GrowthFactorAI: StoreView

Krystal Wu, Lucas Chen, Hyeokjin Jin

February 2025

1 Approach and Assumptions

Our approach aims to evaluate the visibility and traffic of a storefront using geospatial data, historical visibility metrics, traffic density information, and pedestrian density information.

Approach:

Geolocation Retrieval: Using OSM to convert addresses into latitude and longitude coordinates, and then creating Geometric coordinates.

Traffic Data Processing: Analyzing road segment data from a CSV dataset and gathering potential visibility points from a 200m radius.

Visibility Computation: Using seasonal visibility data from the Visual Crossing API to adjust the visibility rating for different seasons and location.

Obstacle Detection: Extracting nearby buildings using OpenStreetMap (OSM) and filtering obstacles that block visibility. Foot Traffic Data Processing: Using sidewalk data from a CSV dataset and determining location pedestrian density.

Assumptions:

The visibility obstruction caused by buildings is approximated using a 10cm buffer for line-of-sight calculations. Traffic volume in road segments is a valid proxy for estimating potential impressions. Historical visibility data accurately approximates seasonal visibility variations. Geospatial coordinates retrieved from OSM are sufficiently precise for storefront evaluation and location approximation.

2 Mathematical Reasoning

Traffic volume for each visible segment is estimated using:

$$T = \sum_{i=0}^{n}(\frac{x}{y} \times 1.6) \times u + (\sum_{i=0}^{n}(p \times l) \times u) \div 10$$

where: n is the number of line segments, 1.6 is the average amount of people per car, x is the trips volume column, or the number of trips observed, y is the trips sample count, or sample size for volume estimation, u is the visibility factor, p is the pedestrian density, l is the line segment length, and 10 is the pedestrian scaling factor. This impression score will vary based on seasonability, so it must

be calculated for one location four times to account for the average visibility per season.

3 Data Preprocessing Steps

Filtering Traffic Data: Extracting relevant road segments for the store's state from the CSV dataset.
Parsing Address for State Identification: Using usaddress or alternative parsing methods to filter the dataset by state.
Converting Geometries: Transforming road segment data from WKT to Shapely geometries for spatial computation.
Spatial Projection: Converting all geometries to EPSG:3857 for accurate distance calculations.

4 Real World Applications and Limitations

Applications:
Retail Site Selection: Businesses can use this model to assess locations with high foot traffic and visibility before leasing a storefront.
Billboard and Advertisement Placement: Helps advertisers determine optimal positions for visibility and exposure.
Urban Planning: City planners can use visibility and traffic data to optimize pedestrian and vehicle flows.
Limitations:
API Rate Limits: External APIs (e.g., Visual Crossing) impose request limits, which can slow down large-scale evaluations.
Data Quality: The accuracy of traffic data depends on the completeness and reliability of the dataset.
Static Obstacle Assumption: Buildings and obstacles are assumed static, without accounting for seasonal construction changes or temporary obstructions, as well as new construction plans that could come into works.
Fixed Correction Factor: The 1.4 multiplier in traffic computation is a heuristic and may need recalibration for different locations.

5 Scalability

Parallel Processing: Large-scale deployment would benefit from parallel computation using libraries such as Dask to process geospatial queries efficiently.
Database Optimization: Storing geospatial data in a PostGIS-enabled database would enhance performance for querying large datasets.
Caching Strategies: Implementing local caching of API responses would reduce redundant API calls and improve runtime efficiency.
Batch Processing for Large Areas: Instead of analyzing one location at a

time, the approach can be extended to compute traffic and visibility scores for multiple locations simultaneously.