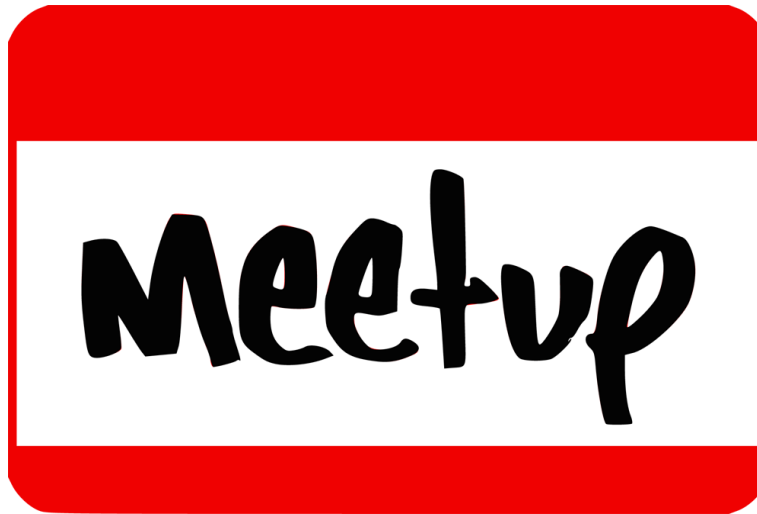



Meetup.com Analysis



Anne Chen

08/14/16

What is Meetup?

**Find**
a Meetup Group

Start
a Meetup Group

English

Log in

Sign up



Meetups are

neighbors getting together to learn something, do something, share something...

Sign me up!

Photo: The V Star Sydney Meetup

All Meetups  within 5 miles of Hoboken, NJ

GroupsCalendar

Sort by Best match

Hudson Valley Hikers

We're 21,842 HV Hikers



NY Tech Meetup

We're 50,121 NYC Technolo...



Dharma Drink: the hangout for Buddhists (& like minded)

We're 4,878 Members



Data Scraped

- Events within 10 miles in these 10 cities*:
New York, San Francisco, Chicago, Washington DC, Palo Alto, Boston, Los Angeles, Mountain View, Seattle, and Austin
- Date range: 08/11/2016 to 10/07/2016
- Total number of event scraped: 12123 events
- This analysis only involved in open group/event

Group information

- Founded date
- Group keywords
- Group name
- Number of members
- Number of past meetups
- Number of upcoming meetups
- Number of Sponsors

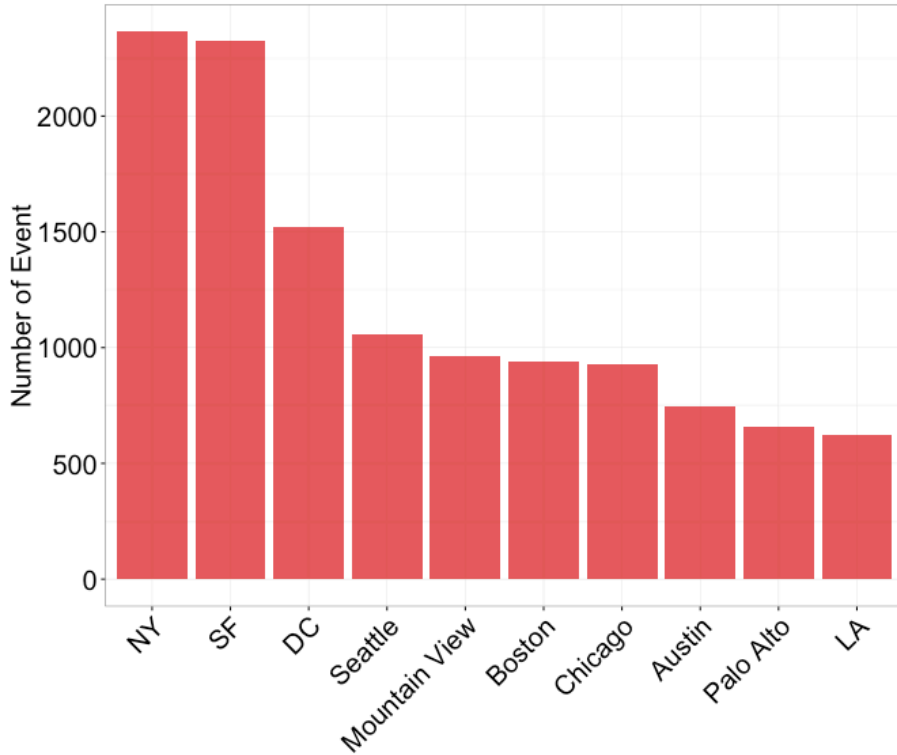
Event Information

- Event date
- Event start time
- Event title
- Number of participants
- Price
- Number of comment and reply

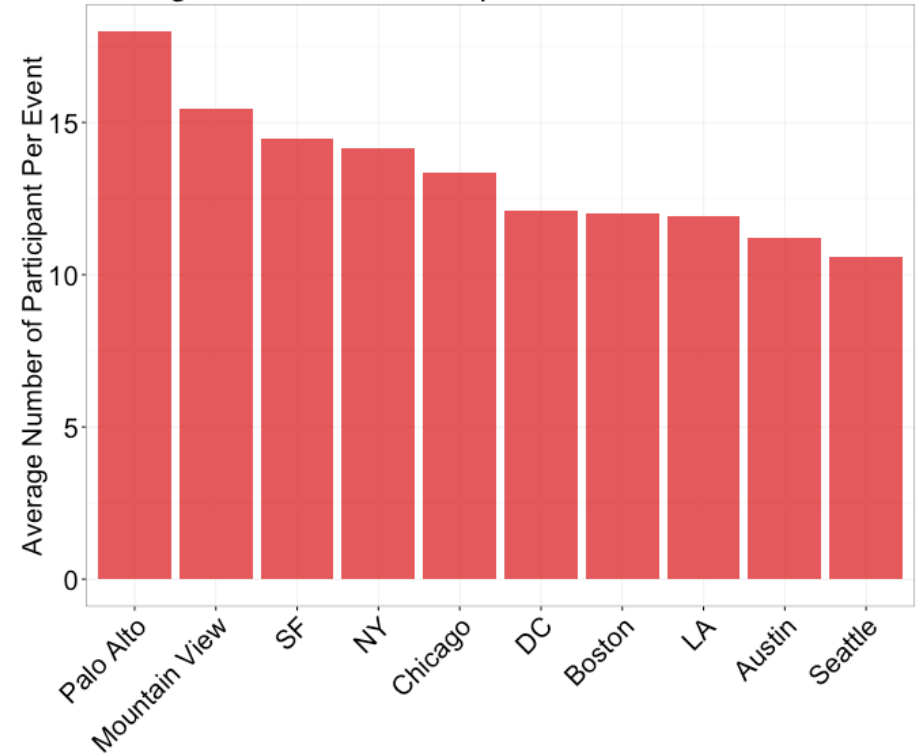
*Top 10 Cities were chosen based on the total number of membership in 2014:
<https://blog.rjmetrics.com/2014/04/23/whos-meeting-up-a-ranking-of-top-startup-cities/>

Exploratory Data Analysis

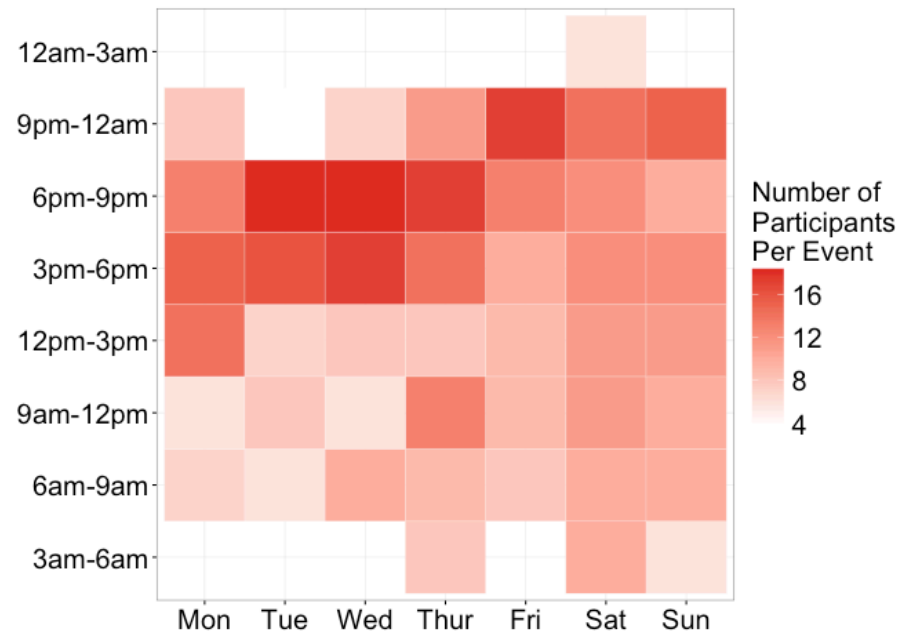
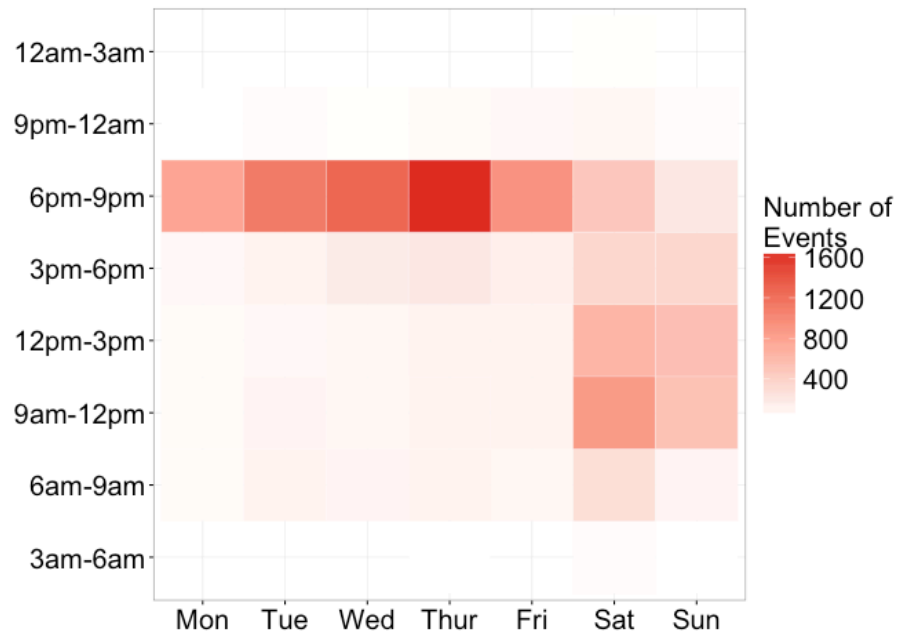
Number of Event Collected in Ten Cities



Average Number of Participants Per Event in Ten Cities



Popular and Best Time to Host an Event



Note: In order to avoid bias, only data with number of event > 3 are shown

Most Common Words Used in Group Keywords



Most Common Words Used in Event Titles

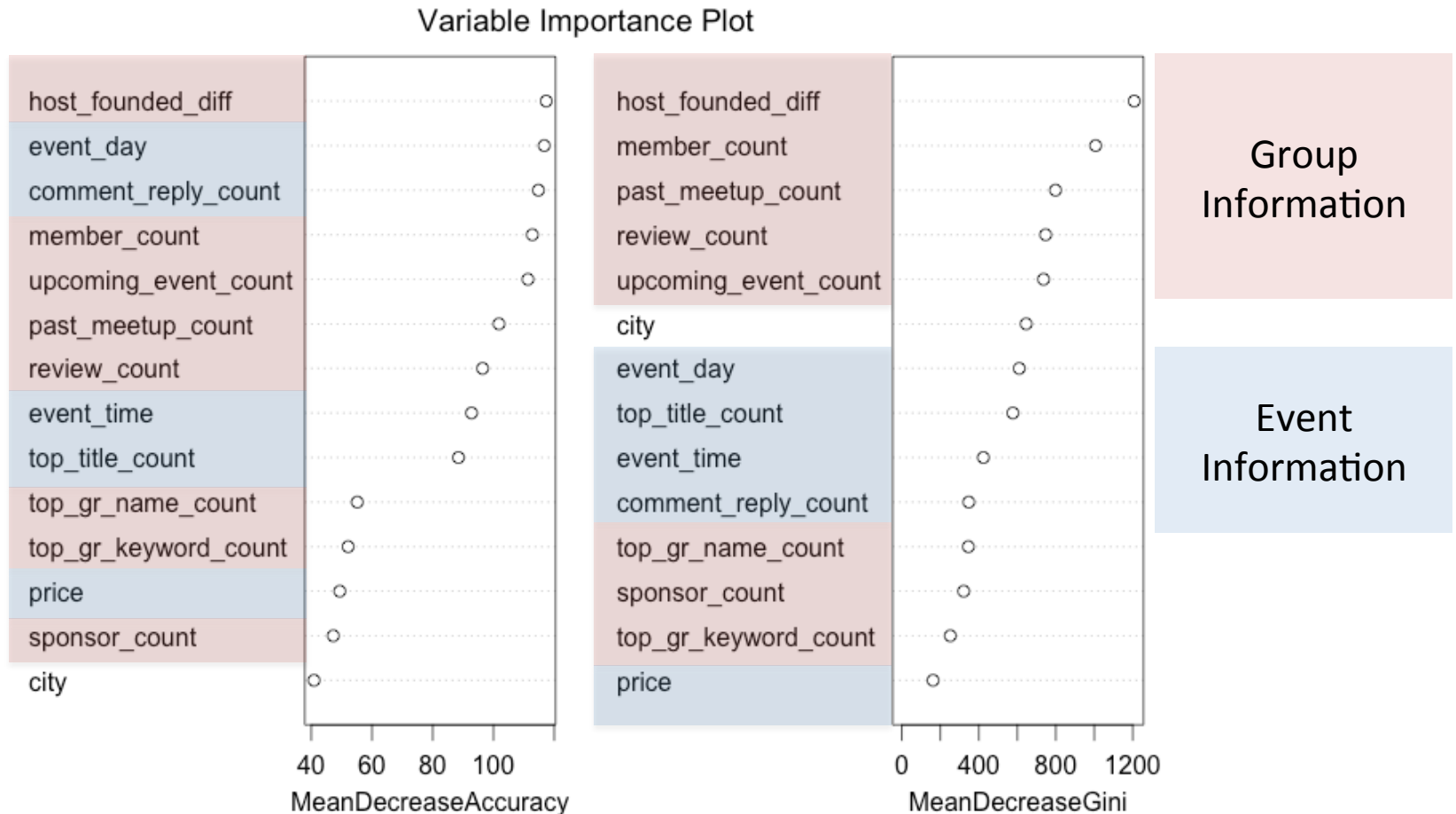


Most Common Words Used in Group Names



Model Building – Random Forest

- Number of variables randomly sampled as candidates at each split: 4
- Number of trees to grow: 500



How Well Did the Model Perform?

Predict number of participants (n) of an event

- Absolute Accuracy: 41.9%
- Relative Accuracy: 75.8%

Groups: " $n \leq 10$ ", " $10 < n \leq 20$ ",
" $20 < n \leq 30$ ", " $30 < n \leq 50$ ",
" $50 < n \leq 100$ ", and " $n > 100$ "

Summary

- Tuesday to Thursday from 6pm to 9pm are popular time to host an event
- Event hosted by a group that
 1. Exists longer
 2. Has more member
 3. Remains highly activemay have more participants

Questions?

APPENDIX SLIDES

Words Removed from Computing Word Cloud

"nyc", "new jersey", "new york", "chicago", "seattle", "bay area", "austin", "boston", "silicon valley", "washington", "los angeles", "mountain view", "san francisco", "central park", "meetup", "meeting", "meet", "event", "group", "club", "events", "north", "south", "east", "west", "area", "city", "brooklyn", "jersey", "manhattan", "hoboken", "queens", "hudson", "2016", "august", "september", "october", "day", "monday", "tuesday", "wednesday", "thursday", "friday", "saturday", "sunday"

Though Thursday 6pm to 9pm is the Most Popular Time to Host an Event Across Ten Cities...

City	Event Day and Time	Max(participants/event)
Austin	Thursday, 6am-9am	29
Boston	Tuesday, 3pm-6pm	20
Chicago	Tuesday, 12pm-3pm	25
DC	Friday, 9pm-12am	18
LA	Thursday, 3pm-6pm	27
Mountain View	Thursday, 9am-12pm	38
NY	Monday, 12pm-3pm	26
Palo Alto	Wednesday, 6pm-9pm	35
Seattle	Monday, 3pm-6pm	14
SF	Monday, 12pm-3pm	27

Note: In order to avoid bias, only data with number of event > 3 were used¹⁵

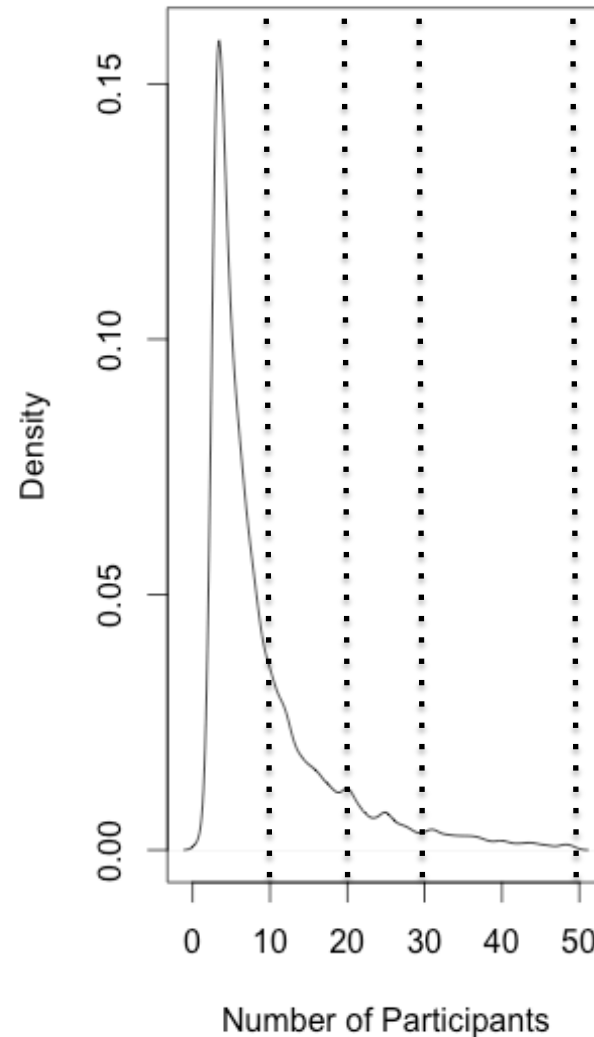
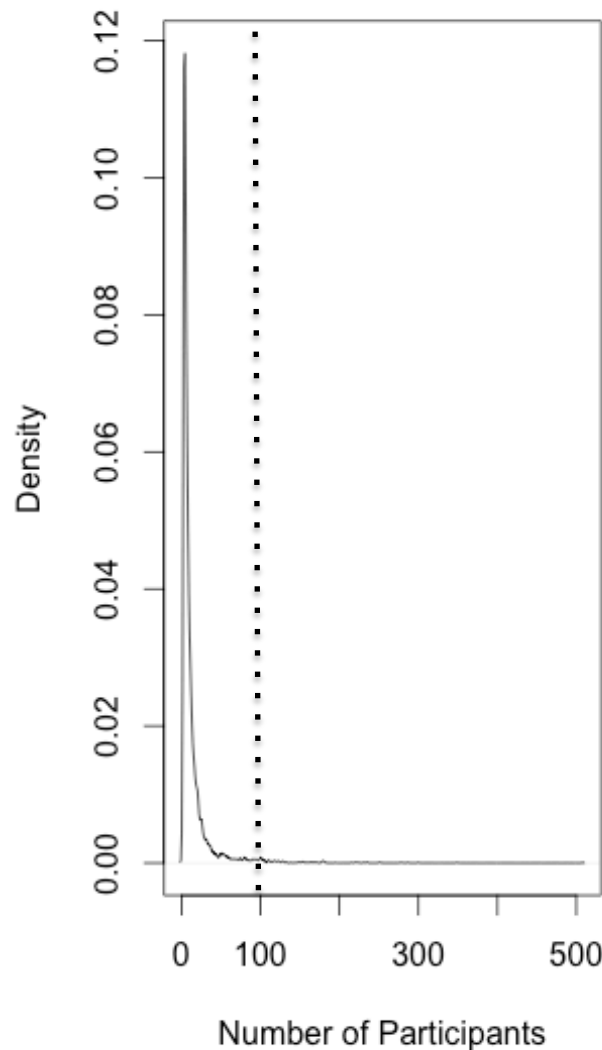
Summary of Variables Used in Model

# of Member	# of Participant	# of Past Meetup	Price	# of Review	# of Sponsor
Min. : 1	Min. : 1.00	Min. : 0.0	Min. : 0.00	Min. : 0.00	Min. : 0.000
1st Qu.: 232	1st Qu.: 4.00	1st Qu.: 12.0	1st Qu.: 0.00	1st Qu.: 3.00	1st Qu.: 0.000
Median : 697	Median : 6.00	Median : 95.0	Median : 0.00	Median : 10.00	Median : 0.000
Mean : 1745	Mean : 13.43	Mean : 405.0	Mean : 2.26	Mean : 32.39	Mean : 1.164
3rd Qu.: 1837	3rd Qu.: 13.00	3rd Qu.: 374.5	3rd Qu.: 0.00	3rd Qu.: 30.00	3rd Qu.: 1.000
Max. : 50098	Max. : 508.00	Max. : 14728.0	Max. : 950.00	Max. : 1081.00	Max. : 64.000
# of Upcoming Event	Group Founded (Days)	Top 1% Title Count	Top 1% Group Name Count	Top 1% Group Keyword Count	# of Comment + # of Reply
Min. : 0.00	Min. : 2	Min. : 0.000	Min. : 0.000	Min. : 0.0000	Min. : 0.0
1st Qu.: 3.00	1st Qu.: 402	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.0
Median : 9.00	Median : 1170	Median : 1.000	Median : 1.000	Median : 0.0000	Median : 0.0
Mean : 25.98	Mean : 1432	Mean : 1.463	Mean : 0.804	Mean : 0.3665	Mean : 1.2
3rd Qu.: 23.00	3rd Qu.: 2270	3rd Qu.: 2.000	3rd Qu.: 1.000	3rd Qu.: 1.0000	3rd Qu.: 1.0
Max. : 2361.00	Max. : 5107	Max. : 9.000	Max. : 6.000	Max. : 6.0000	Max. : 141.0

Model Building Workflow

- Impute 0 to missing values
- 20% data → testing set
80% data → training set
- Check both sets have similar distribution by creating density plots for number of participants
- Find the optimal mtry for randomForest()
- Run randomForest()
- Predict number of participants for testing set
- Calculate absolute and relative accuracy

How Did I Decide the Groups to Determine Relative Accuracy



Another Model Ran – Lasso Regression

(Intercept)	4.37
price	-0.02
event_dayMonday	0.57
event_daySaturday	1.14
event_daySunday	0.91
event_dayThursday	3.47
event_dayTuesday	3.78
event_dayWednesday	3.81
event_time12pm-3pm	.
event_time3am-6am	-0.33
event_time3pm-6pm	1.75
event_time6am-9am	-2.52
event_time6pm-9pm	3.26
event_time9am-12pm	-0.68
event_time9pm-12am	0.84
top_title_count	0.25
top_gr_name_count	0.09
top_gr_keyword_count	2.36
comment_reply_count	0.61

member_count	0.00
past_meetup_count	0.00
review_count	-0.01
sponsor_count	0.83
upcoming_event_count	-0.01
host_founded_diff	.
cityboston	.
citychicago	1.06
citydc	-0.29
cityla	-0.74
citymountain_view	3.42
cityny	1.18
citypalo_alto	5.07
cityseattle	-0.91
citysf	1.39