

Topic: SumAll.org Assignment
Candidate: Reshama Shaikh
Date: 4/27/14

Q1:

These are the 10 that are closest to the SumAll office

Rank, Name, Distance

1	GRAND_LINDSAY	1.1180906353526509e-05
2	SKREPICK_KEITH_RON_MICHAEL	3.4391892949710935e-05
3	LIAN FENG TRADING_INC.	3.9638682498204743e-05
4	DAVID;NORMAN INVESTO_RS; INC	4.0349290432427324e-05
5	LOW_HA_POH HA_LOW POH	6.663023528566444e-05
6	CONTEST PROMOTIONS-_NY LLC	6.918278087996733e-05
7	BUTTERWORTH_WILLIAM	7.049615578549744e-05
8	SAGAR_LOU	7.097935355484579e-05
9	HODGSON_HALL_JOHN_GEORGE	7.780603898675918e-05
10	PETRY_SPIRO_GUSTAV_LAURIE	8.93705736336282e-05
11	ZHENG_JUN LI	9.085521362166687e-05

1 620732,GRAND,LINDSAY,,8,CENTRE MARKET PLACE,1,NY 10013
2 363953,SKREPICK/KEITH,RON/MICHAEL,,66,CROSBY STREET,6C,NY 10012
3 369638,LIAN FENG TRADING,INC.,,128-,138 MOTT STREET,STORE,NY 10012
4 1079124,DAVID & NORMAN INVES,TORS INC,,43,FRONT SPRING ST,,NY (not really NYC address)
5 106186,LOW/HA,POH HA/LOW POH,,88,BAXTER STREET,B4,NY 10013
6 1101412,CONTEST PROMOTIONS-,NY LLC,,95,BOWERY,,NY 10002
7 140952,BUTTERWORTH,WILLIAM,,163,CHRYSTIE STREET,4,NY 10002
8 383094,SAGAR,LOU,,250,MOTT STREET,3,NY 10012
9 626862,HODGSON/HALL,JOHN/GEORGE,,151,CANAL STREET,2,NY 10002
10 115289,PETRY/SPIRO,GUSTAV/LAURIE,,41,WHITE STREET,2A,NY 10013
11 355802,ZHENG,JUN LI,,50,BAYARD STREET,5S, 10013

#4 was geocoded, but does not seem like a valid NYC address.

Considerations: There are about 200 entries in the name database that are non-unique

Q2: Data visualization of tweet counts

link to visualization

http://vleda.stern.nyu.edu/~rs2715/langone2014/07_sortable_bar_chart_edit/sortable_histogram_title.html

Q3: File is: insults_test_prediction_2000.csv