Where Next? A Behavioral and Explainable Framework for City and Neighborhood Recommendation

[1]Gustavo H. Santos, [1]Myriam Delgado, [2]Daniel Silver, and [1,2]Thiago H. Silva
[1]Federal University of Technology - Parana, Curitiba, Brazil
[2]University of Toronto, Toronto, Canada
gustavohenriquesantos@alunos.utfpr.edu.br,myriamdelg@utfpr.edu.br,{dan.silver,th.silva}@utoronto.ca

ABSTRACT

Understanding why individuals choose to visit particular cities and specific neighborhoods within them is essential for advancing both urban mobility research and personalized tourism technologies. This paper proposes a novel multi-level (city and neighborhood levels), explainable recommendation framework that models user interest based on area similarities across geographic, demographic, cultural, and venue-category dimensions. Our approach predicts user interest through a behaviorally informed, interpretable machine learning model. Using large-scale review data from Google Places, enriched with U.S. Census, political, and cultural indicators, we analyze mobility through the lens of high-interest and low-interest divisions and two behavioral archetypes: returners (who repeatedly visit familiar areas) and explorers (who seek out new destinations). Results show that explorers are more interested in geographically clustered cities, suggesting a search for new experiences in nearby locations. In contrast, returners attach to areas that align with their past experiences (e.g., venue categories). Beyond good predictive performance, our system provides natural-language explanations for each recommendation, offering actionable insights into user behavior. A demonstration system illustrates how our approach enables transparent, behavior-informed travel recommendations. This work bridges gaps in urban AI by integrating spatial granularity, behavioral segmentation, and explainability.

KEYWORDS

Human Mobility, Place Recommendation, Returners and Explorers, Explainable Machine Learning, Urban Behavior Modeling

1 INTRODUCTION

Exploring key aspects of individuals in complex urban mobility systems is essential for developing smart city applications and plays a central role in shaping urban dynamics, including tourism, infrastructure, and service delivery [4]. As individuals increasingly travel across cities and neighborhoods for leisure or exploration, understanding *where* they go becomes a key question for urban planners and recommendation system designers. Existing studies show that mobility patterns are not random, but often reflect stable yet nuanced individual preferences [3, 8]. However, most location-based recommender systems suffer from two key limitations: i- Oversimplification: They focus either on venue-level suggestions or generic city-level rankings, ignoring the multi-scale nature of urban travel decisions. ii- Black-box nature: Few provide explainable recommendations tied to socio-demographic, cultural, or behavioral factors (e.g., why a neighborhood suits a user's preferences).

In this work, we address these gaps by modeling user preferences through the lens of high-interest and low-interest division and *returners vs. explorers* dichotomy: returners are individuals who frequently revisit a small number of familiar places, while explorers actively seek out novel destinations, i.e., their mobility patterns are less routine-driven [17]. While this segmentation has been influential in mobility studies, it remains unclear which spatial, cultural, or socio-demographic factors drive these behaviors and how they can be harnessed to generate more meaningful and personalized recommendations.

Improving the explainability of machine learning models has become a central focus in artificial intelligence research. When users understand why a system produces a particular output, their trust in the system increases. Recommender systems, in particular, have received heightened attention, with significant efforts aimed at making their outputs more interpretable [18]. However, efforts in this direction are less common in place recommendation systems.

To address these challenges, we propose an explainable, multi-level place recommendation framework that operates at the city and neighborhood levels. Using large-scale review data from Google Places, a broadly used social media for places review, enriched with socio-demographic, cultural, and political indicators, we predict user interest across U.S. Core-Based Statistical Areas (CBSAs) and zip codes. Our framework moves beyond black-box modeling: leveraging explainable machine learning techniques (SHAP [13] and LIME [19]), we provide natural-language justifications for each recommendation, grounded in quantifiable similarities between places.

This paper makes four main contributions: i- A multi-level behavioral model that outperforms baselines (e.g., 0.7 recall for city recommendations); ii- Explainable recommendations grounded in geographic, cultural, and socioeconomic features; iii- Empirical insights into returner/explorer divergence (e.g., explorers favor proximity; returners prioritize venue categories); iv- An interactive demo system (Section 7) that translates model outputs into actionable, interpretable advice.

By integrating behavioral insights, spatial granularity, and model explainability, our approach helps develop next-generation urban recommender systems. Our study also helps explain the interest in urban areas, information that could be useful beyond the recommendation context.

The rest of the paper is organized as follows: Section 2 reviews related work; Section 3 describes the dataset; Sections 4–6 present the methodology and results; Section 7 introduces our demonstration system; and Section 8 concludes the work.

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2025). Rio de Janeiro, Brazil. Porto Alegre: Brazilian Computer Society, 2025.

ISSN 2966-2753

2 RELATED WORK

Human Mobility Behavior Modeling. The returner and explorer dichotomy introduced by Pappalardo et al. [17] has been validated and extended in numerous studies. Schläpfer et al. [21] identified fundamental scaling laws in visitation patterns across diverse urban settings, while Xu et al. [24] demonstrated how mobility behavior shapes urban growth with fine-grained resolution. Wang et al. [23] synthesize a decade of research, underscoring how digital traces have redefined urban movement analysis. Senefonte et al. [22] revealed clusters of tourists with similar mobility profiles, aiding predictive tasks. Challenging the notion of universal mobility laws, Napoli et al. [16] highlight the influence of contextual factors in behavioral modeling. The push for transparent AI has led to greater adoption of explainability in mobility studies. Chen et al. [4] call for equitable and transparent systems in smart cities, while Liang et al. [11] show how large language models improve mobility predictions during public events. In this direction, Santos et al. [20] model urban behavior across platforms (Google Places and Foursquare) by analyzing users' regional interests, largely reflecting resident behavior, and show that geographic factors influence preferences. Building on these insights, our work incorporates multi-level place characteristics, such as geographic and cultural, to explain how preferences form in human mobility across cities and neighborhoods.

Place Recommendation Systems. Place recommendation systems have evolved to incorporate both user preferences and contextual signals such as time and location [5, 7]. While most focus on point-of-interest (POI) suggestions [1, 2, 5, 7], few address recommendations at the city or neighborhood level. Cheng et al. [5] achieved early success with spatially- and socially-aware matrix factorization. Abbasi and Alesheikh [1] used word embeddings and conceptual spaces to better match human geographic understanding. Baral et al. [2] introduced a hierarchical model to more effectively capture locality and user preferences in POI sequence recommendations. Closer to our setting, Gao et al. [7] designed a system for travelers exploring new areas, combining time- and location-based social media data for personalized POI recommendations. In contrast, our approach operates across multiple levels—predicting interest in both cities (CBSAs) and sub-regions (ZIP codes)—while integrating socio-demographic, political, and cultural signals to model the distinct preferences of returners and explorers.

This foundation supports our contribution: an explainable, multi-level recommendation system that links behavioral patterns to spatial, cultural, and socioeconomic features, providing transparent justifications for its predictions. While explainability is not new in recommendation systems (such as [6, 18]), this is less common in the context of area recommendations Note that explainability in this study serves not only to inform users but also to better understand user interest in urban areas.

3 DATASETS AND PREPROCESSING

Google Places is a location-based social network (LBSN) that enables users to explore and share information about local venues, landmarks, and points of interest, including restaurants, museums, train stations, and recreational areas. Originally introduced in [10, 25], it comprises 666,324,103 geo-localized reviews on 4,963,111 venues, written by 113,643,107 users across the United States up to September 2021. The majority of this data (98.9%) spans the years 2015 to 2021.

To study tourist behavior, we filtered this dataset by selecting users who reviewed venues in at least six Core Based Statistical Areas (CBSAs). CBSAs are geographic regions defined by the U.S. Office of Management and Budget, consisting of one or more counties anchored by an urban center with a substantial population nucleus, and surrounding areas with strong economic and social integration. Setting the threshold at six reviewed CBSAs allows us to differentiate between cities with higher and lower user engagement (measured by review volume), while still preserving a significant portion of the data (37% of all reviews). Our final dataset comprises 245,322,994 reviews of 4,613,409 users. Approximately 64.8% of the filtered users submit all their reviews within four years, with an average of 53 reviews per user (median: 35, std: 59, skewness: 4.5, kurtosis: 58), indicating a long-tailed activity distribution with a minority of highly prolific contributors, as expected.

In addition to city-level analysis (CBSAs), we also consider neighborhood data using U.S. Zip Codes. Our goal is to understand the factors associated with higher user interest, not only across cities but also within specific areas of those cities.

We define a user's interest in a region by the number of reviews they posted there, with higher review counts indicating stronger interest. For our analyses (see Section 4), we rank CBSAs and zip codes based on the number of reviews. In case of ties, we use the average rating the user gave to the region, where higher ratings indicate stronger interest and, therefore, a lower rank. For simplicity, we assume that review volume is a stronger indicator of interest than ratings, as it reflects repeated engagement — users who review multiple venues are likely more engaged with the area. Thus, we prioritize review count and only use average ratings to break ties, favoring regions where users are both active and satisfied. Table 1 summarizes key statistics by CBSA rank, including the number of reviews per user, zip codes visited, and review distributions across top zip codes within each CBSA.

Since our study focuses on tourism-oriented behavior, we also explore users' travel patterns across regions. On average, users have visited 5.3 different U.S. states, with a standard deviation of 2.8. This supports the assumption that the users have a tourist profile.

The geospatial definitions for the 935 CBSAs, and 33,791 Zip Codes - of which 926 CBSAs (99%) and 32,041 zip codes (94.5%) had their reviews addressed in the present work - have been sourced from the U.S. Census Bureau's TIGER geographic database[1].

To provide a richer context, we enrich each region (CBSA or Zip Code) with socio-demographic indicators, including race composition (White, Black, Asian, Hispanic), median household income, population size, percentage of residents with higher education, and employment rate. These variables are obtained from the National Historical Geographic Information System (NHGIS) [15], based on the 2021 American Community Survey, which aggregates data over the period 2017–2021.

We also incorporate political landscape features. Following the methodology proposed in [12], we estimate regional political orientation by calculating the percentage of voters supporting the

[1] https://www.census.gov/cgi-bin/geo/shapefiles/index.php

Table 1: Mean Summary Statistics by CBSA Rank per user (±STD)

CBSA Rank	Mean Reviews	Zipcodes Visited	Reviews per Zipcode	Reviews Top Zipcode	Reviews 2nd Zipcode	Reviews 3rd Zipcode
1	30.7 ± 40.2	19.4 ± 15.8	3.2 ± 2.8	8.7 ± 10.5	5 ± 5.8	3.6 ± 4.1
2	7.7 ± 10.2	7.1 ± 6.5	2.1 ± 1.8	3.1 ± 3.7	1.8 ± 1.8	1.5 ± 1.3
3	3.9 ± 4.7	4.1 ± 3.8	1.7 ± 1.3	1.9 ± 1.9	1.3 ± 0.9	1.2 ± 0.7
4	2.5 ± 2.9	2.8 ± 2.6	1.5 ± 0.9	1.5 ± 1.2	1.2 ± 0.6	1.1 ± 0.5
5	1.9 ± 1.9	2.2 ± 1.9	1.3 ± 0.8	1.3 ± 0.9	1.1 ± 0.5	1.1 ± 0.4
6	1.6 ± 1.5	1.8 ± 1.6	1.2 ± 0.6	1.2 ± 0.7	1.1 ± 0.4	1.1 ± 0.3

Democratic candidate in the 2020 U.S. presidential election. This data has been obtained from "An Extremely Detailed Map of the 2020 Election"[2], which provides precinct-level shapefiles and vote counts. We aggregate vote totals for all precincts intersecting each zip code and compute the corresponding percentage of Democratic support. A similar procedure is applied to estimate political leanings at the CBSA level.

Lastly, we evaluate the cultural dimensions of each region using the Scenes Theory [9], which characterizes cultural profiles across 15 dimensions based on the types of venues present. We apply the methodology described in [9] to compute scene vectors for each region. Additionally, we consider a simpler proxy—venue category frequency distribution, which offers a complementary way to describe regional culture. Both approaches are used in our modeling.

4 METHODOLOGY

This section is structured into four main subsection: in Section 4.1 we identify representative regions in two different spatial granularities (Cities and Zip Codes); Section 4.2 presents explanatory features, in Section 4.3 we establish comparison baselines, while Section 4.4 discusses how we identified returners and explorers. These steps collectively allow us to assess urban interest patterns at scale and uncover the key factors driving visitation behaviors.

4.1 Defining Top/Bottom Regions (Cities and Zip Codes)

4.1.1 Procedures for cities (CBSAs). Our analysis aims to understand what attracts users to the cities[3] they like the most. Our approach involves observing the cities where users have shown the greatest interest, i.e., where they have made the most reviews (top), and comparing them to the cities where users have shown the least interest, i.e., where they have made the fewest reviews (bottom).

Therefore, we define a value k that separates the user's top-interest CBSAs from the rest. CBSAs visited by the user are ranked based on their review counts. The top k CBSAs, form the high-interest group, while the remaining are considered the bottom group.

Formally, assuming C as the set of CBSAs, and

$$C^u = \{c \in C \mid c \text{ was reviewed by } u\}$$

as the set of CBSAs reviewed by user u; $\text{rev}(u, c)$ as the number of reviews a user u has made in c; $\overline{\text{rat}}(u, c)$ as the average rating given by user u in c; and $|\{.\}|$ as set cardinality, we compute the top k CBSAs as follows:

- **Ranking of CBSAs:** For user u, rank all CBSAs $c \in C^u$ by their review count and average rating:

$$\text{rank}_u(c) = \Big|\Big\{c' \in C^u \;\Big|\; \text{rev}(u, c') > \text{rev}(u, c)\Big\} \cup \Big\{c'' \in C^u \;\Big|\; \text{rev}(u, c'') = \text{rev}(u, c) \wedge \overline{\text{rat}}(u, c'') \geq \overline{\text{rat}}(u, c)\Big\}$$

  Low rank values mean high interest by the user.

- **Top-k city partition:** $C^u_k = \{c \in C^u \mid \text{rank}(c) \leq k\}$

4.1.2 Procedures for areas (Zip Codes). In addition to identifying cities of interest to tourists, we propose a method to pinpoint specific areas within these cities that users are most likely to visit. For a user's top k cities (C^u_k), we rank the zip codes based on review counts. The top m zip codes in each city—those with the highest ranks—form the top set, while the rest form the bottom set. For example, with $m = 1$ and $k = 3$, we select the most reviewed zip code from each of the top-3 cities (handling ties through ratings). We focus on the top k cities to better understand which features of the most frequented areas appeal to the user.

Formally, for each top CBSA $c \in C^u_k$; assuming Z as the set of zip codes, and $\mathcal{Z}^u = \{z \in Z \mid u \text{ visited } z\}$:

- **Extract relevant zip codes:**

$$\mathcal{Z}^u_k = \{z \in \mathcal{Z}^u \mid z \text{ is within } c \in C^u_k\}$$

- **Compute local ranking for zip codes:**
  For each $z \in \mathcal{Z}^u_k$, we rank zip codes within the same CBSA c based on review count, breaking ties using the average rating:

$$\text{rank}_u(z) = \Big|\Big\{z' \in \mathcal{Z}^u_k \;\Big|\; \text{rev}(u, z') > \text{rev}(u, z)\Big\} \cup \Big\{z'' \in \mathcal{Z}^u_k \;\Big|\; \text{rev}(u, z'') = \text{rev}(u, z) \wedge \overline{\text{rat}}(u, z'') \geq \overline{\text{rat}}(u, z)\Big\}\Big|$$

- Select **top-m** zip codes:

$$\mathcal{Z}^u_{k,m} = \{z \in \mathcal{Z}^u_k \mid \text{rank}(z) \leq m\}$$

4.2 Feature Engineering

For each region $r \in R$ representing either a CBSA or a zip code and assuming $\mathcal{T}^u_k$ representing the top-k partitions of CBSAs (C^u_k) or top-m zip codes ($\mathcal{Z}^u_{k,m}$), R^u as C^u or $\mathcal{Z}^u$, depending on which type of region r is assigned with, we define the feature vector $\mathbf{f}$ for the

[2] https://github.com/TheUpshot/presidential-precinct-map-2020
[3] For simplicity, we use the word city as a synonym for CBSA in this work.

region r and a set of features $f \in \mathcal{F}$ as the concatenation of top and bottom feature vectors:

$$\mathbf{f}^{r,u} = \left[\mathbf{f}^{r,u}_{\text{top}} \parallel \mathbf{f}^{r,u}_{\text{bottom}}\right] \in \mathbb{R}^{2|\mathcal{F}|}$$

with

$$\mathbf{f}^{r,u}_{\text{top}} = \left(\text{stat}\left\{s_f(r, r_t) \mid r, r_t \in \mathcal{T}^u_k,\ r \neq r_t\right\}\right)_{f\in\mathcal{F}} \quad (1)$$

$$\mathbf{f}^{r,u}_{\text{bottom}} = \left(\text{stat}\left\{s_f(r, r_b) \mid r, r_b \in \mathcal{R}^u \setminus \mathcal{T}^u_k,\ r \neq r_b\right\}\right)_{f\in\mathcal{F}}. \quad (2)$$

The operator stat aggregates, per feature, the similarities s_f of r and other regions (see Section 5.1 for more details on each feature).

For a user u and a region r, the similarity sets of r regarding u can be calculated based on the similarity vector $\mathbf{s}(r, r') = \left(s_f(r, r')\right)_{f\in\mathcal{F}}$ as:

$$\mathcal{S}^u_{\text{top}}(r) = \{\mathbf{s}(r, r_t) \mid r, r_t \in \mathcal{T}^u_k,\ r \neq r_t\}$$
$$\mathcal{S}^u_{\text{bottom}}(r) = \{\mathbf{s}(r, r_b) \mid r, r_b \in \mathcal{R}^u \setminus \mathcal{T}^u_k,\ r \neq r_b\}$$

For example, we define the similarity set for the zip codes as:

$$\mathcal{S}^u_{\text{top}}(z) = \{\mathbf{s}(z, z_t) \mid z \text{ within } r, z_t \text{ within } r_t\}$$

with r and $r_t \in \mathcal{Z}^u_{k,m}$, $r \neq r_t$,

$$\mathcal{S}^u_{\text{bottom}}(z) = \{\mathbf{s}(z, z_b) \mid z \text{ within } r, z_b \text{ within } r_b,\}$$

with r and $r_b \in \mathcal{Z}^u \setminus \mathcal{Z}^u_{k,m}$, $r \neq r_b$.

Note that we compute the similarity only between zip codes from different CBSAs. This ensures that, when evaluating the probability of a zip code in a new city being of high interest to the user (i.e., being ranked at the top), we compare it solely to zip codes the user has visited in the past, since no other zip code within this new city is considered.

Having the similarity set of a region, we can compute, for each feature, its statistics regarding other top or bottom regions and use them in our recommendation task, which is addressed in this work as a binary classification problem $(\mathbf{f}^r, y^r)$, with $y^r \in \{0, 1\}$, $r = 1, ..., |R|$, using a multi-level approach. On one level, we predict whether a CBSA belongs to a user's top-k preferred CBSAs ($y^r = \mathbb{I}[r = c \in C^u_k]$, a binary top-$k$ membership). On another level, we predict whether a zip code belongs to a user's top-m favored zip codes ($y^r = \mathbb{I}[r = z \in \mathcal{Z}^u_{k,m}]$).

In both cases, the predictions are based on the similarity sets $\mathcal{S}^u_{\text{top}}(r)$ and $\mathcal{S}^u_{\text{bottom}}(r)$, and their associated top and bottom feature vectors. The proposed classifiers are evaluated using explainable AI techniques to understand the importance of each feature and how its values affect the prediction.

4.3 Baselines

For comparison, we first define a **popularity-based baseline** that selects regions based on popularity (total reviews across users). Let r be a region identifying either a city or a zip code, and the global review count for the region r be given by $\text{popularity}(r) = \sum_{u\in U} \text{rev}(u, r)$. For user u, the baseline defines:

$$\text{pop } C^u_k = \left\{c \in C^u \mid \text{popularity}(c) \geq \text{popularity}(c^*_k)\right\},$$
$$\text{pop } \mathcal{Z}^u_{k,m} = \left\{z \in \mathcal{Z}^u_k \mid \text{popularity}(z) \geq \text{popularity}(z^*_m)\right\}$$

with z and z^*_m within the same c, $c^*_k \in C^u$ as the city with the k-th largest popularity value and z^*_m as the zip code with the m-th largest popularity value in city c.

We also define an **Item Collaborative Filtering (ICF) baseline** -widely adopted personalized recommendation method [5, 7]. For the set of users $\mathcal{U} = \{u_1, u_2, \ldots, u_{|\mathcal{U}|}\}$ and the set of regions $\mathcal{R} = \{r_1, r_2, \ldots, r_{|\mathcal{R}|}\}$, each region as CBSAs or zipcodes, we define the user-region interaction matrix $\mathbf{X} \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{R}|}$, where each entry x_{u_i,r_j} represents the number of reviews of user u_i in region r_j and $\mathbf{x}_{r_j}$ is a column vector with the number of reviews of all users in region r_j. Finally, the interaction between two regions r_a and r_b is the cosine similarity between their corresponding user interaction vectors, that is, $\text{interaction}(r_a, r_b) = \frac{\mathbf{x}_{r_a} \cdot \mathbf{x}_{r_b}}{\|\mathbf{x}_{r_a}\|\|\mathbf{x}_{r_b}\|}$. With this, we have $|\mathcal{F}| = 1$, i.e. there is only one feature $\{f\} = \{\text{interaction}\}$ in the set $\mathcal{F}$, and then we can apply the same experimental setup as defined in Section 5.

These baselines have been selected to represent a generic standard (popularity) and a personalized, yet feature-agnostic standard (ICF), thus ensuring that our model's improvements are fairly validated.

4.4 Identifying Returners and Explorers

Based on [17], we calculate the radius of gyration of the user $\left(\text{radius}_g\right)$, using the centroids of the CBSAs visited by the users, as well as their radius of gyration using only the top k most visited CBSAs $\left(\text{radius}^{(k)}_g\right)$. Then, we define a user as a k-returner if $\text{radius}^{(k)}_g > \text{radius}_g/2$ or a k-explorer otherwise.

5 EXPERIMENTAL SETUP

In the experiments, we assume the statistic operator stat $\in$ {mean, median, std}, where std considers both mean and standard deviation calculations, to provide the feature vector. Moreover, we consider $\mathcal{F} = \{f\} = \{$geographic, population, income, education, employment, race, voting, scenes, categories $\}$. We also test the inclusion of a feature 'is_returner' (see section 6), which indicates whether a user is a returner or an explorer given k.

5.1 Similarity Metrics

In Eqs. 1 and 2, we calculate differently the similarity s_f of two regions $(r, r') \in C$ or $(r, r') \in Z$, depending on the addressed feature f:

- **Geographic Similarity** (f is geographic distance):

  $$s_{\text{geographic}}(r, r') = \|\mathbf{centroid}(r) - \mathbf{centroid}(r')\|_2$$

  (in ESRI:102005 projection)

- **Demographic Similarity** (|.| is absolute value):

  – f is population:

  $$s_{\text{population}}(r, r') = |\text{population}(r) - \text{population}(r')|$$

  – f is income:

  $$s_{\text{income}}(r, r') = |\text{median\_income}(r) - \text{median\_income}(r')|$$

– f is education:

$$s_{\text{education}}(r, r') = |\text{bachelor}(r) - \text{bachelor}(r')|$$

– f is employment:

$$s_{\text{employment}}(r, r') = |\text{employment}(r) - \text{employment}(r')|$$

– f is race:

$$s_{\text{race}}(r, r') = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{P_i(r)}{P(r)} - \frac{P_i(r')}{P(r')} \right|$$

where $P(r)$ represents the population size of region r, and $P_i(r)$ is the population size of the i-th racial category in region r. In this work, we use the racial categories: White, Black, Asian, and Hispanic.

- **Political Similarity** (f is voting):

$$s_{\text{voting}}(r, r') = |\text{votes\_democrat}(r) - \text{votes\_democrat}(r')|$$

- **Cultural Similarity** (f is scenes):

$$s_{\text{scenes}}(r, r') = \|\mathbf{scenes}(r) - \mathbf{scenes}(r')\|_2,$$

where $\mathbf{scenes}(\cdot)$ is the vector of 15 cultural dimensions.

- **Categorical Similarity** (f is categories (Frequency Cosine)):

$$s_{\text{categories}}(r, r') = \frac{\mathbf{v}(r) \cdot \mathbf{v}(r')}{\|\mathbf{v}(r)\| \|\mathbf{v}(r')\|}$$

where $\mathbf{v}(\cdot)$ is the frequency vector of the venue categories.

We leverage geographic distance to detect spatial travel patterns (e.g., preference for nearby destinations), population size to distinguish urban versus rural appeal, and income-employment metrics to assess the economic compatibility of destinations (e.g., luxury vs. budget tourism). Education levels (bachelor's degree prevalence) proxy for cultural-educational interests, racial composition captures Demographic affinity, and voting data reveal political polarization's influence on destination choices. Cultural features (Scenes) and venue categories (e.g., restaurants, parks) further refine understanding by identifying shared experiential preferences, collectively modeling how socioeconomic, spatial, and cultural factors shape the users' behavior.

These similarity calculations can be used to provide the elements of the feature vector. For example, assuming stat = mean, f = race, and r as CBSAs, the mean race similarity of a particular top CBSA ($r = c \in C_k^u$) and the remaining top CBSAs ($r_t = c_t \in C_k^u$) visited by user u is given by:

$$f_{\text{top}}^{r,u}(\text{race}) = \text{mean}\left\{ s_{\text{race}}(r, r_t) \mid r, r_t \in \mathcal{T}_k^u,\ r \neq r_t \right\}$$

5.2 Predictive Modeling

For our predictive model, we use a LightGBM classifier, which has performed best in comparison to Logistic Regression and XGBoost in the experiments (Sec 6.3). To address class imbalance, we apply class weighting by downweighting the majority class: the weight assigned to the top regions is proportional to the ratio of bottom regions to top regions for each user. We show the results obtained using the default hyperparameters provided by the Python LightGBM library. We use random 5-fold cross-validation, where users from the train set don't appear in the validation set. This is applied to our approach and the baselines. To evaluate the interpretability and behavior of the model, we apply both global and local explanation techniques. On the global level, we use feature permutation importance to assess the overall influence of each feature on the predictions and mean absolute SHAP values [14] to examine how feature values affect the output. We also implement a demonstration system that integrates the local LIME explanations [19] with natural language summaries generated by a large language model (see Section 7).

6 RESULTS

This section presents the results considering the multi-level approach: a city-level analysis in Section 6.1 and a neighborhood-level analysis in Section 6.2.

6.1 CBSA Recommendation Performance

We trained the model on the CBSA data using the final feature vector $\mathbf{f}^r$, varying k from 2 to 5 and stat set as mean in $\mathbf{f}_{\text{top}}^r$ and $\mathbf{f}_{\text{bottom}}^r$ calculations. Our city-level analysis reveals distinct patterns in user preferences across CBSAs. The proposed model achieves strong performance with a recall of approximately 0.7 for identifying top-ranked cities (Figure 1), significantly outperforming the popularity baseline, and being comparable, at higher k, with both baselines. Our model's better performance is particularly pronounced for lower values of k (2-4), where the distinction between top and bottom CBSAs is more evident (Table 1). While precision remains around 0.5, indicating some false positives, these recommendations still represent cities where users have demonstrated measurable interest. From a recommendation system perspective, this is not a major issue, as these cities still represent places where the user showed some interest, albeit to a lesser extent.

Regarding the explainability analysis, Figure 2 shows the permutation feature importance (normalized by min-max) of the top 10 most important features, ordered by the mean importance over the k scenarios. According to this figure, the category feature ($f_{\text{top}}^r(\text{categories})$) emerges as the strongest predictor, indicating users prefer cities offering experiences similar to their past high-interest locations. Geographic proximity ($f_{\text{top}}^r(\text{geographic})$) also plays a significant role, demonstrating that users tend to visit nearby cities. As k increases, the feature importance of their bottom counterparts, $f_{\text{bottom}}^r(\text{categories})$ and $f_{\text{bottom}}^r(\text{geographic})$, grows. This pattern emerges because as k increases, the review-count differences between top-ranked and bottom-ranked CBSAs diminish. Consequently, for marginal cases (CBSAs near the k-threshold), similarity to bottom-ranked CBSAs suggests they likely belong to the lower end of the top-k. Also, for clear bottom-ranked CBSAs, dissimilarity from top-ranked CBSAs helps confirm their classification. After that, $f_{\text{top}}^r(\text{population})$, shows a tendency of the users to visit cities based on a preference for big or small cities, with its bottom counterpart providing a complementary analysis to it. Then, $f_{\text{top}}^r(\text{education})$, $f_{\text{top}}^r(\text{scenes})$, $f_{\text{top}}^r(\text{voting})$ and $f_{\text{bottom}}^r(\text{voting})$ help provide an overview of the characteristics of the city that are relevant to the problem under study. Finally, other features, such as income, employment and race, along with the bottom equivalents

of the previously mentioned overview-like features, proved less important for this task.

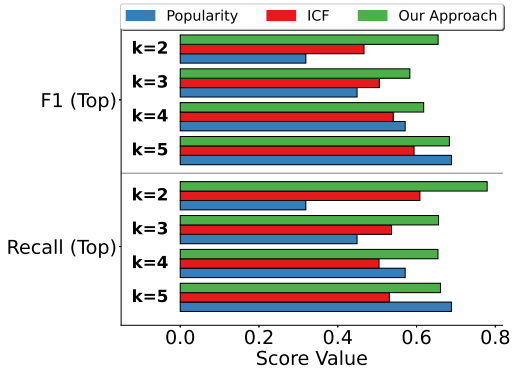


Figure 1: City-Level Model Performance

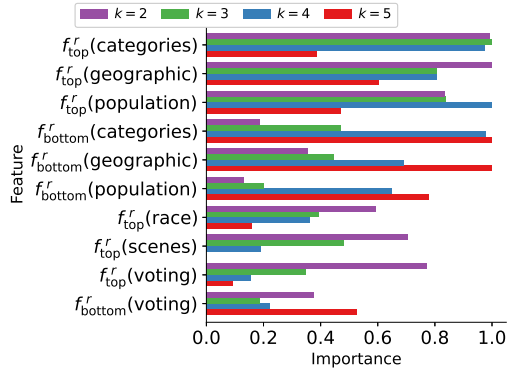


Figure 2: City-Level Feature Importance

6.1.1 Returners vs. Explorers: Divergent Patterns. Similar to [17], we found in our dataset a clear separation between Returners and Explorers. As expected, the percentage of Returners increases with higher values of k: 46.2% for $k = 2$, 57.2% for $k = 3$, 81.6% for $k = 4$, 89.5% for $k = 5$. This trend aligns with the theoretical expectation that users exhibiting higher revisit rates (Returners) become more prevalent as the scope of analysis expands to include more cities. To investigate behavioral differences between these groups, we examine metrics such as the number of CBSAs visited and the distribution of reviews across top CBSAs. Surprisingly, no statistically significant differences emerge in these dimensions. This absence of variation suggests that quantitative measures of visitation frequency or review volume alone may not fully capture the fundamental distinction between Returners and Explorers. Instead, the divergence likely stems from qualitative factors, such as preference heterogeneity or contextual influences, that extend beyond the analyzed data.

With the split of returners and explorers, the model is trained with only the subset of returners and, also, only with the explorers, using the stat=mean in $\mathbf{f}^r$ calculation. The behavioral dichotomy between returners and explorers reveals striking differences (Figures 3 and 4). Explorers demonstrate better predictability (higher recall, except for $k = 6$) and show strong geographic clustering of preferred cities, with f^r_{top}(geographic) as their top feature. In contrast, returners exhibit less geographic dependence but stronger category-based preferences, with f^r_{top}(categories) showing approximately twice the importance compared to explorers. These findings align with their fundamental behavioral differences: *explorers seek novel but geographically concentrated experiences, while returners show consistent category preferences.*

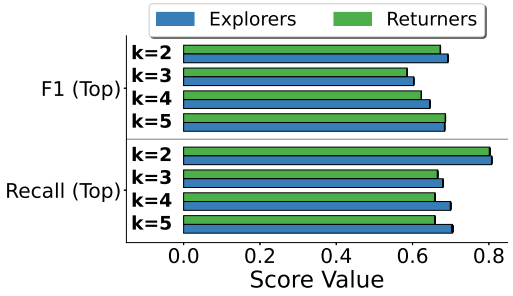


Figure 3: Returners *vs* Explorers Performance (city-level analysis)

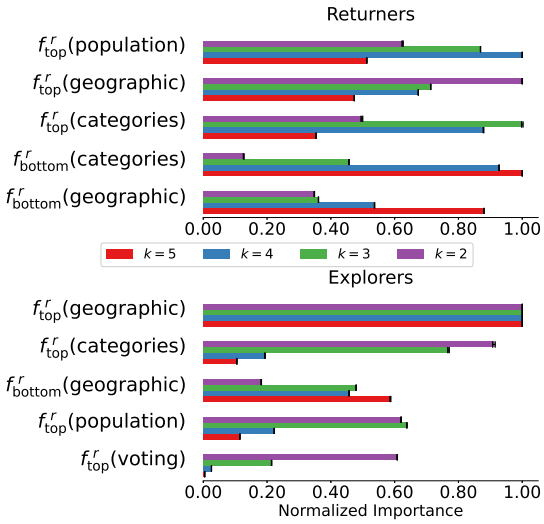


Figure 4: Returners *vs* Explorers Feature Importance (city-level analysis)

6.2 Neighborhood-level Analysis

6.2.1 Zip Code Recommendation Performance. Our analysis at neighborhood-level examines recommendation performance using the feature vector $\mathbf{f}^r$, r as zip codes, across parameters k ranging from 2 to 4 and m from 1 to 3. The results demonstrate that our approach consistently outperforms the baselines, as shown in Figure 5. The only exception occurs when $k = 2$ and m equals either 2

or 3, where users appear to predominantly visit the most popular areas within cities, making the baseline approach competitive in these specific cases.

The feature importance analysis, min-max normalized mean SHAP values, presented in Figure 6 reveals several key insights about neighborhood preferences. The geographic distance features $f^r_{\text{top}}(\text{geographic})$ and $f^r_{\text{bottom}}(\text{geographic})$ emerge among the most significant predictors. The model reveals an unexpected pattern: preferred neighborhoods in one city tend to be geographically closer to lower-rated neighborhoods in other cities. This inverse relationship, though seemingly paradoxical, becomes clearer when viewed through the demonstration system (see Section 7). For instance, in both New York and Los Angeles, preferred neighborhoods are often located in the western parts of the cities. Due to the cities' geography (on opposite sides of the map), some highly rated neighborhoods (west side) in New York are geographically closer to lower-rated neighborhoods (east side) in Los Angeles. However, in the model prediction, the reverse is not necessarily true: highly rated neighborhoods in Los Angeles are not geographically closer to poorly rated ones in New York. The asymmetry (west-NY → east-LA *vs.* west-LA → east-NY) indicates that, although proximity to poorly rated areas in other cities can serve as a proxy for identifying appealing neighborhoods, the relationship is not trivial. The model exploits this complex, directional spatial effect, revealing how geographic relationships interact asymmetrically with neighborhood quality across cities.

Beyond geographic factors, the analysis shows that the feature $f^r_{\text{top}}(\text{categories})$ remains a crucial predictor, indicating that users consistently prefer neighborhoods offering experiences similar to those they've enjoyed in the past. Interestingly, economic compatibility ($f^r_{\text{top}}(\text{income})$) shows greater importance at the neighborhood level than observed in city-level analysis, suggesting users pay closer attention to economic matching when selecting specific areas within cities. Political alignment ($f^r_{\text{top}}(\text{voting})$) and urban character ($f^r_{\text{top}}(\text{population})$) also play significant roles in neighborhood preferences. Additional discriminative power comes from education levels, racial composition, and cultural scene characteristics, which help identify neighborhoods similar to those the user has previously shown interest in.

6.2.2 Behavioral Uniformity at Neighborhood Scale. Unlike city-level patterns, returners and explorers show no significant behavioral differences in neighborhood preferences. This suggests that while users may choose cities differently based on their explorer or returner tendencies, their engagement patterns within cities follow similar dynamics regardless of this classification.

6.3 Robustness Analysis

We conduct several experiments to evaluate the robustness of our model across different hyperparameter configurations and feature representations. First, using median instead of mean in the similarity aggregation yielded comparable performance, though slightly lower interpretability due to loss of distribution sensitivity. We also experimented with adding standard deviation to capture user preference variability, though we restricted this analysis to $k \in \{3, 4, 5\}$

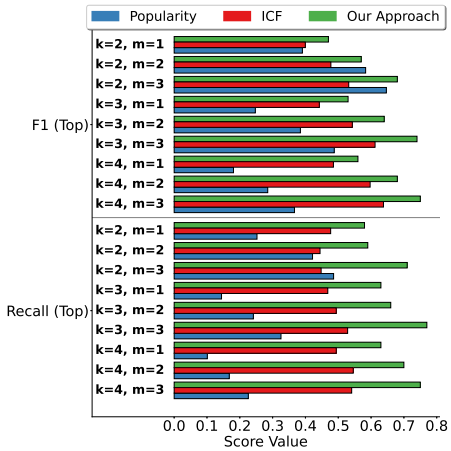


Figure 5: Neighborhood-Level Performance

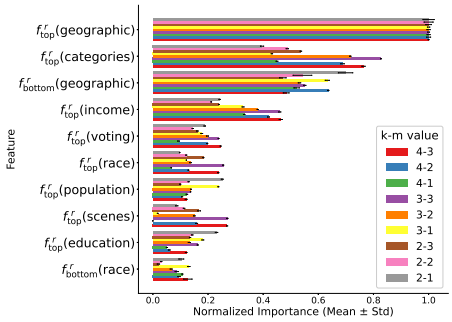


Figure 6: Neighborhood-Level Feature Importance

to ensure meaningful standard deviation comparisons between regions. This improves city-level performance (up to 0.75 recall, see Figure 7), but not at the neighborhood level. This small performance improvement comes at the cost of hindering results interpretability. When comparing with the other models, XGBoost gives similar results, with mean recall (over k-fold CV) of the top class being 0.8, 0.67, 0.63, 0.65 for k from 2 to 5 using the features with mean, and Logistic Regression, gives worse results, with recall, for the top class, being 0.58, 0.57, 0.56, 0.55 for k from 2 to 5, in the city-analysis. In the neighborhood analysis, results follow a similar pattern, with comparable results from XGBoost and worse results from logistic regression. We, also, employ a randomized grid search to fine-tune hyperparameters and enhance model performance, testing ranges for number of trees (100–300), learning rate (0.01–0.1), tree depth (3–10), number of leaves (20–50), minimum child samples (10–30), and L1/L2 regularization (0–0.5). While this tuning leads to modest gains, typically around a 0.2 increase in recall, it introduces complexity with limited benefit in weaker configurations. As these exploratory results are not shown in the main results section, we note that further improvements will likely depend more on feature space refinement than hyperparameter tuning.

To assess minimal data requirements, we train separate models using only top-rated ($\mathbf{f}^r = \mathbf{f}^r_{\text{top}}$) or only bottom-rated ($\mathbf{f}^r = \mathbf{f}^r_{\text{bottom}}$)

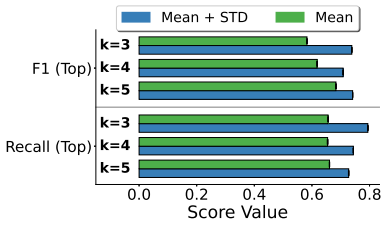


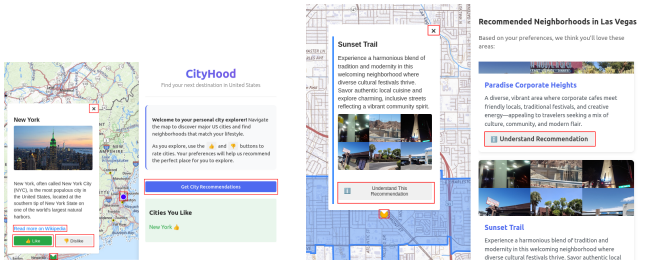
Figure 7: Mean vs Mean+STD Performance Comparison

city features. As expected from our feature importance analysis, the top-only model performed similarly to the full model, though recall degrades at higher k values. The bottom-only model initially shows poor precision but gradually approximates full model performance by $k = 5$. This demonstrates that meaningful recommendations can be generated even with limited user history data.

Finally, we check whether adding an 'is_returner' feature would help predictions. The modest ~ 0.03 recall improvement, consistent across different k values despite the increasing ratio of returners to explorers, along with unchanged precision, suggests that binary returner classification offers limited additional value. Feature importance rankings place 'is_returner' below geographic, categorical, and population factors, indicating that returner behavior is better captured through preference patterns than through explicit labeling. This pattern holds consistently across city and neighborhood analyses, with no notable performance differences between median and standard deviation approaches.

7 TOURIST RECOMMENDATION SYSTEM

To showcase the capabilities of our city- and neighborhood-level tourist recommendation approach, we present CityHood, an interactive and explainable travel recommender system available at [4], which provides personalized suggestions at both city and neighborhood levels based on users' stated preferences and regional characteristics. Users begin by selecting U.S. cities (Core-Based Statistical Areas) they liked or disliked via a map interface (Fig. 8a). Based on this input, the LightGBM model recommends new cities by comparing geographic, socio-demographic, political, and cultural features. Each recommendation includes a natural-language explanation. To generate them, we structure a prompt for a large language model (e.g., GPT-4). The prompt includes: (1) the user's input cities (liked/disliked), (2) the recommended city, and (3) the top contributing features from LIME, along with their raw similarity scores and a brief definition of each feature. There are also options to see these raw values directly in the recommendation interface. After selecting a recommended city, the user proceeds to neighborhood-level refinement, where ZIP codes within previously liked cities can be labeled. Each neighborhood includes a text summary generated using a local LLM (deepseek-r1:32b) and curated images filtered through Places365 (outdoor scenes) and YOLOv8 (filter out selfies and cars) for visual context. The system then recommends neighborhoods in the selected destination city (fig 8b), again offering LIME-backed natural-language explanations and feature-level insights. This two-stage design supports fine-grained interest modeling, explainability, and a more transparent exploration of urban destinations.



(a) City Exploration **(b) Neighboorhod Recommendation**

Figure 8: CityHood (clickable elements in red borders)

8 CONCLUSIONS

In this work, we have proposed a multi-level model for personalized city and neighborhood recommendations that integrates spatial, cultural, demographic, and political features from users' visited places. Our framework outperformed baseline approaches while providing explainable insights into user behavior. Some of the key findings revealed by the analysis: (1) geographic proximity and category similarity are consistently strong predictors of user preferences across all configurations, and (2) the returner-explorer dichotomy shows distinct patterns - with explorers favoring novel but geographically concentrated experiences and returners prioritizing familiar venue categories.

Our study has focused on users with substantial travel histories (visiting at least six CBSAs), which may limit generalizability to casual tourists with sparser data. Future work could address this by developing models that infer preferences from limited observations. Additionally, incorporating Natural Language Processing to analyze user reviews may improve the identification of high- and low-interest areas and enhance recommendation quality. Exploring temporal dynamics (e.g., seasonal trends, life-stage changes) and trip purposes (leisure vs. business) could further refine personalization. Cross-platform validation would also help to ensure robustness across diverse urban contexts. Applying this approach to other countries to assess recommendation performance and feature importance is also an important future step. Also, a user survey on the demo could further evaluate its usability, particularly the LLM-based explanations. While our baseline methods were chosen for their simplicity and interpretability, adapting more complex state-of-the-art POI recommenders to our setting—focused on aggregated geographic areas—would require substantial modifications and is thus left as promising future work.

9 ACKNOWLEDGMENTS

This research is partially supported by the SocialNet project (process 2023/00148-0 of FAPESP), by CNPq (proc. 314603/2023-9, 441444/2023-7, 409669/2024-5, and 444724/2024-9) and the INCTs ICoNIoT and TILD-IAR funded by CNPq (proc. 405940/2022-0 and 408490/2024-1) and CAPES FinanceCode 88887.954253/2024-00.

[4] https://cityhood.vercel.app/

REFERENCES

[1] Omid R. Abbasi and Ali A. Alesheikh. 2023. A Place Recommendation Approach Using Word Embeddings in Conceptual Spaces. *IEEE Access* 11 (2023), 11871–11879.

[2] Ramesh Baral, S. S. Iyengar, Xiaolong Zhu, Tao Li, and Pawel Sniatala. 2019. HiRecS: A Hierarchical Contextual Location Recommendation System. *IEEE Trans. on Comp. Social Systems* 6, 5 (2019), 1020–1037.

[3] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. 2006. The scaling laws of human travel. *Nature* 439, 7075 (2006), 462–465.

[4] Yong Chen, Xiqun Michael Chen, and Ziyou Gao. 2024. Toward equitable, transparent, and collaborative human mobility computing for smart cities. *The Innovation* 5, 5 (2024).

[5] Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Z Sui. 2012. Fused matrix factorization with geographical and social influence in location-based social networks. *Proc. of the AAAI* (2012).

[6] Luis M De Campos, Juan M Fernández-Luna, and Juan F Huete. 2024. An explainable content-based approach for recommender systems: a case study in journal recommendation for paper submission. *User Modeling and User-Adapted Interaction* 34, 4 (2024), 1431–1465.

[7] Keyan Gao, Xu Yang, Chenxia Wu, Tingting Qiao, Xiaoya Chen, Min Yang, and Ling Chen. 2020. Exploiting Location-Based Context for POI Recommendation When Traveling to a New Region. *IEEE Access* 8 (2020), 52404–52412.

[8] Marta C Gonzalez, Cesar A Hidalgo, and Albert-László Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–782.

[9] Fernanda R. Gubert, Gustavo H. Santos, Myriam Delgado, Daniel Silver, and Thiago H. Silva. 2025. Culture Fingerprint: Identification of Culturally Similar Urban Areas Using Google Places Data. In *Proc. of ASONAM*. Rende, Italy, 286–297.

[10] Jiacheng Li, Jingbo Shang, and Julian McAuley. 2022. UCTopic: Unsupervised Contrastive Learning for Phrase Representations and Topic Mining. In *Proc. of the ACL'22*. ACL, Dublin, Ireland.

[11] Yuebing Liang, Yichao Liu, Xiaohan Wang, and Zhan Zhao. 2024. Exploring large language models for human mobility prediction under public events. *Comp., Envir. and Urban Systems* 112 (2024), 102153.

[12] Xi Liu, Clio Andris, and Bruce A. Desmarais. 2019. Migration and political polarization in the U.S.: An analysis of the county-level migration network. *PLOS ONE* 14, 11 (Nov. 2019), e0225405.

[13] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. (2017), 4768–4777.

[14] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Proc. of NIPS*. Red Hook, NY, USA, 4768–4777.

[15] Steven Manson, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2024. National Historical Geographic Information System: Version 19.0.

[16] Ludovico Napoli, Márton Karsai, and Esteban Moro. 2024. One rule does not fit all: deviations from universality in human mobility modeling. *arXiv preprint arXiv:2410.23502* (2024).

[17] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. 2015. Returners and explorers dichotomy in human mobility. *Nature Communications* 6, 1 (Sept. 2015).

[18] Niloofar Ranjbar, Saeedeh Momtazi, and MohammadMehdi Homayoonpour. 2024. Explaining recommendation system using counterfactual textual explanations. *Machine Learning* 113, 4 (2024), 1989–2012.

[19] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proc. of KDD*. San Francisco, USA.

[20] Gustavo H. Santos, Fernanda R. Gubert, Myriam Delgado, and Thiago H. Silva. 2025. Modeling Interest Networks in Urban Areas: A Comparative Study of Google Places and Foursquare Across Countries. *Journal of Internet Services and Applications* 16, 1 (Mar. 2025), 25–42.

[21] Markus Schläpfer, Lei Dong, Kevin O'Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, Carlo Ratti, and Geoffrey B West. 2021. The universal visitation law of human mobility. *Nature* 593, 7860 (2021), 522–527.

[22] Helen C. Mattos Senefonte, Myriam Regattieri Delgado, Ricardo Lüders, and Thiago H. Silva. 2022. PredicTour: Predicting Mobility Patterns of Tourists Based on Social Media User's Profiles. *IEEE Access* 10 (2022), 9257–9270.

[23] Ruoxi Wang, Xinyuan Zhang, and Nan Li. 2022. Zooming into mobility to understand cities: A review of mobility-driven urban studies. *Cities* 130 (2022), 103939.

[24] Fengli Xu, Yong Li, Depeng Jin, Jianhua Lu, and Chaoming Song. 2021. Emergence of urban growth patterns from human mobility behavior. *Nature Computational Science* 1, 12 (2021), 791–800.

[25] An Yan, Zhankui He, Jiacheng Li, Tianyang Zhang, and Julian McAuley. 2023. Personalized Showcases: Generating Multi-Modal Explanations for Recommendations. In *Proc. of the SIGIR*. Taipei, Taiwan, 5 pages.