Mining Points-of-Interest Data to Predict Urban Inequality: Evidence from Germany and France

Manuel Ganter[1], **Malte Toetzke**[1], **Stefan Feuerriegel**[1 2]
[1] ETH Zurich
[2] LMU Munich
mganter@student.ethz.ch, mtoetzke@ethz.ch, feuerriegel@lmu.de

Abstract

Reducing inequality is a major goal of the Sustainable Development Goals. Inequality is many-sided and often appears across geographic boundaries. Urban inequality refers to inequality between urban neighborhoods. Despite close distances, it reveals considerable disparities in income level, unemployment rates, and other socio-economic indicators and is highly dangerous for democratic societies. However, little is known about determinants indicating urban inequality. Here, we propose to explain urban inequality based on point-of-interest (POI) data from the online platform OpenStreetMap. For this, we leverage machine learning to predict three major indicators of urban inequality, namely, unemployment rate, income level, and foreign national rate. We evaluate our machine learning approach using POI data for neighborhoods in Paris, Lyon, Marseille, Berlin, Hamburg, and Bremen. We find: (1) POIs are highly predictive of intra-city inequality explaining up to 75% of out-of-sample variance of urban inequality. (2) POIs generalize across cities and, thereby, can help to explain urban inequality in other cities, where no socio-economic data is available. (3) Important POIs for the prediction model are, e.g., banks and playgrounds. To the best of our knowledge, our work is the first to show urban inequality through POIs. As such, POIs can be used to infer granular mappings of urban inequality and thereby provide cost-effective evidence for policy-makers.

Introduction

Inequality is a global issue, limiting prosperity and opportunities of individuals. Despite widespread public awareness, differences between rich and poor continue to increase dramatically across the globe, especially within developed countries. The destabilizating forces of inequality put modern societies at risk. Hence, there is great urgency to reduce inequality, which has become a top priority for public policy and constitutes a major objective of the Sustainable Development Goals (UNICEF Office of Research - Innocenti 2017; UN General Assembly 2015; Sachs et al. 2021).

Inequality at the macro level (e.g., between countries and global regions) is evident and well-measured. However, inequality also appears at the micro level between urban neighborhoods (Bourguignon and Scott-Railton 2015; Elgar, Stefaniak, and Wohl 2020). This type of inequality is commonly referred to as **urban inequality**. Urban inequality exposes considerable within-city variation in socio-economic indicators such as unemployment rates, income levels, and foreign national rates (Cassiers and Kesteloot 2012). As an example, compare Manhattan and Bronx in New York City: both neighborhoods are in close proximity; yet, the average household income in Bronx is 54 % lower than it is in Manhattan. Of note, urban inequality is a particularly critical issue in Western parts of the world (Musterd and Ostendorf 2013). Urban inequality drives higher crime rates and lower population growth and can destabilize democratic societies (Glaeser, Resseger, and Tobio 2008). Hence, reducing urban inequality is of utmost importance to policy-makers.

Providing evidence on urban inequality is crucial to develop effective policies for reducing urban inequality. However, urban inequality is often not measured and thus not directly visible. Furthermore, little is known about determinants explaining urban inequality. Therefore, it is important to improve our understanding on how inequality is manifested through characteristics of different neighborhoods. By identifying common determinants of urban inequality, one can develop counterbalancing measures aiming at mitigating disadvantages in specific neighborhoods. For instance, if public transport is linked to higher levels of urban inequality, policy-makers can act by expanding public transport networks in particularly poor neighborhoods. Such associations with public transport have been found for other social phenomena such as crime (Kadar et al. 2020), and it might thus be likely that one can leverage similar patterns for urban inequality.

In this paper, we propose to explain urban inequality based on point-of-interest (POI) data from the online platform OpenStreetMap. Generally, POIs refer to locations of interest for individuals or businesses (e.g., museum, pharmacy, hotel). Here, we hypothesize that POIs are associated with urban inequality represented by various socio-economic indicators. This association can be hypothesized due to diverse underlying causal effects. For example, a lack of kindergartens and playgrounds in specific neighborhoods may complicate child care and increase the burden of residential parents, which negatively impacts family planning and employment possibilities for them. A lack of kindergartens and playgrounds may further prevent young families

from moving to specific neighborhoods.

Our results show that a large share of urban inequality can be explained at very granular levels of urban neighborhoods using machine learning and POI data from the online platform OpenStreetMap. More specifically, we predict various socio-economic indicators characteristic for urban inequality, namely *unemployment rate*, *income level*, and *foreign national rate* in six different cities (Paris, Lyon, Marseille, Berlin, Hamburg, and Bremen).

Our results help to map urban inequality and provide important evidence for policy makers addressing underlying issues of urban inequality. All results are corroborated by extensive robustness checks (e.g., different machine learning models and different feature engineering). In summary, we make the following **contributions**[1]:

1. We show the predictive power of POIs from OpenStreetMap for estimating micro-level variation in urban inequality across different cities and different socio-economic indicators. To the best of our knowledge, this is the first work to explain urban inequality through POIs.
2. We confirm that POIs as inequality predictors – to some extent – generalize across cities and explain inequality for out-of-sample cities.
3. We establish *which* POIs have the largest predictive power and thus are important determinants characterizing urban inequality.

Related Work

Inequality: Reducing inequality is a major objective of the United Nations' Sustainable Development Goals (UN General Assembly 2015). In the literature, urban inequality is typically measured via the following three socio-economic indicators: (1) unemployment rate (Morrison 2005; Todaro 1969; Xue and Zhong 2003), (2) income level (Chakravorty 1996; Glaeser, Resseger, and Tobio 2009; Schaffar 2008), and (3) foreign national rate (Benassi, Lipizzi, and Strozza 2019; O'Loughlin 1980; Strozza et al. 2016). Other dimensions of inequality (e.g., health or education) are typically correlated with one of these indicators (and thus not included in this study).

Policy-makers require evidence with sufficient granularity (cf. Toetzke, Banholzer, and Feuerriegel 2022), such as granular mappings that show the spatial variation in inequality (Puttanapong, Martinez, and Addawe 2020; Zhao et al. 2019), in order to reduce urban inequality. Such mappings are commonly available at macro-level (e.g., between countries). However, there is a lack of inequality mappings at micro-level. One work reports a mapping of inequality at a regional level (Dong, Ratti, and Zheng 2019), yet this is – by far – of much lower resolution than a neighborhood level. To fill this gap, we later deliver a high-resolution mapping of urban inequality that is at the neighborhood level (= administrative post code unit, or smaller).

[1]The code of all analyses is publicly available via https://github.com/ManuelGanter/urban-inequality.

POI data: POIs have become a widespread source of geo-tagged data, especially thanks to the success of online platforms such as Google Maps or OpenStreetMap. POIs refer to locations, which one might find useful or important (Rae et al. 2012). Typical examples are restaurants, bars, pharmacies, hospitals, or public buildings. POIs are known as a decisive factor for explaining urban phenomena (Cranshaw et al. 2012; Hidalgo, Castañer, and Sevtsuk 2020; Hristova et al. 2016; Noulas et al. 2012; Taylor, Lim, and Chan 2018; Tschernutter and Feuerriegel 2021; Yuan, Zheng, and Xie 2012). In particular, they are used for modeling social science phenomena, such as education (Miller 2012), racial segregation (Zenk et al. 2005), housing prices (Fu et al. 2019; Tang et al. 2018; Xiao et al. 2017), or crime (Kadar and Pletikosa 2018; Kadar, Maculan, and Feuerriegel 2019; Wang et al. 2016). In other context, they are used to infer human activities from GPS data (Furletti et al. 2013) or to plan touristic tours within cities (Brilhante et al. 2013). However, to the best of our knowledge, no paper has used POIs for modeling urban inequality (i.e., modeling the strong variations in unemployment rates, income levels, and foreign national rates within cities).

POI modeling: POI modeling typically makes use of feature engineering, which can be grouped according to (1) distance-based features and (2) density-based features as follows:

1. *Distance-based features* make inferences based on the distance to the closest POI of the same type (e.g., distance to closest hospital or closest bar). The intuition behind distance-based features is that the walking time until reaching a respective type of POI can be relevant for individuals (e.g., pharmacies).
2. *Density-based features* aggregate POIs in a certain area of interest (e.g., within a spatial radius), for instance, by counting the overall frequency (e.g., Kadar and Pletikosa 2018; Kadar, Maculan, and Feuerriegel 2019; Karamshuk et al. 2013; Puttanapong, Martinez, and Addawe 2020; Tingzon et al. 2019). The intuition behind density-based features is that some areas benefit from having many POIs of the same type in close vicinity (e.g., bars in nightlife districts). Here, a common choice is to count the number of POIs within a specific radius when making inferences in a city context (e.g., Hummler, Naumzik, and Feuerriegel 2022; Naumzik, Zoechbauer, and Feuerriegel 2020).

In line with existing literature, we adopt both ways of feature engineering – i.e., (1) distance-based and (2) density-based POI features – in our analysis.

Inferences from geo-tagged data have two key degrees of freedom. (1) An appropriate spatial resolution has to be chosen. For instance, many urban phenomena are recorded based on data from official administrative units (e.g., Zenk et al. 2005). Analogously, we make use of neighborhoods as defined by official administrative units. (2) An appropriate machine learning (ML) model must be selected. When fed with features from POI data, granular associations between POIs and urban phenomena are learned.

Country	City	#Neighborh.	#Buildings	#POIs
France	Paris	868	88,701	106,908
	Lyon	177	36,880	33,959
	Marseille	345	154,434	52,159
Germany	Berlin	122	404,430	156,908
	Hamburg	125	278,350	72,215
	Bremen	129	165,448	29,710

Table 1: Number of neighborhoods, buildings, and POIs per city.

A common limitation of previous works is that the above feature engineering is computed with respect to the "center" of a spatial unit. For example, let us assume a setting where inferences are made at neighborhood level as the spatial unit of interest. For this, let n denote a neighborhood where the center is given by $center(n)$ such as the centroid of n. Then the distance-based feature for a POI p is $\left\|(p_{\text{lat}}, p_{\text{lon}})^T - center(n)\right\|$. However, this calculation scheme ignores where people reside. As a remedy, we later develop a tailored calculation scheme in which we compute the feature with respect to **all** buildings in each neighborhood n. That is, we extract all buildings b located in n, and, then, compute the distance between each POI p and each building b.

Methods

Data

We focus on France and Germany, the two countries with the highest population in the European Union. Within these countries, we choose three of the major cities, namely Paris, Lyon, and Marseille (in France), and Berlin, Hamburg, and Bremen (in Germany). Our dataset includes all neighborhoods of the selected cities. Table 1 gives an overview of the number of neighborhoods and other summary statistics.

Target variables: We collected data on urban inequality at neighborhood level with respect to (1) unemployment rate, (2) income level, and (3) foreign national rate as follows:[2] (1) The unemployment rate measures the share of dependent civilian labor force without a job in a neighborhood. (2) The income level refers to the median income (in euro per month) per neighborhood. (3) The foreign national rate reports the share (in percent) of non-national people in each neighborhood. Summary statistics for the three inequality indicators can be found in Appendix . For notation, we refer to the neighborhoods via $n = 1, \ldots, N$ and to an inequality indicator in neighborhood n via y_n.

Predictors: We collected POI data from OpenStreetMap (OpenStreetMap 2021). Our data includes 118 types t of POIs (e.g., bar, restaurant, hospital). A POI p is a 3-tuple $(p_{\text{lat}}, p_{\text{lon}}, p_{\text{type}})$, where p_{lat} is the latitude, where p_{lon} is the longitude, and where p_{type} refers to a type $t \in \mathcal{T}$ (i.e., whether the POI is a school, bar, pharmacy, etc.). For simplicity, we write $\forall p \in \mathcal{P}(t)$ when indexing over all POIs of type t. The POI data were filtered for the six selected cities using the geographic polygons. To assure that the we avoid a structural bias at the borders of the cities, we buffer the individual polygons with a buffer of 3000 m to allow POIs at borders to be included in our dataset.

For our feature engineering, we additionally collected a comprehensive dataset of all buildings.[3] Specifically, we retrieved all residential or commercial buildings with a ground area of more than $25\,\mathrm{m}^2$ in each neighborhood n (we removed buildings with a smaller ground area since these typically do not house residents but represent, e.g., phone boxes and bus stops). We represent each building b by its geographic coordinates (b_{lat}, b_{lon}). For simplicity, we write $\forall b \in \mathcal{B}_n$ when indexing over all buildings within the spatial boundaries of neighborhood n.

Standard preprocessing was applied as follows. (1) We followed the recommendations in European Environment Agency (2013) in order to compute distances in the unit of meter. That is, we transform the POI and building locations from the projection EPSG 4326 to the projection EPSG 3035. (2) We merged some POI types for better interpretability. For example, we replaced both "college" and "university" by the latter; we replaced "beergarden" by "restaurant", "food court" by "fast food", etc. Details are in the appendix.

ML Approach

In our ML approach, we combine (1) feature engineering and (2) a prediction model. We later compare the prediction performance across different variants of the ML approach (i.e., we vary both feature engineering and the underlying prediction model). Key to our ML approach is a strong regularization, which is owed to the fact that we have many POI features relative to the number of neighborhoods.

(1) Feature engineering: We apply feature engineering to the POI data as follows. In line with the above literature on POI modeling (see Sec.), we compute density- and distance-based POI features (see Fig. 1 for an example).

Distance-based POI features: The distance features are used to reflect that, for some POIs (e.g., hospitals), the distance from buildings in a neighborhood to the closest POI is important. For this, we follow a two-step approach: (i) For each building b, we compute the distance between building b and the next POI p of type t. (ii) We then aggregate the distances per building in each neighborhood by taking the mean.

Formally, we first compute the shortest distance $\delta_{b,t,n}$ between building b and a POI p of type t, for all buildings b and POI types $t \in \mathcal{T}$. It is important to note that we evaluate the distance to all POIs within the city and not just the POIs within the neighborhood of the respective building. That is, a POI p which is not located in the neighborhood n of the

[2]Source: PriceHubble AG: www.pricehubble.com; data originally collected from the National Institute of Statistics and Economic Studies (INSEE) in France and from Infas360 in Germany. For Germany, the income level is inferred by the weighted average of the underlying income ranges. Data is from 2015 (the most recent year for which such census data are available).

[3]Source: PriceHubble AG: www.pricehubble.com

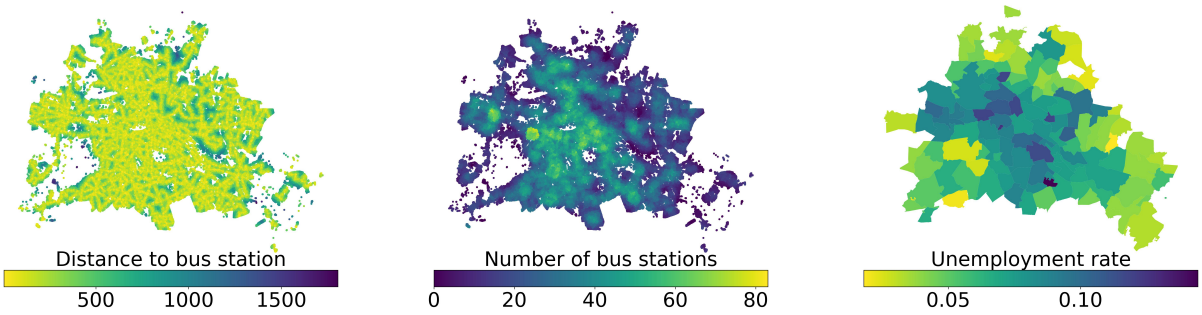

Figure 1: Example for POIs of type"bus stop" in the city of Berlin. *Left:* Spatial point plot where each dot is a building, colored by the distance to the next bus stop (in meters). *Center:* Spatial point plot where each dot is a building, colored by the number of bus stops within 1000 m distance. *Right:* For comparison, urban inequality as measured by unemployment rate per neighborhood.

respective building b can still be the one with the shortest distance. We use the Euclidean distance for travel distances as best-practice (Luo and Wang 2003). We thus yield

$$\delta_{b,t,n} = \underset{\forall p \in \mathcal{P}(t)}{\arg\min} \left\| \begin{pmatrix} p_{\text{lat}} \\ p_{\text{lon}} \end{pmatrix} - \begin{pmatrix} b_{\text{lat}} \\ b_{\text{lon}} \end{pmatrix} \right\|_2 \tag{1}$$

for all buildings b and all POI types $t \in \mathcal{T}$, where $\mathcal{P}(t)$ represents a set of POIs of type t, i.e.,

$$\mathcal{P}(t) = \left\{ p \,\middle|\, p = (p'_{\text{lat}}, p'_{\text{lon}}, p'_{\text{type}}), t = p'_{\text{type}} \right\}. \tag{2}$$

Afterwards, we compute the distance-based POI feature for each neighborhood by averaging over all buildings b, i.e.,

$$d_{t,n} = \frac{1}{|\{b \,|\, b \in \mathcal{B}_n\}|} \sum_{b \in \mathcal{B}_n} \delta_{b,t,n} \tag{3}$$

for all POI types t and all neighborhoods n, where $\mathcal{B}_n$ is the set of buildings b in the respective neighborhood n. The feature $d_{t,n}$ is the input to the prediction model, thereby capturing the average distance to the next POI of type t.

Density-based POI features: For some POIs, it is likely that not the distance but the frequency in a spatial area is of relevance (e.g., number of bars within close spatial proximity). Density-based features are computed by the number of POIs within a radius of $\theta = 1000$ meters around each building. (In our robustness check, we repeat our analysis also with other radii θ but arrive at consistent conclusions.). Formally, we proceed as follows. First, we determine whether a POI p is within the area of interest via

$$a_{b,p,n} = \begin{cases} 1, & \text{if } \left\| \begin{pmatrix} p_{\text{lat}} \\ p_{\text{lon}} \end{pmatrix} - \begin{pmatrix} b_{\text{lat}} \\ b_{\text{lon}} \end{pmatrix} \right\|_2 \leq \theta, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

where $a_{b,p,n}$ equals 1 if POI p is within a distance θ of building b. Otherwise, $a_{b,p,n}$ is zero. Afterwards, we compute the frequency of all POIs of type t around each building via

$$f_{b,t,n} = \sum_{p \in \mathcal{P}(t)} a_{b,p,n}, \tag{5}$$

for all buildings b. Similar to the distance-based feature, we finally compute the average score of the buildings within each neighborhoods as

$$s_{t,n} = \frac{1}{|\{b \,|\, b \in \mathcal{B}_n\}|} \sum_{b \in \mathcal{B}_n} f_{b,t,n} \tag{6}$$

for all POI types t and all neighborhoods n, where $\mathcal{B}_n$ is the set of buildings b in the respective neighborhood n. $s_{t,n}$ is the density-based POI feature that we put in the prediction model.

Distance-to-city-hall POI: We further control for whether a neighborhood is close or far away from the city center by adding a feature that measures the average distance to the city center. Here, we approximate the latter by the POI that represents the city hall. Formally, we compute

$$\psi_n = \frac{1}{|\{b \,|\, b \in \mathcal{B}_n\}|} \sum_{b \in \mathcal{B}_n} \delta_{b,\texttt{City Hall}}, \tag{7}$$

where $\delta_{b,\texttt{City Hall}}$ is the distance from building shape b to the city hall in the respective city and ψ_n is the resulting POI-based feature.

As a result, we have 237 features for each neighborhood (i.e., 118 distance-based, 118 density-based, and one distance-to-city-hall POI feature).

(2) Prediction model: We choose predictions models that come with comparatively strong regularization to prevent overfitting (since we have more features than observations (neighborhoods)). These are (i) lasso regression (Tibshirani 1996), where a L1-norm penalty on the coefficients performs an implicit feature selection; and (ii) regularized gradient boosting (XGBoost) (Chen and Guestrin 2016), where the underlying trees are regularized by taking impurity and the model complexity into account. By comparing lasso and XGBoost, we can further assess the relative gain in predictive power due to handling non-linearities.

Variants: We compare the following variants of the above ML approach:

- **M1**: a standard lasso regression is applied to reduce the feature space via its regularization.
- **M2**: first performs feature selection via a lasso; then feds the predictors into XGBoost.
- **M3**: first performs dimensionality reduction via a principal component analysis (PCA); then makes predictions from the principal components via a lasso.
- **M4**: first performs dimensionality reduction via a lasso that is applied to the PCA output; then makes predictions from the selected principal components via XGBoost.
- **M5**: an ensemble learner combining the models M2 and M4 using the mean to make predictions.

The implementation details as well the details on the parameter tuning we performed during the training process of our model pipeline can be found in Appendix .

Evaluation Setting

Our evaluation setting is a two-fold, namely (1) an intra-city evaluation and (2) a cross-city evaluation, as detailed in the following.

(1) Intra-city evaluation: We perform an **intra-city evaluation** where we test how well POIs predict urban inequality within a given city. Here, we train and evaluate the ML approach separately with data from each city. For this, we apply a 80/20 train/test split to the neighborhoods in each city.

(2) Cross-city evaluation: We perform a **cross-city evaluation** to test how well predictions based on POIs can be generalized across cities. Here, we evaluate three different strategies:

(i) *City-to-city*: For each of the six cities, we train the ML approach on one city and then evaluate it on another city. This allows us to quantify the generalizability between two specific cities of interest (e.g., how POIs transfer from Paris to Berlin).

(ii) *Leave-one-city-out*: We train the ML approach using 5 out of the 6 cities and evaluate it on the hold-out city. We repeat this for each of the different cities. Eventually, we macro-average the performance metrics for each indicator.

(iii) *Leave-one-city-out-within-country*: We train the ML approach using 2 out of the 3 cities from a given country and evaluate it on the hold-out city. We repeat this for each of the different cities and for both countries. Eventually, we macro-average the performance metrics, for each indicator. Different from the above, we explicitly account for the different countries and thus examine to what extent POIs generalize *within* countries.

Baseline: For better comparability, we report the performance improvement in mean squared error (MSE) compared to a naïve baseline without POIs, which is given by the in-sample mean. As such, we can measure the predictive power due to POIs by computing $\frac{\text{MSE}_{\text{baseline}} - \text{MSE}_{\text{variant}}}{\text{MSE}_{\text{baseline}}} \cdot 100\,\%$.

Results

Intra-City Prediction

The prediction performance for the intra-city setting is reported in Table 2 and 3. Our results show that the ML approaches (M1–M5) outperform the naïve baseline, thus establishing the predictive power of POIs. For most cities, the ensemble learner (M5) achieves the best prediction results for urban inequality indicators. It reaches an average out-of-sample R^2 across all cities between 0.395 and 0.528. (Hence, we use M5 later for answering RQ2). The corresponding improvements in MSE compared to the naïve baseline amount to 43.46% to 55.08%. The overall improvement across all cities is statistically significant ($p < 0.001$; based on t-tests on the residuals). Importantly, similar patterns can be observed for the other variants (M1–M4).

We make further observations: (1) For all city-indicator combinations, ML is able to explain a considerable amount of variance in the target variable. For some cities, the results are very good with an explained variance of over 75% and consistent improvements over the naïve baseline. (2) We find that the ensemble learner (M5) achieves the overall best results. This may be expected due to the fact that it has access to other ML variants. (3) We cannot infer a clear ranking among the other ML variants. Some variants, like M4, are not as strong as others, which can be attributed to the excessive dimensionality reduction in the feature space. (4) The prediction performance for unemployment rate is of similar magnitude across cities. The explained variance of the best models range from 35.9% to 55%. This implies that it can be successfully modeled on the basis of POIs for all cities in our study. (5) In contrast, a larger variability is observed for the other target variables, namely income level and foreign national rate. For instance, for foreign national rate, the explained variance of M5 ranges from 26% (for Hamburg) to 74% (for Marseille). This suggests that, for these indicators, the role of POIs may differ across cities.

Cross-City Evaluation

City-to-city: Here, we use M5 due to its best average performance for the intra-city predictions. Fig. 2 reports the prediction results when training on one city and evaluating on another to study the generalizability of POIs from one city to another. Promising results are seen in the top left and the bottom right quadrant. This indicates that a transfer of POIs between cities of the same country is more effective. For example, using the model from Hamburg to predict the unemployment rate in Bremen allows us to explain 32% of variance. However, in few cases, a transfer also works well across country borders. For example, the prediction of the foreign national rate in Berlin based on the model of Marseille can explain up to 40% of variance in the data.

Leave-one-city-out: For better comparison of the results, we again use model M5 in the following. The results are shown in Table 4. On average, the trained models explain between 12% and 20% of the out-of-sample variance in the three socio-economic indicators representing urban inequality. While this is still an improvement compared to the baseline of using the in-sample average. Using t-tests on the

Inequality indicator	City	Naïve baseline			M1				M2			
		MSE	RMSE	R^2	MSE	RMSE	R^2	Impr.	MSE	RMSE	R^2	Impr.
Unemployment rate	Paris	0.002	0.041	-0.014	0.001	0.034	0.322	33.179	0.001	0.033	0.359	36.797
	Lyon	0.003	0.051	-0.002	0.002	0.045	0.227	22.838	0.002	0.039	0.412	41.320
	Marseille	0.011	0.104	-0.001	0.006	0.080	0.416	41.680	0.007	0.084	0.351	35.214
	Berlin	0.001	0.029	0.000	0.001	0.024	0.305	30.558	0.001	0.024	0.321	32.081
	Hamburg	0.000	0.019	-0.124	0.000	0.015	0.326	40.074	0.000	0.013	0.441	50.268
	Bremen	0.002	0.044	-0.043	0.001	0.034	0.377	40.244	0.001	0.029	0.528	54.756
Income level	Paris	840140	916	-0.006	241064	490	0.711	71.307	206929	454	0.752	75.370
	Lyon	244401	494	-0.001	139584	373	0.428	42.887	116461	341	0.523	52.348
	Marseille	361248	601	-0.001	91614	302	0.746	74.639	92147	303	0.745	74.492
	Berlin	731046	855	-0.001	603192	776	0.174	17.489	439543	662	0.398	39.875
	Hamburg	667237	816	-0.043	407304	638	0.364	38.957	418910	647	0.345	37.217
	Bremen	626827	791	-0.205	420229	648	0.192	32.959	372419	610	0.284	40.587
Foreign national rate	Paris	0.003	0.054	-0.022	0.002	0.045	0.288	30.350	0.002	0.043	0.333	34.698
	Lyon	0.002	0.050	-0.002	0.002	0.043	0.265	26.653	0.002	0.043	0.249	25.032
	Marseille	0.009	0.096	-0.026	0.003	0.052	0.696	70.372	0.002	0.048	0.746	75.275
	Berlin	0.004	0.063	-0.034	0.001	0.038	0.615	62.772	0.002	0.040	0.583	59.713
	Hamburg	0.005	0.072	-0.261	0.003	0.055	0.264	41.669	0.005	0.068	-0.116	11.529
	Bremen	0.007	0.082	-0.013	0.005	0.073	0.188	19.853	0.005	0.068	0.290	29.928

Table 2: Prediction performance across different cities and indicators capturing urban inequality. Here, the naïve baseline is without any access to POIs, while M1 and M2 leverage POIs for prediction.

Inequality indicator	City	M3				M4				M5			
		MSE	RMSE	R^2	Impr.	MSE	RMSE	R^2	Impr.	MSE	RMSE	R^2	Impr.
Unemployment rate	Paris	0.001	0.033	0.347	35.577	0.001	0.035	0.281	29.095	0.001	0.033	0.354	36.335
	Lyon	0.002	0.045	0.228	22.908	0.002	0.039	0.407	40.835	0.002	0.039	0.423	42.387
	Marseille	0.006	0.078	0.444	44.439	0.006	0.076	0.463	46.359	0.006	0.076	0.469	46.930
	Berlin	0.001	0.024	0.304	30.441	0.001	0.023	0.377	37.697	0.001	0.023	0.384	38.399
	Hamburg	0.000	0.016	0.254	33.660	0.000	0.015	0.347	41.923	0.000	0.013	0.441	50.257
	Bremen	0.001	0.032	0.459	48.184	0.001	0.032	0.445	46.799	0.001	0.029	0.551	56.941
Income level	Paris	237427	487	0.716	71.740	248646	498	0.702	70.404	213180	461	0.745	74.626
	Lyon	125048	353	0.488	48.835	114525	338	0.531	53.140	103497	321	0.576	57.653
	Marseille	86080	293	0.762	76.171	117659	343	0.674	67.430	89192	298	0.753	75.310
	Berlin	933634	966	-0.278	-27.712	896577	946	-0.228	-22.643	513697	716	0.297	29.731
	Hamburg	422752	650	0.339	36.641	401822	633	0.372	39.778	352175	593	0.450	47.219
	Bremen	430742	656	0.172	31.282	375476	612	0.278	40.099	338858	582	0.348	45.941
Foreign national rate	Paris	0.002	0.045	0.286	30.146	0.002	0.045	0.285	30.013	0.002	0.043	0.339	35.326
	Lyon	0.002	0.043	0.259	26.076	0.002	0.043	0.268	26.926	0.002	0.042	0.275	27.669
	Marseille	0.002	0.050	0.722	72.938	0.003	0.058	0.621	63.060	0.003	0.050	0.717	72.405
	Berlin	0.002	0.041	0.552	56.644	0.001	0.037	0.637	64.860	0.001	0.036	0.665	67.604
	Hamburg	0.004	0.063	0.050	24.711	0.005	0.074	-0.317	-4.445	0.004	0.063	0.059	25.432
	Bremen	0.005	0.073	0.184	19.423	0.005	0.069	0.272	28.074	0.005	0.067	0.315	32.321

Table 3: Prediction performance across different cities and indicators capturing urban inequality for models M3 – M5

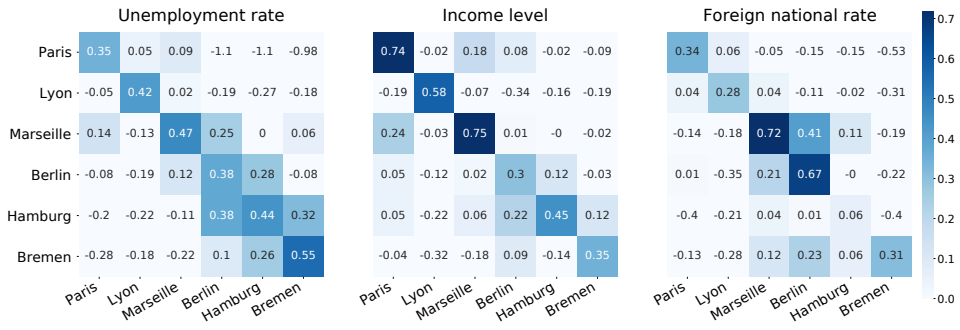

Figure 2: Out-of-sample explained variance (R^2) of city-to-city prediction.

	Unemployment rate	Income level	Foreign nat. rate
Paris	0.13	0.10	0.12
Lyon	–0.00	–0.06	–0.01
Marseille	0.25	0.19	0.30
Berlin	0.34	0.20	0.32
Hamburg	0.33	0.20	0.19
Bremen	0.23	0.13	0.01
Mean	0.21	0.13	0.16

Better: larger values

Table 4: Out-of-sample R^2 for the *leave-one-city-out* evaluation across different target variables (i.e., columns).

	Unemployment rate	Income level	Foreign nat. rate
Paris	0.12	–0.00	–0.04
Lyon	0.08	0.04	–0.02
Marseille	0.17	0.26	–0.10
Berlin	0.32	0.23	0.24
Hamburg	0.34	0.16	0.18
Bremen	0.14	0.12	–0.33
Mean	0.19	0.16	–0.01

Better: larger values

Table 5: Out-of-sample R^2 for the *leave-one-city-out-within-country* evaluation across different target variables (i.e., columns).

residuals, the overall improvement across all cities is statistically significant ($p < 0.05$) for all three indicators. Yet, the results are not as good as for the intra-city prediction approach. Again, the prediction performance is better and more stable for unemployment rate as compared to the other urban inequality indicators, which could indicate that unemployment has more similar patterns across different cities than income and foreign national rates.

Leave-one-city-out-within-country: This evaluation strategy yields good results for income level and unemployment rate (see Table 5). For instance, for the income level, POIs explain up the 16% of the variance. Again, results are based on M5. The overall improvement across all cities statistically significant for both indicators ($p < 0.05$ using t-tests on the residuals). This suggests that POIs for both unemployment rate and income level are transferable indicators of urban inequality within countries. In contrast, the average out-of-sample R^2 for the foreign national rate is around zero, suggesting that POIs lack predictive power for foreign national rates.

Important POI Predictors

To understand which POIs are particularly important for explaining urban inequality, we make use of SHAP (SHapley Additive exPlanations) values (Lundberg and Lee 2017). For this, we train one model for each target variable on all six cities as this allows to understand the overall feature importance per urban inequality indicator. The results are based on variant M2. The reason is that similar analyses are infeasible for M5 due to the fact that the latter uses principal component, whereas M2 uses raw POIs as predictors.

The ten most important features based on the SHAP values for all three urban inequality indicators are shown in Fig. 3. Across all indicators, Muslim places of worship (distance feature) has the most predictive power. Interestingly, the four most important features for the target variables unemployment rate and income level are largely identical. A decisive POI for modeling unemployment rate are banks, due to the fact that both the distance- and the density-based feature for banks are among the top-10 predictors. A similar pattern can be detected for the importance of Muslim place of worship for the foreign national rate. Other POIs which have substantive predictive power for more than one target variable are playgrounds (distance feature) and beverage shops (distance feature). This answers research question RQ3.

Discussion

Main findings: Our results demonstrate that urban inequality can be modeled on the basis of POI data at a highly granular and accurate level. POIs are able to explain a large portion of the out-of-sample variance for different indicators of urban inequality (with an out-of-sample R^2 of up to 75%).

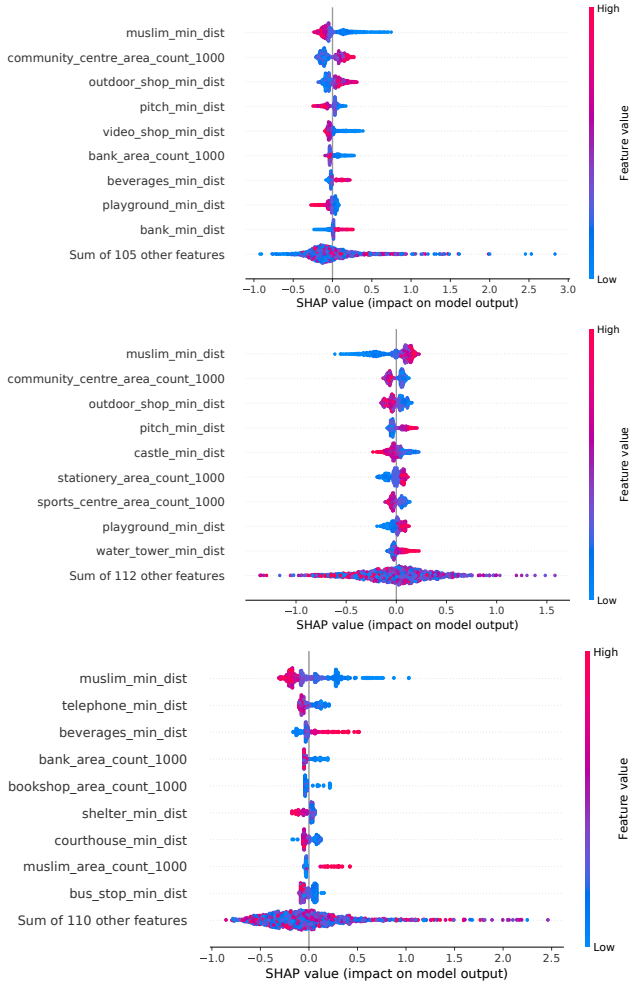

Figure 3: Most important SHAP values (top-10) for the three urban inequality indicators. *Top:* Unemployment rate; *Center:* Income level; *Bottom:* Foreign national rate.

Across all three indicators, the findings are robust to different variants of our approach. Moreover, the estimated relationships among POIs are further transferable across cities, especially if the cities are from the same country.

Interpretation: The identified association at the POI level may have different explanations. For example, POIs related to leisure (e.g., cafes, restaurants) are relevant for modeling urban inequality, as they may attract young people or people with high income. Schools and playgrounds, on the other hand, may attract young families with a different income structure. Therefore, some POIs related to the attractiveness of neighborhoods for certain population groups, while others may not. This establishes that POIs are important determinants of urban inequality.

The performance of the cross-city evaluation (i.e., training the ML approach on one city and evaluating it on a different, out-of-sample city) varies, which is best seen in the following, interesting examples. (1) The roles of specific types of POIs are subject to heterogeneity across cities. For example, the shortest distance to restaurants is an important feature for predicting urban inequality in Paris, but not in the case of Berlin and Bremen. This explains why the prediction performance of a model trained on one city can drop significantly when it is used to predict inequality in another city. (2) We find that POIs as predictors entail comparatively low transferability across cities for foreign national rate. This can be attributed to the nature of the target variable itself. In some cities, the foreign national rate is strongly negatively correlated with income (e.g., Marseille), whereas, in other cities, it is only weakly negatively correlated with income (e.g., Berlin), implying that the residence of foreigners is less segregated.

Limitations and generalizability: Our work is subject to limitations, similar to related research involving POI data (Kadar et al. 2020). First, we build upon a comprehensive dataset of POIs. The availability and accuracy of POI data has been subject to discussion (e.g., Thebault-Spieker, Hecht, and Terveen 2018). Nevertheless, our results show that POIs are strong predictors of urban inequality. Second, we study inequality in six European cities. This choice was done because urban inequality is a critical issue in Western parts of the world (Musterd and Ostendorf 2013). Third, our findings are associative rather than causal in nature as we are interested in modeling urban inequality. Here, it is important to further investigate the causal nature of highly predictive POIs for urban inequality. Forth, the socioeconomic data were collected in 2015 while POI data have been downloaded in 2021. However, we believe that urban inequality most likely has not changed substantially over these years, and, thus, results may only change marginally. Fifth, all buildings are weighted the same, regardless of their size. This may provide opportunities for future research. Sixth, we made use of SHAP value method to generate insights into our predictions but acknowledge that there are issues regarding interpretability (Kumar et al. 2020; Weerts, van Ipenburg, and Pechenizkiy 2019).

For end-users, care is needed when deploying our machine learning framework in practice in order to mitigate

risks of feedback loops. One example is that data quality is lower in areas with large urban inequality (e.g., fewer data on POIs), which may result perpetuate bias in the prediction and thus lead to errors in downstream decision-making. Hence, interpretations and, more importantly, interventions based on the predictions must be carefully inspected to reduce risk of feedback loops during decision-making.

Implications: Our results offer new insights to policy-makers aiming at reducing urban inequality. By identifying predictors of urban inequality, our results inform targeted policy-making at the micro level. For instance, a widespread assumption in computational social science is that public transport infrastructure is directly associated with many urban phenomena (e.g., Kadar et al. 2020). Interestingly, we find that the corresponding association (1) between public transport infrastructure and unemployment rate and (2) between public transport infrastructure and income level is not among top-10 features in our model.

For policy-makers, it is of direct value to analyze the most predictive POI as this will help to understand certain urban phenomena. Based on evidence from POI data, policy-makers can implement effective measures that mitigate population dynamics that lead to segregation and, eventually, urban inequality. For example, building schools, playgrounds, and kindergartens in disadvantaged urban neighborhoods could facilitate family planning or attract young families to the respective neighborhood. Our results of how POI relate to urban inequality can thus represent important factors for such decision-making to reduce urban inequality.

Finally, our ML approach could serve another use case: Policy-makers can use machine learning to generate inequality mappings at a granular level. This is directly needed in practice, since there is widespread lack of granular data on inequality in many cities. Oftentimes, obtaining census data at micro level is highly costly (The Economist 2011). As a remedy, policy-makers could train our ML approach using POIs from one city where such urban inequality indicators are available at neighborhood level, and then generate predictions for a different city of interest with similar characteristics. However, before applying such an approach in practice, further research is needed to investigate the transferability of the trained models between cities and potential biases. Combining our ML approach with other methods for inferring inequality (e.g., from satellite or mobile phone data) could further improve the robustness.

Appendix

Encoding of POI Types

Table 6 presents the manual encoding to adjust POI types for interpretability reasons.

Original POI type	↦	**Replacement**
'college'	↦	'university'
'bus_station'	↦	'bus_stop'
'railway_halt'	↦	'railway_station'
'biergarten'	↦	'restaurant'
'pub'	↦	'bar'
'food_court'	↦	'fast_food'
'christian_anglican'	↦	'christian'
'christian_catholic'	↦	'christian'
'christian_evangelical'	↦	'christian'
'christian_lutheran'	↦	'christian'
'christian_methodist'	↦	'christian'
'christian_orthodox'	↦	'christian'
'christian_protestant'	↦	'christian'
'christian_baptist'	↦	'christian'
'muslim_sunni'	↦	'muslim'
'muslim_shia'	↦	'muslim

Table 6: Manual encoding to adjust POI types for interpretability reasons.

Target Distribution

Figure 4 shows the Kernel density estimation for the distribution for the different indicators of urban inequality (i.e., target variables) in the six cities).

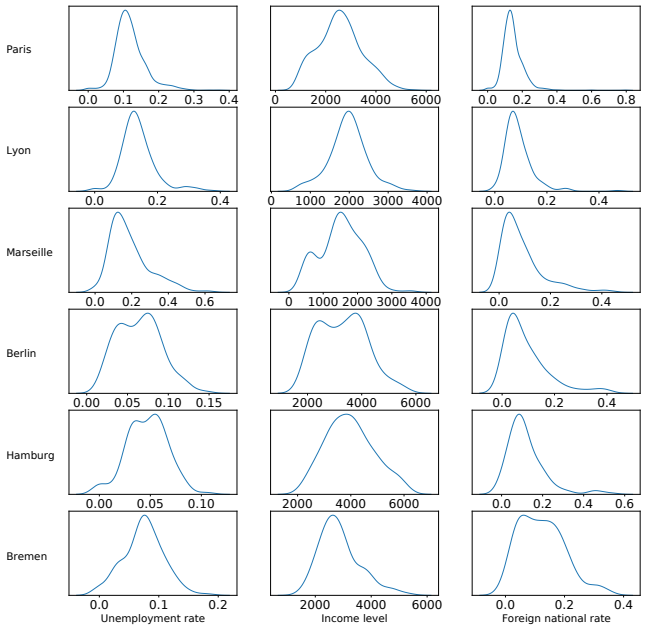

Figure 4: Kernel density estimation for the distribution for the different indicators of urban inequality (i.e., target variables) in the six cities).

Model	Tuning parameter	Values
Lasso	Regularization strength α	0.0001, ..., 5
XGBoost	Learning rate η	0.01, 0.1, 0.3, 0.5
	Max. tree depth	1, 2, 4, 8

Table 7: Grid-search for hyperparameter tuning.

Implementation

Hyperparameter tuning: Hyperparamter tuning is performed via a grid-search (as part of a five-fold cross-validation). The search grid is reported in Table 7.

Implementation details: Training was done using an L2 loss. We implemented all the ML approach in SCIKIT-LEARN (Pedregosa et al. 2011). Both features and target variables were scaled using the RobustScaler from SCIKIT-LEARN. The scaler was chosen over the standard scaler for an important reason: neighborhoods with particularly large/small values should not be treated as outliers but are particularly relevant for policy-makers.

Robustness Checks

Distance-based POI features: We vary the radius used for the density-based features for POI modeling. Here, we test two alternatives. (1) We set the radius to 500 m (i.e., half of the radius from the main analysis). (2) We set it to 2000 m (twice the radius from the main analysis). The prediction results of the models using the two alternatives are robust.

Density-based POI features: We compare the impact of other aggregations when computing the the density-based POI features. Rather than simply aggregating all POIs with equal weight, we now introduce a "score", so that POIs closer to buildings receive a larger value. This should capture a possible preference for POIs that are in closer proximity. For this, we use a kernel function as follows. In a first step, every POI p in the area of interest receives a score

$$a'_{b,p,n} = \begin{cases} 1 - \frac{1}{\theta} \left\| \begin{pmatrix} p_{\text{lat}} \\ p_{\text{lon}} \end{pmatrix} - \begin{pmatrix} b_{\text{lat}} \\ b_{\text{lon}} \end{pmatrix} \right\|_2, & \text{if } \left\| \begin{pmatrix} p_{\text{lat}} \\ p_{\text{lon}} \end{pmatrix} - \begin{pmatrix} b_{\text{lat}} \\ b_{\text{lon}} \end{pmatrix} \right\|_2 \leq \theta, \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

where $a'_{b,p,n}$ is the weight of a POI with respect to building b. If the distance is larger than the radius θ, zero will be assigned to the weight as before. In a second step, all scores for each building b and all POIs of the same type t are summed up to provide the final score via

$$f'_{b,t,n} = \sum_{p \in P(t)} a'_{b,p,n}, \quad (9)$$

where $f'_{b,t,n}$ is the density-based POI score for each building b. We finally also average the score of all buildings within each neighborhood according to

$$s'_{t,n} = \frac{1}{|\{b \mid b \in \mathcal{B}_n\}|} \sum_{b \in \mathcal{B}_n} f'_{b,t,n} \quad (10)$$

which is then used as a input to the ML approach. The prediction results based on the above weighting are robust and largely overlap with our main results from Table 2 and 3.

Meta learner: We also repeated the leave-one-city-out evaluation using a meta learner which learns to combine M5 of five cities via a regression to predict the left-out city. Thereby, the different city models are assigned different weights. However, this approach led to overfitting due to the sample size but, nevertheless, the overall qualitative conclusions are confirmed.

Prediction models: We also experimented with different neural networks; however, even with a high dropout rate, we experienced overfitting. We also considered the use of (Naumzik, Zoechbauer, and Feuerriegel 2020) but requires a spatial distribution as target variable, not data at neighborhood level, because of which this was inapplicable.

Ethical Statement

This research did neither involve interventions with human subjects nor individualized human data. Thus, no approval from the Institutional Review Board was required by the author institutions.

Acknowledgements

We thank PriceHubble AG for supporting this research by providing the dataset used in the analysis. Furthermore, Manuel Ganter was supported by a fellowship within the IFI program of the German Academic Exchange Service (DAAD).

References

Benassi, F.; Lipizzi, F.; and Strozza, S. 2019. Detecting foreigners' spatial residential patterns in urban contexts: Two tales from Italy. *Applied Spatial Analysis and Policy*, 12(2): 301–319.

Bourguignon, F.; and Scott-Railton, T. 2015. *The globalization of inequality*. Princeton: Princeton University Press.

Brilhante, I.; Macedo, J. A.; Nardini, F. M.; Perego, R.; and Renso, C. 2013. Where shall we go today? Planning touristic tours with Tripbuilder. In *CIKM*.

Cassiers, T.; and Kesteloot, C. 2012. Socio-spatial inequalities and social cohesion in European cities. *Urban Studies*, 49(9): 1909–1924.

Chakravorty, S. 1996. A measurement of spatial disparity: The case of income inequality. *Urban Studies*, 33(9): 1671–1686.

Chen, T.; and Guestrin, C. 2016. XGBoost: A scalable tree boosting system. In *KDD*.

Cranshaw, J.; Schwartz, R.; Hong, J.; and Sadeh, N. 2012. The livehoods project: Utilizing social media to understand the dynamics of a city. *ICWSM*.

Dong, L.; Ratti, C.; and Zheng, S. 2019. Predicting neighborhoods' socioeconomic attributes using restaurant data. *PNAS*, 116(31): 15447–15452.

Elgar, F. J.; Stefaniak, A.; and Wohl, M. J. A. 2020. The trouble with trust: Time-series analysis of social capital, income inequality, and COVID-19 deaths in 84 countries. *Social Science & Medicine (1982)*, 263: 113365.

European Environment Agency. 2013. EEA reference grid. https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2. Accessed: 2021-07-17.

Fu, X.; Jia, T.; Zhang, X.; Li, S.; and Zhang, Y. 2019. Do street-level scene perceptions affect housing prices in Chinese megacities? An analysis using open access datasets and deep learning. *PLOS ONE*, 14(5): e0217505.

Furletti, B.; Cintia, P.; Renso, C.; and Spinsanti, L. 2013. Inferring Human Activities from GPS Tracks. In *KDD*.

Glaeser, E.; Resseger, M.; and Tobio, K. 2008. Urban inequality. *National Bureau of Economic Research*, (14419).

Glaeser, E. L.; Resseger, M.; and Tobio, K. 2009. Inequality in cities. *Journal of Regional Science*, 49(4): 617–646.

Hidalgo, C. A.; Castañer, E.; and Sevtsuk, A. 2020. The amenity mix of urban neighborhoods. *Habitat International*, 106: 102205.

Hristova, D.; Williams, M. J.; Musolesi, M.; Panzarasa, P.; and Mascolo, C. 2016. Measuring urban social diversity using interconnected geo-social networks. In *WWW*.

Hummler, P.; Naumzik, C.; and Feuerriegel, S. 2022. Web mining to inform locations of charging stations for electric vehicles. In *WWW Companion*.

Kadar, C.; Feuerriegel, S.; Noulas, A.; and Mascolo, C. 2020. Leveraging mobility flows from location technology platforms to test crime pattern theory in large cities. *ICWSM*.

Kadar, C.; Maculan, R.; and Feuerriegel, S. 2019. Public decision support for low population density areas: An imbalance-aware hyper-ensemble for spatio-temporal crime prediction. *Decision Support Systems*, 119: 107–117.

Kadar, C.; and Pletikosa, I. 2018. Mining large-scale human mobility data for long-term crime prediction. *EPJ Data Science*, 7(1).

Karamshuk, D.; Noulas, A.; Scellato, S.; Nicosia, V.; and Mascolo, C. 2013. Geo-spotting: Mining online location-based services for optimal retail store placement. In *KDD*.

Kumar, I. E.; Venkatasubramanian, S.; Scheidegger, C.; and Friedler, S. A. 2020. Problems with Shapley-value-based explanations as feature importance measures. In *ICML*.

Lundberg, S. M.; and Lee, S.-I. 2017. A unified approach to interpreting model predictions. *NIPS*.

Luo, W.; and Wang, F. 2003. Measures of spatial accessibility to health care in a GIS environment: Synthesis and a case study in the Chicago region. *Environment and Planning B: Planning and Design*, 30(6): 865–884.

Miller, P. M. 2012. Mapping educational opportunity zones: A geospatial analysis of neighborhood block groups. *The Urban Review*, 44(2): 189–218.

Morrison, P. S. 2005. Unemployment and urban labour markets. *Urban Studies*, 42(12): 2261–2288.

Musterd, S.; and Ostendorf, W. 2013. *Urban segregation and the welfare state: Inequality and exclusion in western cities*. New York: Routledge.

Naumzik, C.; Zoechbauer, P.; and Feuerriegel, S. 2020. Mining points-of-interest for explaining urban phenomena: A scalable variational inference approach. In *WWW*.

Noulas, A.; Scellato, S.; Lambiotte, R.; Pontil, M.; and Mascolo, C. 2012. A tale of many cities: Universal patterns in human urban mobility. *PLOS ONE*.

O'Loughlin, J. 1980. Distribution and migration of foreigners in German cities. *Geographical Review*, 70(3): 253.

OpenStreetMap. 2021. https://www.openstreetmap.org. Accessed: 2021-05-05.

Pedregosa, F.; et al. 2011. Scikit-learn: Machine learning in Python. *JMLR*, 12: 2825–2830.

Puttanapong, N.; Martinez, A.; and Addawe, M. 2020. Predicting poverty using geospatial data in Thailand. *ADB Economics Working Paper Series*, (630).

Rae, A.; Murdock, V.; Popescu, A.; and Bouchard, H. 2012. Mining the web for points of interest. In *SIGIR*.

Sachs, J.; Kroll, C.; Lafortune, G.; Fuller, G.; and Woelm, F. 2021. *Sustainable Development Report 2021*. New Work: Cambridge University Press.

Schaffar, A. 2008. Income inequality, urbanisation and regional development in China. *Région et Développement*, 28: 132–156.

Strozza, S.; Benassi, F.; Ferrara, R.; and Gallo, G. 2016. Recent demographic trends in the major Italian urban agglomerations: The role of foreigners. *Spatial Demography*, 4(1): 39–70.

Tang, J.; Liu, Z.; Wang, Y.; Yang, J.; and Wang, Q. 2018. Using geographic information and point of interest to estimate missing second-hand housing price of residential area in urban space. In *IEEE International Smart Cities Conference*.

Taylor, K.; Lim, K. H.; and Chan, J. 2018. Travel itinerary recommendations with must-see points-of-interest. In *WWW Companion*.

The Economist. 2011. Costing the count: Old style censuses are cumbersome and costly. Reform is coming. *The Economist*.

Thebault-Spieker, J.; Hecht, B.; and Terveen, L. 2018. Geographic biases are 'born, not made': Exploring contributors' spatiotemporal behavior in OpenStreetMap. In *ACM Conference on Supporting Groupwork (GROUP)*.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1): 267–288.

Tingzon, I.; Orden, A.; Sy, S.; Sekara, V.; Weber, I.; Fatehkia, M.; García-Herranz, M.; and Kim, D. 2019. Mapping poverty in the Philippines using machine learning, satellite imagery, and crowd-sourced geospatial information. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-4/W19: 425–431.

Todaro, M. 1969. A model for labor migration and urban unemployment in less developed countries. *The American Economic Review*, (59): 138–148.

Toetzke, M.; Banholzer, N.; and Feuerriegel, S. 2022. Monitoring global development aid with machine learning. *Nature Sustainability*.

Tschernutter, D.; and Feuerriegel, S. 2021. A latent customer flow model for interpretable predictions of check-in counts. In *IEEE BigData*.

UN General Assembly. 2015. Transforming our world: The 2030 Agenda for Sustainable Development.

UNICEF Office of Research - Innocenti. 2017. Growing inequality and unequal opportunities in rich countries. *Innocenti Research Briefs*, 2017/16(16): 1–5.

Wang, H.; Kifer, D.; Graif, C.; and Li, Z. 2016. Crime rate inference with big data. In *KDD*.

Weerts, H. J. P.; van Ipenburg, W.; and Pechenizkiy, M. 2019. A human-grounded evaluation of SHAP for alert processing. In *KDD workshop on Explainable AI 2019*.

Xiao, Y.; Chen, X.; Li, Q.; Yu, X.; Chen, J.; and Guo, J. 2017. Exploring determinants of housing prices in Beijing: An enhanced hedonic regression with open access POI data. *ISPRS International Journal of Geo-Information*, 6(11): 358–370.

Xue, J.; and Zhong, W. 2003. Unemployment, poverty and income disparity in urban China. *Asian Economic Journal*, 17(4): 383–405.

Yuan, J.; Zheng, Y.; and Xie, X. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *KDD*.

Zenk, S. N.; Schulz, A. J.; Israel, B. A.; James, S. A.; Bao, S.; and Wilson, M. L. 2005. Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. *American Journal of Public Health*, 95(4): 660–667.

Zhao, X.; Yu, B.; Liu, Y.; Chen, Z.; Li, Q.; Wang, C.; and Wu, J. 2019. Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh. *Remote Sensing*, 11(4): 375–393.