Geocoding Truck Stops Update

William Co

2025-07-27

This update provides an analysis of geocoding discrepancies when identifying truck stops using data from Yelp and Yellow Pages. The report highlights the issues of false positives and outlines a proposed post-processing approach to filter inaccurate matches.

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 incorrect matches. Fortunately, this issue appears to be relatively small and solvable.

The complete dataset can be accessed here, which includes the latitude and longitude values used for geocoding.

2 Discrepancy Analysis

2.1 Without Yelp and Yellow Pages

If we remove the problematic Yelp and Yellow Pages entries, we observe geolocation 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 often so large—some exceeding 10 miles—that the HTML visualization becomes difficult to interpret due to the overwhelming number of false matches.

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.

3 Next Steps

To improve the accuracy of our dataset, we plan to implement a post-processing step to filter out false positives. A simple programmatic comparison of coordinate columns—similar to the method used in earlier iterations—can help us reject mismatched entries based on predefined thresholds or manual review.

This filtering should significantly improve the reliability of our geocoded truck stop dataset and better support downstream applications.