UCL DEPARTMENT OF GEOGRAPHY

YEAR 2024-25

EXAM CANDIDATE ID:	**LTKB8**
MODULE CODE:	**GEOG0051**
MODULE NAME:	**Mining Social and Geographic Datasets**
COURSE PAPER TITLE:	**Human Mobility and Text-Based Spatial Analysis: Case of Cambridge, UK and Calgary, Canada**
WORD COUNT:	**1499 (excluding tables & figures contents and references)**
CODE REPOSITORY LINK:	**https://drive.google.com/drive/folders/1hlqKbv9tOPwq9RKZbZITdBRDr6X2jEZU?usp=sharing**

I acknowledge the use of Artificial Intelligence (AI), specifically ChatGPT (https://chatgpt.com/), to support both writing refinement and code troubleshooting in the completion of this work. I used ChatGPT to enhance the clarity and conciseness of my writing after drafting initial ideas, while ensuring that the core message and intent remained intact. Additionally, I utilized ChatGPT to help troubleshoot specific coding issues. For example, when I encountered difficulties adding a proposed restaurant location to the centrality map, I used the prompt: *"I want to add this point '51°03'25.1"N 114°03'47.9"W' as my proposed location for the new restaurant."* To ensure accuracy, I manually verified the coordinates using Google Maps. All ChatGPT-assisted code sections were clearly credited within the submitted Python script. My use of ChatGPT was limited to improving clarity and resolving technical challenges, with all outputs carefully reviewed for correctness and transparency.

Your essay, appropriately anonymised, may be used to help future students prepare for assessment. Double click this box to opt out of this ☐

1. Overall Introduction

Many contemporary datasets include links between geographic locations and natural language texts, commonly referred to as geotext data. The rise of social media platforms has significantly contributed to the growth of this data type, as users often share content tagged with location information. Such data offers valuable insights into human mobility patterns as well as public perceptions and opinions about specific places. These insights can inform urban planning efforts, contributing to more effective infrastructure design, equitable resource distribution, and overall improvements in city development (Hu, 2018; Carvalho, Ferreira & Dias, 2021; Zhang et al., 2021). In this study, the author contributes by outlining techniques for analyzing mobility and textual data, using mobility datasets from Gowalla in Cambridge and amenity review texts from Calgary, Canada, to inform future research.

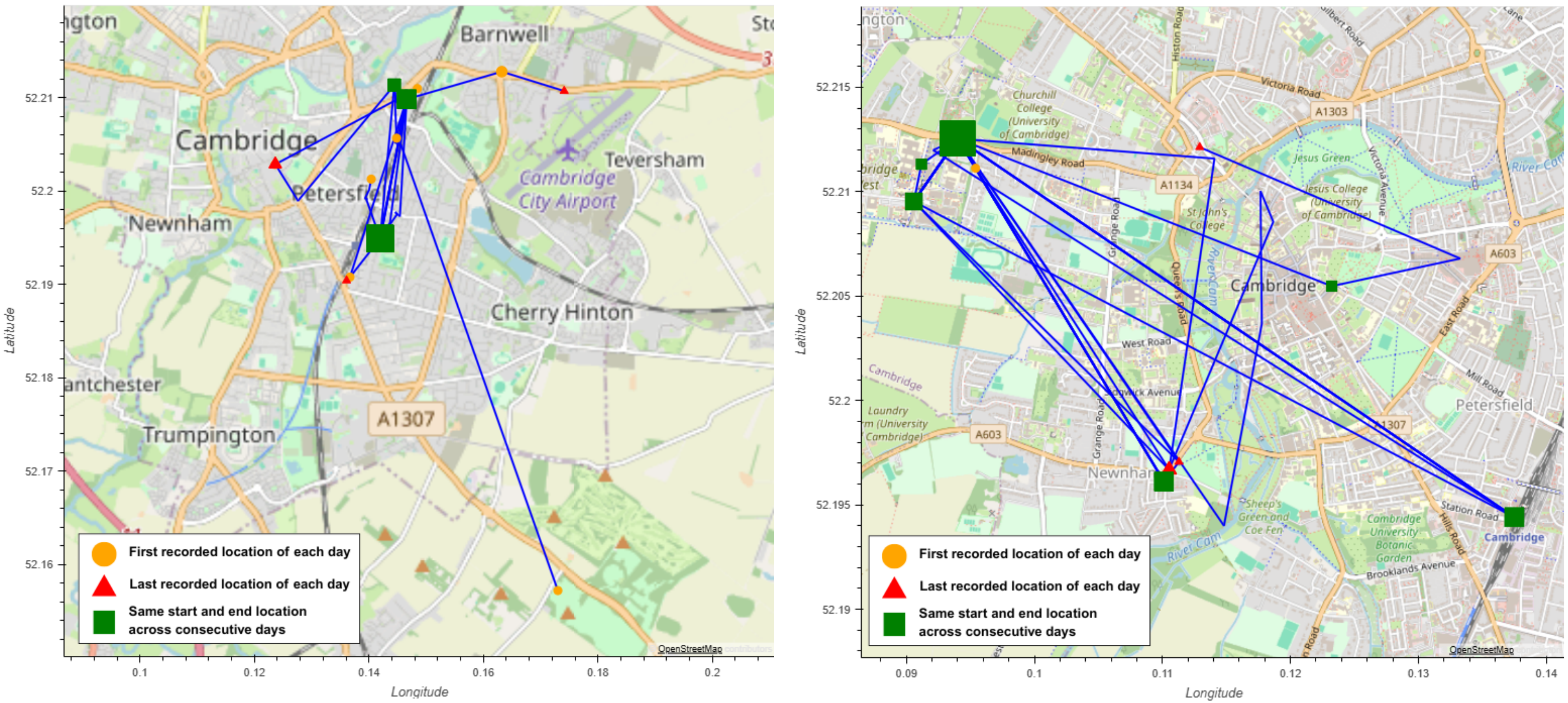
2. Mobility Analysis

2.1. Introduction

In this first part of the analysis, the author uses the 2010 Cambridge individual mobility dataset from Gowalla to analyze individual mobility patterns and characterize check-in locations. The study also proposes a new amenity, such as a community park, based on insights derived from the Gowalla dataset and OpenStreetMap data.

2.2. Analysis

For the mobility analysis, this study focuses on two specific user IDs: 75027 and 102829. Figure 1 illustrates their movement patterns between April 17 and May 27, 2010. In the visualizations, orange circles denote the first recorded location of each day, red triangles indicate the last, and green squares indicate instances where the final location of one day matches the first location of the following day. Marker size reflects the frequency of recorded coordinates, with green squares potentially identifying areas of daily activities or the users' residence. Following a review of overall movement patterns, the analysis narrows to two specific trips: user 75027 on January 30, 2010, and user 102829 on May 24, 2010. For each trip, the shortest path between recorded coordinates was assessed, along with calculations of maximum, average, and total displacement.



(a) Mobility of User 75027

(b) Mobility of User 102829

Figure 1: Users' Mobility between April 17 and May 27, 2010.

Prior to analysis, relevant data cleaning procedures were applied to ensure accuracy. Outlier points were removed where user speed exceeded 100 km/h or was more than 1.5 standard deviations above the median trajectory speed (Law et al., 2025). The next step involves point node simplification. Initially, the Douglas–Peucker algorithm will be applied, which simplifies polylines by merging points that fall within a specified threshold, which is 0.001 degrees in this case (Law et al., 2025).

However, because the Douglas–Peucker algorithm prioritizes line simplification, it often results in less efficient paths, leading to longer travel distances despite the cleaner overall shape. To explore alternatives, the author compared this method with point node simplification using DBSCAN, a clustering algorithm frequently applied in trajectory analysis (Bai et al., 2023). The results showed that DBSCAN produced routes with shorter total distances based on driving network analysis. Table 1 presents a comparison of the Douglas–Peucker algorithm and DBSCAN, focusing on user 75027, for whom the two methods produced different route distances. In contrast, no distance difference was observed between the two algorithms for user 102829. Moreover, the paths generated by DBSCAN followed a more consistent direction toward the target coordinates, unlike the Douglas–Peucker algorithm, which sometimes required backtracking along the same route, as shown in Figure 2.

Algorithm	**User**	**Total Distance (Meters)**	**Average Displacement (Meters)**	**Maximum Displacement (Meters)**
Douglas–Peucker	75027	12,690.28	2,115.05	7,583.29
DBSCAN	75027	12,139.56	2,023.26	6,781.73
Douglas–Peucker	102829	15,603.39	2,229.06	5,023.12
DBSCAN	102829	15,603.39	2,229.06	5,023.12

Table 1: Mobility Summary.

(a) Shortest Path Route of User 75027 on 30/01/2010 (Douglas-Peucker)

(b) Shortest Path Route of User 75027 on 30/01/2010 (DBSCAN)

(c) Shortest Path Route of User 102829 on 24/05/2010 (Douglas-Peucker & DBSCAN)

Figure 2: Users' Shortest Path Routes.

Using the Gowalla dataset, the author also attempted to propose the development of a new community park in Cambridge. This recommendation aligns with the ongoing initiatives of the Cambridge City Council, which is actively working to enhance green spaces across the city (Greater Cambridge Planning Service, 2021). A key motivation for this initiative is that green infrastructure not only contributes to Cambridge's natural landscape identity but also offers significant economic benefits to the city (Cohen, Burrowes, & Gwam, 2022).

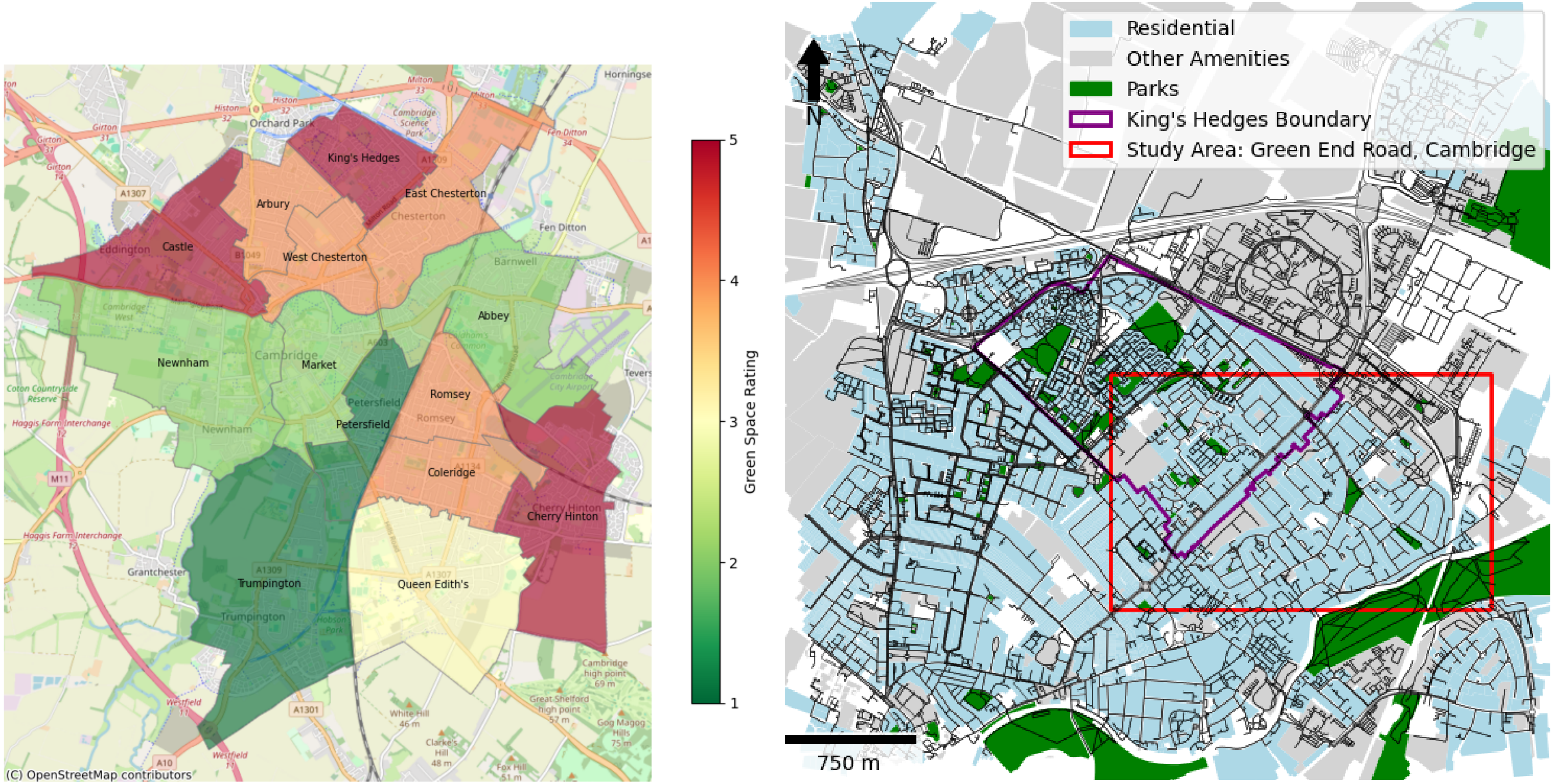
To identify a suitable location for a new community park, this study first examines the Green Space Deprivation Index across wards in Cambridge, as reported by Friends of the Earth (2020), as shown in Figure 3a. The index is based on the proportion of the population living within a five-minute walk of at least two hectares of public green space, as well as the total green space per area in 2020. The analysis reveals notable deprivation in the northern and southeastern wards of the city.

This study further focuses on the area surrounding King's Hedges ward in Cambridge as a proposed site for a new community park. King's Hedges was selected due to its notably poor access to green spaces, as indicated by its low green space rating. Additionally, the ward contains the highest number of LSOAs ranked among Cambridge's most deprived in the 2019 Index of Multiple Deprivation. Given the well-documented socioeconomic and monetary benefits of community parks, this location presents a strong case for targeted green space development.

Upon closer examination of King's Hedges ward, two green spaces were identified in the northern part of the area. As a result, the proposed location for a new community park was selected in the southern part of the ward, near the border with East Chesterton. This southern section is predominantly residential, and residents would greatly benefit from improved access to green spaces, as shown in Figure 3b.

Although the map indicates several small green areas in this southern region, verification against official community park location data from Cambridge City Council (2025) revealed that most of these are either private gardens or very small playgrounds.The green space data used in this study were sourced from OpenStreetMap, which may contain inaccuracies. Therefore, a key limitation is the inability to verify the status, quality, or maintenance of these green spaces.

Given the limitations and also the evident gap in accessible public green space, the author decided that a suitable location for a new community park would be between the King's Hedges and East Chesterton wards. This placement would improve accessibility for residents in the southern part of King's Hedges and benefit East Chesterton, which also exhibits high levels of green space deprivation.



(a) Green space ratings by ward across Cambridge

(b) King's Hedges Ward (2 Km Radius)

Figure 3: Spatial Context of Cambridge's Green Space Accessibility.

In this study, centrality analysis was employed to identify optimal locations for new community parks. When planning land acquisition for public green spaces, centrality measures can be used to optimize park placement and size in order to enhance network efficiency (Gyurkovich, 2023; Yin et al., 2022). Following the betweenness and closeness centrality analysis using a walking network, Green End Road emerged as the most central route, exhibiting the highest values in both metrics (See Figure 4). This suggests that Green End Road serves as a key connector across the area. Based on these findings, this location is considered well-suited for a community park that could support healthier and socially cohesive communities within and around King's Hedges.

(a) Green End Road: Closeness Centrality

(b) Green End Road: Betweenness Centrality

Figure 4: Proposed Site for a New Community Park.

2.3. Limitation

While these analyses support the development of intelligent urban applications, they also present important limitations. In the study of individual mobility patterns and check-in behaviors, there is a risk of re-identifying individuals, as personal identities can often be inferred from frequently visited locations (Pellungrini et al., 2018; Krishnamurthy & Pelechrinis, 2013; Mattos et al., 2022; Andrienko, Andrienko & Fuchs, 2013). Similarly, in the context of proposing new community parks, simple centrality analysis should not be used as standalone tools for urban planning and should further consider key variables such as park quality, socioeconomic conditions, legal constraints, and precise spatial context (Ojeda-Revah, Bojorquez & Osuna, 2017; Moore, Boyle & Lynch, 2022; Yiliu, Hussain & Shukor, 2025; Gyurkovich, 2023). Future research should address these limitations by incorporating these additional spatial and contextual factors to improve the robustness and applicability of such analyses.

3. Text-Spatial Analysis

3.1. Introduction

This study analyzes venue review data in Calgary, Canada, with a focus on restaurants as the most frequently reviewed amenity in this dataset. Supervised machine learning models are applied to predict review scores based on textual review content. In addition, spatial and centrality analyses are conducted to identify optimal locations for new restaurant development, leveraging insights extracted from the review data.

3.2. Analysis

Prior to analysis, the dataset was cleaned by filtering for reviews related specifically to restaurants and food-related amenities, as these categories contained the highest volume of reviews. Text preprocessing steps included converting text to lowercase, removing stopwords, lemmatization, and word filtering to ensure consistency and improve analytical accuracy. The dataset also included user-generated tags—'useful,' 'funny,' and 'cool'—which reflect peer validation of reviews. These tags were deemed valuable indicators of perceived review quality by the author and were therefore incorporated into the machine learning analysis. Finally, TF-IDF vectorization was used instead of a Countvectorizer, as prior studies have shown that TF-IDF often yields better model performance in text classification tasks (Suryaningrum, 2023; Rakhmanov, 2020).

This study employed three machine learning algorithms: Logistic Regression (Le, 2022; Sun, 2022; Shihab et al., 2018), Multinomial Naive Bayes (Reddy et al., 2017), and Random Forest (Reddy et al., 2017; Haq et al., 2022). These models are selected for their established effectiveness in predicting review star ratings from textual data. To properly evaluate model performance, macro-averaged F1 scores were used. This is especially relevant given the class imbalance in the dataset, where 67% of restaurant reviews were positive.

Notably, Reddy et al. (2017) found that a Bigram-Trigram Multinomial Naive Bayes model performed particularly well in this type of task. Because of this insight, the author evaluated each model using both unigram and bigram-trigram text structures, resulting in six total model configurations. Furthermore, non-textual features were incorporated only in the Logistic Regression and Random Forest models, as the assumptions of the Multinomial Naive Bayes model, which is its reliance on discrete word frequency inputs, make it unsuitable for handling continuous data from the non-textual data (Xu, Li, & Wang, 2017; GeeksforGeeks, 2023).

Overall, unigram models outperformed their bigram–trigram counterparts, with an average macro F1 score of 0.82 compared to 0.75 (See Table 2). The best-performing model was the unigram logistic regression, achieving a macro F1 score of 0.86, indicating strong overall performance in terms of precision and recall despite class imbalance (Mathew et al., 2023; Buhl, 2023). In contrast, the Bigram-Trigram Multinomial Naive Bayesmodel yielded a lower macro F1 score of 0.71, which contradicts findings by Reddy et al. (2017), suggesting this model may not generalize well to the Calgary review dataset. Furthermore, the confusion matrix in Figure 5 further supports this, as the logistic regression with unigram data is showing high accuracy in both true positives and true negatives while Bigram-Trigram Multinomial Naive Bayes showed a high number of false positives.

Data Type	**Model**	**Unigram**		**Bigram-Trigram**	
		Macro Avg	**Weighted Avg**	**Macro Avg**	**Weighted Avg**
Text & Non-Text Data	Logistic Regression	0.86	0.88	0.78	0.81
Text Data	Multinomial Naive Bayes	0.78	0.81	0.71	0.76
Text & Non-Text Data	Random Forest	0.82	0.84	0.77	0.80

Table 2: Model Performance Comparison.

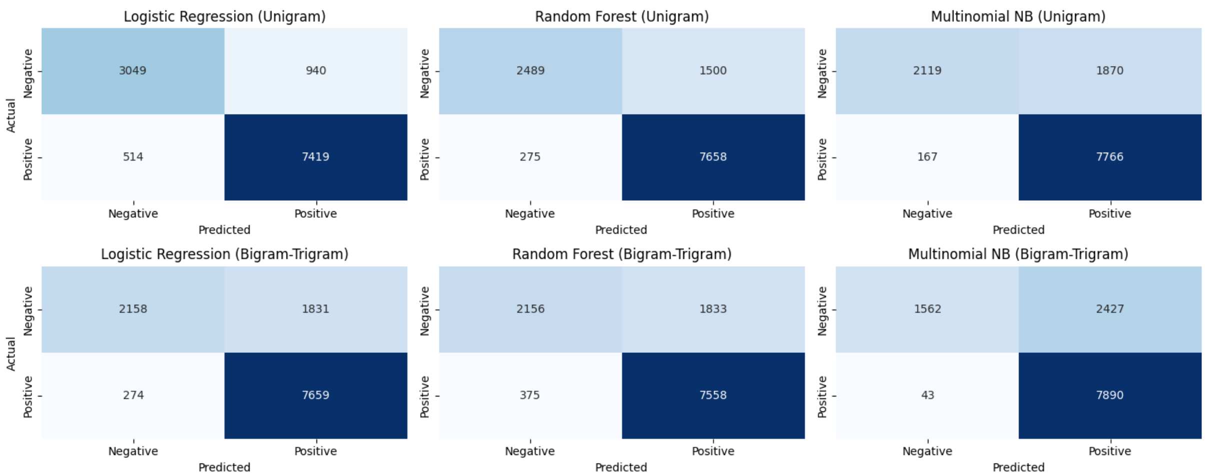


Figure 5: Confusion Matrices.

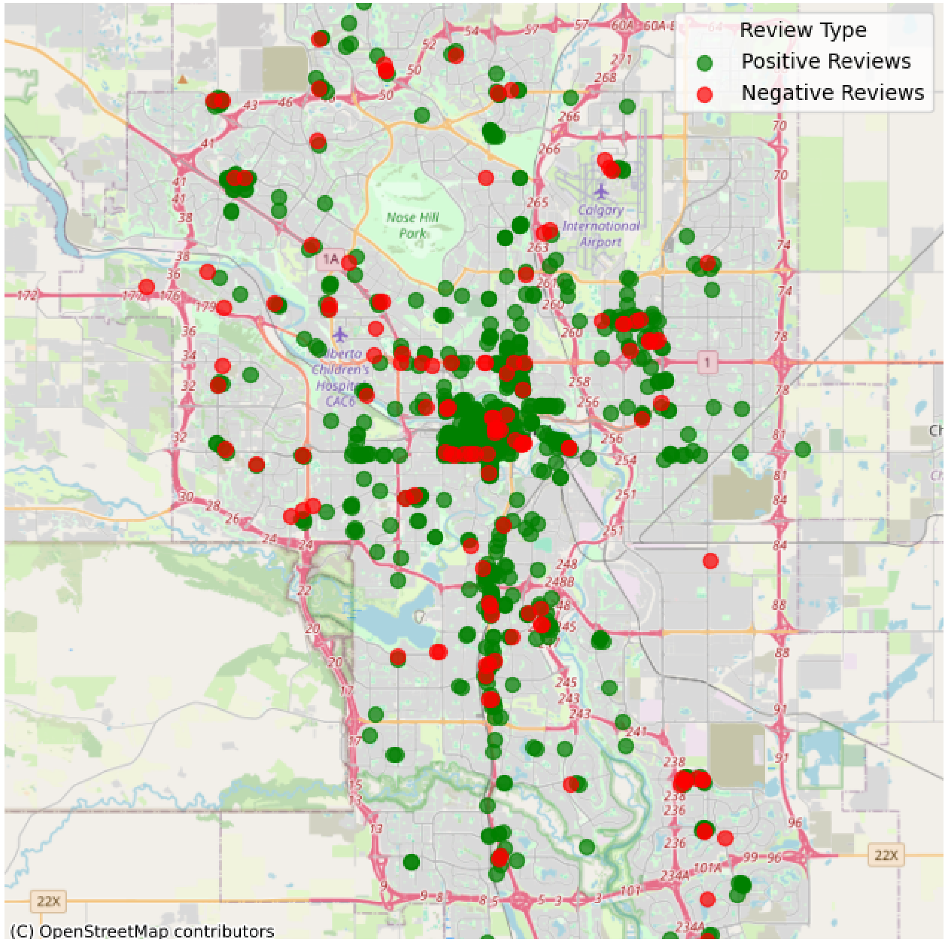
After conducting supervised machine learning analysis, the author also attempted to use the review dataset to identify promising locations for potential new restaurant development through combined text and spatial analysis. To ensure the robustness of the analysis, a minimum threshold of 15 reviews, three above the dataset's median of 12, was applied to focus the analysis on restaurants with sufficient representative samples of customer feedback. After mapping the data, Figure 6a showed that most restaurants are concentrated in Calgary's downtown area, which also hosts the majority of positively reviewed restaurants. Additional clusters are observed along Macleod Trail and near the Horizon neighbourhood.

In order to inform the selection of potential sites for a new restaurant, the analysis examined customer reviews containing the phrase "great location." This approach aimed to identify whether a restaurant's location was perceived positively. Negative reviews were also included in this stage, based on the rationale that some poorly rated restaurants might still be situated in desirable locations, with the low ratings attributed to other factors such as service or food quality (See Figure 6b).

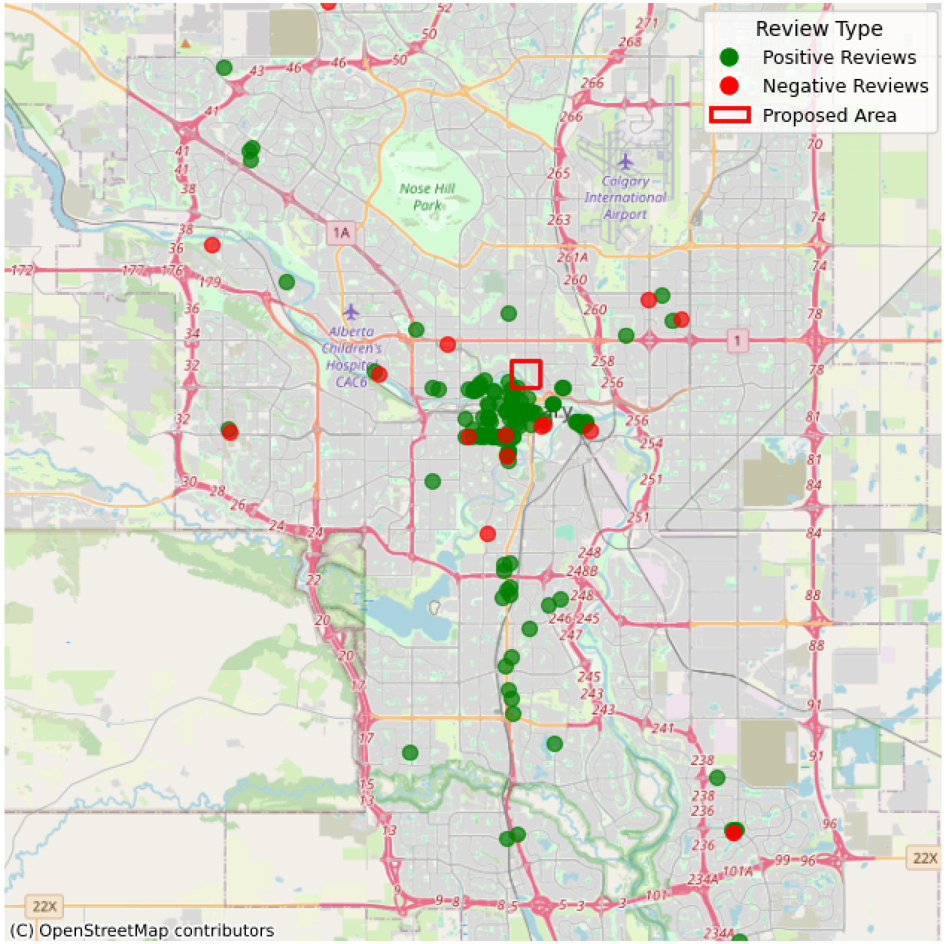
To better understand what constitutes a "great location" from the customer's perspective, the author examined keywords associated with place descriptions in reviews that mentioned a "great location." These keywords were categorized into three themes: natural features (e.g., river, park, hills), urban structures (e.g., city center, downtown, main street), and abstract qualities or ambiance (e.g., quiet, bustling, vibrant). This analysis aimed to better understand the specific spatial or experiential elements contributing to positive location-based reviews and the typical locations of such restaurants. A similar keyword-based approach was employed by Zhang et al. (2022) in their study on spatial perception in restaurant reviews.

The filtered text analysis revealed several frequently mentioned location-related terms. The top 15 terms included: park (42), view (24), mall (22), green (12), downtown (10), city (9), river (9), avenue (6), central (6), bustling (6), upscale (5), plaza (4), modern (4), hill (4), and nearby (3). Based on the prominence of these keywords, combined with the spatial clustering of positively and negatively reviewed restaurants, Crescent Heights, located just north of downtown, was identified as a promising location for a new restaurant development.

The rationale for this site selection also draws on existing research on restaurant clustering. High-performing restaurants often cluster near one another due to imitation behaviors and the perceived quality spillover from well-established competitors (Sun & Paule, 2017; Motoyama & Usher, 2020; Hotelling, 1929; Mossay, Shin, & Smrkolj, 2020). However, differentiation remains critical, particularly in terms of location, to establish a unique market position (Miller & Friesen, 1986; Leonardi & Moretti, 2023). For these reasons, the selected area in this study is situated near existing clusters of positively rated restaurants to benefit from quality spillover, while remaining sufficiently distant to retain spatial distinction from direct competitors.



(a) Overall Positive and Negative Reviewed Restaurants



(b) Restaurants with Reviews Mentioning a “great location”

Figure 6: Calgary’s Restaurant Distributions.

Specifically, the proposed location for the new restaurant is at coordinates -114.063306, 51.056972, situated on Crescent Road NW, as shown in Figure 7. This site is considered highly suitable due to its proximity to features frequently associated with the phrase "great location" in user reviews. Notably, it is near parks such as Sunny Bank Park and Rotary Park, the Bow River, and elevated terrain in Crescent Heights. The location provides access to green spaces and panoramic views of both natural and urban landscapes, while remaining close to the city center (See Figure 8). Additionally, this location lies near a road segment with the highest betweenness centrality in the area, indicating strong accessibility and integration within the city's street network and thus can be a really promising location for a brand new restaurant.

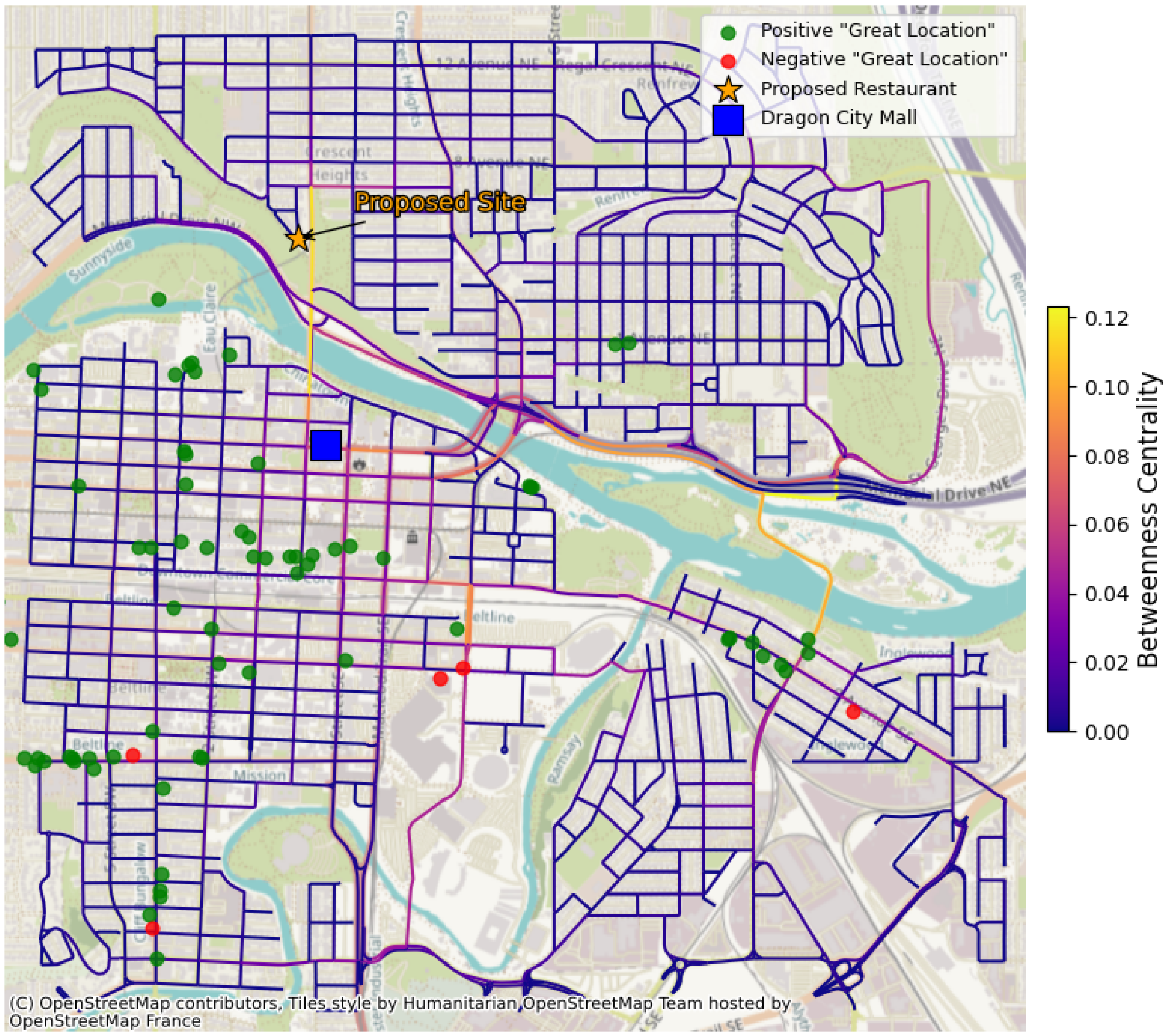


Figure 7: Proposed Site for a New Restaurant.

Figure 8: View from the Proposed Site. Courtesy: Google Maps (2019).

3.3. Limitation

This text-spatial analysis has several limitations. The machine learning component could be strengthened by evaluating additional models, such as Support Vector Machines (Suryaningrum, 2023; Shihab et al., 2018) or more advanced architectures like BERT-based classifiers (Bilal Almazroi, 2022), to benchmark performance across similar datasets. Regarding restaurant location analysis, key contextual factors remain unaddressed. For instance, in Calgary, new restaurant developments must comply with zoning regulations (City of Calgary, 2025), which should be integrated into future studies. Public concerns, such as noise in residential areas (Calgary, 2023), and the potential for gentrification and displacement following restaurant openings (Glaeser, Kim, Luca, 2018), also warrant consideration. Overall, this study serves as an initial exploration intended to inform further research in human mobility and text-based spatial analysis

References

Andrienko, N.V., Andrienko, G.L. and Fuchs, G. (2013). Towards Privacy-Preserving Semantic Mobility Analysis. EuroVis. https://doi.org/10.2312/pe.eurovast.eurova13.019-023.

Bai, X., Xie, Z., Xu, X. and Xiao, Y. (2023). An adaptive threshold fast DBSCAN algorithm with preserved trajectory feature points for vessel trajectory clustering. Ocean Engineering, 280. https://doi.org/10.1016/j.oceaneng.2023.114930.

Bilal, M. and Almazroi, A.A. (2022). Effectiveness of Fine-tuned BERT Model in Classification of Helpful and Unhelpful Online Customer Reviews. Electronic Commerce Research, 23, pp.2737–2757. https://doi.org/10.1007/s10660-022-09560-w.

Buhl, N. (2023). F1 Score in Machine Learning. [online] encord.com. Available at: https://encord.com/blog/f1-score-in-machine-learning/.

Calgary (2023). Noise on residential properties. [online] Calgary. Available at: https://www.calgary.ca/bylaws/residential-noise.html.

Calgary (2025). Calgary Land Use Bylaw 1P2007. [online] Calgary. Available at: https://www.calgary.ca/planning/land-use.html.

Cambridge City Council (2019). Indices of multiple deprivation 2019 report. Cambridge City Council.

Cambridge City Council (2025). Map of parks and playgrounds. [online] Cambridge City Council. Available at: https://www.cambridge.gov.uk/map-of-parks-and-playgrounds

Cohen, M., Burrowes, K. and Gwam, P. (2022). The Health Benefits of Parks and their Economic Impacts: A Review of the Literature. Urban Institute.

Friends of the Earth (2020). England's green space gap: How to end green space deprivation in England. Friends of the Earth.

GeeksfoGeeks (2023). Multinomial Naive Bayes. [online] GeeksforGeeks. Available at: https://www.geeksforgeeks.org/multinomial-naive-bayes/.

Glaeser, E.L., Kim, H. and Luca, M. (2018). Nowcasting Gentrification: Using Yelp Data to Quantify Neighborhood Change. AEA Papers and Proceedings, 108, pp.77–82. https://doi.org/10.2139/ssrn.3123733.

Google Maps (2019). Google Maps. [online] Google Maps. Available at: https://www.google.com/maps/@51.0570499

Greater Cambridge Planning Service (2021). North East Cambridge Area Action Plan Proposed Submission. Greater Cambridge Planning Service.

Gyurkovich, K.D. (2023). Study of Centrality Measures in the Network of Green Spaces in the City of Krakow. Sustainability, 15(18). https://doi.org/10.3390/su151813458

Haq, B., Khan, S., Umer, R. and Uddin, N. (2022). Machine Learning Algorithms for Predicting the Impact of Business Attributes Over Business Stars of Yelp Data-Set. International Review of Basic and Applied Sciences, 10(2).

Hotelling, H. (1929). Stability in Competition. The Economic Journal, 39(153), pp.41–57.

Hu, Y. (2018). Geo-text data and data-driven geospatial semantics. Geography Compass, 12(11). https://doi.org/10.1111/gec3.12404

Krishnamurthy, P. and Pelechrinis, K. (2013). Location-Based Social Networks. In: Advanced Location-Based Technologies and Services. Taylor Francis.

Law, S., Neira, M., Tanu, N., Keel, T., Jie, G., Tang, J. and Hu, D. (2025). Lab Notebook 6: Introduction to Human Mobility Analysis. UCL Department of Geography.

Le, J.P. (2022). *Rating Prediction from Review Text with Regularization — Linear Regression vs Logistic Regression.* [online] Medium. Available at: https://medium.com/mitb-for-all/rating-prediction-from-review-text-with-regularization-linear-regression-vs-logistic-regression-df0181fe9c07

Leonardi, M. and Moretti, E. (2023). The Agglomeration of Urban Amenities: Evidence from Milan Restaurants. American Economic Review: Insights, 5(2), pp.141–157. https://doi.org/10.1257/aeri.20220011

Mathew, J., Kshirsagar, R., Abidin, D.Z., Griffin, J., Kanarachos, S., James, J., Alamaniotis, M. and Fitzpatrick, M.E. (2023). A comparison of machine learning methods to classify radioactive elements using prompt-gamma-ray neutron activation data. Scientific Reports, 13(1). https://doi.org/10.21203/rs.3.rs-2518432/v1

Mattos, E.P. de, Domingues, A.C.S.A., Santos, B.P., Ramos, H.S. and Loureiro, A.A.F. (2022). The Impact of Mobility on Location Privacy: A Perspective on Smart Mobility. IEEE Systems Journal, 16(4), pp.5509–5520. https://doi.org/10.1109/jsyst.2022.3147808

Miller, D. and Friesen, P.H. (1986). Porter's (1980) Generic Strategies and Performance: An Empirical Examination with American Data. Organization Studies, 7(3), pp.255–261. https://doi.org/10.1177/017084068600700303

Moore, A., Boyle, B. and Lynch, H. (2022). Designing for Inclusion in Public playgrounds: a Scoping Review of definitions, and Utilization of Universal Design. Disability and Rehabilitation. Assistive Technology, 18(8), pp.1453–1465. https://doi.org/10.1080/17483107.2021.2022788

Mossay, P., Shin, J.K. and Smrkolj, G. (2020). Quality Differentiation and Spatial Clustering among Restaurants. Munich Personal RePEc Archive.

Motoyama, Y. and Usher, K. (2020). Restaurant Reviews and Neighborhood Effects. Papers in Applied Geography, 6(4), pp.386–401. https://doi.org/10.1080/23754931.2020.1791942

Ojeda-Revah, L., Bojorquez, I. and Osuna, J.C. (2017). How the legal framework for urban parks design affects user satisfaction in a Latin American city. Cities, 69, pp.12–19. https://doi.org/10.1016/j.cities.2017.05.016

Pellungrini, R., Pappalardo, L., Pratesi, F. and Monreale, A. (2018). Analyzing privacy risk in human mobility data. Software Technologies: Applications and Foundations: STAF 2018 Collocated Workshops, Toulouse, France, June 25-29, 2018, Revised Selected Papers, pp.114–129.

Rakhmanov, O. (2020). A Comparative Study on Vectorization and Classification Techniques in Sentiment Analysis to Classify Student-Lecturer Comments. Procedia Computer Science, 178, pp.194–204. https://doi.org/10.1016/j.procs.2020.11.021

Reddy, Ch.S.C., Kumar, K.U., Keshav, J.D., Prasad, B.R. and Agarwal, S. (2017). Prediction of star ratings from online reviews. TENCON 2017 - 2017 IEEE Region 10 Conference. https://doi.org/10.1109/tencon.2017.8228161

Shihab, I.F., Oishi, M.M., Islam, S., Banik, K. and Arif, H. (2018). A Machine Learning Approach to Suggest Ideal Geographical Location for New Restaurant Establishment. In: IEEE Xplore. pp.1–5. https://doi.org/10.1109/R10-HTC.2018.8629845

Sun, Y. (2022). Prediction of Yelp Score from Reviews with Machine Learning Model. University of California, Los Angeles.

Sun, Y. and Paule, J.D.G. (2017). Spatial analysis of users-generated ratings of Yelp venues. Open Geospatial Data, Software and Standards, 2(1). Available at: https://doi.org/10.1186/s40965-017-0020-9

Suryaningrum, K.M. (2023). Comparison of the TF-IDF Method with the Count Vectorizer to Classify Hate Speech. Engineering, Mathematics and Computer Science Journal (EMACS), 5(2), pp.79–83. https://doi.org/10.21512/emacsjournal.v5i2.9978

Xu, S., Li, Y. and Wang, Z. (2017). Bayesian Multinomial Naïve Bayes Classifier to Text Classification. Lecture Notes in Electrical Engineering, pp.347–352. https://doi.org/10.1007/978-981-10-5041-1_57

Yiliu, W., Hussain, N. and Shukor, S.F.A. (2025). Safety perceptions related to park environments: a scoping review. World Leisure Journal. https://doi.org/10.1080/16078055.2025.2469778

Yin, G., Liu, T., Chen, Y. and Hou, Y. (2022). Disparity and Spatial Heterogeneity of the Correlation between Street Centrality and Land Use Intensity in Jinan, China. International Journal of Environmental Research and Public Health, 19(23). https://doi.org/10.3390/ijerph192315558

Zhang, J., Li, X., Yao, Y., Hong, Y., He, J., Jiang, Z. and Sun, J. (2021). The Traj2Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types. International Journal of Geographical Information Science, 35(1), pp.193–211. https://doi.org/10.1080/13658816.2020.1726923

Zhang, S., Ly, L., Mach, N. and Amaya, C. (2022). Topic Modeling and Sentiment Analysis of Yelp Restaurant Reviews. International Journal of Information Systems in the Service Sector, 14(1). https://doi.org/10.4018/ijisss.295872