Clara
San Mateo
Contra Costa
Sonoma
Marin
Solano
Napa
County
0
10
20
30
40
50
60
70
80
90
100
Score

CONCLUSION

If the county has the most variety of restaurants with the highest ratings then a foodie will live/eat there.

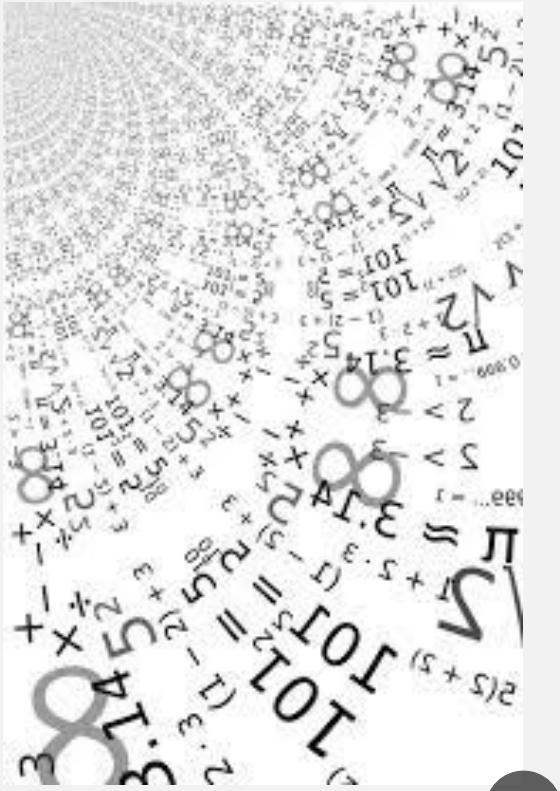
- Most popular food in the Bay Area was AMERICAN
Followed closely by Mexican & Chinese

-74% of Bay Area restaurants have a rating of 3.0 and up

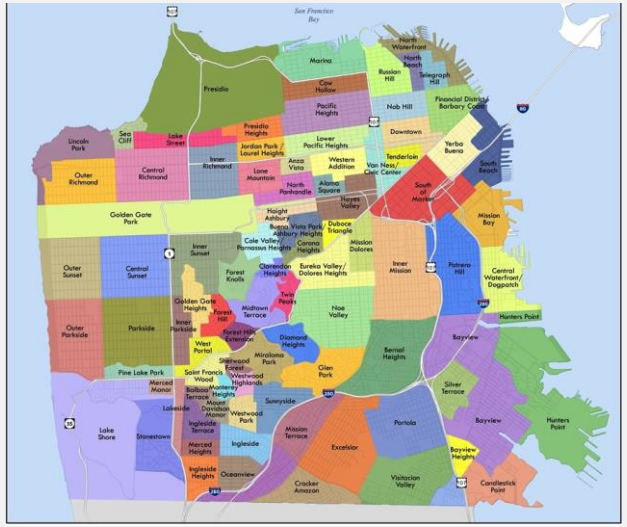
-SF had the best ratings in all grading categories, scoring highest overall, Followed by Alameda & Santa Clara

AREAS TO IMPROVE

Verify Data
Research Preference Factors
Tweak Formula



TIPS FOR REPEAT STUDIES



Compare on Local Level
Combine with Alternate Sources
Study Other Major Cities