

Yelp Trendsetters



Karissa Nanetta

Agenda

- 1. Why** try to identify trendsetters?
- 2. Who** is a trendsetter?
- 3. How** did I identify trendsetters?
- 4. What** do the results look like?
- 5. What next?**

Purpose

Identifying **the trendsetters**
among Yelp users

Why try to identify trendsetters?

- Trendsetters create valuable new content
- Trendsetters increase other users' engagement
- Gain insight on how Yelp can create more trendsetters in the future

Why try to identify trendsetters?

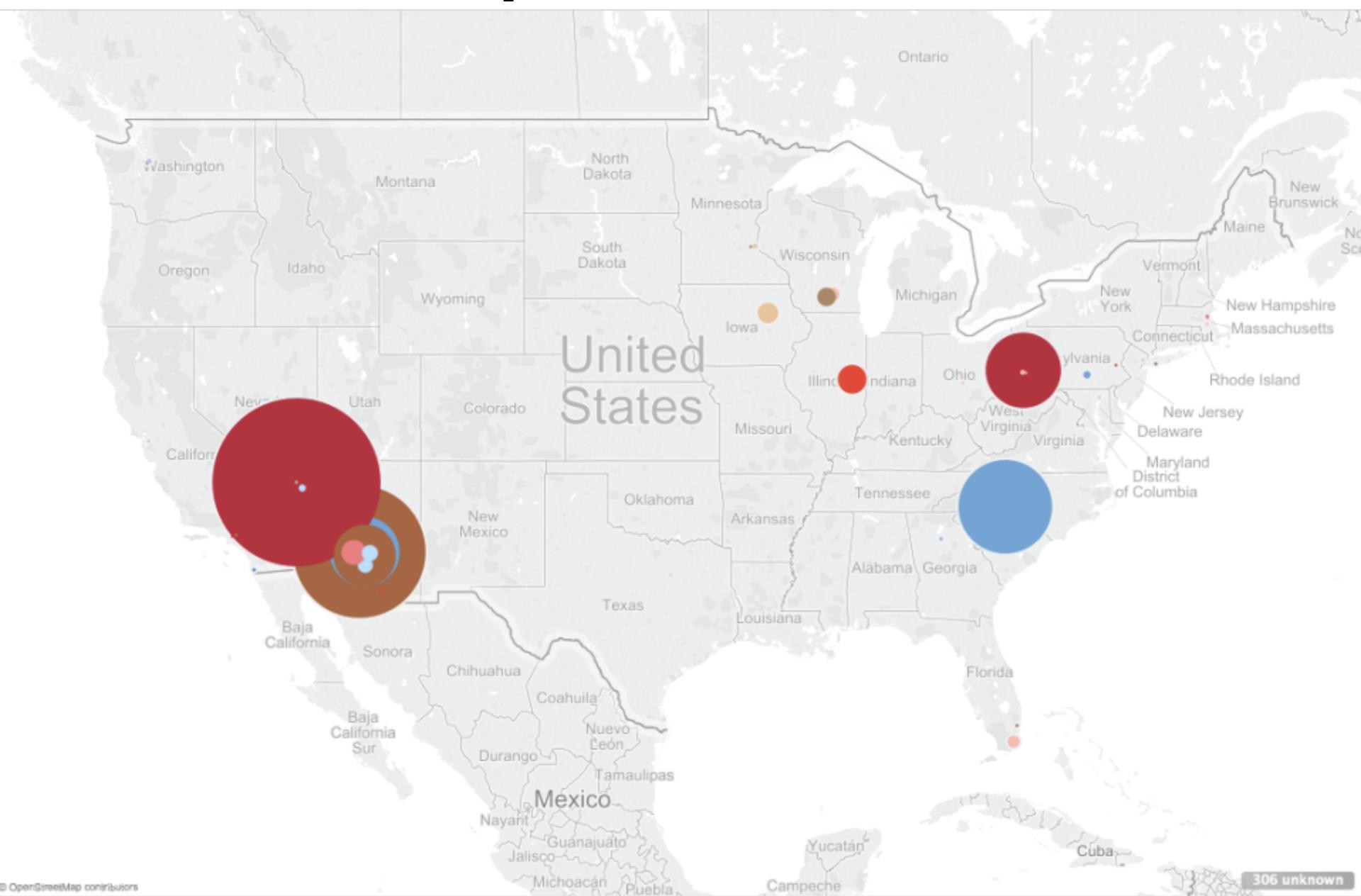
- Trendsetters create valuable new content
- Trendsetters increase other users' engagement
- Gain insight on how Yelp can create more trendsetters in the future

= more people will use Yelp

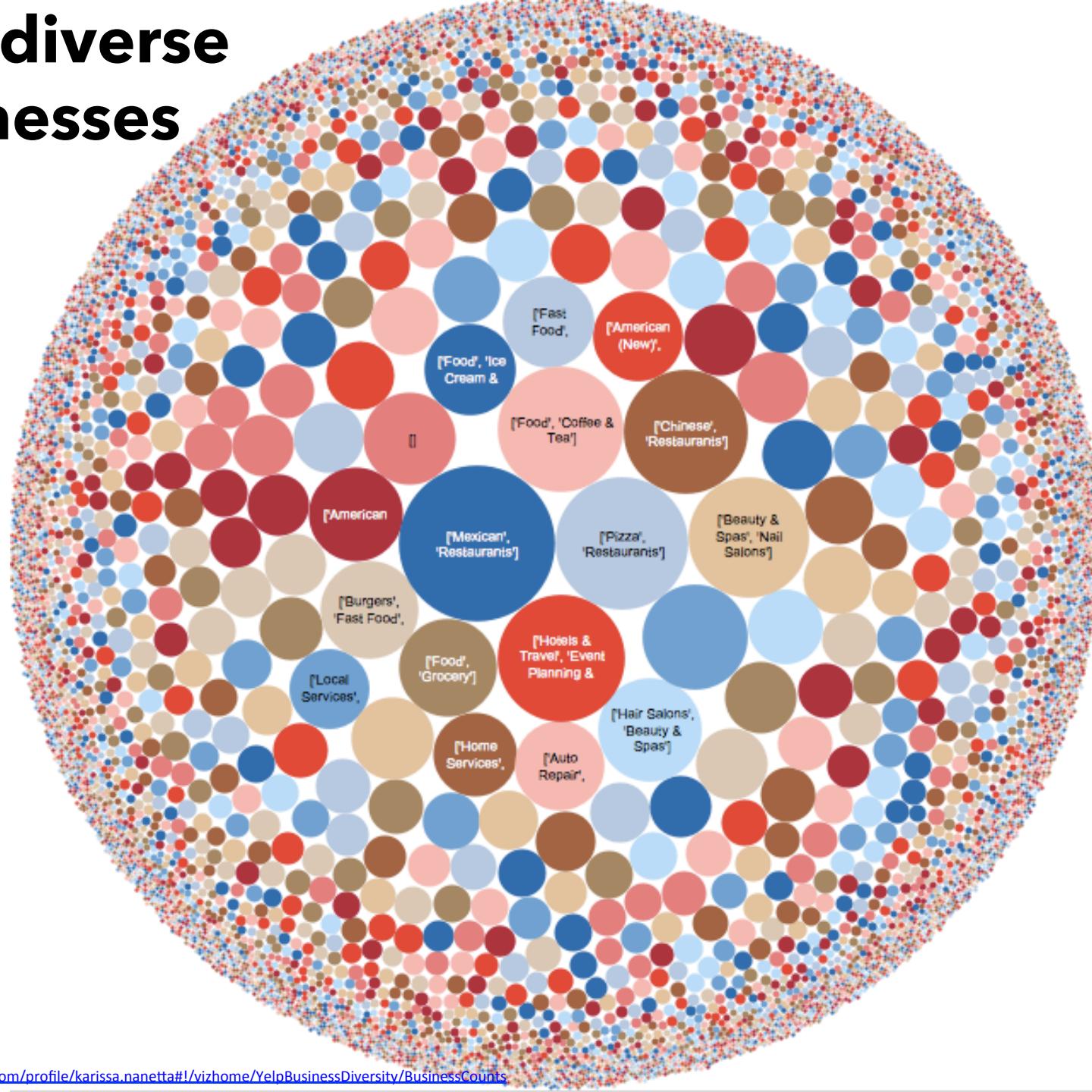
= ad revenue!

\$\$\$

What's in the Yelp data? Ten different cities



Very diverse businesses

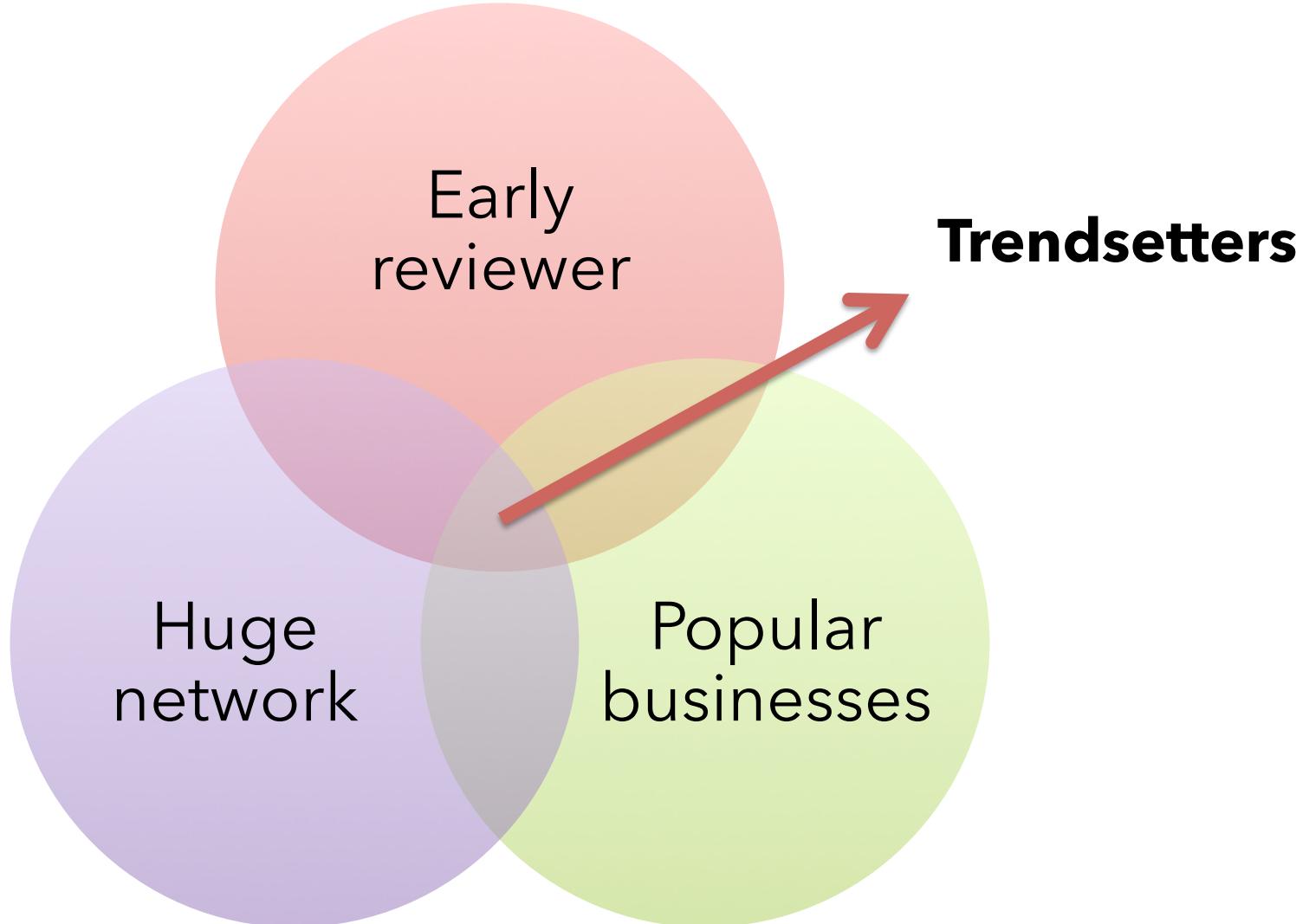


What's in the Yelp data?

5 separate datasets:

- Business (60K+)
- Reviews (1M+)
- User (366K+)
- Tip
- Check-in

Who is a trendsetter?

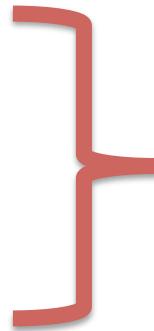


Who is a trendsetter?

- **Early Reviewer**
 - How soon does a Yelp user leave a review?
- **Popular Businesses**
 - How well are the businesses they review doing?
- **Huge Network**
 - How many people are listening to the trendsetters?

Who is a trendsetter?

- Early Reviewer
- Popular Businesses
- Huge Network



Feature
engineering

Who is a trendsetter?

- Early Reviewer
- Popular Businesses

}

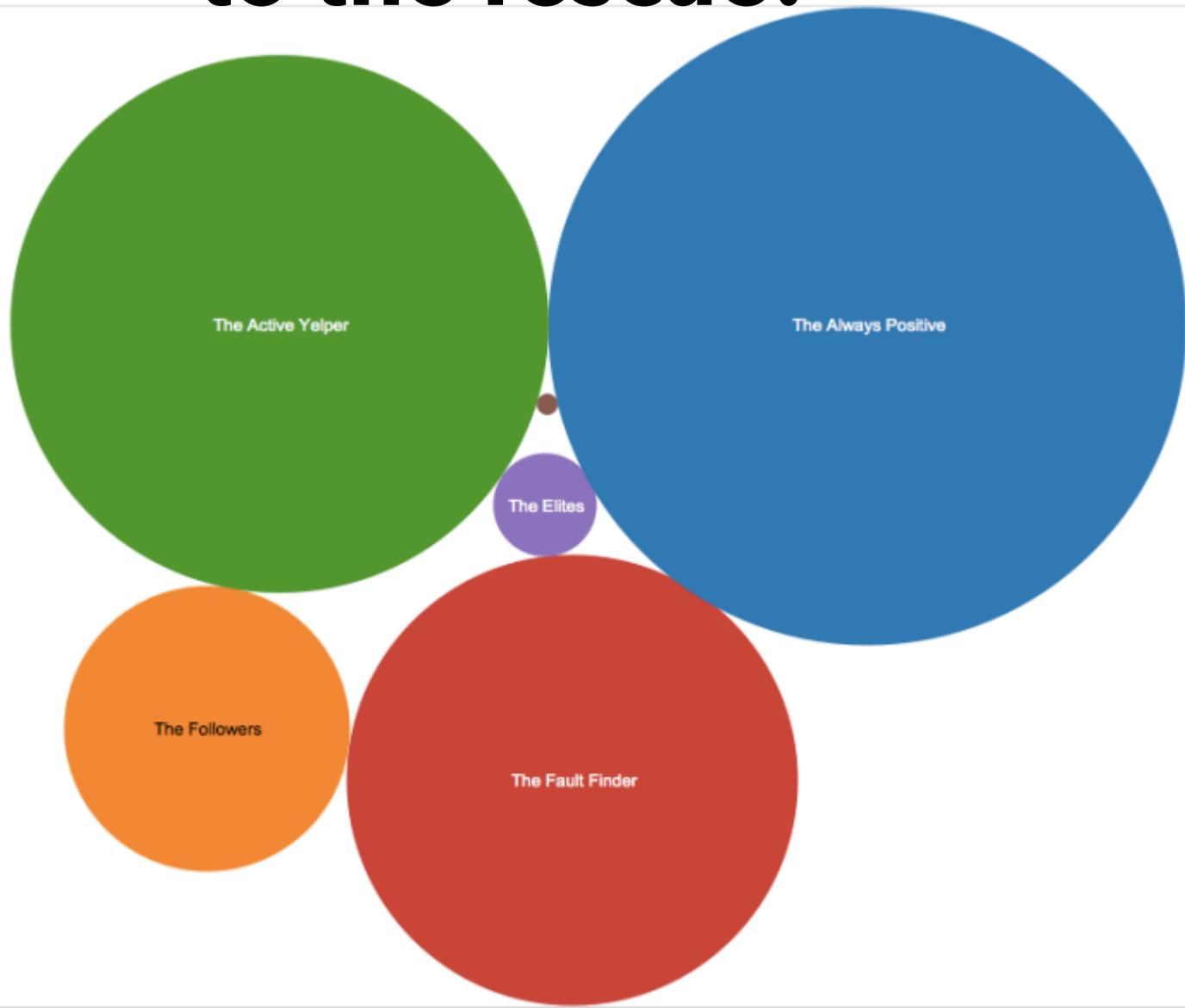
Feature
engineering

```
full2b = full2.copy()
i = 0
for index, row in full2b.iterrows():
    # row['business_id'] # '3ZVKmuK217uXPE6lXY4Dbg'
    # business_id_to_review_dates['3ZVKmuK217uXPE6lXY4Dbg']
    num_reviews_before_user_review = sum(1 for i in business_id_to_review_dates[row['business_id']] if i < (row['date_original'].value / 10**9))
    num_reviews_before_user_review = 0
    i += 1
    if i % 1000 == 0:
        print 'Processed %d rows' % i
    for date in business_id_to_review_dates[row['business_id']]:
        if date < (row['date_original'].value / 10**9):
            num_reviews_before_user_review += 1
        else:
            break
    full2b.loc[index, 'num_reviews_before_user_review'] = num_reviews_before_user_review
full2b.head()
```

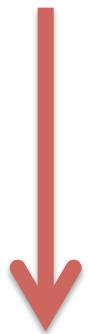
How did I identify trendsetters?

K-means Clustering

to the rescue!



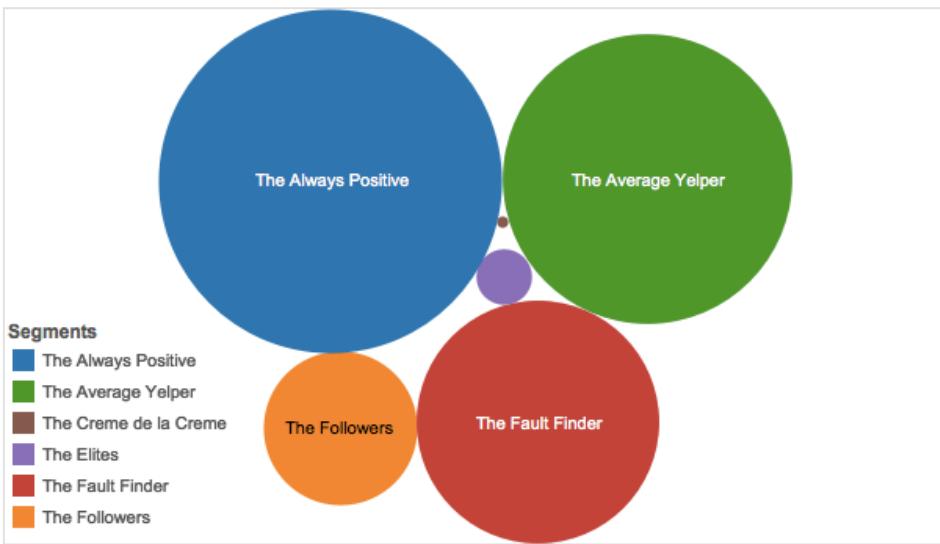
K-means Clustering



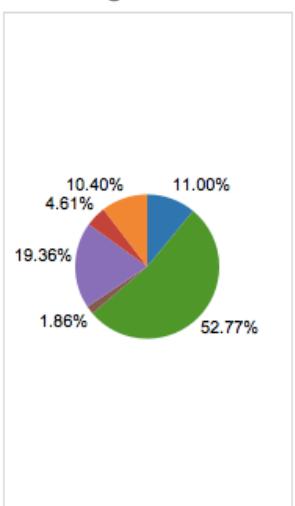
**Further analysis to determine
the subset of trendsetters**

K-means gave 6 distinct clusters

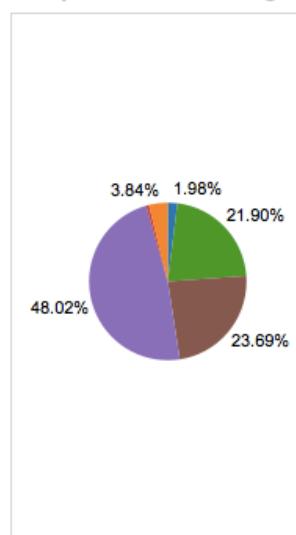
Yelp Segment Size



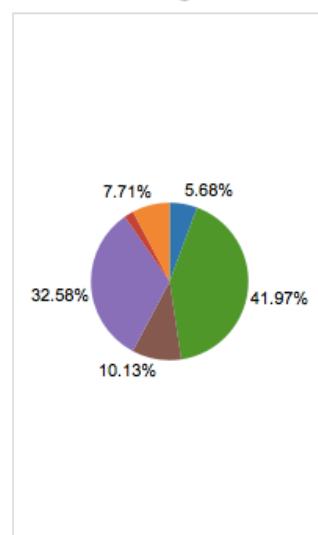
Review Count Percentage



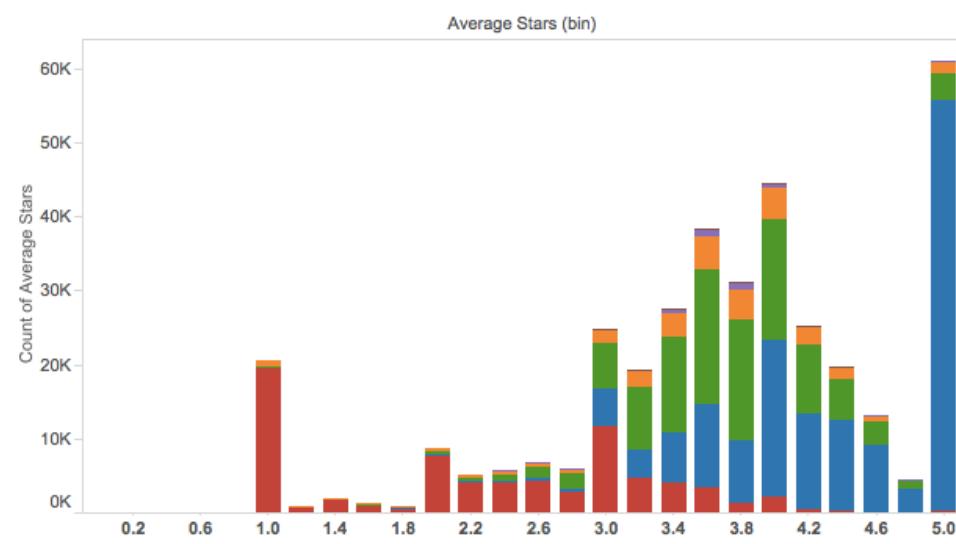
Compliment Percentage



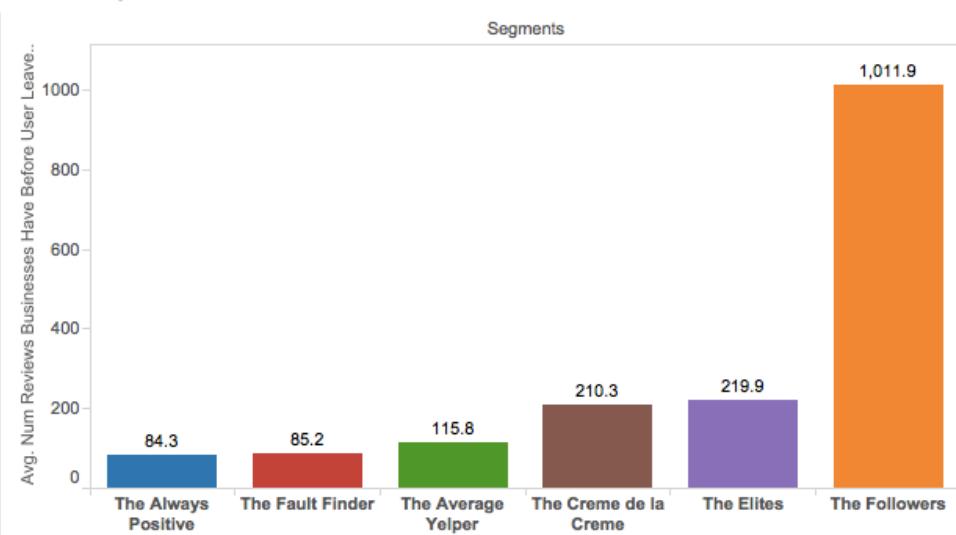
Fans Percentage



Yelp Rating Distribution



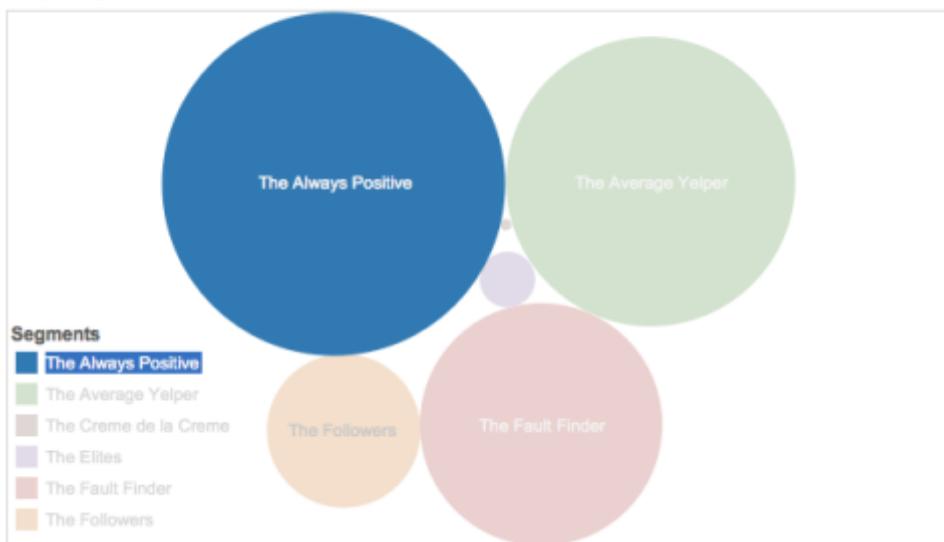
Review Speed



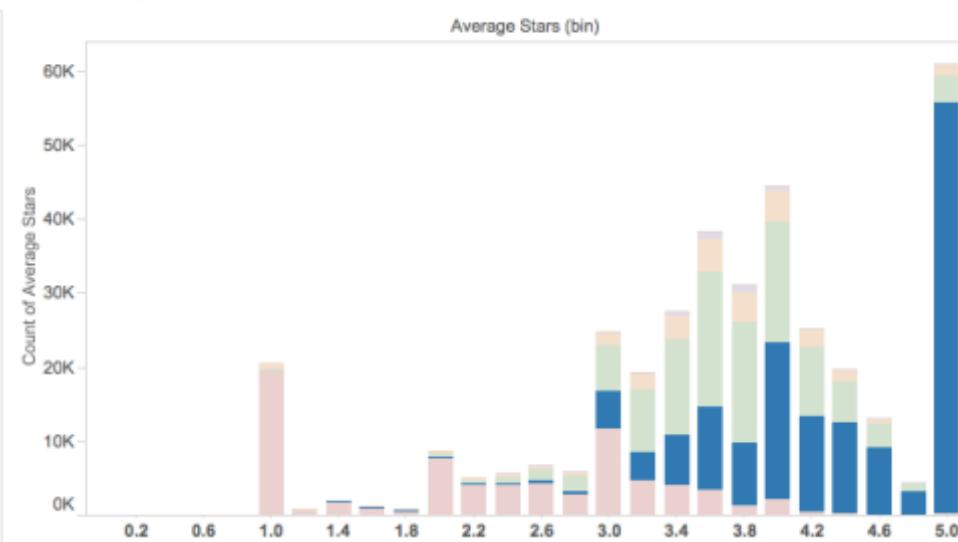
The “Always Positive” might be friends & family of the businesses... OR BOTS!

Reviews very early, always positive reviews, unpopular businesses

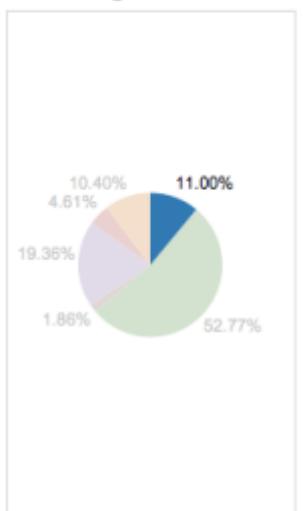
Yelp Segment Size



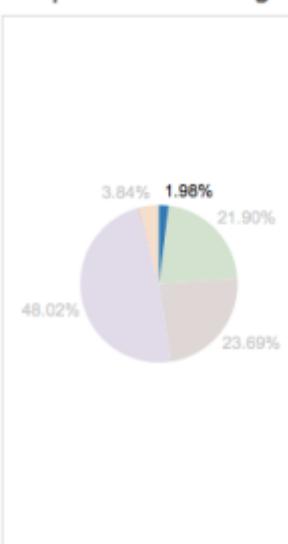
Yelp Rating Distribution



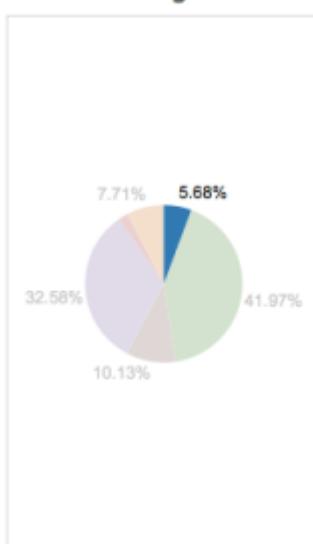
Review Count Percentage



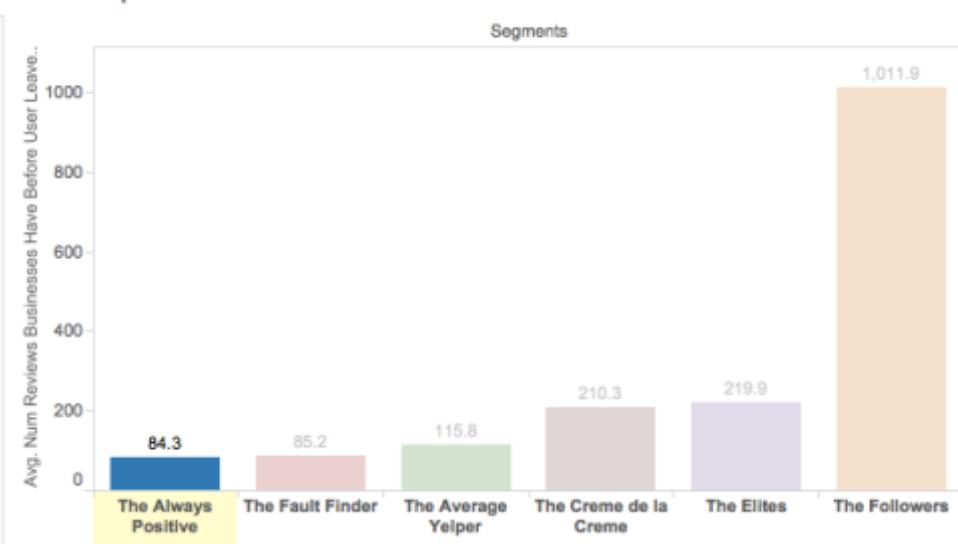
Compliment Percentage



Fans Percentage



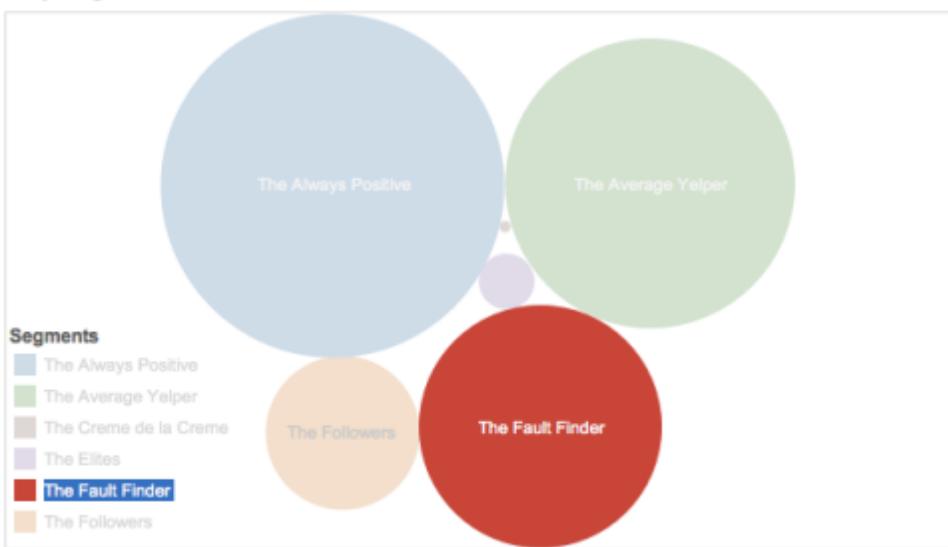
Review Speed



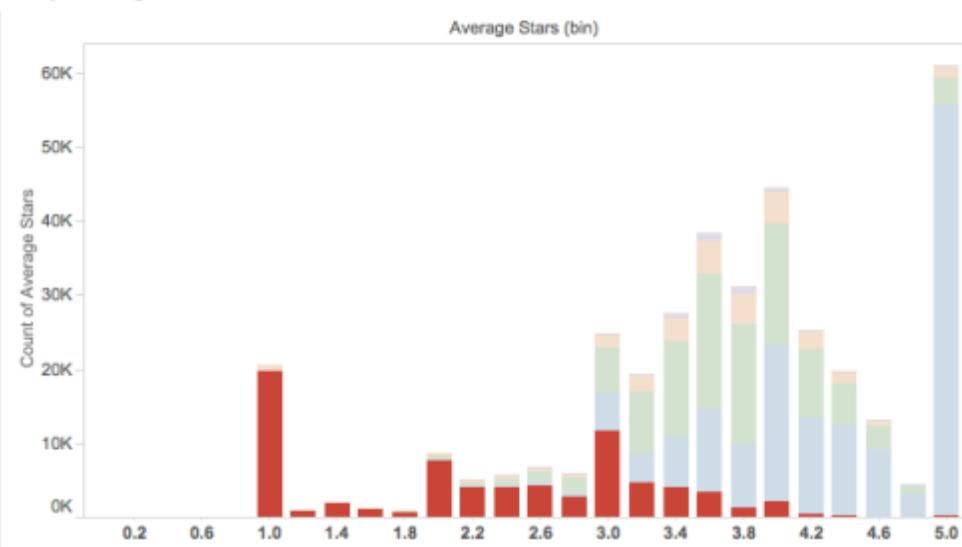
A large group in Yelp are “Fault-Finders”

Reviews very early, only gives negative reviews

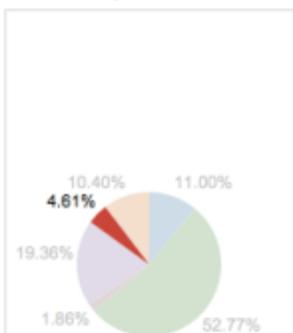
Yelp Segment Size



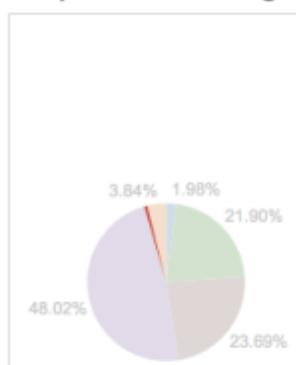
Yelp Rating Distribution



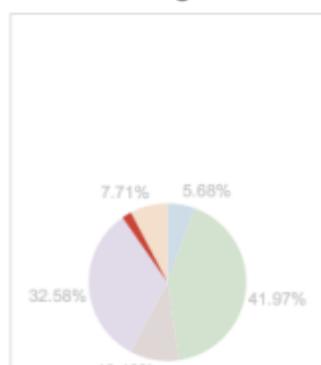
Review Count Percentage



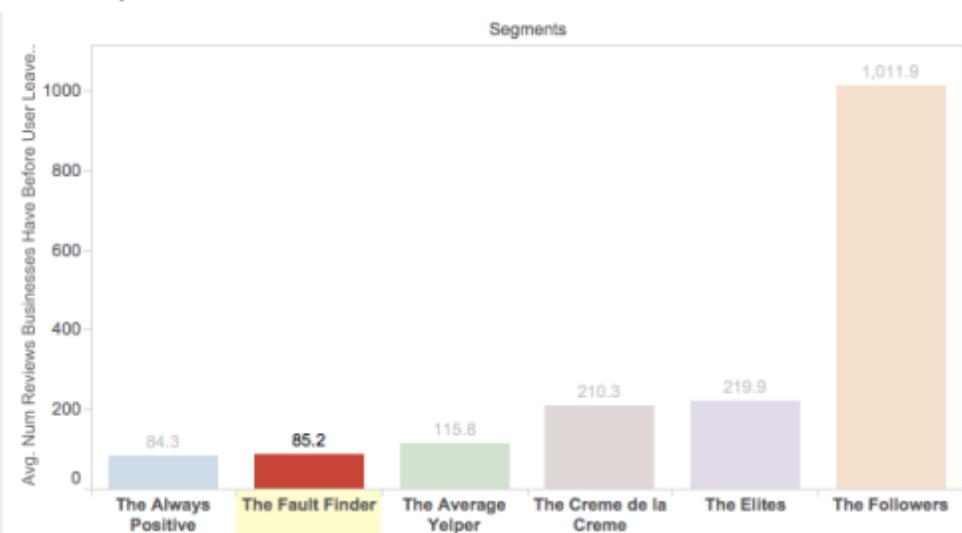
Compliment Percentage



Fans Percentage



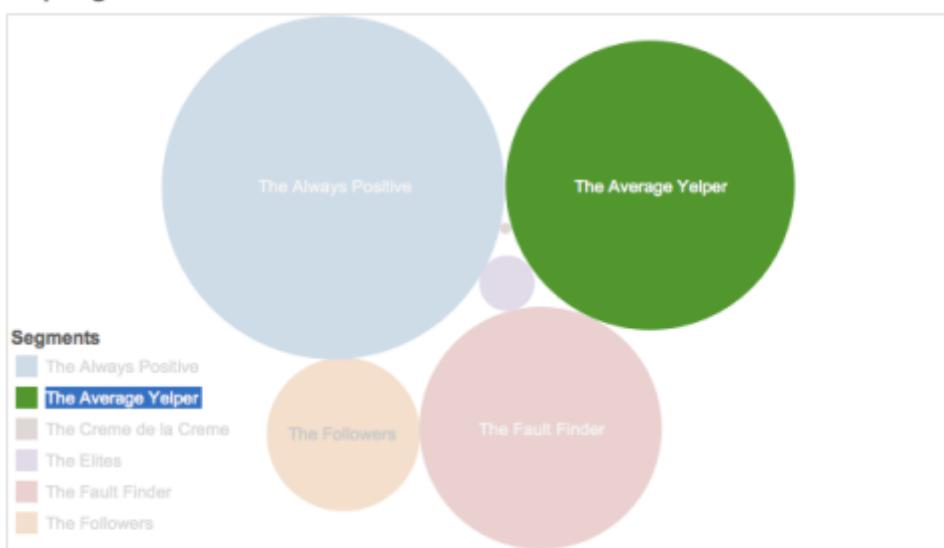
Review Speed



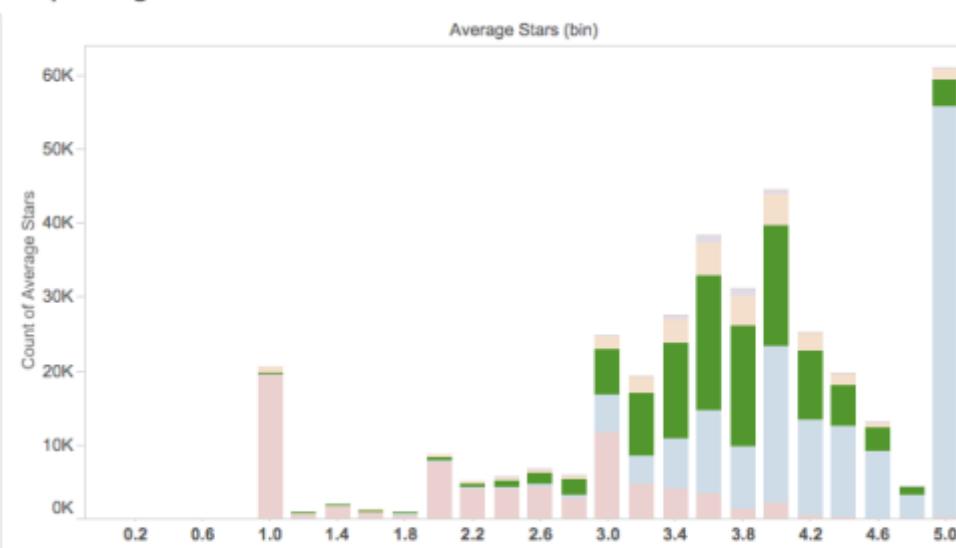
The "Average Yelper" are the closest to overall average population

Review ~50 businesses fairly, have several friends, average speed

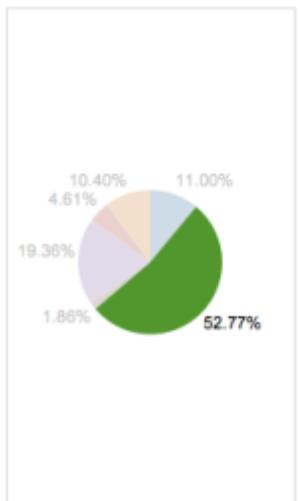
Yelp Segment Size



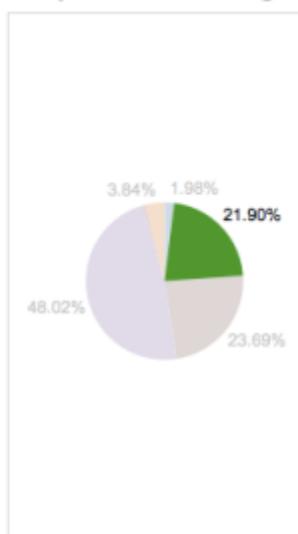
Yelp Rating Distribution



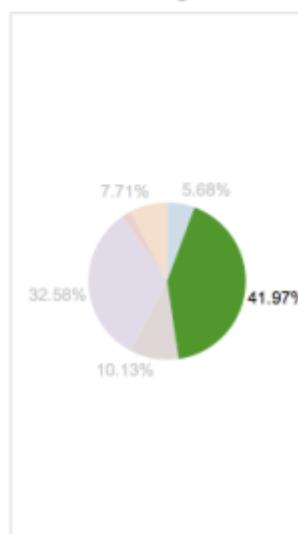
Review Count Percentage



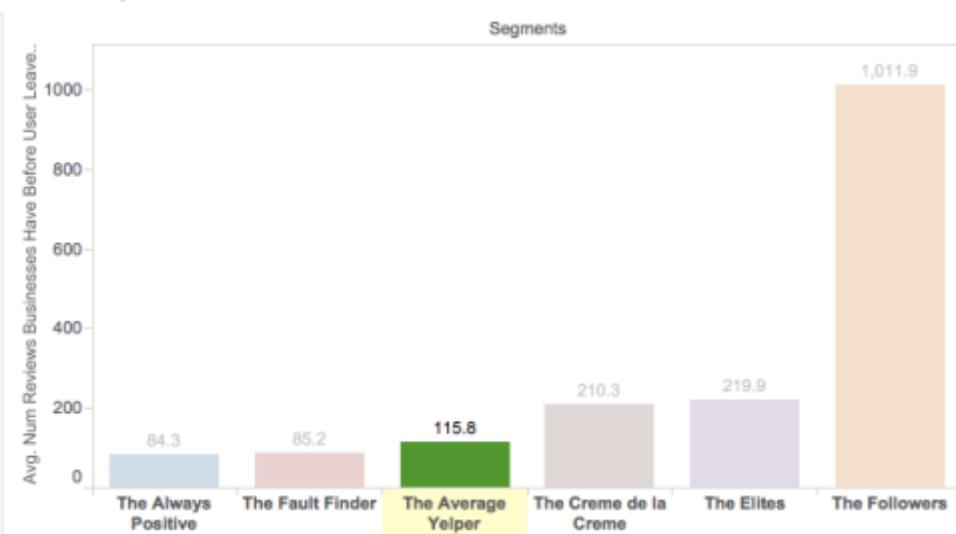
Compliment Percentage



Fans Percentage



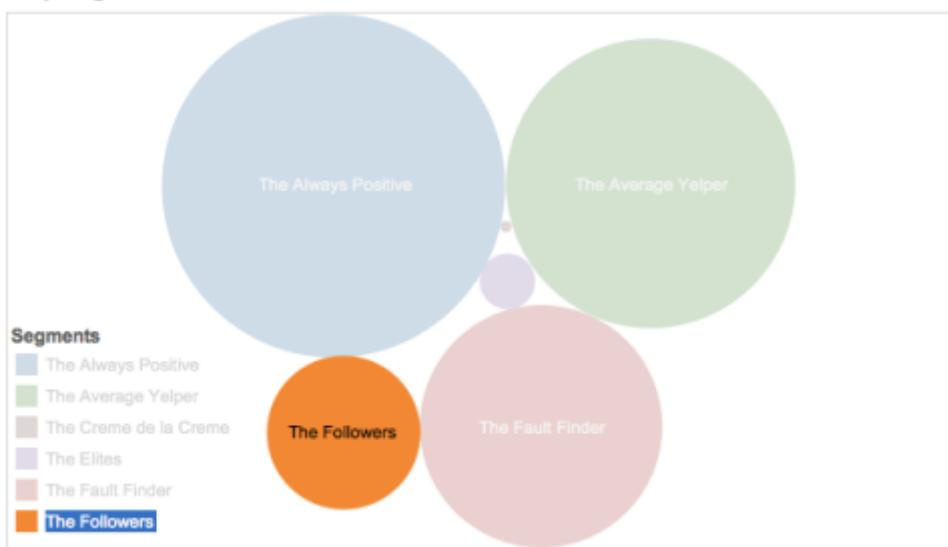
Review Speed



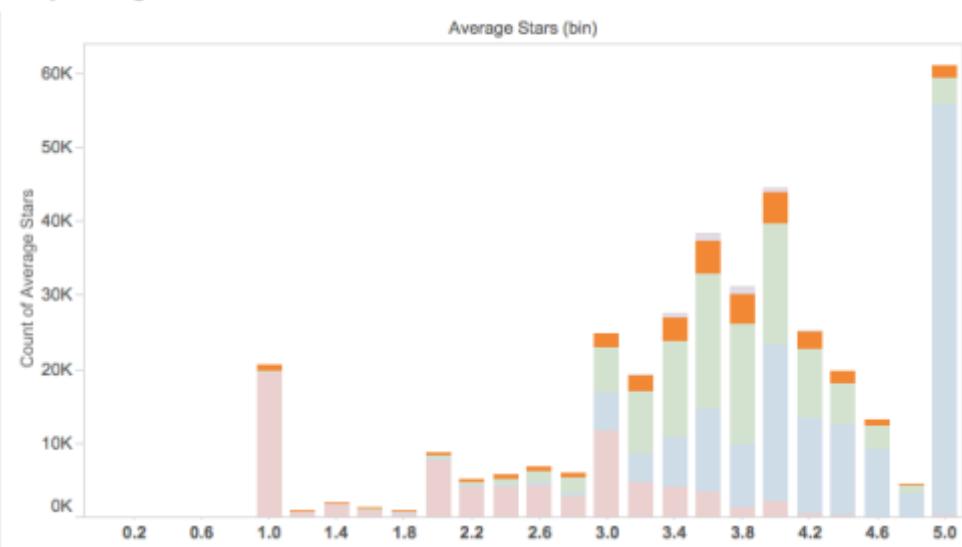
I fall under “The Followers”

Only visit businesses with 4+ stars and hundreds of reviews

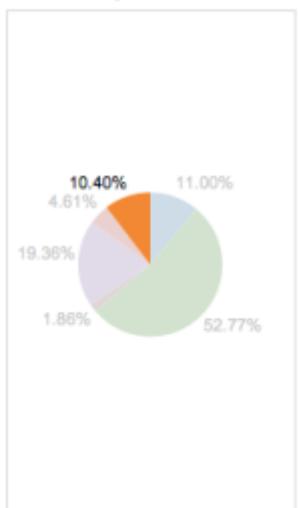
Yelp Segment Size



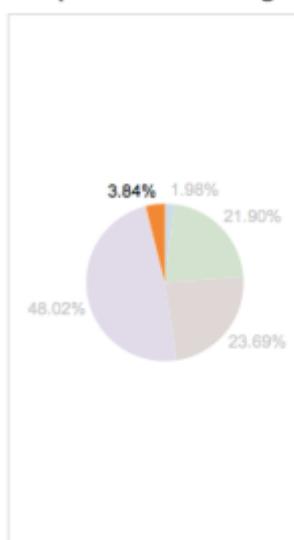
Yelp Rating Distribution



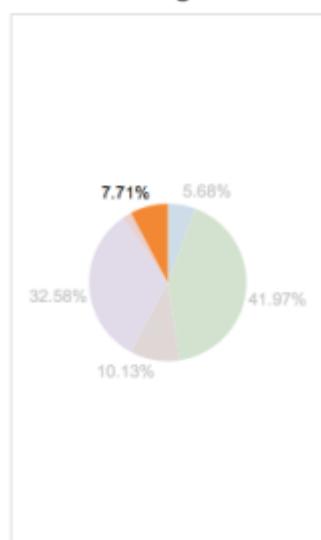
Review Count Percentage



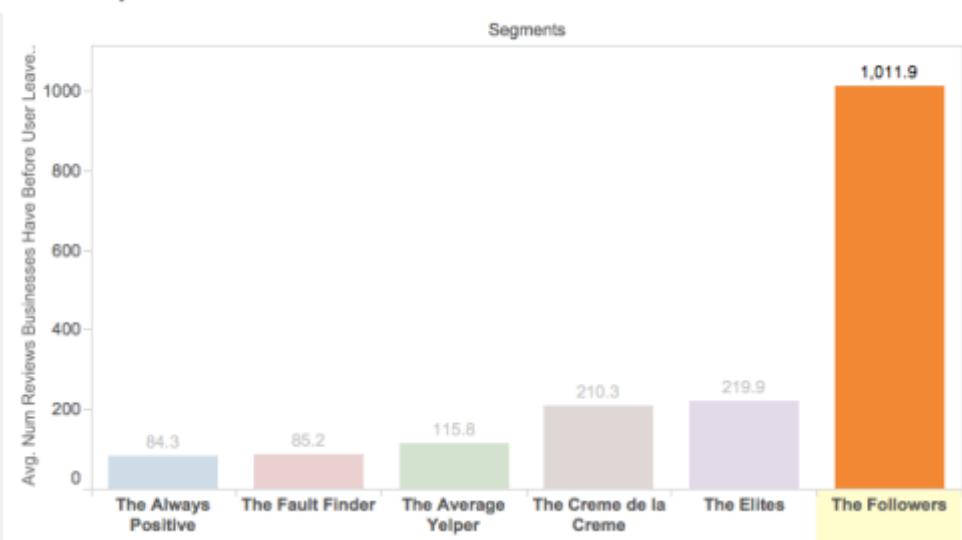
Compliment Percentage



Fans Percentage



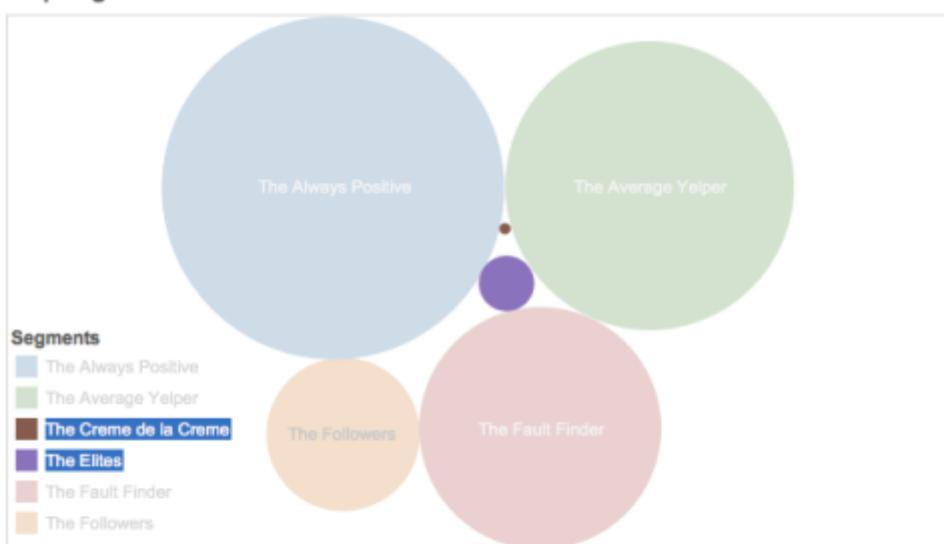
Review Speed



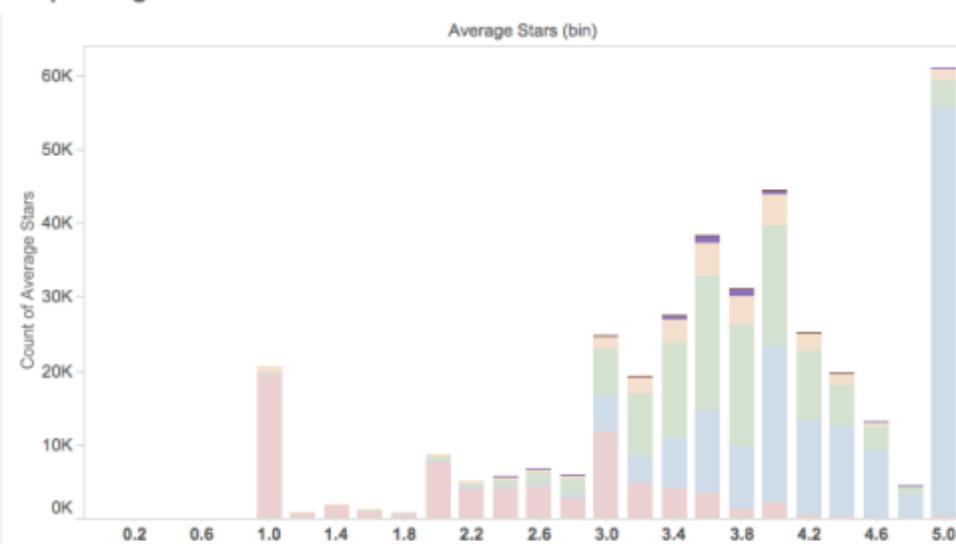
The “Crème de la Crème” & “Elites” are good candidates to be the trendsetters

Large following, popular businesses, reviews relatively early

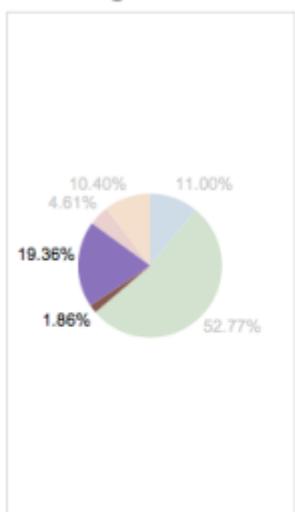
Yelp Segment Size



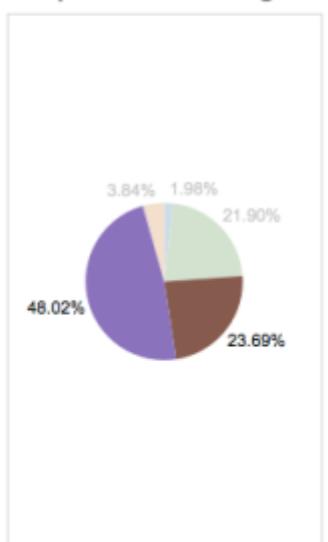
Yelp Rating Distribution



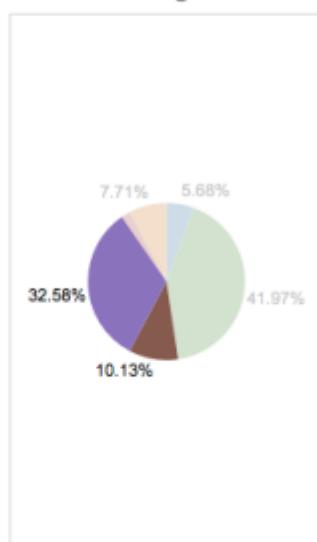
Review Count Percentage



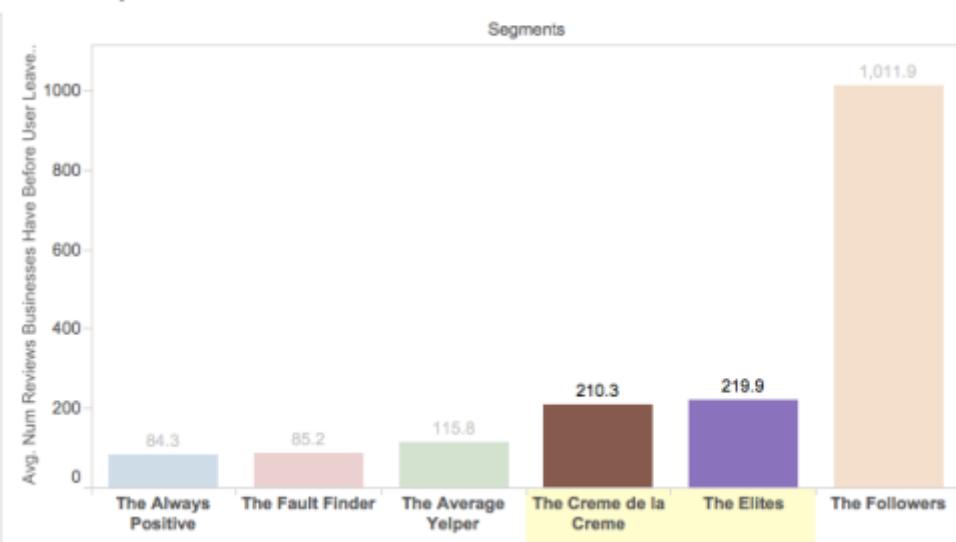
Compliment Percentage



Fans Percentage



Review Speed

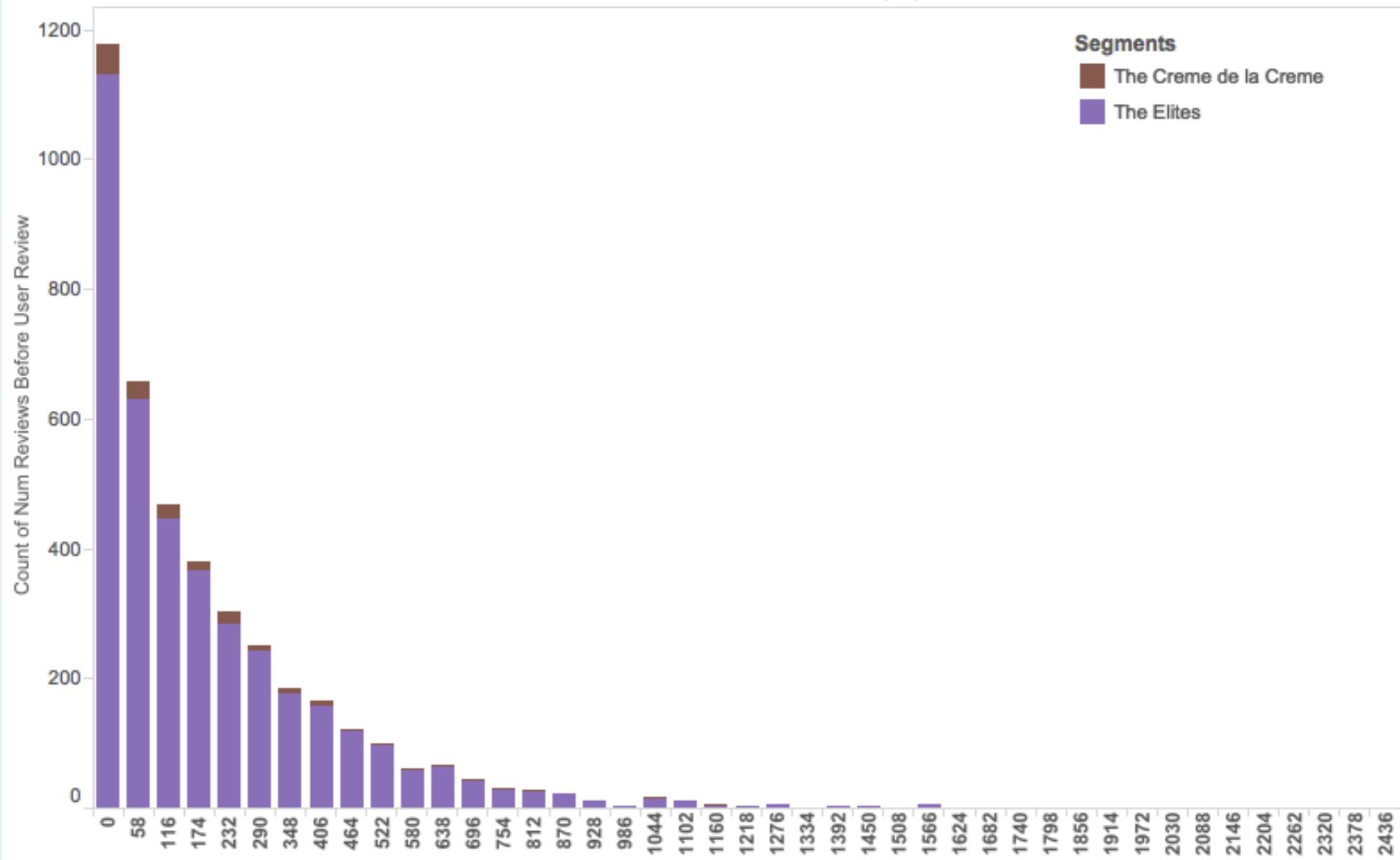


So, who are the trendsetters?

Trendsetters



Num Reviews Before User Review (bin)

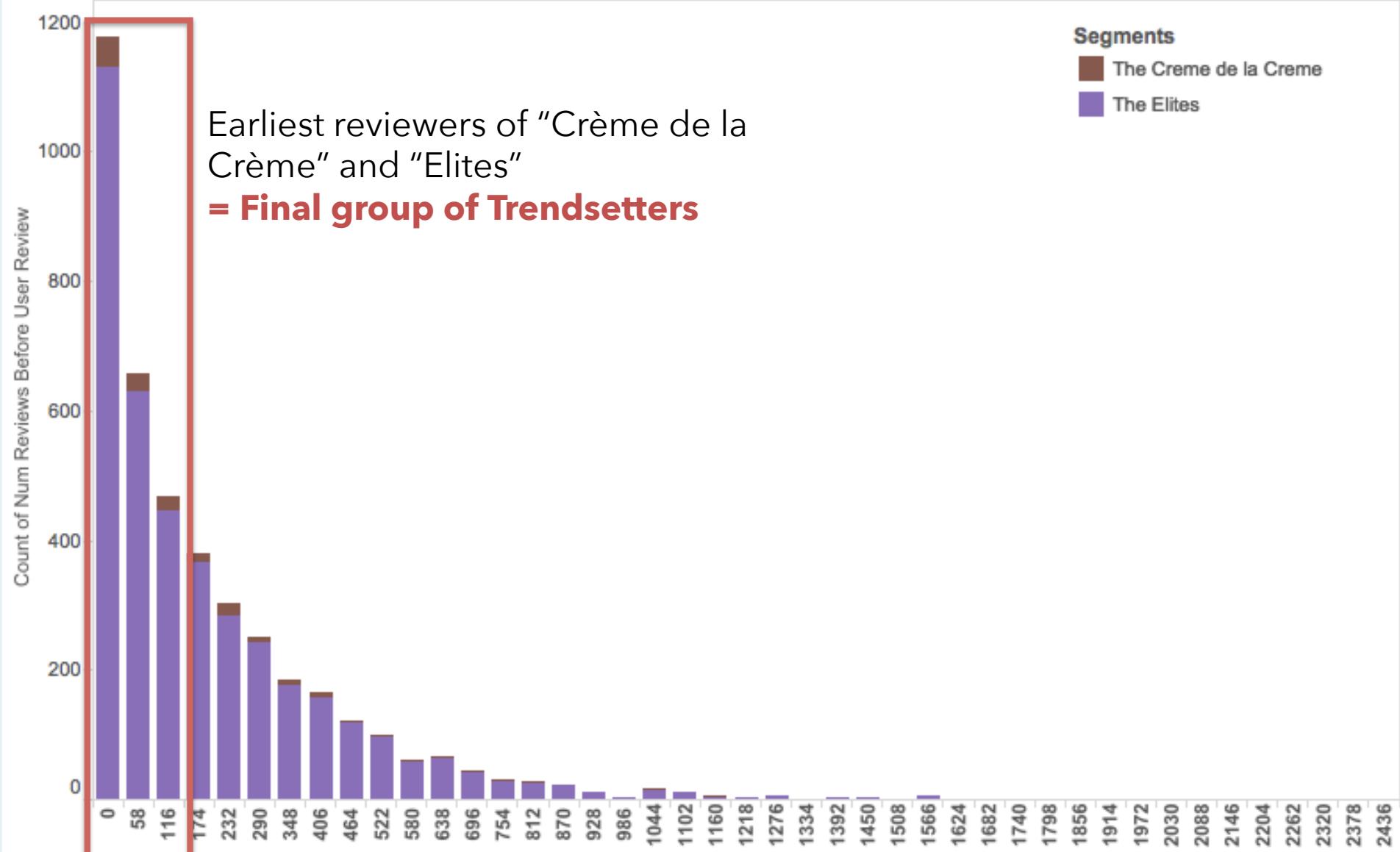


So, who are the trendsetters?

Trendsetters



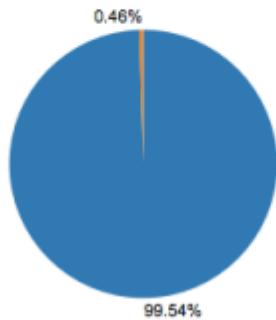
Num Reviews Before User Review (bin)



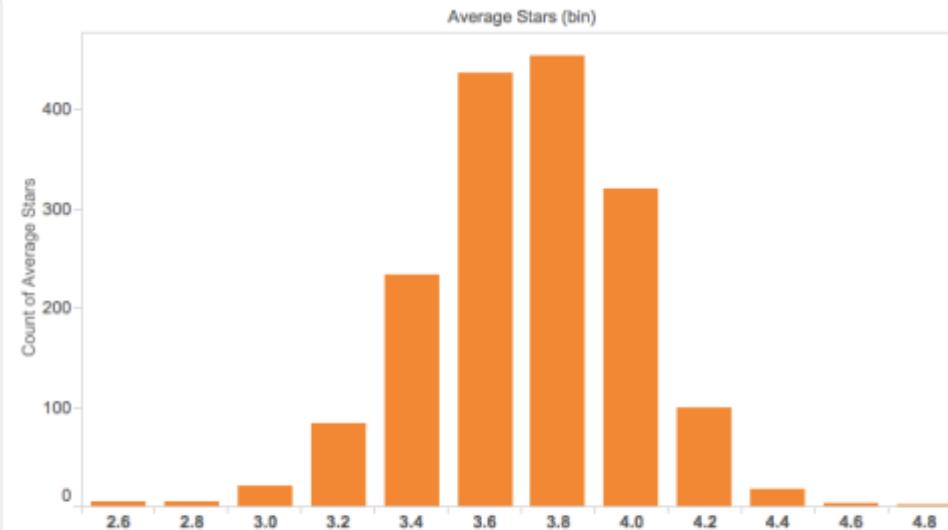
So, who are the trendsetters?

Size

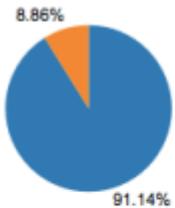
Trendsetter
Non-Trendsetters
Trendsetters



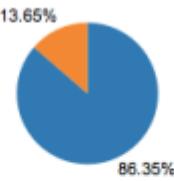
Rating Distribution



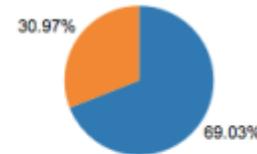
Review Count



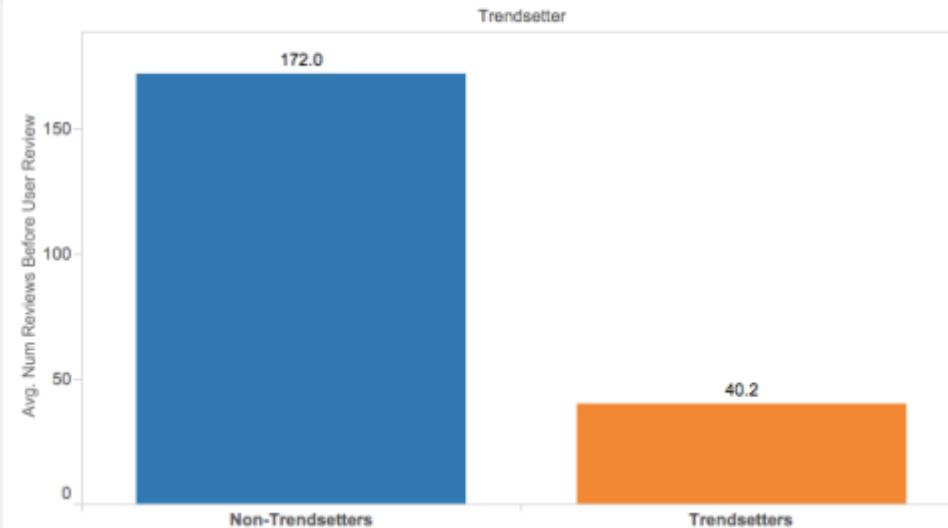
Friends



Compliment



Review_Speed



What can Yelp do next?

- Give a badge to the trendsetters, so other users know who they are and follow them
- Give trendsetters gift cards to try out new businesses as a form of marketing



What else can we do?

- Predict whether a new user will become a trendsetter (supervised learning)
- Analyze effects on types of businesses (restaurants, laundromats, doctors, etc.)
- Given more data...
 - Perform longitudinal analysis
 - Analyze user profile texts to detect bots

