Fine-Grained Mobility Prediction for Retail Demand using Graph Neural Networks

Partial excerpt from a Master's thesis conducted at Locatium.AI

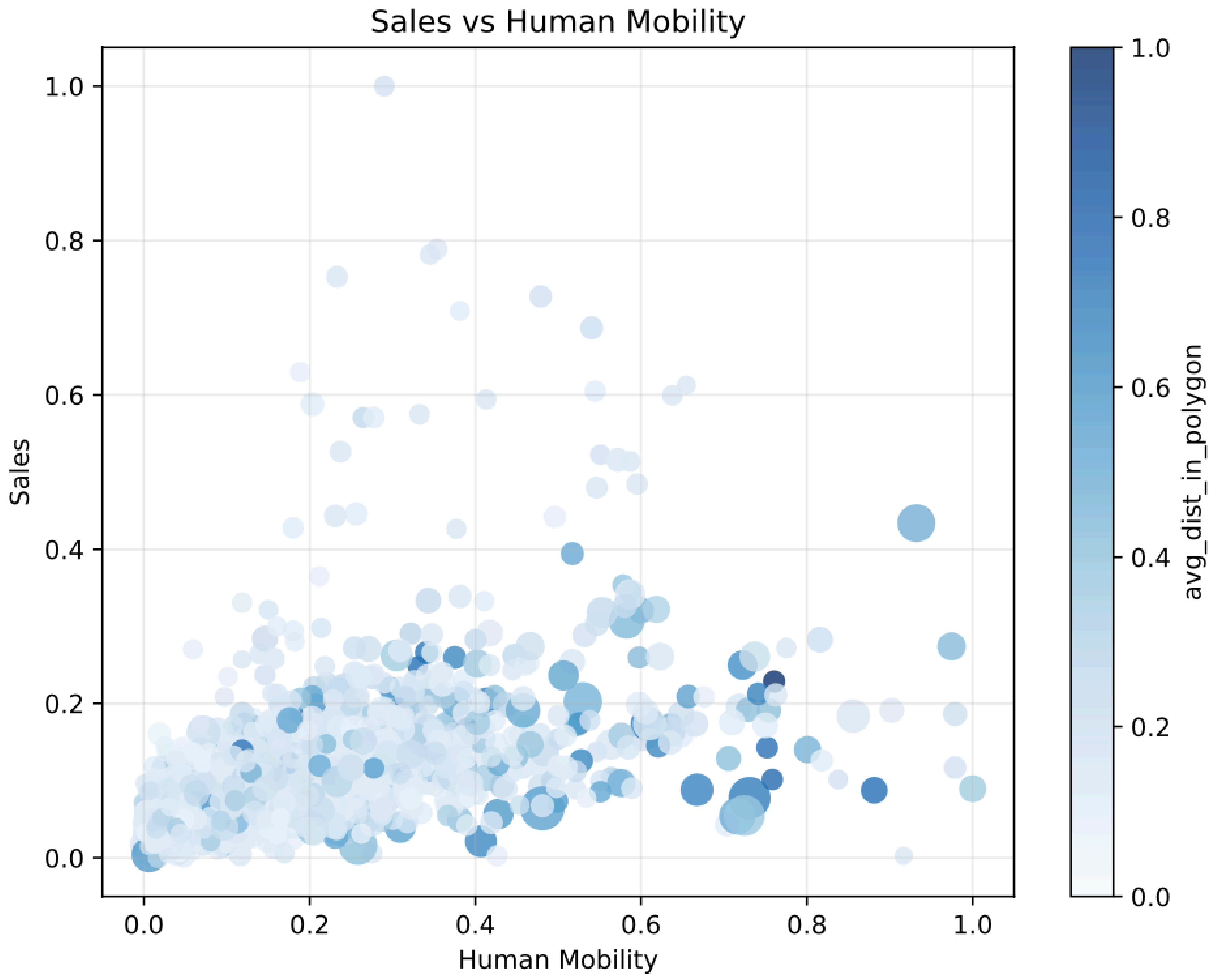
Francesco Gabriele

Thesis Supervisor: Professor Gian Antonio Susto, University of Padua

Due to internal confidentiality constraints, this represents only a limited subset of the company's full internal dataset, and the mobility data provider cannot be disclosed.

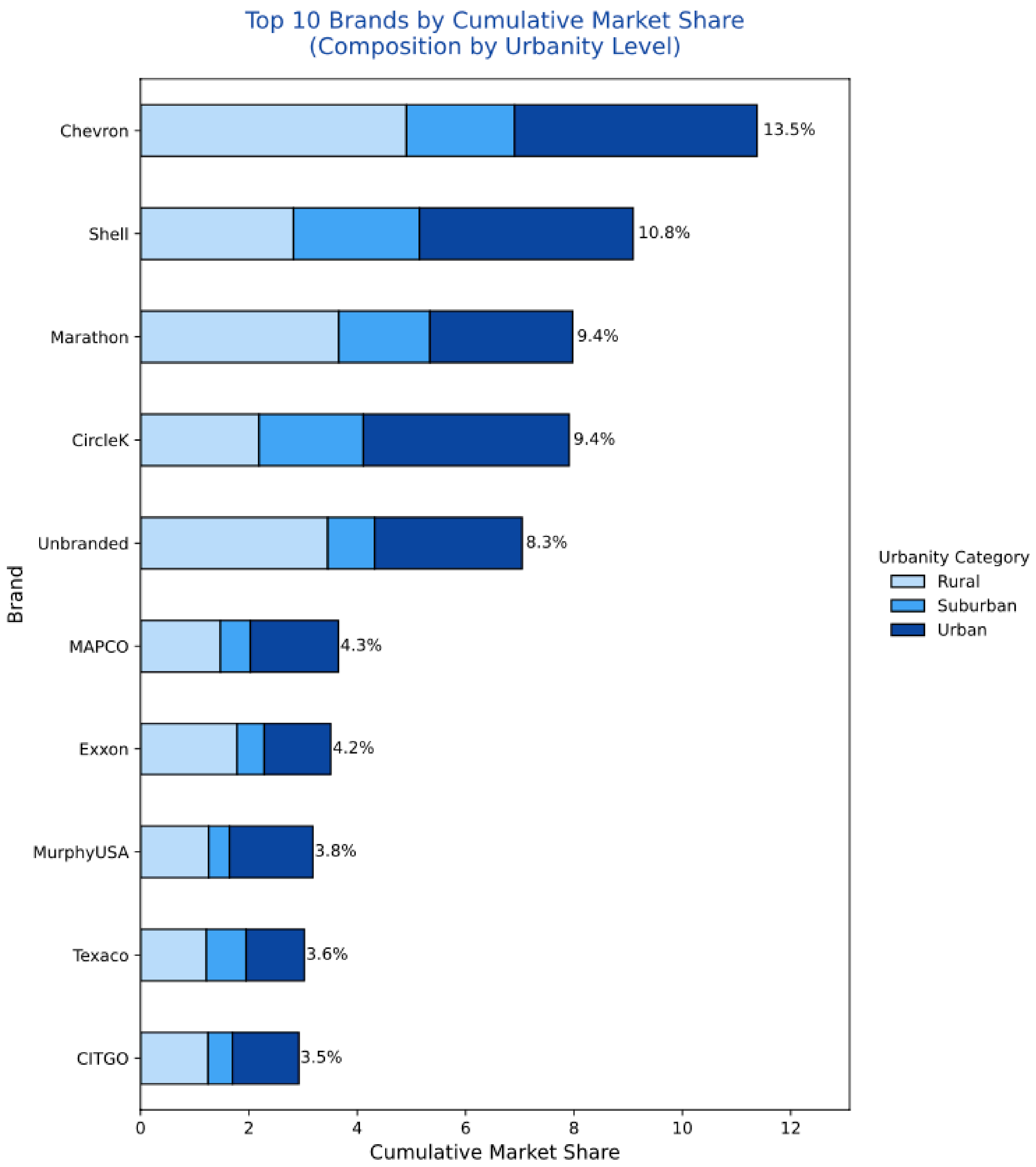
Mobility as a proxy for sales

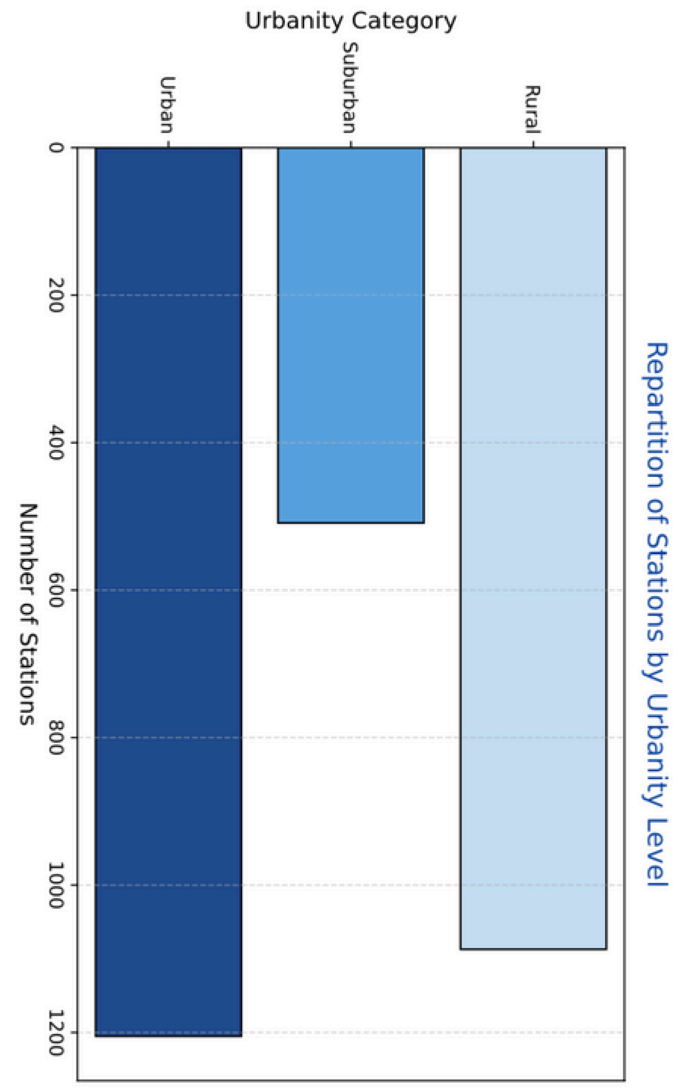
The relationship is noisy, as expected from behavioural data, but there is a clear positive association: the Pearson correlation between mobility and sales is approximately 0.7, indicating a strong linear relationship and supporting the use of mobility as a predictive feature.

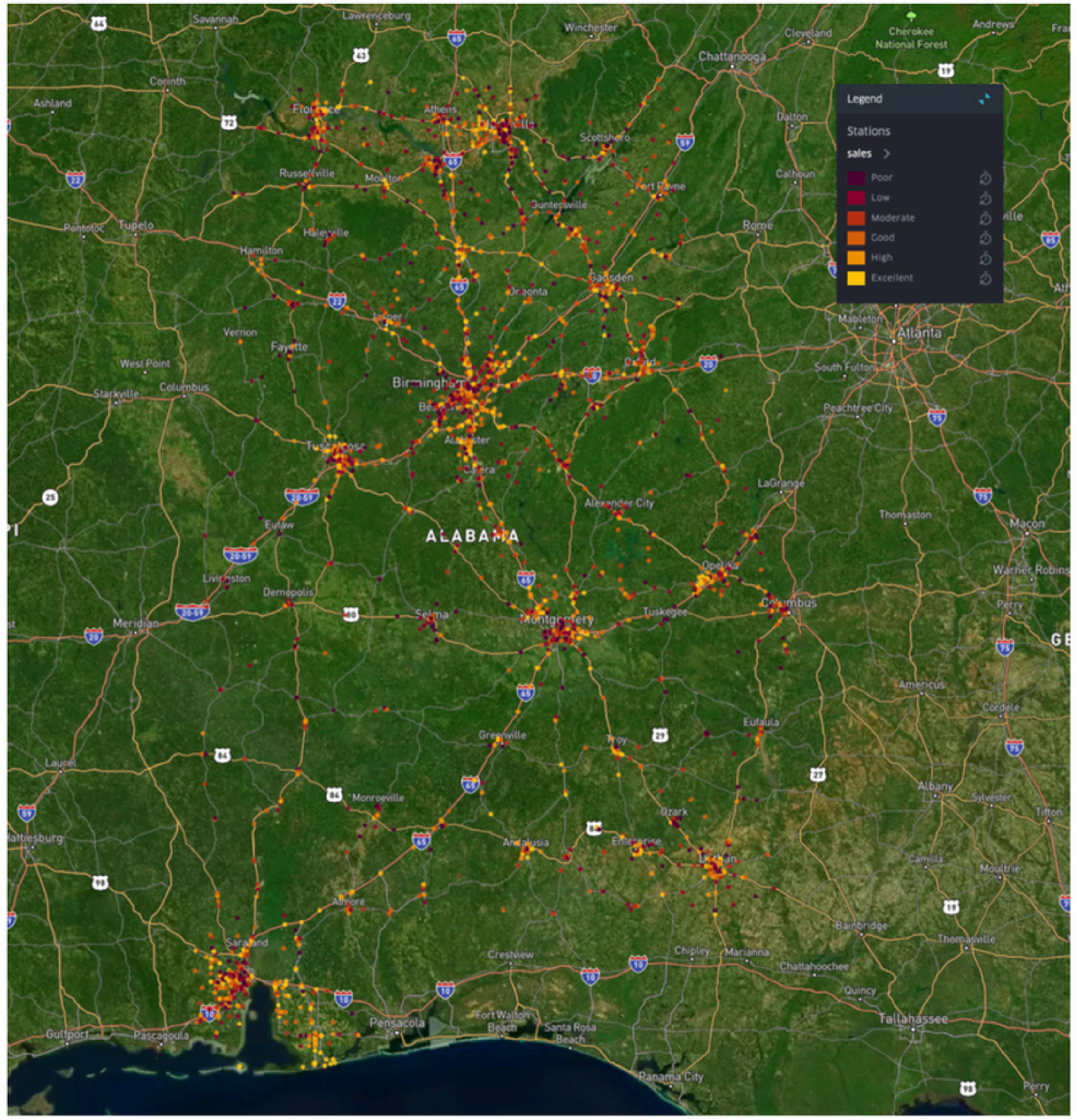


Gas Stations in Alabama

The dataset contains 2,801 fuel stations across Alabama.

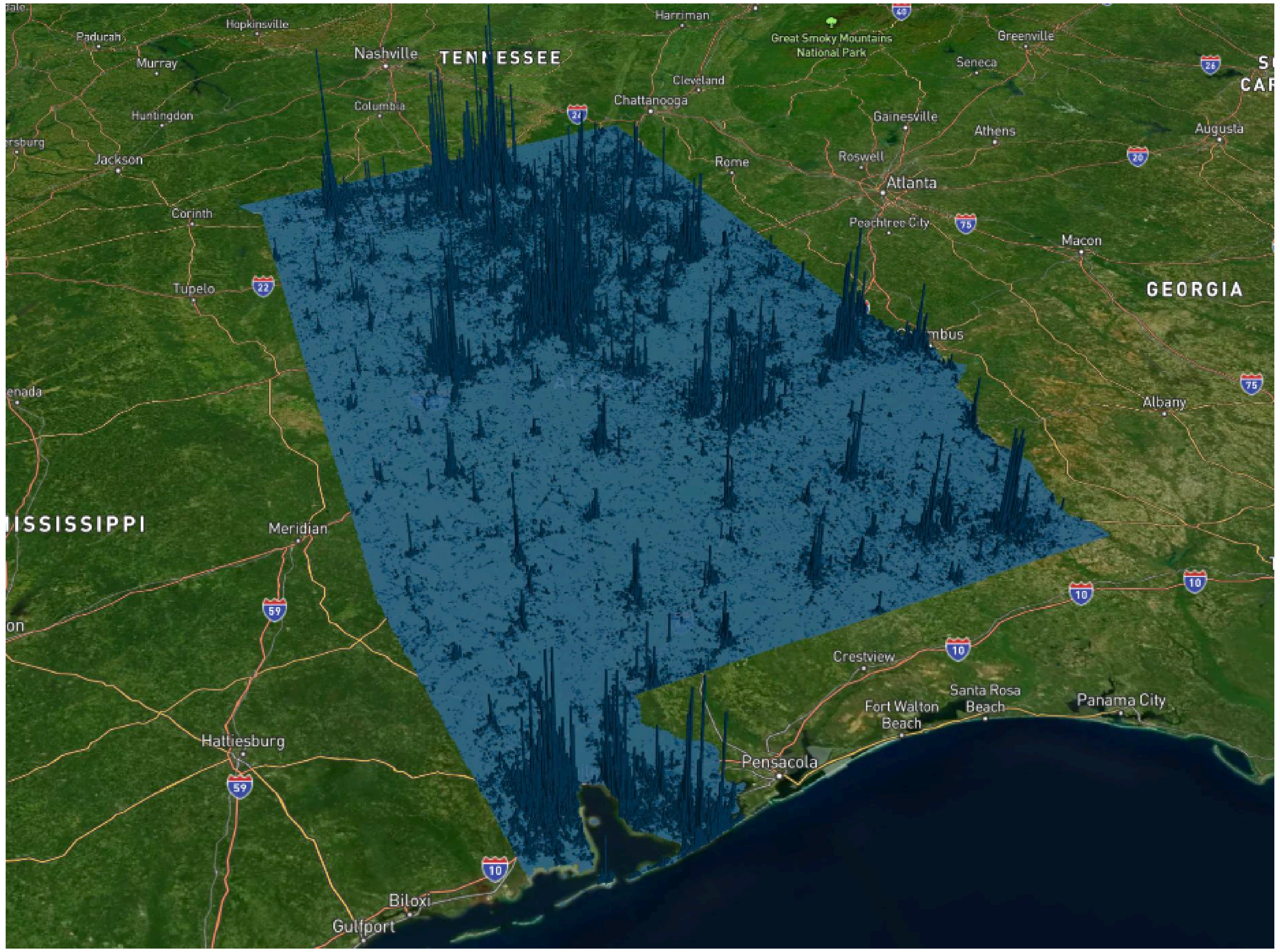






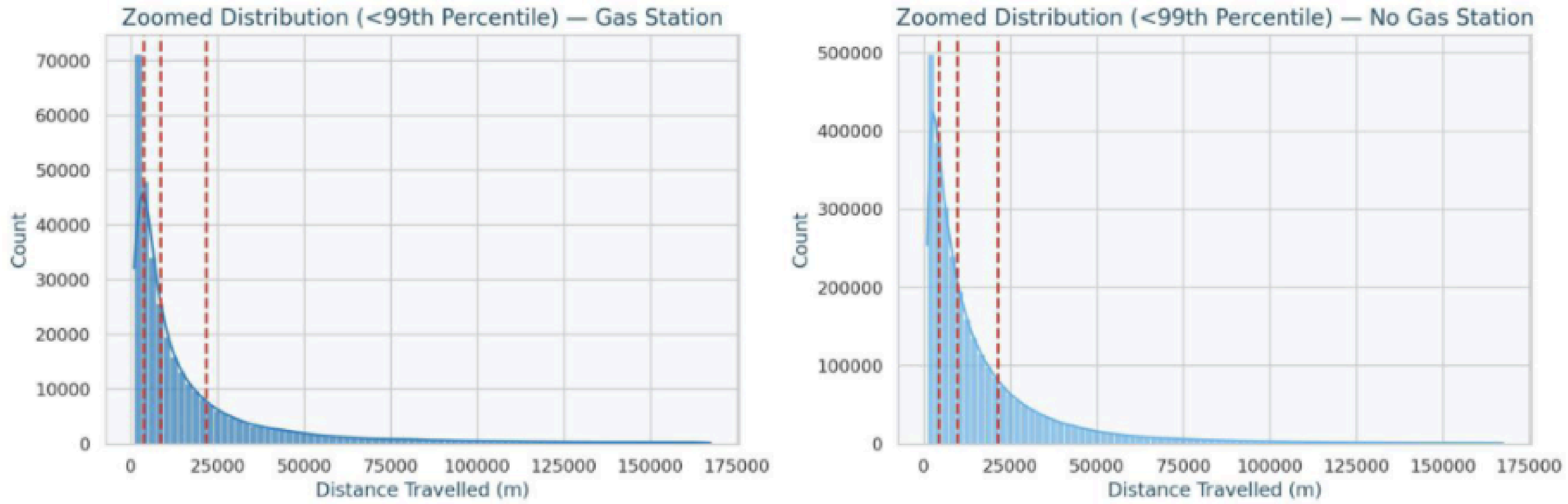
Mobility Data

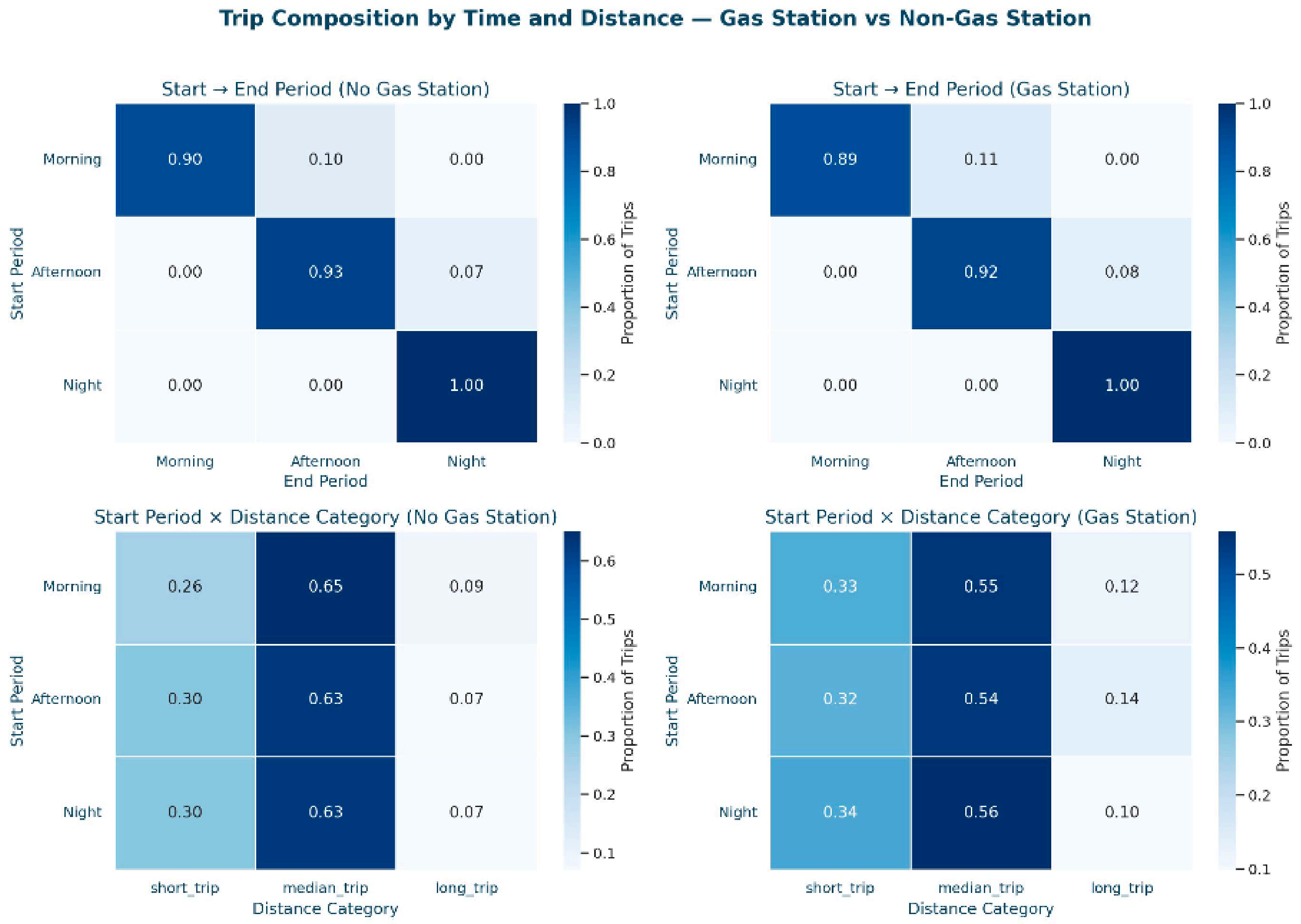
The human mobility dataset consists of car-location pings collected throughout 2024. The data is trajectory-based, but trajectories do not represent full continuous trips: a new trajectory begins every time a vehicle's engine is turned off, which fragments movement patterns.



Mobility Data

The distribution of distances traveled by trajectories that end in a gas station is similar to that of trajectories that do not.





There is also no evident difference in the timing of the trips.

Human Mobility Modelling

Review of Existing Models for Mobility Prediction

- **Zigh (Intercity Movement of Persons):** Models mobility as a function of attractors and deterrents. [2]
  - takeaway: mobility can be captured through attraction vs. resistance.

- **Gravity Model:** Predicts flows between locations proportional to their "mass" and inversely to distance [3].
  - takeaway: distance decay and mass effects are key.

$$T_{ij} = a_i O_i b_j D_j g(c_{ij}), \qquad a_i^{-1} = \sum_k b_k D_k g(c_{ik}), \quad b_j^{-1} = \sum_k a_k O_k g(c_{kj}).$$

- **Radiation Model:** Considers flows based on intervening opportunities between origin and destination [4].
  - takeaway: intermediate locations influence mobility.

$$T_{ij} = \frac{O_i}{1 - m_i/M} \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}.$$

- **Deep Gravity Model:** Uses a neural network to learn spatial interactions from data [5].
  - takeaway: contextual training improves predictive accuracy.

$$s_{ij} = \mathrm{MLP}_\theta\big(\mathbf{z}_i \,\|\, \mathbf{z}_j \,\|\, \phi(d_{ij})\big) \qquad \hat{y}_{ij} = \frac{\exp(s_{ij})}{\sum_{k=1}^{n} \exp(s_{ik})}, \qquad \sum_{j=1}^{n} \hat{y}_{ij} = 1 \qquad \widehat{\mathrm{Flow}}_{ij} = O_i \, \hat{y}_{ij}$$

Human Mobility Modelling

Motivation for a Graph-Based Approach

Building on insights from the most relevant literature, an effective model for predicting mobility at a station should combine attraction-like features, contextual information from neighboring stations, and strong predictive power.

This motivates a neural network that leverages spatial context, naturally leading to a graph-based approach where station relationships are encoded and a readout layer generates the final predictions.

Challenges of Fine-Scale Mobility Prediction

Predicting flux at the store level cannot rely on traditional origin-destination approaches due to the fine granularity. Classical models work at larger scales between administrative units, whereas we target mobility to individual gas station polygons.

At this scale, trajectories ending at a station are hard to distinguish by length, speed, or timing, and aggregating fluxes in 1 km^2 grids shows little correlation with distance, exposing the limits of conventional OD methods.

Human Mobility Modelling

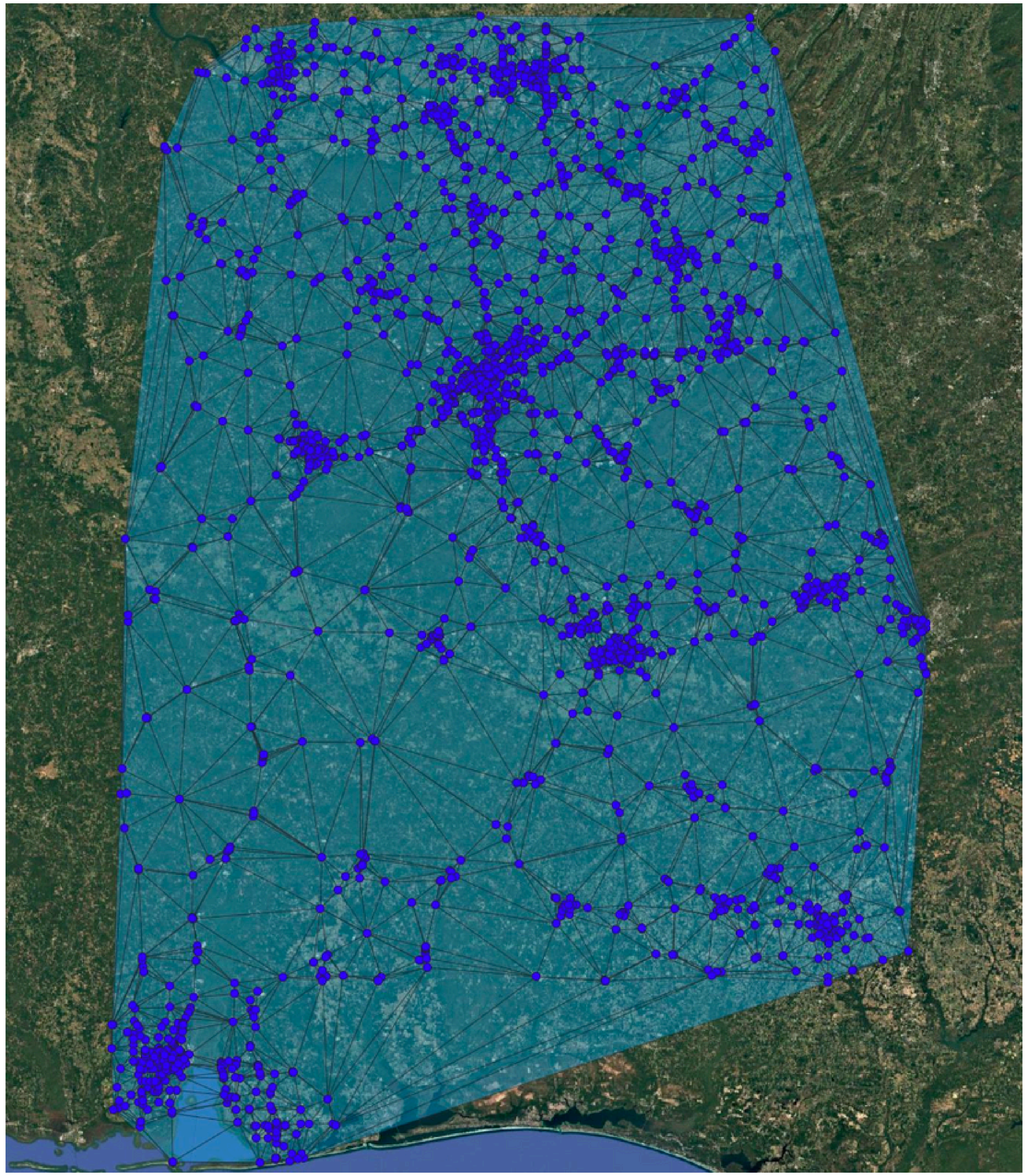
Gas station features

- **Parcel geometry:** area, orientation, rectangularity, elongation
- **Street-network context:** number of neighboring streets, network density, average road level, degree of urbanity
- **Spatial context:** gravity pull of neighboring cities (neighbor_city_gravity_d25000_e15), mean distance to k-nearest neighbors (mean_knn_dist)

Graph Construction

We construct the spatial graph using Delaunay triangulation, chosen for its geometric fidelity and practical interpretability.

- **Competition:** Delaunay triangulation produces a planar, sparse graph that naturally connects stations likely to compete locally while keeping computational complexity low.
- **Edge Weight:** Weighted edges encode interaction strength via distance-decay, ensuring closer stations exert stronger influence than distant ones.
- **Well known:** The method is widely used in geospatial and urban analytics, making the resulting graph intuitive and easy for practitioners to interpret.



Human Mobility Modelling

Graph-Convolutional Architectures for Gas Station Mobility Modeling

We implement two complementary graph-based models to predict station-level mobility:

Graph Convolutional Network (GCN)[6]:

- Every neighbor contributes with fixed importance determined purely by graph structure.
- If edge weights exist (e.g., your distance kernel), they modulate the aggregation but they do not allow the model to learn how important each neighbor is.

$$\mathbf{X}' = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-1/2}\mathbf{X}\Theta, \quad \alpha_{j,i} = \frac{1}{\sqrt{\hat{d}_j \hat{d}_i}}, \qquad \hat{d}_i = 1 + \sum_{j \in \mathcal{N}(i)} e_{j,i}$$

$$\mathbf{x}'_i = \Theta^\top \sum_{j \in \mathcal{N}(i) \cup \{i\}} \alpha_{j,i}\, e_{j,i}\, \mathbf{x}_j.$$

Graph Attention Network (GAT)[7][8]:

- Each neighbor gets a learned importance weight.
- Weights depend on its features, the target node's features, AND the edge features.
- Neighbors can be emphasized or ignored dynamically.

$$\mathbf{x}'_i = \sum_{j \in \mathcal{N}(i) \cup \{i\}} \alpha_{i,j}\, \Theta \mathbf{x}_j,$$

$$\alpha_{i,j} = \frac{\exp\left(\mathbf{a}^\top \mathrm{LeakyReLU}\left(\Theta_s \mathbf{x}_i + \Theta_t \mathbf{x}_j + \Theta_e \mathbf{e}_{i,j}\right)\right)}{\sum_{k \in \mathcal{N}(i) \cup \{i\}} \exp\left(\mathbf{a}^\top \mathrm{LeakyReLU}\left(\Theta_s \mathbf{x}_i + \Theta_t \mathbf{x}_k + \Theta_e \mathbf{e}_{i,k}\right)\right)}.$$

A final linear readout layer maps the last hidden representation to per-node predictions:

$$\hat{\mathbf{y}} = \mathbf{X}^{(L)} \mathbf{w}_{\text{out}} + b_{\text{out}}$$

GCN vs. GAT: Training and Test Performance

Graph Convolutional Network (GCN):

Model	Train R^2	Val R^2	Test R^2
Shallow 64, ReLU	0.150	0.015	0.014
Shallow 128, ELU	0.121	0.031	0.018
Medium 128–64, ReLU + Res	0.412	0.334	0.240
Medium 128–64, GELU	0.185	0.074	0.064
Deep 128–128–64, ReLU + Res	0.435	0.350	0.242
Deep 128–128–64, ELU	0.149	0.050	0.030
Wide 256–128, SELU	0.109	0.062	0.016
Compact 64–32, ReLU	0.152	0.072	0.058

Metric	Value
R^2_{overall}	≈ 0.35
RMSE	≈ 32
MAE	≈ 21

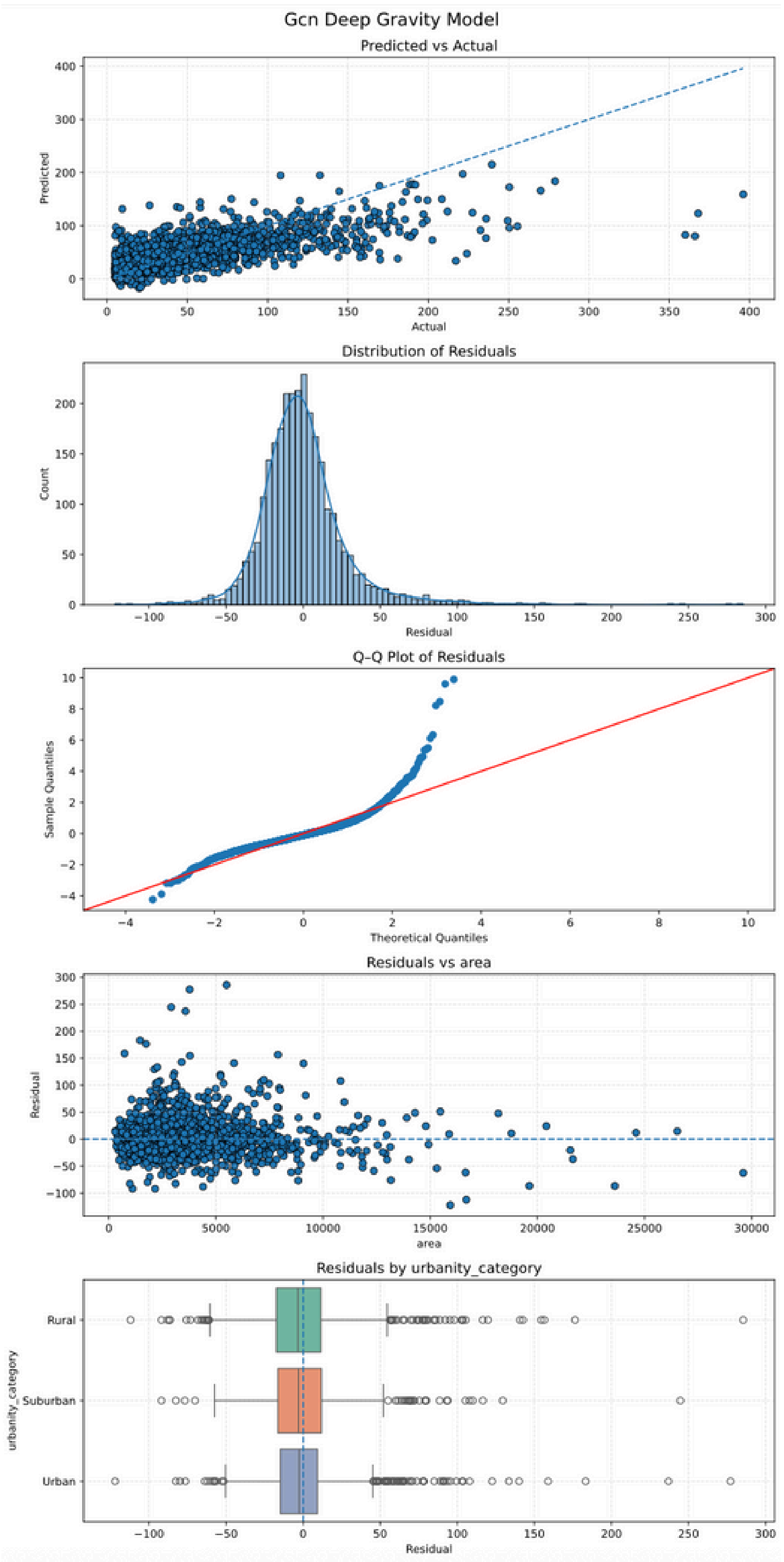
Graph Attention Network (GAT):

Model	Train R^2	Val R^2	Test R^2
Shallow 64, ReLU	0.382	0.339	0.244
Shallow 128, ELU	0.551	0.378	0.263
Medium 128–64, ReLU	0.493	0.368	0.262
Medium 128–64, GELU	0.448	0.374	0.260
Deep 128–128–64, ReLU	0.428	0.389	0.267
Deep 128–128–64, ELU	0.512	0.371	0.235
Wide 256–128, SELU	0.342	0.345	0.241
Compact 64–32, ReLU	0.508	0.348	0.236
Compact 64–32, GELU	0.443	0.338	0.243
Super-wide 512–256–128	0.616	0.380	0.236

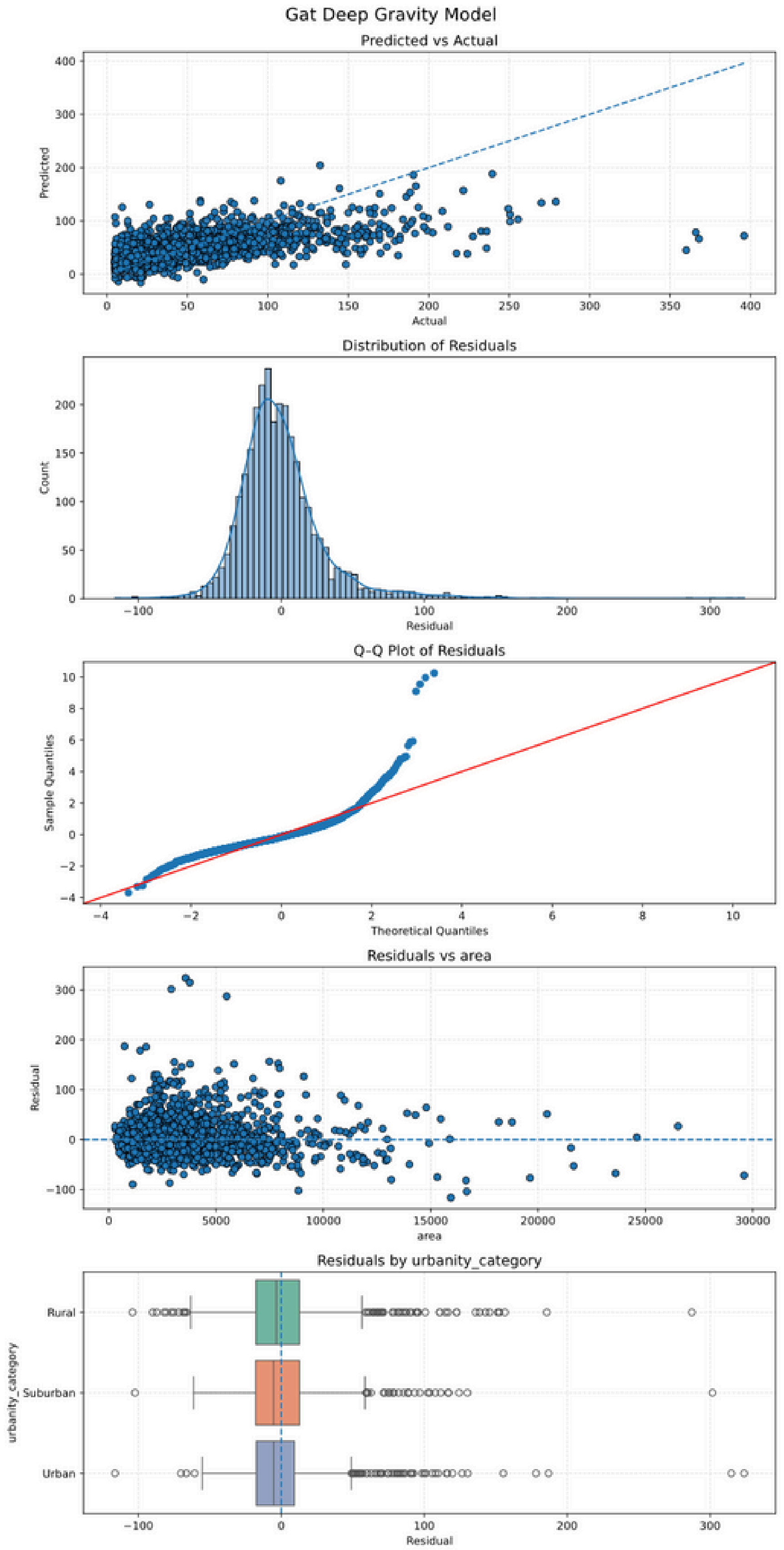
Metric	Value
R^2_{overall}	≈ 0.39
RMSE	≈ 28
MAE	≈ 19

GCN vs. GAT: Residuals

Graph Convolutional Network (GCN):

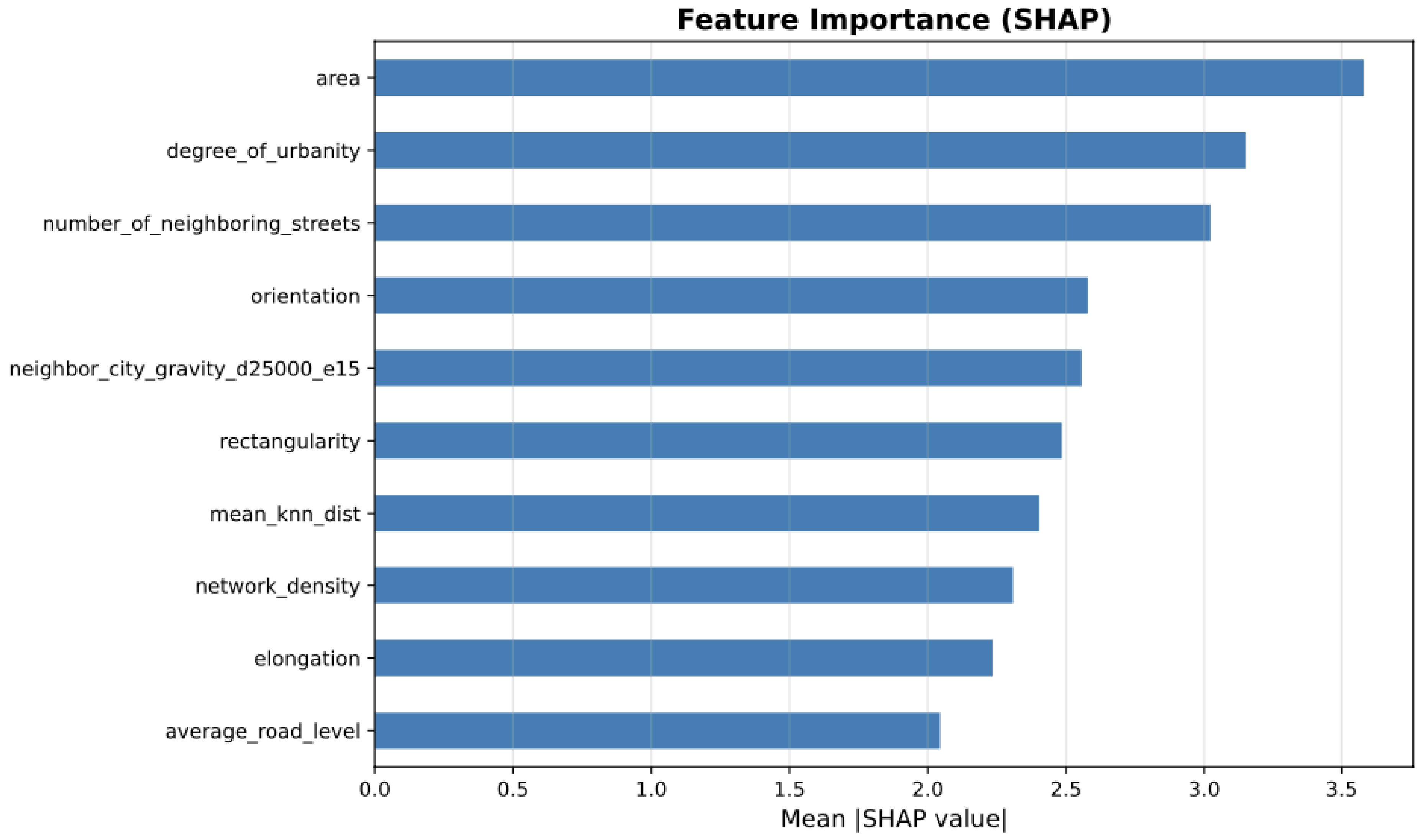


Graph Attention Network (GAT):

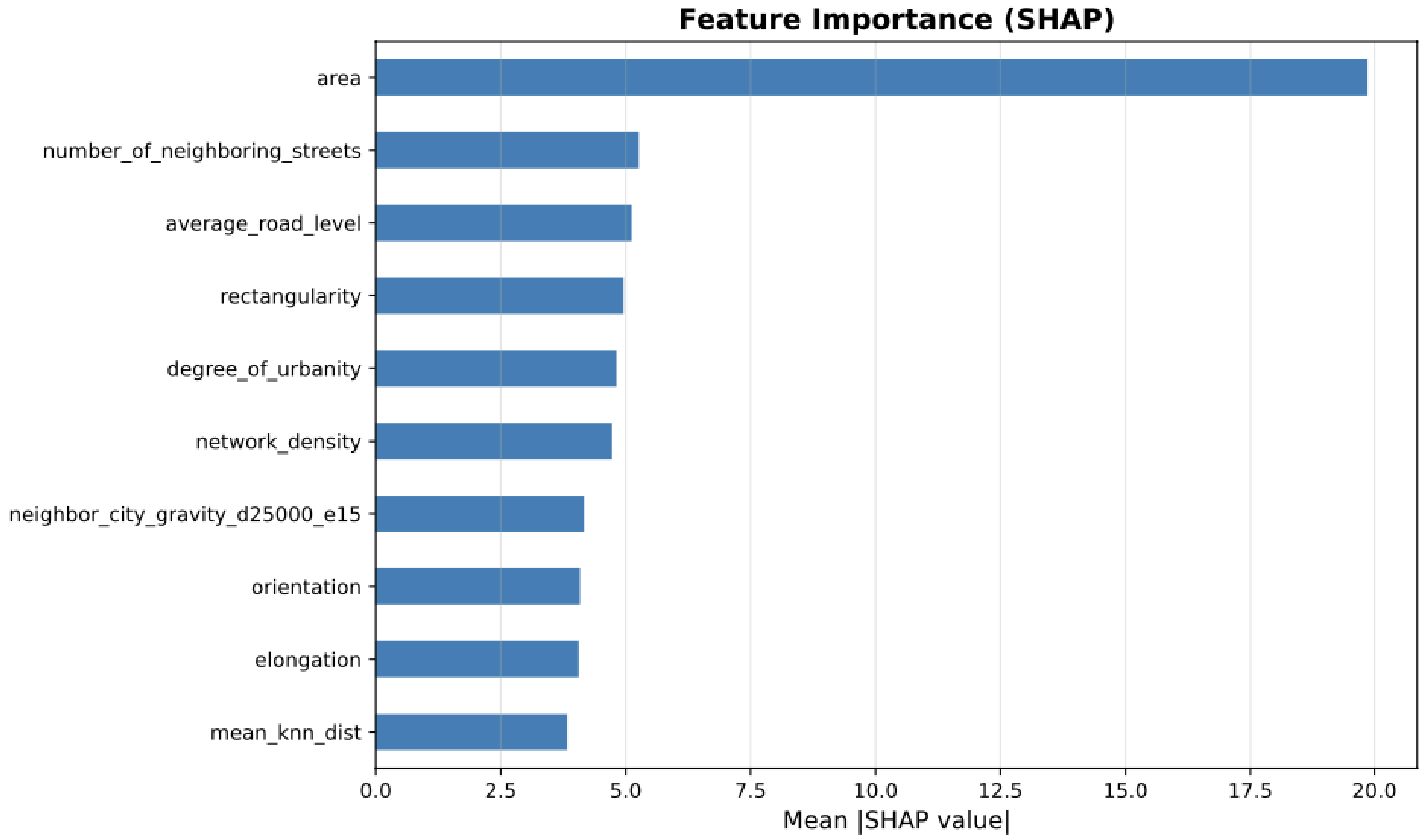


GCN vs. GAT: SHAP values

Graph Convolutional Network (GCN):



Graph Attention Network (GAT):



Future Work

- **Graph type:** Explore alternative graph constructions, including road-network-based, distance-weighted, or traffic-aware connectivity schemes.
- **More features:** Enrich the set of fuel-station features with additional spatial, demographic, economic, and mobility-derived descriptors.
- **More mobility:** Incorporate mobility data from multiple years to capture seasonal variation and long-term behavioral patterns.
- **Driver profiling:** Characterize different classes of users based on home location, trip distribution, and other socio-economic attributes to better understand demand heterogeneity.

Bibliography

1. George Kingsley Zipf. The p 1 p 2/d hypothesis: on the intercity movement of persons. American sociological review, 11(6):677–686, 1946.
2. Alan Geoffrey Wilson. A family of spatial interaction models, and associated developments. Environment and Planning A, 3(1):1–32, 1971.
3. A Paolo Masucci, Joan Serras, Anders Johansson, and Michael Batty. Gravity versus radi- ation models: On the importance of scale and heterogeneity in commuting flows. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 88(2):022812, 2013.
4. Filippo Simini, Gianni Barlacchi, Massimilano Luca, and Luca Pappalardo. A deep gravity model for mobility flows generation. Nature communications, 12(1):6576, 2021.
5. TN Kipf. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.
6. Petar Veliˇckovi´c, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. arXiv preprint arXiv:1710.10903, 2017.
7. Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? arXiv preprint arXiv:2105.14491, 2021.