# Geocoding Truck Stops Update

William Co

2025-07-29

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

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:

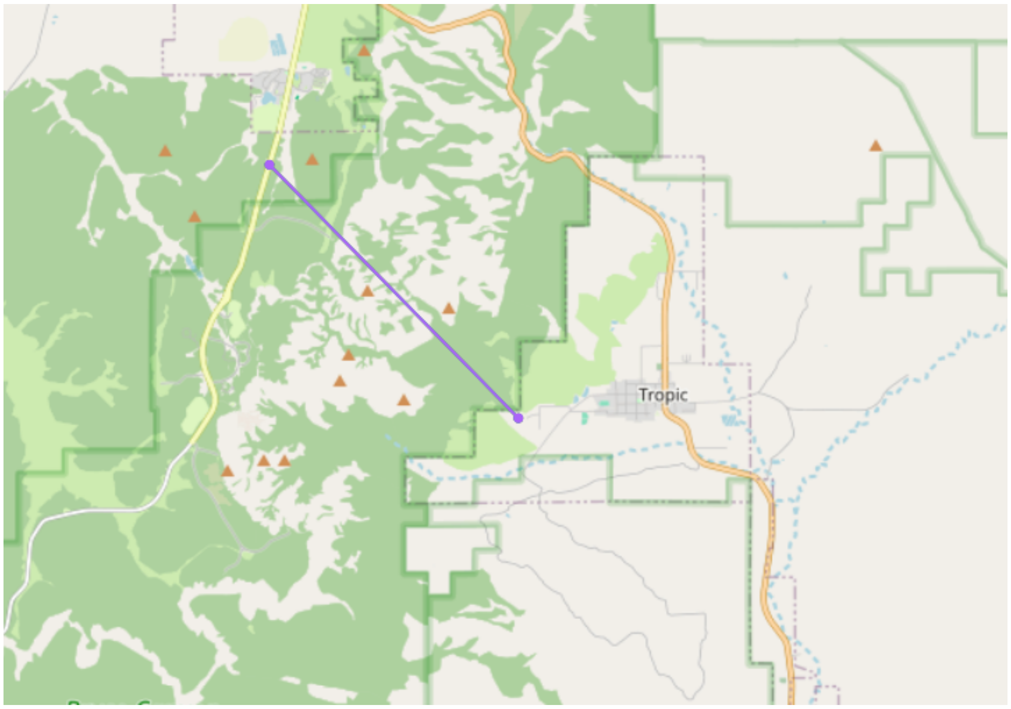


Figure 1: Map showing expected small discrepancies.

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

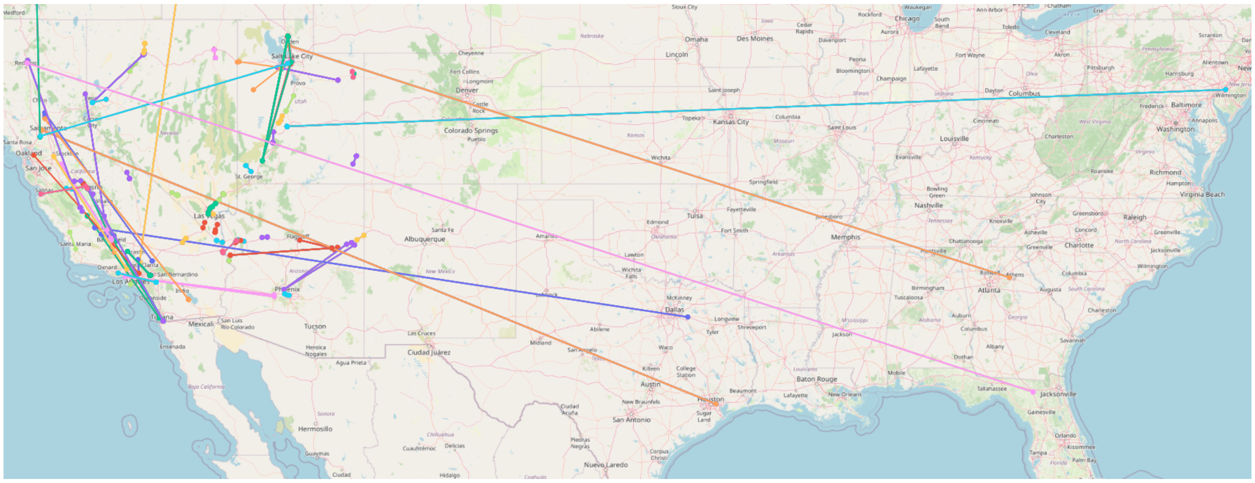


Figure 2: Map showing large discrepancies, including some crossing state lines.

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.

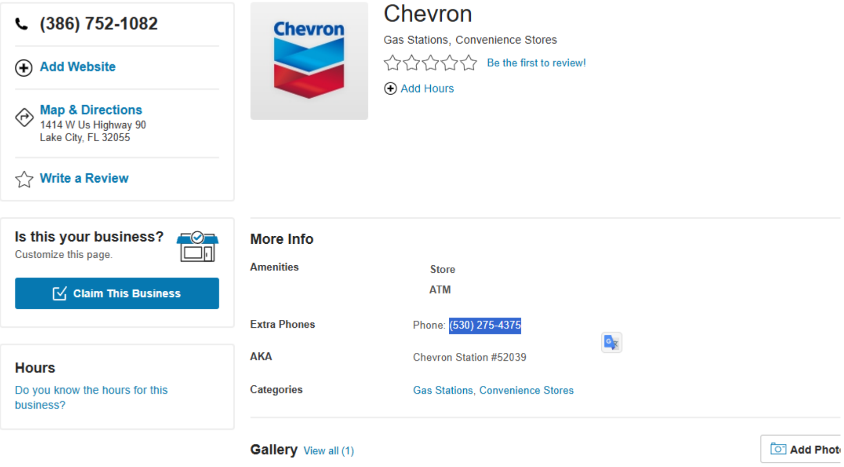
Another potential step is to perform string matching between all phone number-based matches against the original data. However, this approach would likely require research judgment, as string matching is not perfect either. String matching could potentially address egregious mismatches, particularly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



☎ (386) 752-1082

Add Website

Map & Directions
1414 W Us Highway 90
Lake City, FL 32055

Write a Review

Is this your business?
Customize this page.

Claim This Business

Hours
Do you know the hours for this business?

Chevron
Gas Stations, Convenience Stores
☆☆☆☆☆ Be the first to review!
Add Hours

More Info

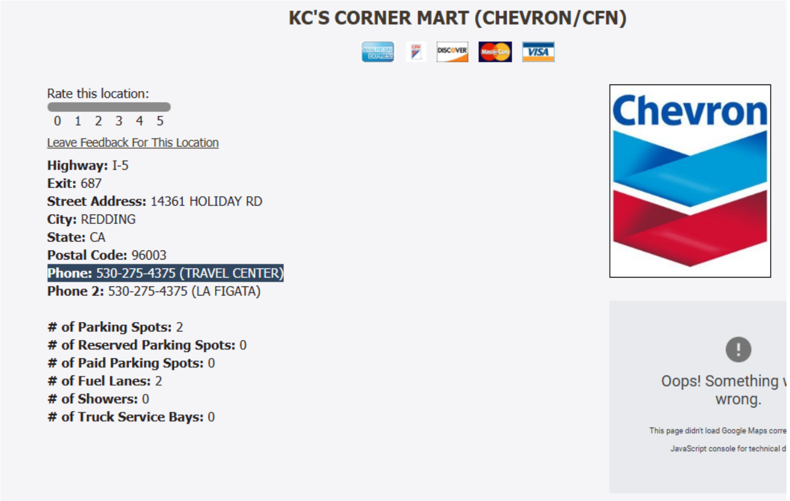
| | |
|---|---|
| Amenities | Store<br>ATM |
| Extra Phones | Phone: (530) 275-4375 |
| AKA | Chevron Station #52039 |
| Categories | Gas Stations, Convenience Stores |

Gallery View all (1)

Add Phot

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:



KC'S CORNER MART (CHEVRON/CFN)

Rate this location:
0 1 2 3 4 5
Leave Feedback For This Location

**Highway:** I-5
**Exit:** 687
**Street Address:** 14361 HOLIDAY RD
**City:** REDDING
**State:** CA
**Postal Code:** 96003
**Phone:** 530-275-4375 (TRAVEL CENTER)
**Phone 2:** 530-275-4375 (LA FIGATA)

**# of Parking Spots:** 2
**# of Reserved Parking Spots:** 0
**# of Paid Parking Spots:** 0
**# of Fuel Lanes:** 2
**# of Showers:** 0
**# of Truck Service Bays:** 0

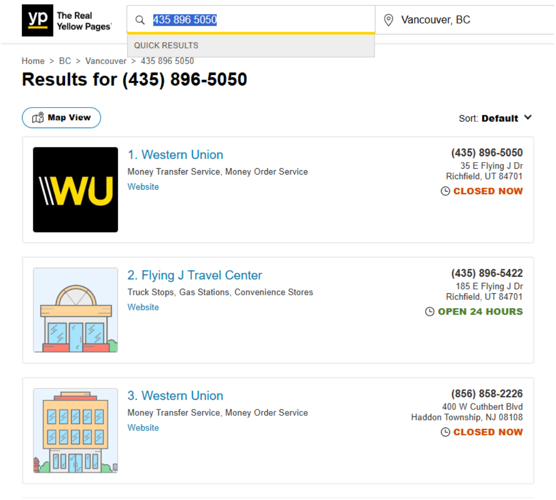
Oops! Something w
wrong.

This page didn't load Google Maps corre
JavaScript console for technical d

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:



Results for (435) 896-5050

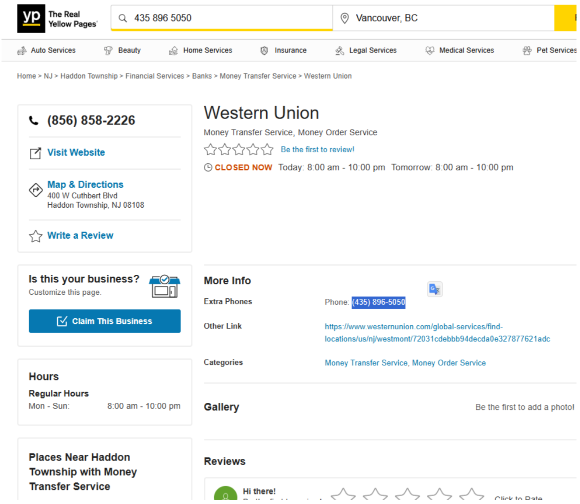
1. Western Union
Money Transfer Service, Money Order Service
Website
(435) 896-5050
35 E Flying J Dr
Richfield, UT 84701
CLOSED NOW

2. Flying J Travel Center
Truck Stops, Gas Stations, Convenience Stores
Website
(435) 896-5422
185 E Flying J Dr
Richfield, UT 84701
OPEN 24 HOURS

3. Western Union
Money Transfer Service, Money Order Service
Website
(856) 858-2226
400 W Cuthbert Blvd
Haddon Township, NJ 08108
CLOSED NOW

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):



Western Union
Money Transfer Service, Money Order Service
CLOSED NOW Today: 8:00 am - 10:00 pm Tomorrow: 8:00 am - 10:00 pm

(856) 858-2226
400 W Cuthbert Blvd
Haddon Township, NJ 08108

Is this your business?
Customize this page.

Hours
Regular Hours
Mon - Sun: 8:00 am - 10:00 pm

Places Near Haddon Township with Money Transfer Service

More Info
Extra Phones — Phone: (435) 896-5050
Other Link — https://www.westernunion.com/global-services/find-locations/us/nj/westmont/72031cdebbb94decda0e327877621adc
Categories — Money Transfer Service, Money Order Service

Gallery
Be the first to add a photo!

Reviews

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.

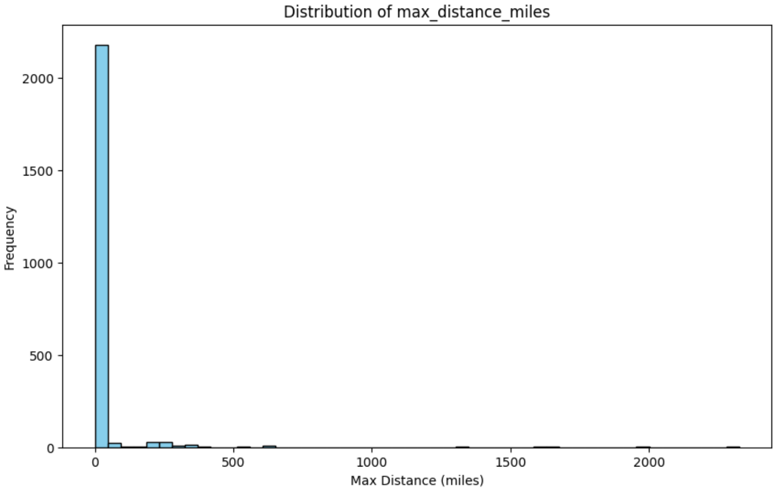


Figure 7: Frequency plot of maximum coordinate discrepancies across truck stops.

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.

# 5 Next Steps

Upon getting feedback we mentioned the strategy is to focus on

1. Redownload the data: The RVers dataset was previously downloaded without geocoordinates and needs to be re-acquired.
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.