



Analytics Insights

Luba Gloukhova for Slalom

A decorative network diagram at the top of the slide, featuring a series of interconnected nodes and lines. The nodes are represented by small circles, some of which are highlighted with a dashed border. The lines are thin and gray, creating a web-like structure. A central node is highlighted with a larger dashed circle and contains the blue double quote symbol.

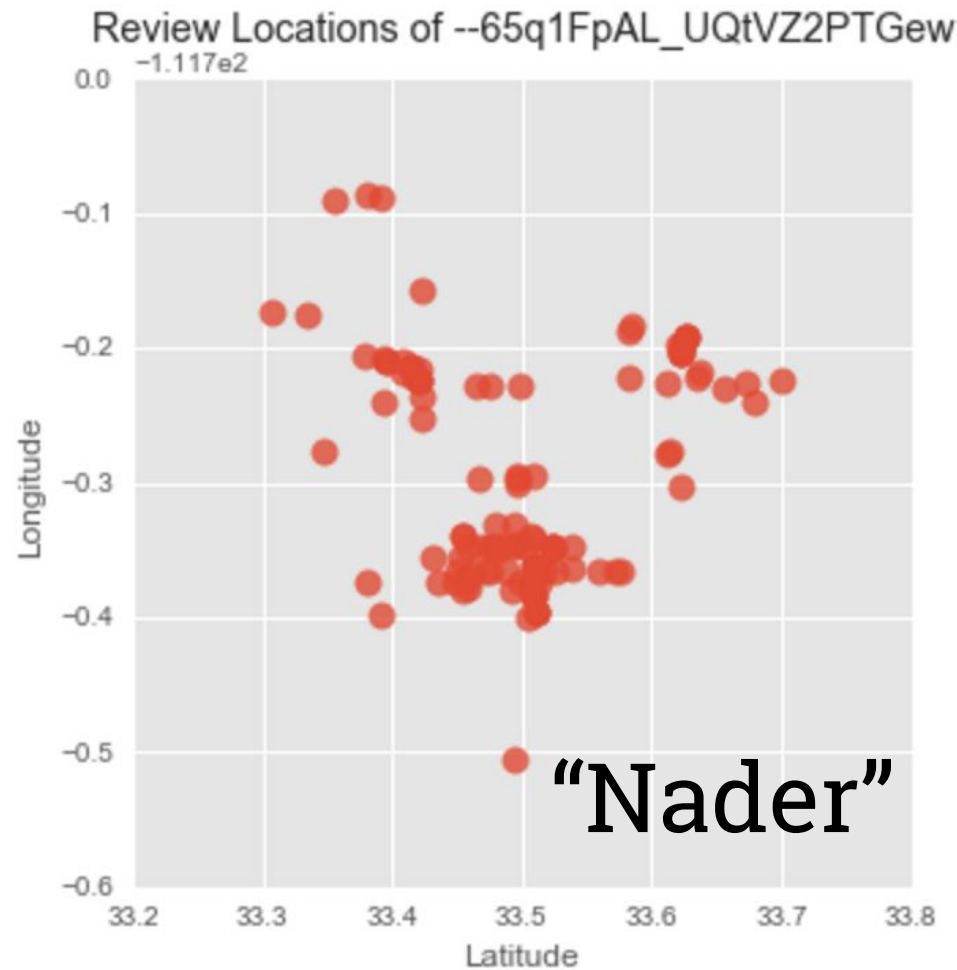
“

*Do users **review** businesses that
are **more expensive** when they are
away from home?*

Identifying a User's "Home"

1. Subset to Users with > 100 reviews
2. Cluster a user's reviews based on Lat/Long
3. Assign "home" to cluster with most reviews
4. All other Reviews = "away"

Identifying a User's "Home"



Identifying a User's "Home"

Review Locations of --65q1FpAL_UQtVZ2PTGew

Longitude

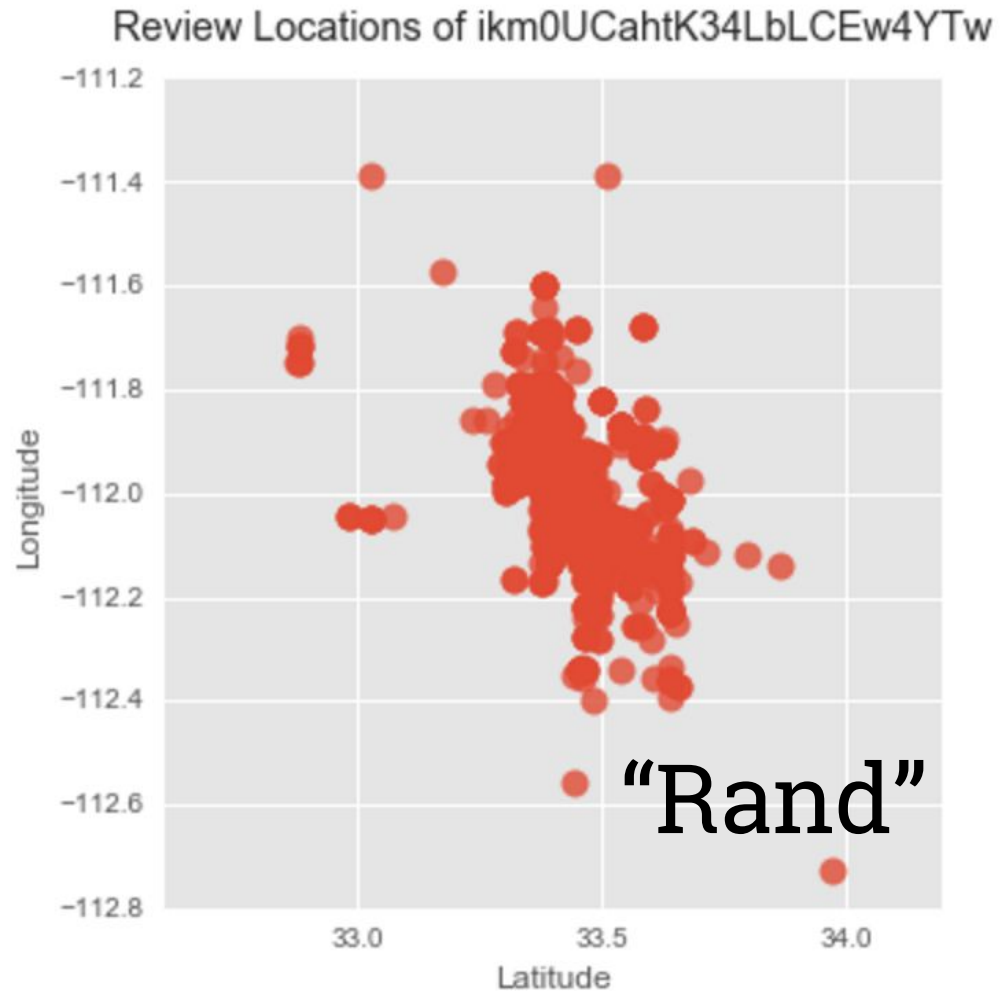
Latitude

Homebase

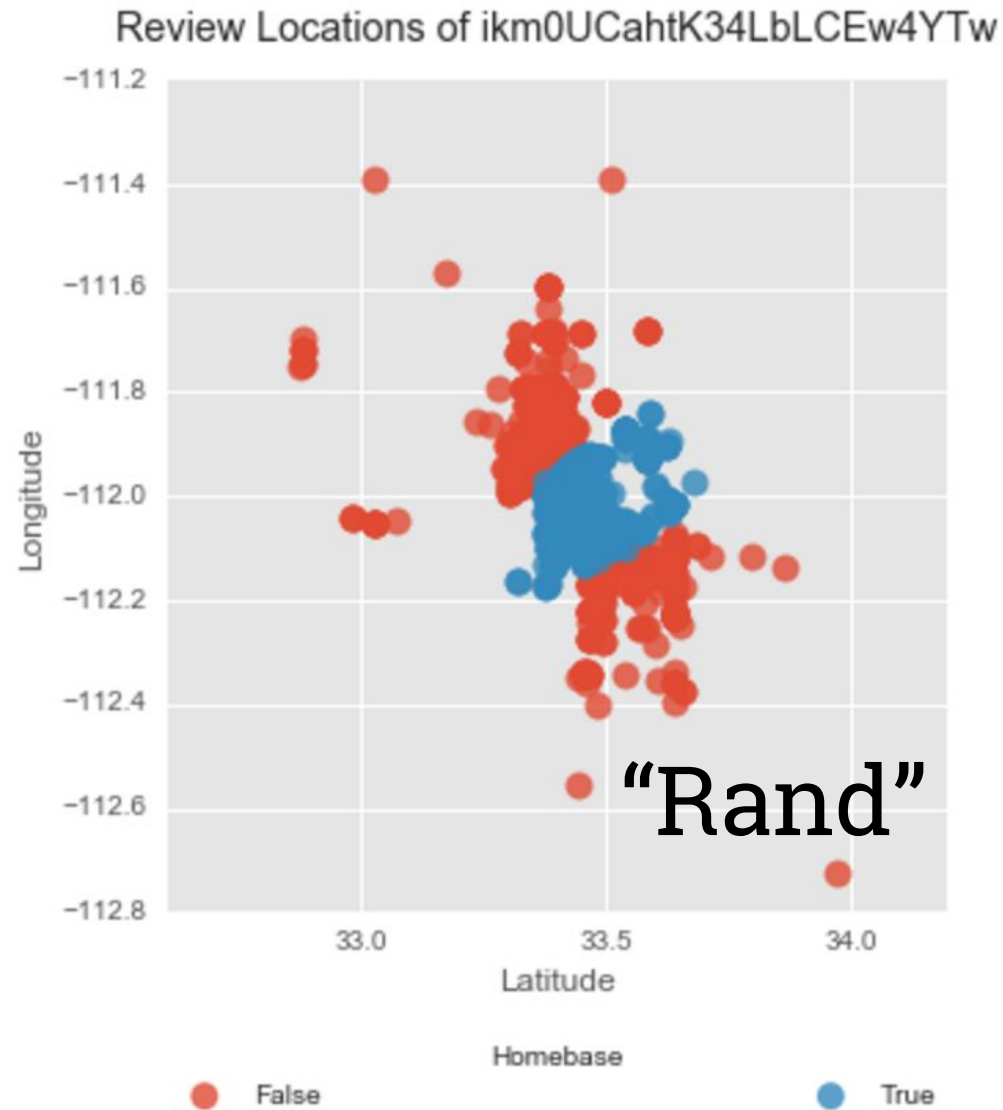
False True

"Nader"

How to identify “Home”



How to identify “Home”



Business Price Range:

“Home” versus “Away”



len	mean	std
11177.0	1.78	0.60

len	mean	std
13884.0	1.72	0.59

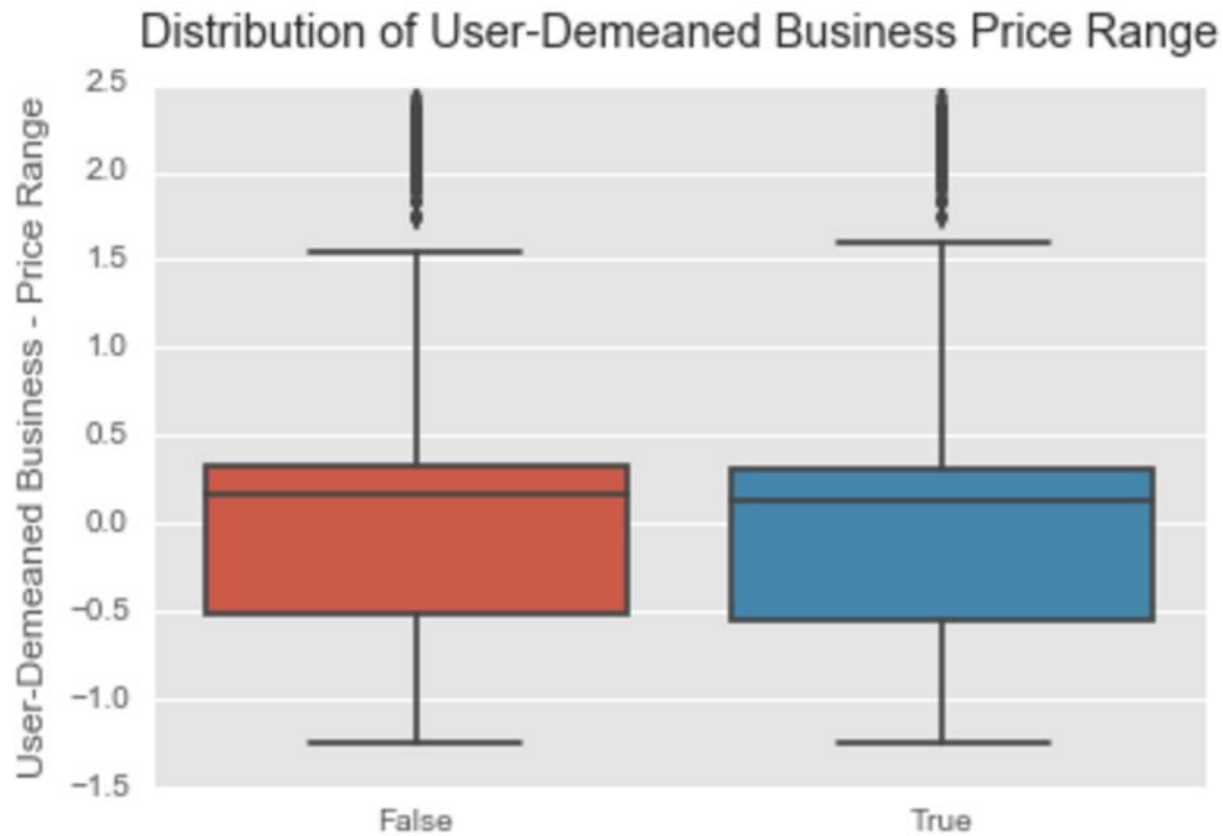
Business Price Range:

“Home” versus “Away”



Business Price Range:

“Home” versus “Away”



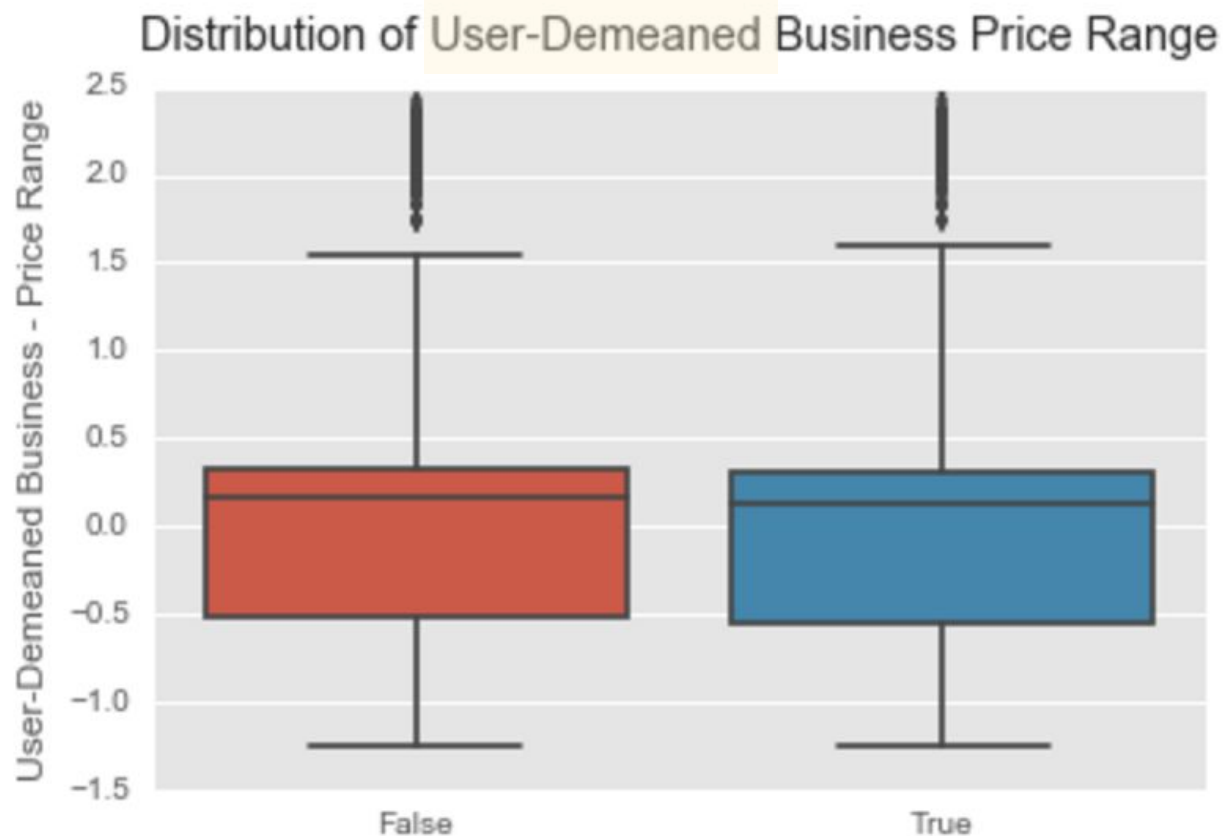
len	mean	std
11177.0	0.03	0.58

homebase

len	mean	std
13884.0	-0.02	0.56

Business Price Range:

“Home” versus “Away”



len	mean	std
11177.0	0.03	0.58

homebase

len	mean	std
13884.0	-0.02	0.56

A decorative network diagram at the top of the slide, featuring a complex web of interconnected nodes and lines. A central node is highlighted with a dashed circle and a solid circle, containing a large blue quotation mark.

“

*Are these user-
demeaned **differences**
statistically significant?*



Are these user-demeaned differences
statistically significant?

H₀: equal pop. means


H_a: unequal pop. means





*Are these user-demeaned differences
statistically significant?*

p-value of 0.0000000034 indicates that
our data provides enough evidence to reject the null
hypothesis for the **entire population**.






*Are these user-demeaned differences
statistically significant?*

99% confidence interval
for the difference in means (home-away)

$[-0.062, -0.024]$





*Are these user-demeaned differences
statistically significant?*

Yes!



*Users **review**
more expensive
venues if they are
further away*

As a business owner:

- rethink local marketing
 - target travelers
- partner accordingly
- cater to non-locals



Next Steps

◎ The Data:

- How do we define Population?
- Data Sample: representative?
- Data Subset: Users w/ > 100 reviews

Next Steps

🎯 Clustering

- $K = 3...$
- distance metric = Euclidean...
- Can we do better with a Kernel?
- Other ways to identify “home”?

Next Steps

- ◎ Continuous distance
- ◎ Geographic considerations
- ◎ Other factors:
 - Business Type
 - Region Type
 - Time of Day
 - Day of Week

Next Steps

- ◎ Can we extend this geographically
 - Business/Pleasure travel
 - Partner with Airlines, Accommodations
- ◎ Is this indicative of
 - user preferences
 - willingness to spend



Thank You!

Any questions?

<http://bit.ly/29fWXgF>

lubagloukhov@gmail.com



