

Simple Presentation

Methods

- Step 1:
Group datasets by location into 200 groups
- Step 2:
Use Fuzzywuzzy calculate string similarity
- Step 3:
Within high similarity pair, calculate their price differences to set threshold

Time Compare

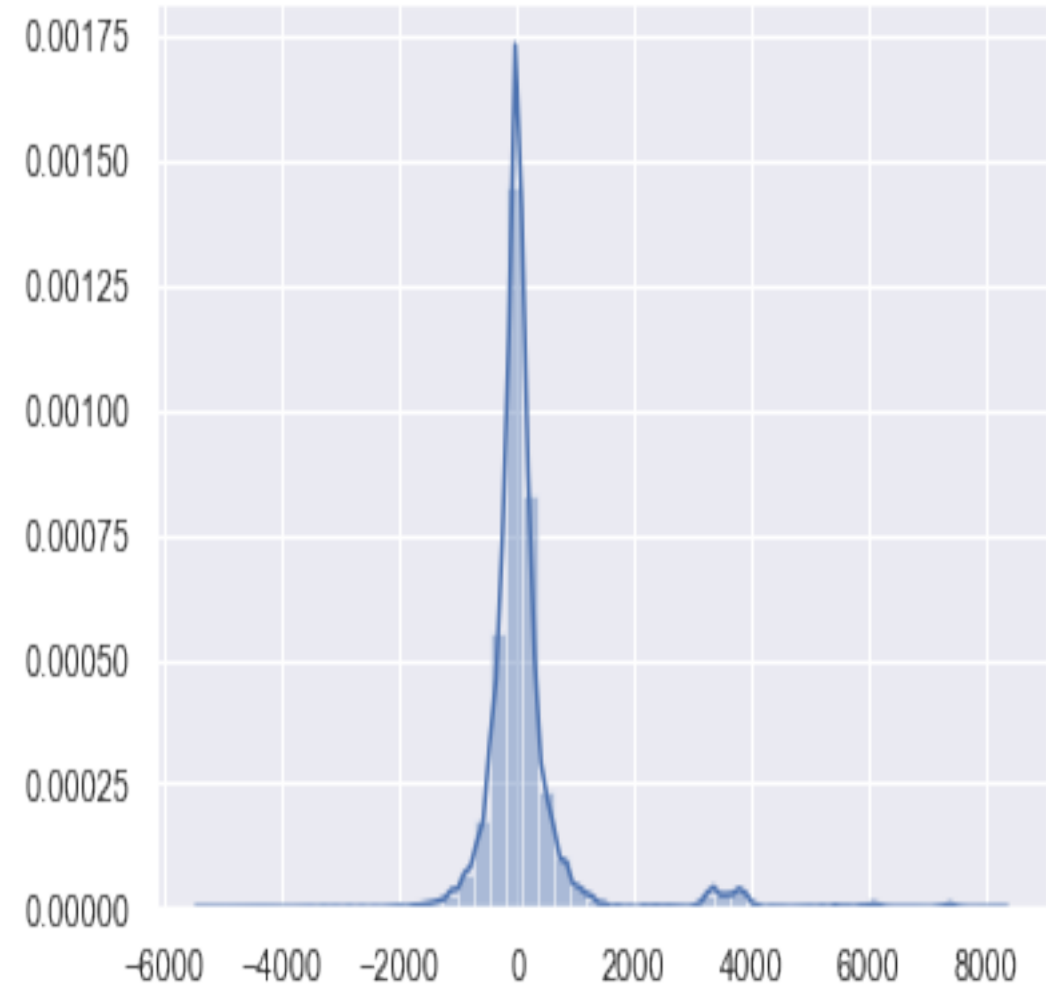
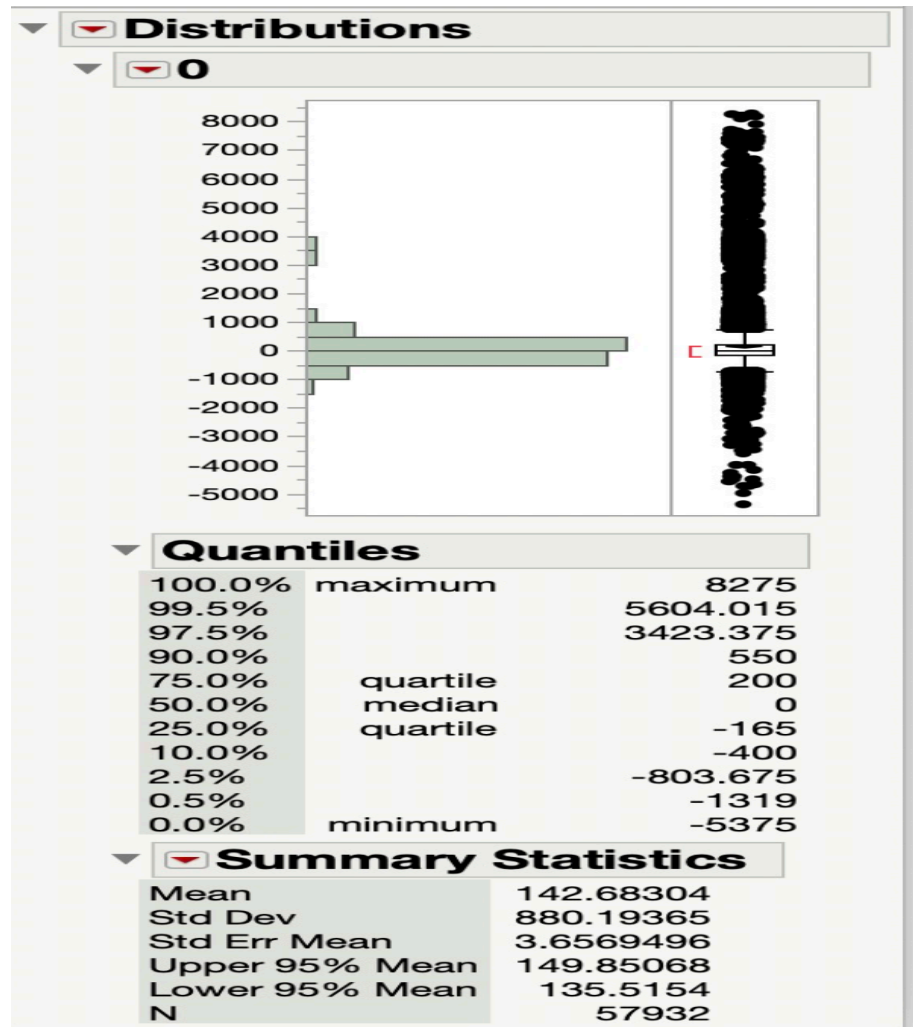
- Old:

277.3550341129303s

- New:

34.967061042785645s

Price Differences Distribution



Data frame

[illegible]

```

index 4 VS 34
SIX BEDROOM COMMONWEALTH AVENUE TWO BATHROOM DISHWASHER DINING ROOM **** SIX BEDROOM COMMONWEALTH AVENUE TWO BATHROOM
DISHWASHER DINING ROOM
index 7 VS 40
SIX ONE MODERN 6BD TUFTS BUS HARVARD DAVIS TWO BATHS LAUNDRY PETS NEG **** SIX MODERN 6BD TUFTS BUS HARVARD DAVIS TWO BATHS
LAUNDRY PETS NEG
index 11 VS 12
JUNE 1ST SINGLE FAMILY SIX BEDROOM TWO BATHROOM PORCHES PARKING LAUNDRY **** BIG SINGLE FAMILY SIX BEDROOM TWO BATHS PORCHES
PARKING LAUNDRY
index 11 VS 13
JUNE 1ST SINGLE FAMILY SIX BEDROOM TWO BATHROOM PORCHES PARKING LAUNDRY **** BIG SINGLE FAMILY SIX BEDROOM TWO BATHS PORCHES
PARKING LAUNDRY
index 12 VS 13
BIG SINGLE FAMILY SIX BEDROOM TWO BATHS PORCHES PARKING LAUNDRY **** BIG SINGLE FAMILY SIX BEDROOM TWO BATHS PORCHES PARKING
LAUNDRY
index 25 VS 26
FOUR ACADEMY HILL ROAD B **** FOUR ACADEMY HILL ROAD C
index 25 VS 27
FOUR ACADEMY HILL ROAD B **** FOUR ACADEMY HILL ROAD D
index 26 VS 27
FOUR ACADEMY HILL ROAD C **** FOUR ACADEMY HILL ROAD D
index 30 VS 31
MODERN SIX BEDROOM THREE BATH SINGLE FAMILY HOUSE IN EXCELLENT CONDITION **** MODERN SIX BEDROOM THREE BATH SINGLE FAMILY
HOUSE IN EXCELLENT CONDITION
index 37 VS 60
NINE ONE AMAZEBALLS SIX BEDROOM THREE BATHS IN UNIT LAUNDRY DAVIS **** NINE ONE AMAZING SIX BEDROOM THREE BATHS IN UNIT
LAUNDRY DAVIS
index 44 VS 57
NINE ONE WHAT 6BED 3BATH LAUNDRY RENOVATED HEART DAVIS **** NINE ONE NO FEE 6BED 3BATH LAUNDRY RENOVATED HEART DAVIS
index 49 VS 59
NO FEE HUGE MODERN SEVEN BEDROOM 3BATH DUPLX STEPS KENDALL CENTRAL SQS **** NO FEE HUGE MODERN SEVEN BEDROOM 3BATH DUPLX
STEPS KENDALL CENTRAL SQS
index 64 VS 65
GARDEN LEVEL STUDIO IN PACKARD S CORNER IN ALLSTON MOVE IN NOW OR FOUR ONE **** GARDEN LEVEL STUDIO IN PACKARD S CORNER IN
ALLSTON MOVE IN NOW
index 69 VS 70
VACANT GARDEN LEVEL STUDIO IN PACKARD S CORNER ALLSTON VIRTUAL TOUR **** VACANT GARDEN LEVEL STUDIO IN PACKARD S CORNER
ALLSTON VIRTUAL TOUR
index 75 VS 1606
MONTH STUDIO SIX ONE PACKARDS CORNER INCREDIBLE VALUE NO FEE **** MONTH STUDIO SIX ONE PACKARDS CORNER INCREDIBLE VALUE NO
FEE

```

New Dataset Ouputs

0	0	1990 0 Commonwealth Ave -71.14154 42.347108	2020-10-07 06:02:15.270817	2020-10-07	162	27 difference:18 96 difference:-690
0	0	3399 0 4 Massachusetts Ave 731 -71.100291 42....	2020-10-07 07:16:21.547766	2020-10-07	150	8 difference:12 9 difference:24 10 differe...
0	0	3387 0 4 Massachusetts Ave 743 -71.100291 42....	2020-10-07 07:17:37.720176	2020-10-07	150	9 difference:12 10 difference:24
0	0	3375 0 4 Massachusetts Ave 788 -71.100291 42....	2020-10-07 07:18:53.854346	2020-10-07	150	10 difference:12

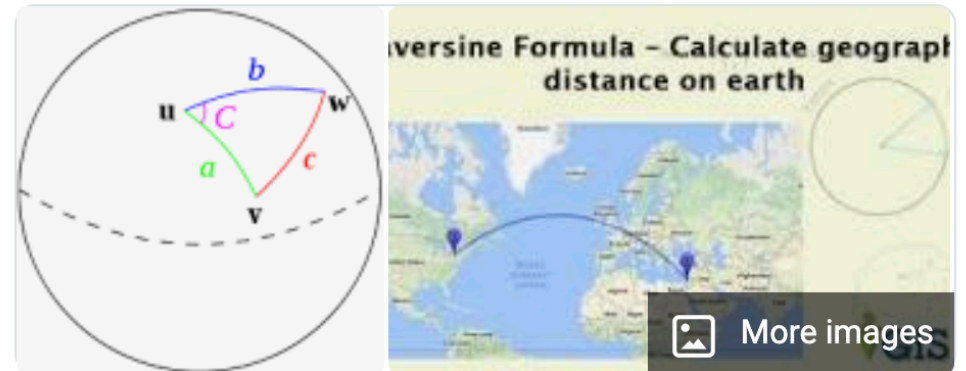
Shapefile Problems

- 1) Inconsistent in CSV & Shapefile
- Missing lat&long in CSV;
- 2) Inconsistent in rental_list & Shapefile
- Missing Zipcode in rental_list
- 3) Missing title information to determine multi-units

Haversine Distance

- Calculate Haversine Distance between those pair whose similarity >85

0.000000	8983
0.118365	35
0.177519	32
0.861100	29
0.966636	28
0.074945	24
0.065951	21
0.113603	20
1.020067	19
0.010028	19
0.191394	18
0.054518	17
0.018970	16
0.139613	16
0.005459	16
0.600409	16
0.368900	16



About

The haversine formula determines the great-circle distance between two points on a sphere given their longitudes and latitudes. Important in navigation, it is a special case of a more general formula in spherical trigonometry, the law of haversines, that relates the sides and angles of spherical triangles. [Wikipedia](#)

Duplicate

- 1) Similarity Ratio=100
- 2) Price Difference=0
- 3) Haversine Distance=0

cleaned_depuplicate_rental

	Index_i	Index_j	Title_Similarity_Ratio	Price_Difference	Location_Haversine_Distance	Title_i	Title_j
78	250	262	100	0	0	FURNISHED BRIGHT FORT POINT STUDIO W WASHER/DRYER IN UNIT	FURNISHED BRIGHT FORT POINT STUDIO W WASHER/DRYER IN UNIT
41	322	436	100	0	0	STANDARD OR STUDIO OR ONE BEDROOM SUITES WK UTILITIES INCLUDED NO MINIMUM	STANDARD OR STUDIO OR ONE BEDROOM SUITES WK UTILITIES INCLUDED NO MINIMUM
44	327	450	100	0	0	STANDARD OR STUDIO OR ONE BEDROOM SUITES WK UTILITIES INCLUDED NO MINIMUM	STANDARD OR STUDIO OR ONE BEDROOM SUITES WK UTILITIES INCLUDED NO MINIMUM
45	331	431	100	0	0	STANDARD OR STUDIO WK UTILITIES INCLUDED NO MINIMUM	STANDARD OR STUDIO WK UTILITIES INCLUDED NO MINIMUM
46	332	432	100	0	0	STANDARD OR STUDIO WK UTILITIES INCLUDED NO MINIMUM	STANDARD OR STUDIO WK UTILITIES INCLUDED NO MINIMUM
66	415	425	100	0	0	STUNNING STUDIO AVAILABLE TWO MONTHS FREE	STUNNING STUDIO AVAILABLE TWO MONTHS FREE
67	419	437	100	0	0	INCLUDES UTILITIES STUDIO APARTMENT SALEM S HISTORIC DISTRICT	INCLUDES UTILITIES STUDIO APARTMENT SALEM S HISTORIC DISTRICT
68	424	445	100	0	0	STUDIO W BEDRM NEW CARPET APPLIANCES ONE MO FREE RENT	STUDIO W BEDRM NEW CARPET APPLIANCES ONE MO FREE RENT
70	435	453	100	0	0	BROOKSIDE STUDIO APARTMENT LOWER LEVEL	BROOKSIDE STUDIO APARTMENT LOWER LEVEL
79	513	4147	100	0	0	QUINCY ONE BEDROOM IN GREAT LOCATION	QUINCY ONE BEDROOM IN GREAT LOCATION
77	600	4381	100	0	0	YOUR HAVEN FROM BUSY LIFESTYLE CALM QUIET ONE BEDROOM ONE BATH	YOUR HAVEN FROM BUSY LIFESTYLE CALM QUIET ONE BEDROOM ONE BATH
79	601	10410	100	0	0	RISE TOP HIGH RISE TWO BEDROOM TWO BATH RESIDENCES TEWKSBURY CO	RISE TOP HIGH RISE TWO BEDROOM TWO BATH RESIDENCES TEWKSBURY CO
96	626	745	100	0	0	APARTMENT RENT	APARTMENT RENT
00	632	5003	100	0	0	LYNN ONE BEDROOM HEAT INCLUDED	LYNN ONE BEDROOM HEAT INCLUDED
50	746	11571	100	0	0	RENOVATED APARTMENT HEAT HOT WATER INCLUDED	RENOVATED APARTMENT HEAT HOT WATER INCLUDED
53	747	4981	100	0	0	LYNN ONE BEDROOM HEAT INCLUDED	LYNN ONE BEDROOM HEAT INCLUDED
62	758	10628	100	0	0	PERFECT TWO BEDROOM IN DANVERS	PERFECT TWO BEDROOM IN DANVERS
74	782	852	100	0	0	CABOT CROSSING COMFORTABLE ONE BEDROOM ONE BATHROOM PRIME LOCATION	CABOT CROSSING COMFORTABLE ONE BEDROOM ONE BATHROOM PRIME LOCATION
90	791	10937	100	0	0	RENOVATED APARTMENT HEAT HOT WATER INCLUDED	RENOVATED APARTMENT HEAT HOT WATER INCLUDED
92	825	11130	100	0	0	WONDERFUL 11130 NORTH ANDOVER TWO BEDROOM	WONDERFUL 11130 NORTH ANDOVER TWO BEDROOM

Potential Multi-Units

- High Similarity & Same Title Length & Different Words

FOUR ACADEMY HILL ROAD B *** FOUR ACADEMY HILL ROAD C
FOUR ACADEMY HILL ROAD B *** FOUR ACADEMY HILL ROAD D
FOUR ACADEMY HILL ROAD C *** FOUR ACADEMY HILL ROAD D
NO BROKER FEE FURNISHED STUDIO NEWBURY STREET AVAILABLE ONE *** NO BROKER FEE FURNISHED NEWBURY STREET STUDIO AVAILA
BLE NOW
NO BROKERS FEE ALL UTILITIES INCLUDED STUDIO IN BACK BAY NINE ONE *** NO BROKERS FEE AWESOME STUDIO IN BACK BAY NINE
ONE UTILS INCLUDED
NO BROKERS FEE AWESOME STUDIO IN BACK BAY NINE ONE UTILS INCLUDED *** NO BROKERS FEE ALL UTILITIES INCLUDED STUDIO I
N BACK BAY NINE ONE
NO BROKERS FEE AWESOME STUDIO IN BACK BAY NINE ONE UTILS INCLUDED *** NO BROKERS FEE STUDIO IN BACK BAY ALL UTILITIE
S INCLUDED NINE ONE
SMALL BASEMENT STUDIO SYMPHONY ROAD ONE OR SIX ONE W KITCHENETTE *** SMALL BASEMENT STUDIO SYMPHONY ROAD NOW OR SIX
ONE W KITCHENETTE
SOUTH END STUDIO NINE FEE YOU SAVE 1789 *** SOUTH END STUDIO NINE FEE YOU SAVE 1664
STUDIO IN SOUTH END NINE FEE SAVE 23 RENT PLACES LESS *** SOUTH END STUDIO ONLY NINE FEE SAVE 18 RENT PLACES LE
SOUTH END STUDIO ONLY NINE FEE SAVE 2184 RENT PLACES LESS *** STUDIO SOUTH END ONLY NINE FEE SAVE 1834 RENT PLACES L
ESS