Spatial Analysis of Train Demand in Greater Melbourne

For GEOM90006 Assignment 4

Lang Chen, Yihan Bao

Research Question and Motivation

Research Question

1. What is the quantitative impact of weather, demographic and non-residential buildings on train demand in Greater Melbourne
2. What locations in Greater Melbourne currently do not have a train station but have high predicted demand?

Motivation: Better public transport promotes sustainability and liveability

- Both team members frequent train travellers, who are curious about what impacts demand at different stations
- Curious about priority of City Tunnel over other potential stations in Melbourne

Data Source

Dataset	Source
Train stop location data and passenger count data	https://discover.data.vic.gov.au/dataset/train-service-passenger-counts
Census Data	https://www.abs.gov.au/census/
Non-residential Facilities Data	- Shopping Centres: https://www.australia-shoppings.com/malls-centres/victoria - Hospitals: https://springernature.figshare.com - Sports Facilities: https://discover.data.vic.gov.au - Schools: https://www.education.vic.gov.au
Rainfall Data	Bureau of Meteorology
SA2 / Greater Capital City Boundary Data	https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/access-and-downloads/digital-boundary-files

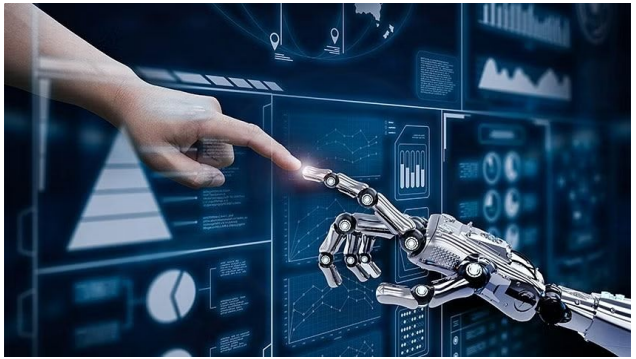
Feature Hypothesis and Representation

Feature	Impact hypothesis	Feature Representation
Mean Rainfall in SA2	Neg - travellers less inclined to use public transport in rain	*(Self explained)*
Census Attributes - Median/Mean	Neg [Wealth related] - those in economically better off would prefer private transport	$V_{\text{Station}} = \sum_{\text{SA2s}} \left(\frac{A_{overlap}}{A_{\text{station ANN-radius proximity}}} \times V_{\text{SA2}} \right)$
Census Attributes - Total	Pos [Total Population] - more potential customers	$V_{\text{Station}} = \sum_{\text{SA2s}} \left(\frac{A_{overlap}}{A_{\text{SA2}}} \times V_{\text{SA2}} \right)$
Existence of Non-Residential Buildings	Pos or insignificant - these buildings may draw people to travel to region	$I_{\text{Building}} = \begin{cases} 1 & \text{if building(s) exist in } \frac{1}{2} \text{ station ANN-radius proximity} \\ 0 & \text{otherwise} \end{cases}$ *(*ANN distance = 1.4 km)*
Weekday	Pos - more travellers for work related purposes	$I_{\text{Weekday}} = \begin{cases} 1 & \text{if weekday} \\ 0 & \text{otherwise} \end{cases}$
Weighted Station Demand	Inconclusive - positive as it should correlate with population but nearby stations could take demand away from each other	$w_{ij} = \begin{cases} 0 & \text{if same station} \\ \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right) & \text{otherwise} \end{cases}$

Model (½): Regression with Weighted Geospatial Features

- Through pre-weighting data, transforms spatial data into ***Tabular Data***

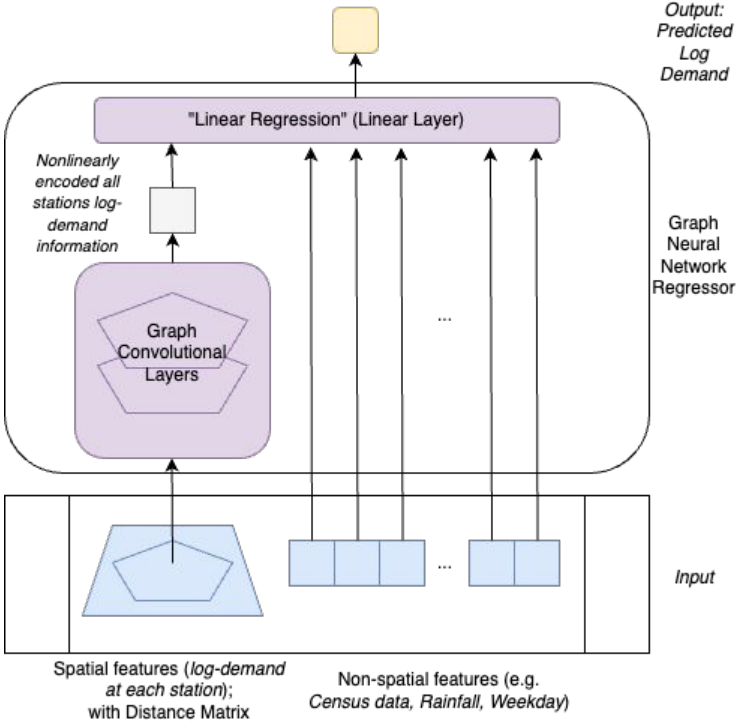
- Allows use of Machine Learning models such as ***Linear Regression***, ***Random Forests*** and ***Boosting Machines*** for accurate prediction



- Using weighted geospatial features (demand) with regular regression serves similar effect to Geographically Weighted Regression

- One model instead of one model per location - better sharing of parameters

Model (2/2): Attention-Graph Neural Network



- Graph Neural Network natural for spatial data
- Better encoder to non-linearly represent spatial information
- ***Attention Mechanism*** in graph automatically mines for relationships between points (stations)

- More sophisticated than Linear Regression at factor analysis
- Not good for inference on new points (Exogenous prediction)

Evaluation (⅕): Model Performance

Model	**Out-of-Sample R2**
Linear Regression	0.3495
Support Vector Machine	0.5697
Random Forest	0.7901
Light Gradient Boosting (LGB)	**0.8624**
Graph Neural Network (GNN)	**0.9952**

Note: LGB is used for forecasting as GNN is only applicable for feature inference

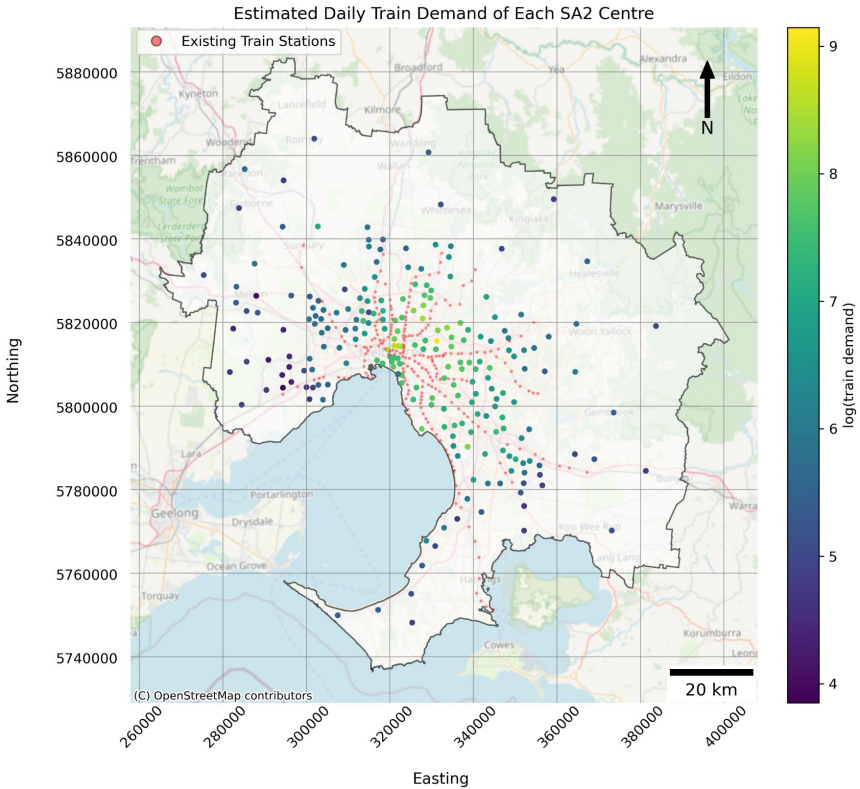
Evaluation (⅖): Feature Analysis

Feature	**Linear Interpretation**	**GNN Interpretation**
Weighted sum demand of other stations	N/A	1% increase in factor=> 0.3% demand inc of this station
Weekday	Weekday has 3.9% less demand than weekends	Weekday has 0.85% more demand than weekends
School	Stations with schools nearby has 70% demand increase	Stations with schools nearby has 0.03% demand increase
Hospital	Stations with hospitals nearby has 41% demand increase	Stations with hospitals nearby has 0.25% demand increase
Sports Facilities	Stations with sports facilities nearby has 6% demand decrease	Stations with sports facilities nearby has 0.2% demand increase
Shopping Centres	Stations with shopping centres nearby has 49% demand increase	Stations with shopping centres nearby has 1.21% demand increase

Evaluation (⅗): Feature Analysis cont.

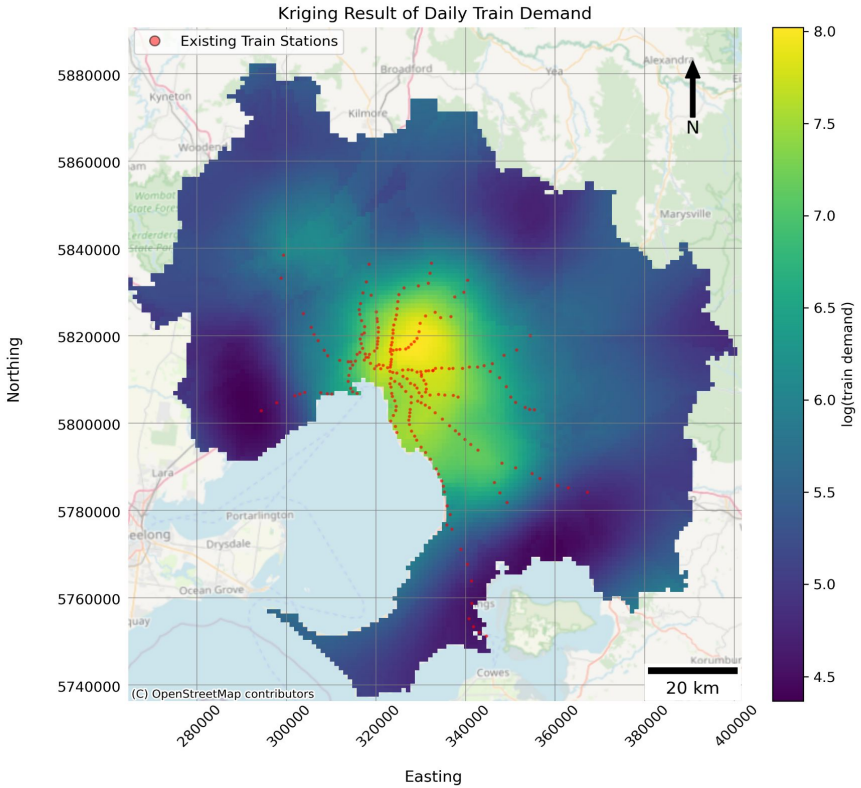
Feature	**Linear Regression**	**GNN**
Mean Rainfall	1 standard deviation (sd) increase in median annual rainfall => 57% decrease in demand	1sd increase => 0.23% demand decrease
Total Population	1sd increase of total nearby population => 48.6% increase in demand	1sd increase => 2.18% demand increase
Median weekly rent	1sd increase of median nearby weekly rent => 41% decrease in demand	1sd increase => 0.27% demand decrease
Median weekly mortgage	1sd increase of median nearby weekly mortgage => 41% increase in demand	1sd increase of median nearby weekly mortgage => 0.74% demand decrease
Median weekly income	1sd increase of median nearby weekly income => 57% increase in demand	1sd increase of median nearby weekly income => 0.47% demand decrease
Median weekly household income	1sd increase of median nearby weekly household income => 77% decrease in demand	1sd increase of median nearby weekly household income => 1.09% demand increase

Evaluation (⅘): Point Estimates



Ranking	SA2	Predicted Daily Demand
1	Balwyn North	9331
2	Carlton	5717
3	Fitzroy	5452
4	West Melbourne	4673
5	Heidelberg West	4519

Evaluation (⅘): Kriging



- East of Melbourne City needs additional train stations
- Some train lines should extend further

Evaluation (5/5): Future Directions

- Research GNN’s encoding ability for prediction
- Spatial-temporal forecasting of daily train demand
- Empirical factors when recommending new stations

References

1. Fotheringham, A. S., Oshan, T. M., & Li, Z. (2023). *Multiscale geographically weighted regression: Theory and practice* (1st ed.). CRC Press.

2. Selby, B., & Kockelman, K. M. (2013). Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression. *Journal of Transport Geography*, 29, 24–32.

3. Song, T., Pu, H., Schonfeld, P., Zhang, H., Li, W., Peng, X., . . . Liu, W. (2021). Gis-based multi-criteria railway design with spatial environmental considerations. *Applied Geography*, 131, 102449.