

Geocoding Truck Stops Update

William Co

2025-07-30

This report documents the geocoding validation process, highlighting discrepancies arising from integrating data sources like Yelp and Yellow Pages. Despite matching entries by phone number, these platforms introduce a high rate of false positives. We compare point discrepancies with and without these sources and visualize the magnitude of spatial discrepancies. The report discusses the potential causes of these inconsistencies, ranging from business relocations to outright geocoding errors, and discusses strategies for post-processing, including centroid averaging and external geocoding API cross-checks.

Table of contents

1	Introduction	2
2	Discrepancy Analysis	2
2.1	Without Yelp and Yellow Pages	2
2.2	With Yelp and Yellow Pages	2
2.3	General Overview	4
3	Possible Steps	4
4	Error Analysis	4
4.1	Are the errors originating from HQ data?	4
4.2	Is phone number matching the source of error?	6
4.3	Magnitude of the Error	6
5	Next Steps	7
6	Appendix	7
6.1	More Examples	7
7	Final Results	12

1 Introduction

During the process of geocoding truck stops, we observed a significant number of false positives. Although Yelp and Yellow Pages entries were matched by phone number, this method still introduced a substantial number of false positive matches. This proves our initial assumption of matching phone numbers, as unique identifiers to be false.

The complete dataset can be accessed [here](#), which includes the latitude and longitude values used for geocoding.

2 Discrepancy Analysis

We study this by plotting coordinate discrepancies. Each truck stop entry in the original truck stop directory has a corresponding match to Yellow Pages, Yelp and Truck Stops and Services.

2.1 Without Yelp and Yellow Pages

If we remove the all Yelp and Yellow Pages entries, we observe point discrepancies illustrated [in this HTML map](#). This file displays only those locations where the geocoded points differ by more than 1 km.

2.2 With Yelp and Yellow Pages

When we include the Yelp and Yellow Pages entries, the discrepancies increase significantly, as shown [in this HTML map](#). The discrepancies are so numerous that the HTML visualization becomes difficult to interpret due to the overwhelming number of false matches. To take this into account, the current visualization has been redone to display only point discrepancies with distances greater than 16 km.

Ideally, we should only observe small discrepancies, as shown below:

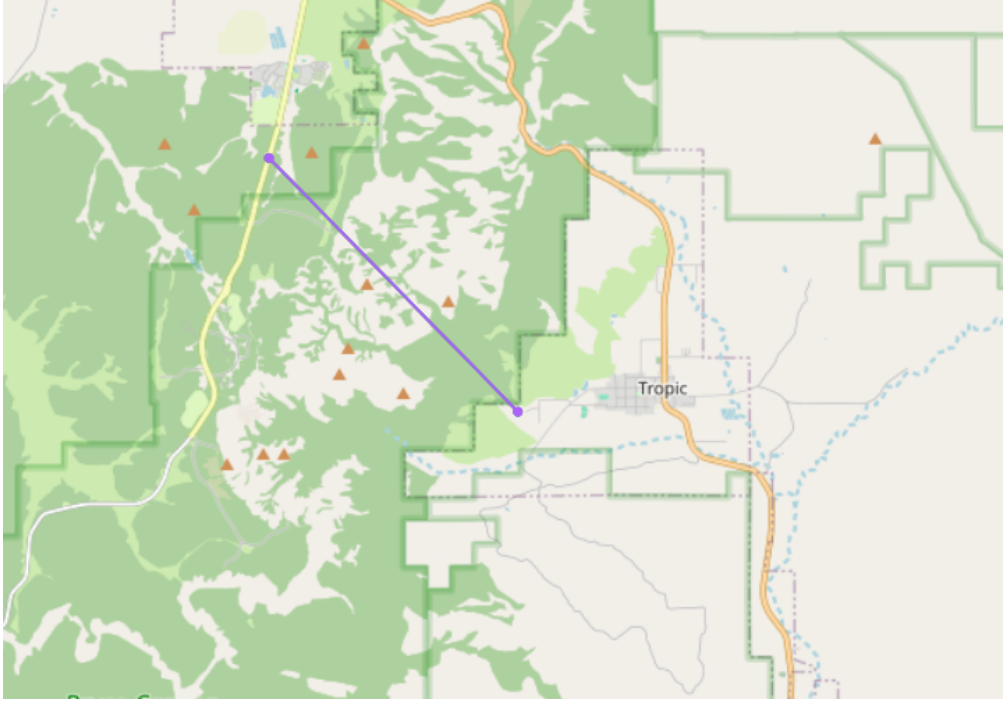


Figure 1: Map showing expected small discrepancies.

Unfortunately, we observe a map similar to the following, which shows very large discrepancies—some more than 16 km, even crossing state lines:

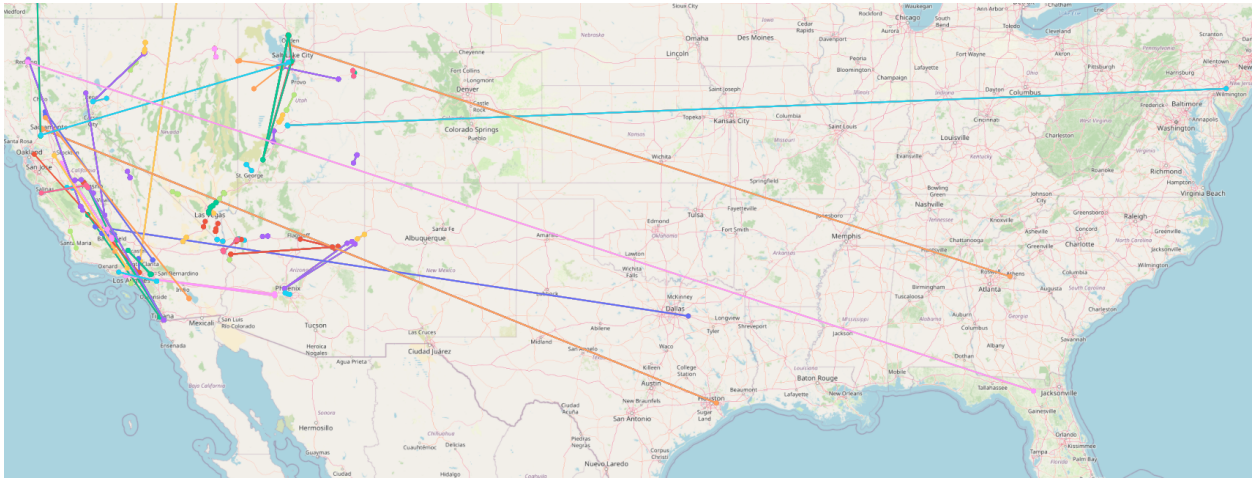


Figure 2: Map showing large discrepancies, including some crossing state lines.

This is a false positive, as Yelp and Yellow Pages should not match locations that are clearly across state lines and not even within our target four states, despite our filtering of the four states.

2.3 General Overview

For a high-level overview of the geocoded points, see [this map](#), which provides a visualization of all the coordinate data regardless of accuracy. Each line corresponds to the coordinate discrepancy, of one location.

3 Possible Steps

This issue could potentially be addressed through **manual review**, although this approach has limitations. Some discrepancies are not clearly identifiable as errors (“splitting hairs”), which complicates the process. At this stage, an **important research judgment** must be made.

The observed differences might simply reflect the **age of the locations**. Perhaps some businesses have moved. Alternatively, the discrepancies could stem from **genuine geocoding errors**.

One possible approach is to compute and use the **centroid** of the available coordinates. Another option is to introduce a **third validation step**. For example, we could use address observations to cross-check the coordinates using **geocoding API** such as Google Maps and I-Exit coordinate data as well. I-Exit validation would be straightforward but using google maps geocoding API, would also run into similar concerns about age discrepancies of locations.

Another potential step is to perform string matching between all phone number-based matches against the original data. However, this approach would likely require research judgment, as string matching is not perfect either. String matching could potentially address egregious mismatches, particularly where the match is clearly not on the same state line.

While we have outlined several potential next steps in this paper, the decision on how to proceed ultimately involves a degree of research judgment, more an interpretive choice than an objective one. As such, any input or guidance would be greatly appreciated.

4 Error Analysis

4.1 Are the errors originating from HQ data?

Upon reviewing the locations with the highest discrepancies, we find that the errors are not due to HQ-related issues, but rather appear to be error noise within the data. For example, consider the entry below:

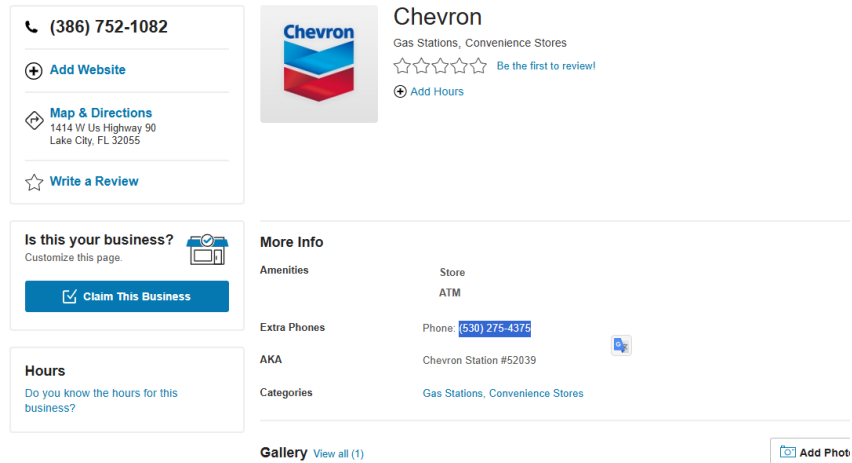


Figure 3: Entry with phone number 530-275-4375 incorrectly points to Florida.

This entry incorrectly points to a location in Florida. However, when compared to another source:

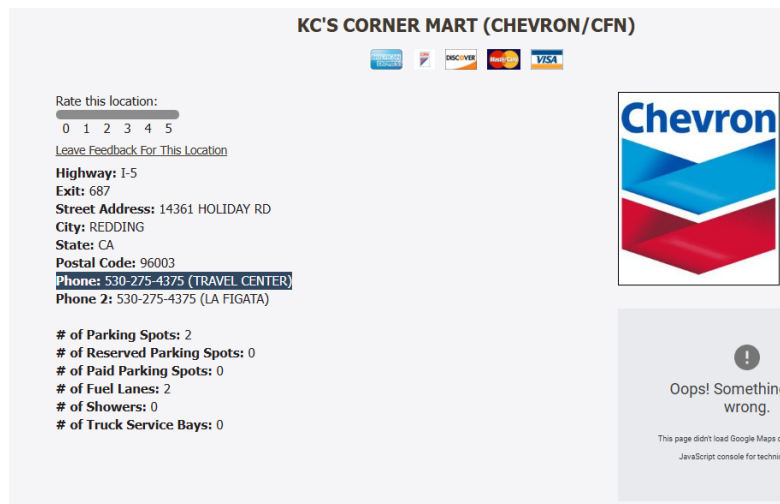


Figure 4: Alternate source showing correct location in California for phone number 530-275-4375.

it is clear that the phone number should correspond to a location in California. Yelp data further confirms the correct location.

Another example involves the following entry:

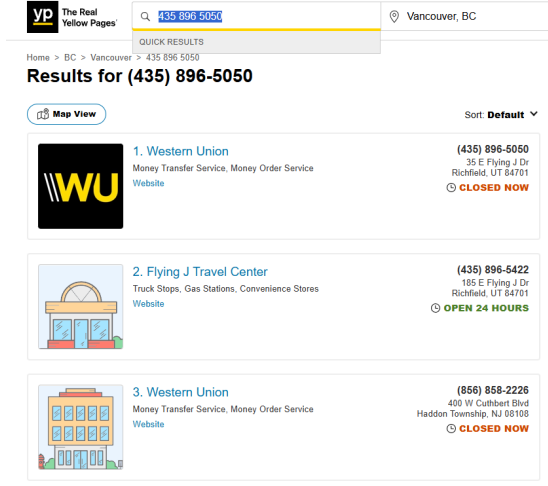


Figure 5: Phone number 435-896-5050 matches three locations: one truck stop and two Western Union branches.

where the phone number 435-896-5050 matches three locations: one is a truck stop, while the others are Western Union branches, with one Western Union location in New Jersey (see below):

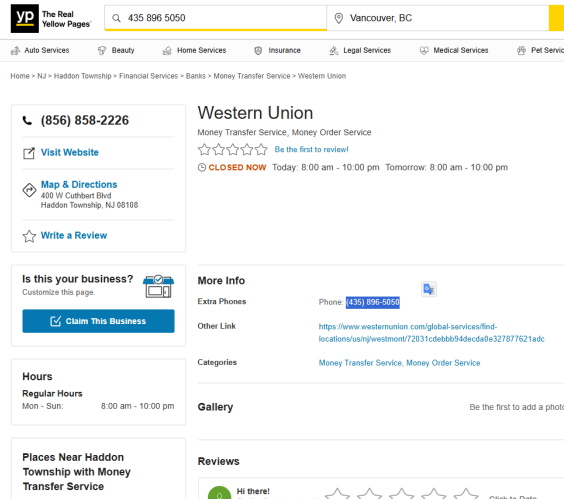


Figure 6: Western Union location in New Jersey matching phone number 435-896-5050.

4.2 Is phone number matching the source of error?

We also investigated potential issues with phone number matching, but this does not appear to be the primary cause. The errors likely stem from internal inconsistencies within the data itself.

4.3 Magnitude of the Error

We measure the maximum distances across every coordinate discrepancy. We then plot the frequency across each truck stop. We see that there is no systematic error with the data.

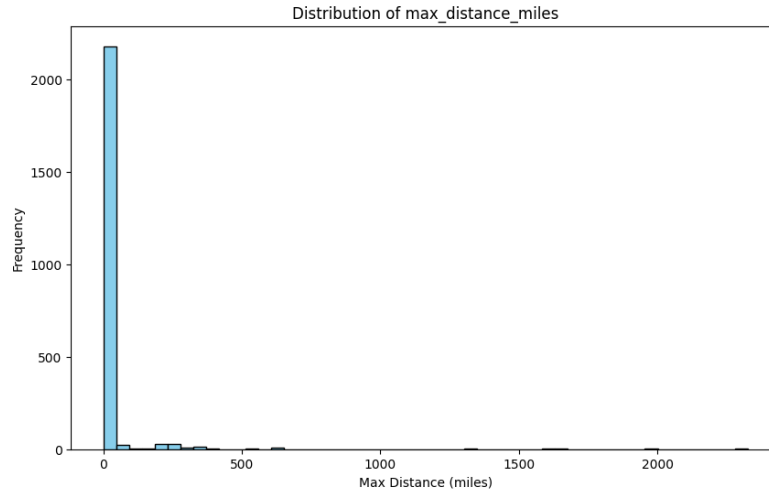


Figure 7: Frequency plot of maximum coordinate discrepancies across truck stops.

The problem may not be as egregious as the map initially suggests. Upon closer inspection, there are 450 rows with discrepancies greater than 10 miles, corresponding to about 75 unique locations.

- 75 False positives
- 30 Unmatched

5 Next Steps

Upon receiving feedback, our primary focus is now on improving accuracy. The next steps are as follows:

1. Redownload the data: The RVers dataset was previously downloaded without coordinates and needs to be re-acquired. (In Progress, the code takes more than 24 hours to run)
2. Apply post filtering: Implement a constraint to exclude non-truck stop entities from the dataset. For example, Western Union locations will be excluded even if they match a phone number. Remove mismatched states.
3. Removed blatantly false states that are not part of AZ, NV, CA, UT
4. Remove phone matches where zipcode does not match.

6 Appendix

6.1 More Examples

We see here that we cannot simply remove any non-truck stop entry. Some phone numbers only match to business locations—not necessarily truck stops.

Home > BC > Vancouver > 661 797 2122; [https://www.yellowpages.com/search?search_terms=661 797 2122](https://www.yellowpages.com/search?search_terms=661+797+2122); <https://www.yellowpages.c>

Results for (661) 797-2122

Map View Sort: Default




- 
1. Western Union
 Money Transfer Service, Money Order Service
 (661) 797-2122
 14808 Warren St
 Lost Hills, CA 93249
 CLOSED NOW
- 
2. Western Union
 Money Transfer Service, Money Order Service
 Website
 (972) 624-2239
 1703 State Highway 34 S
 Terrell, TX 75160
 OPEN 24 HOURS
- 
3. Cinnabon
 Bakeries, Restaurants
 Website
 (661) 797-2122
 14808 Warren St
 Lost Hills, CA 93249
 From Business: Bakery with a limited menu, located in Pilot Travel Center off of exit 278 on I-5 and SR 46.

Figure 8: Phone number match to business location, not necessarily truck stop.

We also see that the top place recommended by the search engine is not necessarily the correct match.

Results for (707) 678-4400

Map View Sort: Default




- 
1. Chevron
 Gas Stations, Convenience Stores
 (801) 967-9217
 6415 W 3500 S
 Salt Lake City, UT 84128
 CLOSED NOW
- 
2. Subway - Closed
 Fast Food Restaurants, Sandwich Shops, Restaurants
 Website
 60 Years
 In Business
 From Business: Your local Dixon Subway® Restaurant, located at 6854 Sievers Rd brings new bold flavors along with old favorites to satisfied guests every day. We deliver these...
- 
3. Chevron
 Gas Stations, Convenience Stores
 Website
 Regular \$4.60 Midgrade \$4.90 Premium \$5.00 Diesel \$5.10
 (707) 678-4400
 6854 Sievers Rd
 Dixon, CA 95620

Figure 9: Top place recommended by search engine is not necessarily the correct match.

The correct match is sometimes presented only after clicking into the result.

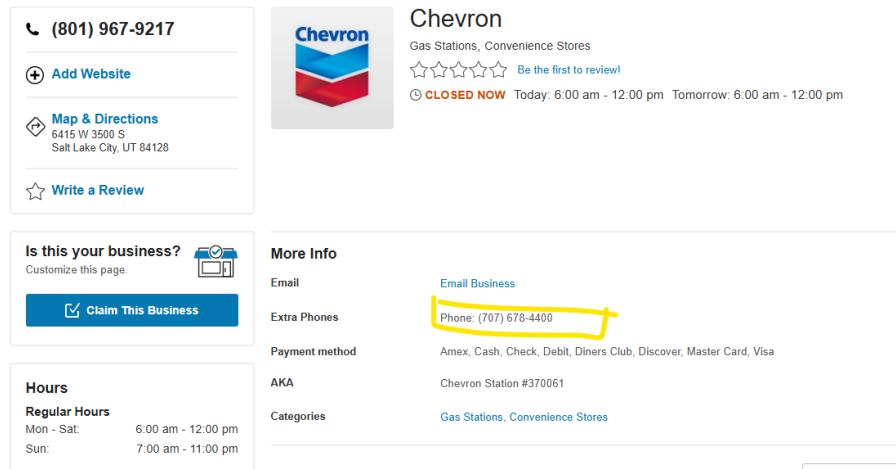


Figure 10: Match is presented when clicking inside.

There are also mismatches associated with place name matches. For example:



Figure 11: Place name mismatch 1

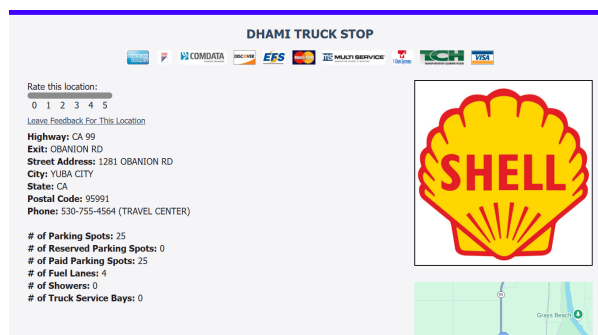


Figure 12: Place name mismatch 2

Both of these examples are a place name match, but they clearly are not.

Once we remove the easily addressable errors, we are left with mismatches that arise from data inaccuracies on the source websites.

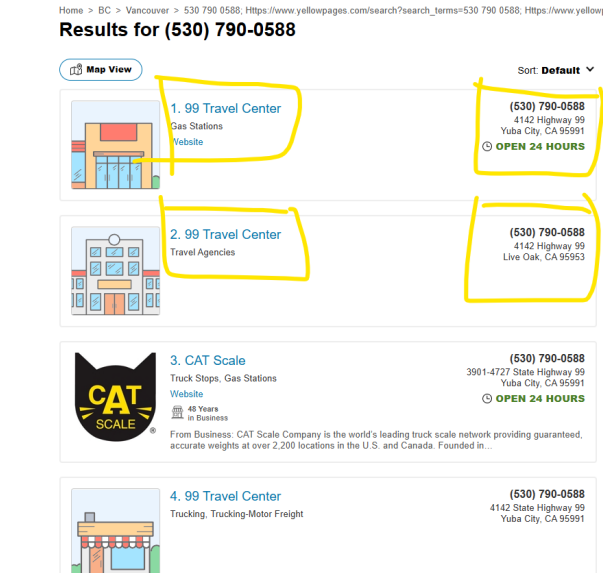


Figure 13: Website data error

In some cases, we face a data ambiguity, such as having to choose between “CAT Scale” and “99 Travel Center.”

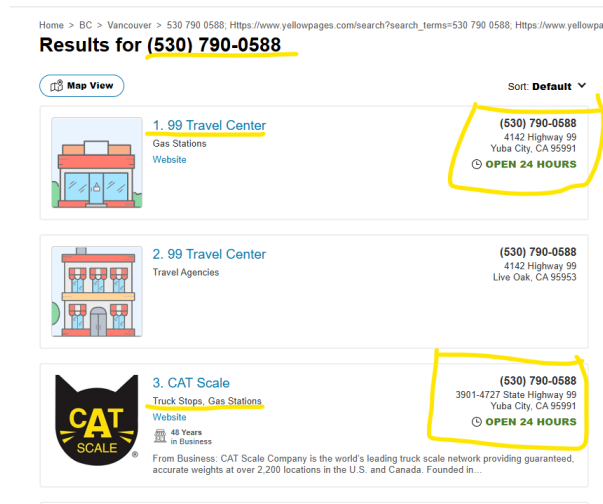


Figure 14: CAT Scale vs 99 Travel Center

Here, **99 Travel Center** is the original (correct) location.

We also find genuine geolocation errors on the *Truck Stops and Services* website. For example, the location we were targeting is shown below:

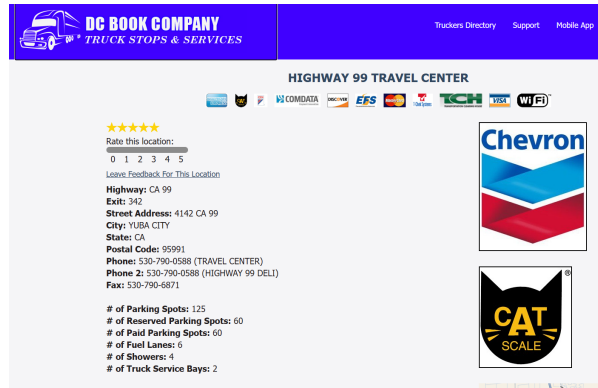


Figure 15: Correct location on map

However, using the coordinates from *Truck Stops and Services* leads us here:

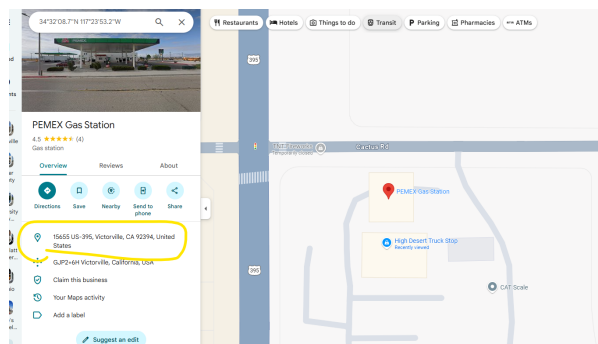


Figure 16: Incorrect geolocation result

Not only is this not the correct 99 Travel Center, but the indicated address also does not match the expected location. The actual, correct address is:

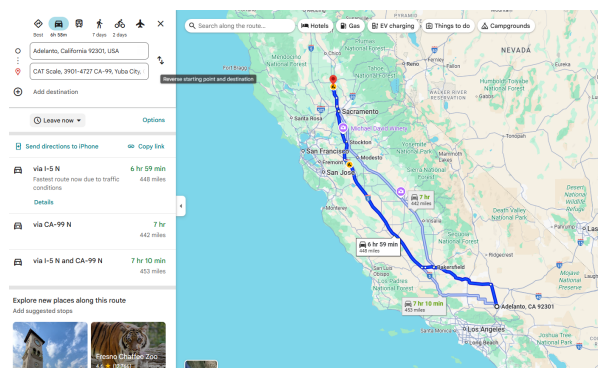


Figure 17: Correct address location

This final location is hundreds of miles away from the incorrectly suggested one.

7 Final Results

100 rows with greater than 10 mile discrepancy 17 places. All of these are phone based matches. We are fixing bugs in place name matching.