

Key determinants of street vendor sales in Dhaka: infrastructure and socioeconomic factors

Journal of
Entrepreneurship
and Public Policy

187

Nurshad Yesmin

*School of Business Administration, East Delta University, Chittagong,
Bangladesh, and*

Beatriz Calzada Olvera

*Institute for Housing and Urban Development Studies,
Erasmus University Rotterdam, Rotterdam, Netherlands and
Maastricht Economic and Social Research Institute on Innovation and Technology,
Maastricht, Netherlands*

Received 23 December 2023
Revised 28 June 2024
Accepted 5 November 2024

Abstract

Purpose – This research identifies the individual socioeconomic and urban infrastructure attributes that explain the sales performance of informal street vendors in Dhaka, Bangladesh, shedding light on areas where public intervention could enhance informal vendors' entrepreneurial activities.

Design/methodology/approach – The study employs a quantitative methodology, focusing on two street vending areas in Dhaka: New Market Area and Mirpur-1. Primary data ($n = 243$) were randomly collected from vendors in these areas. The importance of various attributes was estimated using non-parametric, non-linear methods (random forests and geographical random forests). These results were compared with those obtained from linear multiple regression and Lasso regression models.

Findings – The study shows that having a designated vending spot is the most important attribute associated with higher sales, confirming the critical role of secure vending locations in urban planning policy. Other significant factors include the width of the sidewalk and the level of odors, indicating the role of urban infrastructure on sales performance. Key individual socioeconomic factors include having a bank account and working experience. The initial economic situation, measured by the log of initial capital, also plays a significant role, especially when accounting for spatial heterogeneity.

Originality/value – This research explores the relationship between individual socioeconomic characteristics, urban infrastructure and street vendors' sales performance using advanced machine learning models. Our findings underscore the significance of adequate street vending infrastructure and access to banking services, highlighting critical areas for public policy intervention to support this vital entrepreneurial activity.

Keywords Entrepreneurial activity, Street vending, Sales performance, Infrastructure attributes, Socioeconomic factors

Paper type Research paper

1. Introduction

Street vending is one of the most common types of informal economic activity in developing nations (Lemessa *et al.*, 2021; Recchi, 2021), as it results from poverty and limited employment opportunities in the formal sector (Chen and Carré, 2020; Elgin, 2020). However, street enterprises are often controversial (Bromley, 2000): street vendors are often seen as “intruders on public property,” causing public nuisance, unsanitary conditions, pollution and traffic congestion, among other issues (Bhowmik, 2012; Chen and Carré, 2020; Elgin, 2020). It is unsurprising, then, that informal street vendors are the most visible and contested domain in the urban informal economy (Bhowmik, 2012).



© Nurshad Yesmin and Beatriz Calzada Olvera. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Journal of Entrepreneurship and Public
Policy
Vol. 14 No. 2, 2025
pp. 187-206
Emerald Publishing Limited
e-ISSN: 2045-211X
p-ISSN: 2045-2101
DOI 10.1108/JEPP-12-2023-0135

Nonetheless, street vending is a crucial source of economic livelihood for the urban poor by creating employment and entrepreneurial opportunities. The entrepreneurial activities of informal vendors can stimulate face-to-face encounters in public spaces and bring vitality to street life (Bromley, 2000). Street vendors come from various backgrounds and provide a diverse array of products and services (Michalak, 2020). Moreover, urban street vending can boost local economies by offering innovative community goods and services, providing a low-cost economic entry point, supplying inexpensive goods to low-income clients and enriching urban streets with diversity (Faruque *et al.*, 2010; Chen and Carré, 2020; Elgin, 2020; Adair, 2020).

Bangladesh's capital and most densely populated city, Dhaka, has the highest concentration of street vendor activity. Poverty, rural-to-urban migration, low education levels, the surplus in labor supply and large family sizes drive street vending in this city (Husain *et al.*, 2015). Over 90,000 street vendors operate in Dhaka, with more than 60% of the city's population using their services (Moniruzzaman *et al.*, 2018). Vendors are concentrated in areas like Motijheel, Baitul Mukarram, Gulistan, Shahbagh, Mirpur and New Market, as well as along major road networks (Husain *et al.*, 2015).

Despite their importance to low- and middle-income residents, urban planners and policymakers do not recognize street vending as a legitimate business, viewing it instead as a source of nuisance, congestion and unhygienic conditions (Uddin, 2021). Law enforcement often harasses vendors, and local government policies emphasize secure and accessible public spaces over the needs of street vendors. Consequently, vendors face exploitation, displacement and forced sales at below-market prices (Hussain, 2019).

Recognizing the informal economy – and street vending by extension – through inclusive and coherent policies is vital for addressing inequality, reducing poverty and promoting long-term sustainable growth. Understanding the factors linked to varying income levels among street vendors (which can also be conceptualized as differences in sales performance [1]) can help identify areas for public policy intervention to enhance their performance and reduce poverty. Street vendors' primary purpose is to earn a living in challenging economic conditions; nonetheless, their revenue fluctuates significantly, with the majority experiencing financial difficulties and a large part living below the poverty line (Michalak, 2020). Recognizing the significance of this often-invisible economy is crucial (Uddin, 2021).

Additionally, the typical orientation of urban policy in developing countries has been characterized by decoupling the right to work from the right to access public space, which has more likely led to more marginal forms of street vending (Aliaga Linares, 2018). Examining how street vendors access, operate and manage their business activities in public spaces – a reflection of existing urban infrastructure and its management – can provide further insights into how policy can reshape urban policy to improve the livelihoods of street vendors.

Current empirical literature – albeit limited – spans across several fields, including development studies, business, urban planning and sociology, from which two general strands can be derived. On the one hand, a handful of quantitative studies focus on the individual socioeconomic characteristics of street vendors, such as age, access to capital, business experience and marital status, to explain their success (e.g. Adhikari, 2011; Tuffour *et al.*, 2022). These studies emphasize the individual's productive capabilities to identify drivers of performance. On the other hand, various aspects linked to urban infrastructure, such as access to water or sanitation facilities, have been highlighted in several qualitative studies (e.g. Sekhani *et al.*, 2019; Martínez *et al.*, 2017; Uddin, 2021). These studies emphasize the central role of policies in recognizing street vending as a legitimate entrepreneurial activity. They suggest that access to public spaces and improvements in the urban infrastructure where vendors operate are key to supporting this form of entrepreneurship and improving the livelihoods of informal vendors.

While the studies above suggest there is a link between physical infrastructure and the performance of street vendors, they do so in an exploratory, descriptive manner and do not integrate individual traits. To address this gap, this study investigates how individual socioeconomic characteristics (such as business experience, age and access to capital) impact street vending performance in comparison to infrastructure – an important public policy

domain. Furthermore, the study employs advanced analytical techniques – like machine learning models – to offer more accurate insights into the relative importance of various individual and infrastructure-related characteristics – including potential variations across space.

Namely, this study examines the factors that explain the sales performance of informal street vendors in two selected street markets in Bangladesh. The main research question is: *What individual and infrastructure features matter the most in explaining the sales performance of Dhaka street vendors?* The study uses primarily collected data with a simple random sampling strategy ($n = 243$) in the New Market Area and Mirpur-1 market areas. Using traditional ordinary least squares (OLS) models as a starting point, the study employs a machine learning model [Least Absolute Shrinkage and Selection Operator (LASSO)] to improve the model specification and variable selection, while avoiding overfitting – a problem where statistical relationships follow noise too closely, compromising the accuracy of the estimations (James *et al.*, 2021). This is followed by the main empirical approach, random forests (RF) algorithms [i.e. RF and geographical random forests (GRF)], which enhance accuracy by controlling for confounding factors and potential spatial heterogeneity, capturing non-linear interactions and ensuring robust identification of the most important variables explaining sales performance.

The structure of the paper is as follows: Section 2 explains the most important economic aspects of informality and street vending from a public policy perspective while presenting relevant empirical studies on sales performance; Section 3 presents the main features of the data and methods; Section 4 discusses the empirical results and Section 5 provides a discussion of the results and concludes.

2. Theoretical background

2.1 Street vending's importance and the case for policy intervention

Despite the evident economic importance of street vending for the livelihood of the urban poor, within urban policy, street vending is often perceived as needing to be upgraded *out of the streets* (Aliaga Linares, 2018). This perception, which sees street vending as an “entrepreneurship failure” or an “employment problem,” significantly influences local policy formulation in developing economies. The failed attempts to formalize or upgrade informal street vending underscore the need for policymakers to revisit how it is conceptualized (Aliaga Linares, 2018).

Traditionally, formal and informal activities remain separated by legal recognition and the regulation of economic activities and property rights, which hinders entrepreneurs and overall economic growth (De Soto, 1989). Other scholars such as Bhowmik (2005), Roy (2005) and Daniels (2004) move away from a dualist perspective: “The divide here is not between formality and informality but rather a differentiation *within* informality” (Roy, 2005, p. 149). Instead, they argue for understanding urban informality as intertwined with formal, regulated spaces within a city's ecosystem, revealing a complex relationship between formal and informal commercial exchanges (Martínez *et al.*, 2017).

Schneider and Williams (2013), Williams and Schneider (2016) and Williams (2023) define informality as the “shadow economy,” encompassing lawful paid activities not reported to the government for tax, social security and labor law reasons. Roy (2017) further describes the shadow economy as economic activities and income that evade government regulation, taxation and monitoring.

Although street vendors operate outside the legal framework and lack social protection and labor legislation, they often respect certain aspects of the law. In this manifestation of the shadow economy, street vendors emerge as a particularly good representation of the “invisible entrepreneurs” (Sekhani *et al.*, 2019; Uddin, 2021). Certainly, street vending, despite being informal, is ultimately an entrepreneurial activity where vendors employ production factors to sell goods or services for profit (Teltscher, 1994). The main difference lies in their necessity

and opportunity, with poor street vendors typically being necessity-based entrepreneurs whose informal activities are their primary source of income (Williams, 2008).

Indeed, informality is both a cause and a consequence of economic and institutional underdevelopment, leading to productive inefficiency and a culture of evasion and noncompliance. Yet, without informality, there would likely be greater unemployment, poverty and crime (Loayza, 2018). Moniruzzaman *et al.* (2018) show that street vendors significantly contribute to economic growth by generating income, creating entrepreneurial endeavors and saving a portion of their earnings. Additionally, while informal activities like street vending reflect protracted and long-standing lower levels of development, productivity, government capacity and business flexibility, in the short term, informal employment can serve as an important safety net in short-term economic shocks (Loayza and Rigolini, 2011).

Rapid urbanization in developing economies has not been accompanied by employment opportunities in key sectors like manufacturing, resulting in a rise in informal, petty services (Gollin *et al.*, 2016), such as street vending, as this provides employment opportunities for low-skilled workers (Sekhani *et al.*, 2019). Therefore, the emergence of informal activities in the urban space has often been the result of less inclusive and sustainable growth patterns. However, paying attention to these activities is essential for sustaining economic growth. Improved production capabilities in informal sectors, aligned with broader economic efforts, have been established as fundamental for fostering enhanced structural change processes in developing contexts (Caldarola, 2022).

2.2 Factors linked to street vendor performance

Overall, there are a limited number of empirical studies examining sales performance and street vending. These studies can be categorized into those focusing on individual characteristics and those focusing on infrastructure.

A handful of papers have studied individual socioeconomic features – including access to capital – and their relationship with informal street vending income or welfare using basic quantitative frameworks and primary data (i.e. Adhikari, 2011; Tuffour *et al.*, 2022; Handoyo and Wijayanti, 2021; Nilakusmawati *et al.*, 2019). Adhikari (2011) and Handoyo and Wijayanti (2021) find that capital and education are positively correlated with the earnings of urban street vendors in Nepal and Semarang, Indonesia, respectively. In Bali, Indonesia, Nilakusmawati *et al.* (2019) highlight the significance of marital status and age, highlighting a positive correlation but do not find any significant correlation with education or capital. Finally, in Accra, Ghana, Tuffour *et al.* (2022) look at female street vendors' performance and find that age and business experience are positively correlated, being married is negatively correlated and education is not correlated with business success.

Other studies have approached this question by focusing on the physical features of street vending. These studies, using a qualitative perspective, have focused on the context of Bangladesh, India and Indonesia, examining the infrastructure-related features of the location (Israt and Adam, 2017; Kesumasari, 2020). These can be classified into internal variables (structural conditions of vending facilities) and external variables (physical environment).

These studies have evaluated the structural setting of street vending by considering factors such as fixed or demarcated vending space, condition of canopies, storage facility and facade of the selling site (Israt and Adam, 2017; Kesumasari, 2020). In particular, Kesumasari (2020) examines how the structural setting relates to customer comfort in Surakarta, Indonesia, finding that these physical settings influence both physical and social comfort, attracting a larger customer base and increasing sales.

Other studies focus on additional physical features, such as sidewalk and street surface conditions, the width of the street/sidewalk and the condition of street furniture or plants (Kesumasari, 2020; Israt and Adam, 2017; Nilakusmawati *et al.*, 2019). Israt and Adam (2017) explore these features in Dhaka, Bangladesh, finding a correlation between the availability of physical features and the level of street vending activities, affecting potential customers' perceptions of these areas.

Moreover, street vendors in developing countries face various environmental challenges such as waterlogging, air pollution, noise pollution, inadequate street lighting, extreme weather conditions and limited access to clean water (Kambara and Bairagya, 2021; Israt and Adam, 2017). Ghatak and Chatterjee's (2018) study on street food vendors in Kolkata, India, highlights the operational challenges posed by severe weather conditions like urban flooding and rainfall. In hot cities like Dhaka, street vendors are particularly vulnerable to urban heat stress. Urban trees and vegetation offer valuable ecosystem services like shading and mitigating the urban heat island effect (Basu and Nagendra, 2020), making their availability crucial for supporting street vending activities by potentially attracting more customers and making vendors more productive under better conditions.

Finally, high levels of air pollution, such as those in Dhaka, present additional challenges. Amegah and Jaakkola (2014) find that street vendors in Accra, Ghana, exposed to high levels of traffic-related air pollution face greater health risks. Barbaresco *et al.* (2019) conclude that street vendors in Uberlândia, Brazil, exposed to excessive noise levels, experience increased stress and health risks, potentially affecting their ability to fully operate. While the challenges related to environmental conditions are not necessarily directly related to the immediate vending infrastructure, they certainly relate to the overall quality of the urban infrastructure.

3. Data and method

3.1 Data

This study employs cross-sectional microdata of street entrepreneurs in the selected market areas in Dhaka, Bangladesh, based on the variables identified in the existing literature reviewed in Section 2.2. The following independent variables (also hereafter referred to as predictors or features) were collected and categorized into individual characteristics and urban infrastructure attributes. Individual socioeconomic characteristics include age, gender, education, experience in years and family situation (i.e. marital status and household members). Capital and banking-related variables were also included, such as access to credit and banking services, as well as initial capital measured in Bangladeshi Taka (BDT). To account for other relevant characteristics, working type (full-time or part-time) and the type of goods and services offered were also included. Urban infrastructure encompasses the existence of street furniture, the presence of fixed or demarcated vending spaces, sufficient vending space, the type of vending equipment and furniture and the presence of canopies for vending. It also includes the street vendor's assessment or rating of the conditions of the street and sidewalk width and surface, street hygiene conditions, the level of waterlogging during wet seasons, street lighting conditions at night, the levels of smell, the level of vegetation, noise levels and the level of heat. Finally, to account for variations in infrastructure along the sampled areas, geolocation information (latitude and longitude) was also collected. The dependent variable is sales performance, which is measured as average monthly sales in BDT.

3.2 Data collection strategy

This study focuses on two distinct regions within Dhaka city, Bangladesh: the New Market Area and Mirpur-1. Previous empirical quantitative studies on street vending activities have used sample sizes ranging from 137 to 300 respondents (Adhikari, 2011; Tuffour *et al.*, 2022; Handoyo and Wijayanti, 2021; Nilakusmawati *et al.*, 2019). This study gathered a total of 243 sample observations from the two chosen research areas (Mirpur-1 and New Market Area). This study concentrates on overall street vending activities in this area, regardless of the types of street vendors, such as those selling food or gender (as done in Tuffour *et al.*, 2022).

The data for this study were collected using a simple random sampling method. To ensure that street vendors in various zones were fairly represented in the sample, field surveyors visited the specified locations in Mirpur-1 and the New Market Area to observe and assess the distribution and spread of street vending activities in these areas. Using randomization, no

specific characteristics such as vendor gender, type of selling activity or specific hours were used for selecting samples. The surveyors meticulously covered all defined locations throughout the day and night over a week. This method ensured that all street vendors in the studied areas had an equal opportunity to be contacted and participate in the research.

To ensure transparency and adherence to ethical research principles, informed consent was obtained from each vendor before administering the pre-designed questionnaire. To gather primary data, a structured questionnaire was administered by trained field surveyors. This method ensures that consistent and reliable data are collected directly from respondents. Alongside, geolocation data were incorporated to enrich the spatial analysis, allowing for precise mapping of street vendor locations. For supplementary spatial context, the Geographic Information System database maintained by Dhaka’s development authority (RAJUK), along with other relevant secondary datasets, was employed. These resources were instrumental in developing a comprehensive research map that enhances the granularity and accuracy of the study. See [Figures 1 and 2](#) for the geographic locations where data were collected in the Mirpur-1 and New Market Area, respectively.

3.3 Method

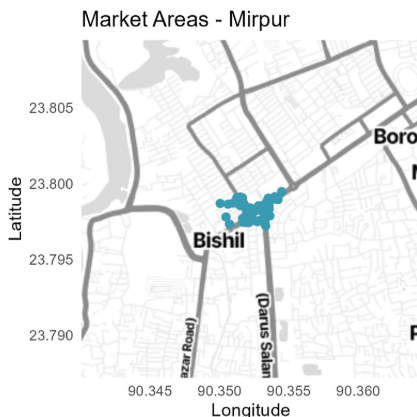
3.3.1 *Traditional and regularized OLS regression.* The empirical method begins with a regression framework, first using an OLS model with three specifications, followed by a preferred LASSO model – a machine learning method.

The OLS estimation ([equation 1](#)) follows the traditional form:

$$Y_i = \alpha + \beta_1 X'_i + \epsilon \tag{1}$$

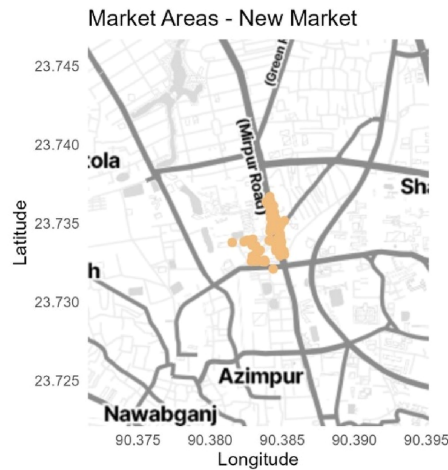
where Y_i refers to the dependent variable, sales performance of street vendors, specifically the log of monthly sales. X' refers to the matrix of covariates: individual variables only (Model 1), infrastructure variables only (Model 2) and all variables combined (Model 3).

The fourth model, and our preferred specification, relies on a regularized OLS model, i.e. LASSO. The LASSO method is itself an adaptation of an OLS model that uses a different cost minimization process (which in this context refers to error minimization) [\[2\]](#), as shown in [Equation \(2\)](#):



Source(s): Authors' own elaboration

Figure 1. Mirpur-1



Source(s): Authors' own elaboration

Figure 2. New Market

$$\text{Minimize} \left(\sum_{i=1}^n (Y_i - \alpha - \beta_1 X_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \right) \quad (2)$$

The LASSO (Model 4 in our regression framework) introduces a penalization element, λ , as a method of regularization. λ penalizes the absolute values of the regression coefficients, as seen in the second element to the right in Equation (2). This penalty helps in two significant ways: it shrinks some coefficients to zero, effectively selecting a subset of the most important predictors (Tibshirani, 1997), and it stabilizes the model by reducing multicollinearity (James et al., 2021). In traditional OLS regression, multicollinearity among predictors can lead to unstable coefficient estimates with high variance. A model with low variance brings stability to the model, i.e. it identifies underlying patterns instead of fitting noise. Thus, the LASSO regularization effectively deals with overfitting – an important issue when dealing with a higher number of variables, p , relative to the number of observations, n , as is the case here.

It is important to note that the primary goal of this approach is to build a robust predictive model rather than to infer the precise relationship between each predictor (independent variable) and the response variable (dependent variable). Therefore, LASSO does not provide traditional OLS metrics (e.g. standard errors or confidence intervals); instead, its focus is on regularization and variable selection. We therefore concentrate on Model 4, reported in Table 2, to discuss results.

It is important to mention that despite the high level of interpretability provided by LASSO, it is still based on the linear model (James et al., 2021). A model that offers more flexibility and the ability to model potential non-linear relationships and better handle confounding factors is RF.

3.3.2 Random forests and geographical random forests. We selected the RF machine learning method (Breiman, 2001) as our preferred approach because it is a non-linear, non-parametric model that can easily learn complex relationships and interactions from data without explicitly modeling them. RF is an ensemble method that combines multiple decision trees to improve predictive accuracy and control overfitting. Unlike traditional parametric models, RF does not assume a fixed structure; each decision tree within the forest grows according to the complexity of the input data, allowing it to capture non-linear relationships.

A key feature of this model is its ability to rank the importance of each variable in the predictive model, providing insights into which features contribute most to the prediction of the response variable – in this case, sales performance.

The “forest” in RF is an ensemble of decision trees usually trained using bootstrap aggregating (bagging). Akin to the explanation provided in similar empirical studies (see Grekousis *et al.*, 2022; Luo *et al.*, 2021), the RF algorithm’s basic steps are:

- (1) Bootstrap sampling: From a given training set, multiple samples are selected randomly with replacement, typically using about two-thirds of the data. The remaining one-third, known as the out-of-bag (OOB) set, is kept out of training and used to estimate the RF’s goodness of fit [e.g. mean squared error (MSE) and R^2 in regression].
- (2) Decision tree construction: For each bootstrap sample, a decision tree is created. At each node of the decision tree, a random subset of predictors ($m < p$) is selected, and the best split is determined using these variables.
- (3) Tree growth: Each tree grows with a constant m to its largest extent without pruning until it cannot be split further.
- (4) Aggregation: The final prediction is made by aggregating the results from all the trees – using the majority vote for classification or the average prediction for regression.

Unlike traditional machine learning methods requiring cross-validation, RF uses the OOB method to validate the model internally (Breiman, 2001). The OOB error is used to estimate the model’s performance and assess the importance of each variable. This method involves randomly permuting the values of each variable in the OOB sample and calculating the increase in the MSE. A higher increase indicates a more important variable. If the OOB error increases with the permuted values, this indicates that the variable is important.

While RF provides a solid foundation for our study, there could potentially be variations in the importance of variables across different segments of the street markets studied, e.g. closeness to the main avenue or a given stream of traffic or noise. To account for these potential differences, we also implement a GRF. GRF is an extension of the traditional RF that addresses spatial heterogeneity (Georganos *et al.*, 2019). Similar to geographically weighted regression, GRF uses local sub-models rather than a single global model. For each location, a local RF is computed using nearby observations defined by a spatial weights matrix. This local calibration allows GRF to handle spatial heterogeneity.

The GRF model follows the basic equation in (3) following Georganos *et al.* (2019):

$$Y_i = a(ui, vi)x_i + e \quad (3)$$

where Y_i is the dependent variable for the i th observation (log of monthly sales), $a(ui, vi)x_i$ is the prediction from the RF model calibrated for location i and (ui, vi) are the coordinates of the spatial unit (in our study the different spatial units across each of the market studies), i.e. a “local model.” Different sub-models are built for each spatial unit, including only its neighboring units, defined by either a distance threshold (fixed kernel) or the number of nearest neighbors (adaptive kernel). Given the variation in the size of spatial units in our case study, we used an adaptive kernel; the ideal distance, or bandwidth, is selected by an algorithm based on OOB accuracy (GRF bandwidth optimization in Georganos and Kalogirou, 2022). Note that a model without location calibration (ax_i) instead of $a(ui, vi)x_i$ would yield the non-linear prediction of RF, i.e. a “global model”.

While GRF and RF are less prone to overfitting by creating multiple trees, each trained slightly differently, we still excluded potential sources of collinearity. For this, we included only variables with a variance inflation factor of less than 7.5 and correlation values of less than 0.70, following Grekousis *et al.* (2022).

An important consideration is how RF and GRF handle confounding variables, especially given the cross-sectional nature of our data. In machine learning experiments, randomization is key to controlling for confounding variables (Brownlee, 2020). Randomization is applied at various levels: each tree is built from randomly drawn samples with replacement (sub-sample level), and at each node, a subset of variables is randomly selected for partitioning (node level). This process reduces the potential for confounding by generating comparable groups for known and unknown confounding variables (Altman and Krzywinski, 2017; Pourhoseingholi *et al.*, 2012). Moreover, GRF selects a different set of variables through randomization during training, testing many variable combinations.

Finally, RF and GRF are less sensitive to outliers. Outliers may be excluded from bootstrap samples, reducing the likelihood of overfitting and improving variable selection. This robustness, combined with the ability to capture complex, non-linear interactions and handle high-dimensional data, makes RF and GRF ideal for our study. To implement RF and GRF analyses, we used the R package “SpatialML” available in the Comprehensive R Archive Network (CRAN) repository (Georganos *et al.*, 2019; Georganos and Kalogirou, 2022).

4. Results

As seen in Table 1, the summary statistics for the variables used in the study are presented. The study’s sample consists of 243 street vendors, with average monthly sales of 41,460 BDT (approximately 490 USD) and a standard deviation of 17,504 BDT (around 210 USD). The range of monthly sales varies widely, from a minimum of 10,000 BDT (about 120 USD) to a maximum of 120,000 BDT (approximately 1,420 USD).

The sample is fairly evenly split between the two locations, with 57% of vendors in the New Market area and the remaining 43% in Mirpur. The type of goods and services offered by the vendors is diverse, with clothing, shoes, food and drinks being the most common products. There is significant variation in several individual characteristics, such as household size, access to finance and years of experience. However, there is little variation in certain variables like gender – reflecting the patriarchal social structures in Bangladesh as documented by Uddin (2021) – as well as education (which tends to be very low) and marital status (most vendors are married).

Regarding infrastructure, a notable majority of vendors reported having sufficient space (84%) and access to canopies and a designated vending spot (69% in both cases). Nearly all vendors use some form of vending furniture or equipment, with only 1% not reporting any. However, levels of vegetation, waterlogging and odors were rated poorly overall, whereas the surface condition, lighting and noise levels received positive ratings.

Having discussed some basic features of the sample, Table 2 presents the results for the OLS models (Models 1, 2, and 3) and the LASSO model (Model 4). Model 4, our preferred model, highlights the selected features. For individual socioeconomic characteristics, three variables are selected in the LASSO model: full-time working ($\beta = 0.170$), having a bank account ($\beta = 0.076$) and working experience ($\beta = 0.001$). The interpretation of these variables in the LASSO model is the same as in an OLS model: Being a full-time street vendor as opposed to a part-time vendor is associated with 17% higher monthly sales, having a bank account is associated with higher monthly sales equivalent to 7.6% and every additional year of business experience is associated with 0.1% higher monthly sales.

Some of the results in this regard are consistent with previous literature such as Tuffour *et al.* (2022), who also found a significant correlation between business experience and sales performance but not with education (also an insignificant variable or with a zero coefficient, as shown in Table 2).

For infrastructure measures, five explanatory variables are identified: having a designated spot ($\beta = 0.227$), having benches and chairs ($\beta = 0.065$), the width of the sidewalk ($\beta = 0.003$) and its condition ($\beta = 0.040$). The level of odors is the only variable with a negative coefficient ($\beta = -0.063$). To illustrate the relationship of some of the most salient explanatory variables

Table 1. Summary statistics

Variables	N	Average	SD	Min	Max
Monthly sales in BDT	243	41,460	17,504	10,000	120,000
Log of monthly sales	243	10.54	0.43	9.21	11.70
<i>Individual socioeconomic characteristics</i>					
Location (1 = New Market area; 0 = Mirpur-1)	243	0.57	0.50	0.00	1.00
Age	243	35.65	9.66	19.00	65.00
Gender (1 = Male; 0 = Female)	243	0.99	0.09	0.00	1.00
Education (1 = Incomplete primary; 2 = Primary; 3 = Secondary; 4 = Higher secondary; 5 = University)	243	1.91	0.78	1.00	5.00
Marital status (1 = Married; 2 = Unmarried; 3 = Widowed; 4 = Divorced)	243	1.23	0.46	1.00	3.00
Household size	243	4.33	1.27	1.00	9.00
Working experience (In years)	243	7.47	6.53	0.17	30.00
Working type (1 = Full-time; 2 = Part-time)	243	0.97	0.18	0.00	1.00
Business = Clothing and shoes	243	0.36	0.48	0.00	1.00
Business = Fruits and vegetable	243	0.16	0.36	0.00	1.00
Business = Aluminum and plastic products	243	0.09	0.28	0.00	1.00
Business = Electrical and mobile accessories	243	0.05	0.23	0.00	1.00
Business = Street food and drinks	243	0.17	0.38	0.00	1.00
Business = Jewelry, bags and accessories	243	0.16	0.37	0.00	1.00
Business = Others	243	0.01	0.09	0.00	1.00
Holds a commercial bank account (1 = Yes; 0 = No)	243	0.53	0.50	0.00	1.00
Access to finance (1 = Yes; 2 = No)	243	0.16	0.36	0.00	1.00
Initial capital amount in BDT	243	57,864	83,746	2,000	600,000
Log of initial capital amount	243	10.48	0.92	7.60	13.30
Capital source = Personal savings/funds	243	0.90	0.30	0.00	1.00
Capital source = Spouse savings/funds	243	0.01	0.09	0.00	1.00
Capital source = Selling own assets	243	0.02	0.14	0.00	1.00
Capital source = Borrowing from relatives and/or friends	243	0.05	0.23	0.00	1.00
Capital source = Support from in-laws	243	0.01	0.11	0.00	1.00
<i>Infrastructure</i>					
Designated or predetermined vending spot (1 = Yes; 0 = No)	243	0.69	0.46	0.00	1.00
Enough space for vending (1 = Yes; 0 = No)	243	0.84	0.36	0.00	1.00
Availability of street vending canopies (1 = Yes; 0 = No)	243	0.69	0.46	0.00	1.00
Availability of street furniture around vending stand (1 = Yes; 0 = No)	243	0.08	0.27	0.00	1.00
Vending furniture and equipment = Stand	243	0.40	0.49	0.00	1.00
Vending furniture and equipment = Floor	243	0.28	0.45	0.00	1.00
Vending furniture and equipment = Table	243	0.56	0.50	0.00	1.00
Vending furniture and equipment = Benches/Chairs	243	0.23	0.42	0.00	1.00
Vending furniture and equipment = Showcase	243	0.12	0.32	0.00	1.00
Vending furniture and equipment = Car/Vehicle	243	0.01	0.11	0.00	1.00
Vending furniture and equipment = Van/Tricycle	243	0.23	0.42	0.00	1.00
Vending furniture and equipment = Drawer	243	0.22	0.42	0.00	1.00
Vending furniture and equipment = Nothing	243	0.01	0.09	0.00	1.00
Rating of the width of the vending area (i.e. street/sidewalk)*	243	3.39	1.03	1.00	5.00
Rating of street/sidewalk surface condition*	243	4.12	0.61	2.00	5.00
Rating of the level of vegetation around the vending stand*	243	1.11	0.46	1.00	4.00
Rating of the hygiene conditions at the vending location*	243	3.65	0.73	2.00	5.00
Level of waterlogging during wet seasons*	243	1.11	0.35	1.00	4.00
Condition of street lighting at night*	243	4.02	0.57	2.00	5.00
Level of odors or smoke at the vending location*	243	1.79	0.86	1.00	4.00
Level of noise at the vending location*	243	4.04	0.95	1.00	5.00
Heat level at the vending site during the summer*	243	4.67	0.52	3.00	5.00
Note(s): *These were reported using the Likert scale, where very bad/very low/not at all hot = 1; bad/low/slightly hot = 2; neutral/moderate hot = 3; good/high/very hot = 4 and very good/very high/extremely hot = 5					
Source(s): Authors' own work					

Table 2. Estimation results: OLS and LASSO

	(1)	St.	(2)	St.	(3)	St.	(4)
	OLS	Error	OLS	Error	OLS	Error	LASSO
<i>Individual characteristics</i>							
Location = Mirpur-1	−0.191***	(0.055)			−0.009	(0.065)	
Age	−0.009***	(0.003)			−0.005	(0.003)	
Gender = Female	−0.312	(0.250)			0.031	(0.247)	
Marital	0.021	(0.070)			0.044	(0.065)	
status = Unmarried							
Marital status = Widowed	0.130	(0.170)			0.056	(0.156)	
Household size	0.033*	(0.019)			0.003	(0.018)	
Working experience	0.022***	(0.004)			0.011**	(0.004)	0.001
Work type = Full-time	0.356***	(0.127)			0.465***	(0.123)	0.170
Business = Fruits and vegetable	−0.117	(0.074)			−0.083	(0.082)	
Business = Aluminum and plastic products	−0.190**	(0.084)			−0.203**	(0.082)	
Business = Electrical and mobile accessories	−0.118	(0.099)			−0.175*	(0.102)	
Business = Street food and drinks	−0.070	(0.075)			−0.024	(0.079)	
Business = Jewelry, bags and accessories	−0.198***	(0.066)			−0.270***	(0.066)	
Business = Others	0.237	(0.259)			0.007	(0.240)	
Education = = 2	0.043	(0.051)			−0.014	(0.047)	
Education = = 3	0.038	(0.067)			0.031	(0.062)	
Education = = 4	0.078	(0.173)			0.038	(0.161)	
Education = = 5	0.761**	(0.350)			0.537	(0.330)	
Capital source = Spouse savings/funds	0.121	(0.234)			0.119	(0.216)	
Capital source = Selling own assets	0.058	(0.167)			0.100	(0.153)	
Capital source = Borrowing from relatives and/or friends	−0.140	(0.103)			−0.085	(0.097)	
Capital source = Support from in-laws	−0.290	(0.194)			−0.431**	(0.179)	
Bank account = Yes	0.119**	(0.054)			0.076	(0.050)	0.076
Access finance = Yes	0.010	(0.067)			−0.020	(0.063)	
Log of initial capital	0.111***	(0.032)			0.091***	(0.031)	
<i>Infrastructure</i>							
Designated vending spot = Yes			0.260***	(0.062)	0.249***	(0.062)	0.227
Enough Vending Space = Yes			−0.020	(0.060)	0.026	(0.058)	
Vending furniture and equipment = Stand			0.040	(0.051)	0.018	(0.051)	
Vending furniture and equipment = Floor			−0.017	(0.054)	−0.109*	(0.058)	
Vending furniture and equipment = Table			0.077	(0.052)	−0.026	(0.055)	
Vending furniture and equipment = Benches/ Chairs			0.181***	(0.056)	0.087	(0.057)	0.065
Vending furniture and equipment = Showcase			0.015	(0.073)	0.024	(0.073)	

(continued)

Table 2. Continued

	(1)	St.	(2)	St.	(3)	St.	(4)
	OLS	Error	OLS	Error	OLS	Error	LASSO
Vending furniture and equipment = Car/Vehicle			−0.130	(0.208)	−0.355*	(0.196)	
Vending furniture and equipment = Van/Tricycle			0.049	(0.063)	−0.065	(0.067)	
Vending furniture and equipment = Drawer			0.099*	(0.053)	−0.016	(0.055)	
Vending furniture and equipment = Nothing			0.157	(0.236)	0.105	(0.219)	
Canopy = Yes			0.066	(0.056)	−0.008	(0.055)	0.003
Rating of the width of the vending area (Street/Sidewalk)			0.072***	(0.026)	0.063**	(0.026)	0.040
Rating of street/sidewalk surface condition			0.009	(0.040)	0.020	(0.038)	
Street furniture = Yes			−0.019	(0.083)	−0.011	(0.079)	
Rating of the level of vegetation around the vending stand			0.051	(0.048)	0.064	(0.046)	
Rating of the hygiene conditions at the vending location			0.019	(0.035)	0.024	(0.033)	
Level of waterlogging during wet seasons			0.127*	(0.066)	0.040	(0.064)	
Condition of street lighting at night			−0.051	(0.039)	−0.038	(0.036)	
Level of odors or smoke at the vending location			−0.144***	(0.033)	−0.105***	(0.034)	−0.063
Level of noise at the vending location			−0.028	(0.025)	−0.007	(0.024)	
Heat level at the vending site during the summer			−0.000	(0.046)	0.008	(0.046)	
Constant	9.423***	(0.478)	10.233***	(0.368)	8.917***	(0.591)	10.135
Observations	243		243		243		243
R ²	0.499		0.505		0.645		
RMSE	0.322		0.318		0.285		0.322
AIC	162.949		153.938		122.793		
Source(s): Authors' own work							

based on Model 4: having a designated spot is associated with 22.7% higher sales compared to those without a designated spot. Similarly, a one-point increase in the rating of odors (worse level of odors) is associated with a 6% decrease in sales. Thus, our findings support [Kesumasari \(2020\)](#), who found that Surakarta street vendors, known as *angkringan*, make customers' experience physically and socially comfortable by arranging them under shade trees, giving them enough space and providing mattresses for them to relax and socialize on, which promotes sales.

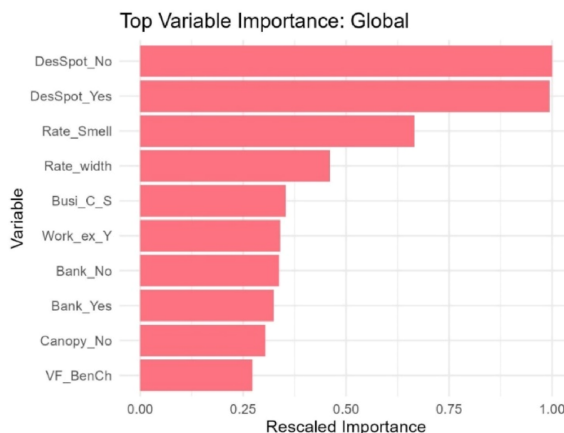
While there is significant overlap in some of the variable coefficients between OLS and LASSO, there are notable differences. Some variables selected in the LASSO model do not show consistent significance across all models. For instance, having a bank account appears significant in Model 1, but not in Model 3 once all variables are introduced. Additionally, the size of the coefficients for working experience and full-time work appears to be upward biased in the OLS models, leading to an overestimation of these individual traits.

These initial results highlight the importance of the urban space and its management in predicting the sales performance of vendors as more variables linked to urban infrastructure – and with relatively bigger coefficients – were selected in the preferred machine learning model (Model 4).

4.1 Random forests

Figure 3 shows that (not) having a designated spot (DesSpot_No; DesSpot_Yes) is the most important factor, with its importance rescaled to 1.00 or 100% for ease of interpretation. The next most important variable is the rate of odors (Rate_Smell), which is only 65% as important as having a designated spot (DesSpot_Yes). Variables of intermediate importance, ranging from 25% to 50% of the importance of having a designated spot (DesSpot_Yes), include the width of the sidewalk (Rate_width), having a business selling clothes and shoes (Busi_C_S), working experience (Work_ex_Y), having a bank account (Bank_Yes), not having a canopy (Canopy_No) and having benches and chairs (VF_BenCh). Except for having a business selling clothes and shoes (Busi_C_S), these variables are consistent with the results of the LASSO model (Model 4). These findings also highlight that the most important variables are predominantly related to urban infrastructure and how vendors access it, including whether they have a designated spot and how much space they have for their vendor activities, as proxied by the width of the sidewalk.

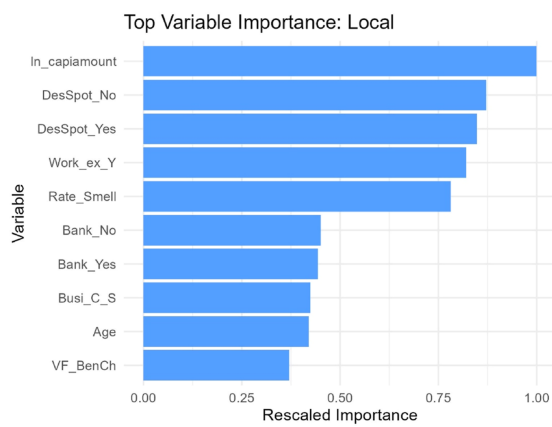
Figure 4 provides a slightly different picture once local models are introduced for each spatial unit, accounting for spatial heterogeneity across vending spots. In this model, the log of capital amount (ln_capiamount) emerges as the most important variable, consistent with the findings of Adhikari (2011) and Handoyo and Wijayanti (2021). The importance of having a designated spot (DesSpot_Yes) remains significant, as seen in the LASSO model (Model 4) and the RF model (Figure 3). Business experience (Work_ex_Y) and its rate are also crucial, being at least 75% as important as the log of capital amount (ln_capiamount). Other variables from the RF model in Figure 3, such as having a bank account (Bank_Yes) and having benches and chairs (VF_BenCh), rank similarly. A new variable, age (Age), also emerges as important, as identified in previous studies like Nilakusmawati *et al.* (2019). This model demonstrates that when controlling for variations linked to specific locations within the market, more individual variables, especially capital-related ones and age, become significant.



Note(s): Values have been rescaled so that 1.00 is the most important feature or variable. Variables with a rescaled importance of higher than 0.25 are reported

Source(s): Authors' own work

Figure 3. Variable importance RF



Note(s): Values have been rescaled so that 1.00 is the most important feature or variable. Variables with a rescaled importance of higher than 0.25 are reported
Source(s): Authors’ own work

Figure 4. Variable importance GRF

4.2 Model assessment

Which RF model is preferred, the global (RF) or the local (GRF)? Based on the metrics in Table 3 [3], both models perform similarly. While metrics in the training data (non OOB) are better in the local (GRF) model, the RF provides better OOB measures: namely, the RF model has a lower OOB prediction error (0.097) compared to the GRF (0.105). Likewise, it also has a higher R^2 (47.27%) compared to the GRF (43.05%). Thus, the RF model is preferred. We place a higher emphasis on OOB metrics, as these provide a more realistic estimate of the model’s performance on unseen data, which is crucial for assessing how well the model generalizes.

Considering this, we can conclude that having a vending spot is the most important factor in vending, as the global RF model explains the best vendors’ sales performance across the New Market area and Mirpur-1. If spatial heterogeneity is taken into account, the initial economic situation of the vendor, proxied by the log of initial capital, is the most important factor next to the vending spots.

In summary, the flexibility of non-linear modeling identifies new, salient variables related to individual characteristics that were not earlier identified in the regression results in Table 2. Regarding infrastructure, the most important feature is having a designated vending spot, followed by the rate of odors and the width of the sidewalk, reflecting aspects related to the quality of infrastructure and vendor access. Important individual variables to explain sales are the type of goods sold (e.g. apparel), having a bank account, work experience, capital and age.

Table 3. RF and GRF model assessment metrics

Metric	Global (RF) model	Local (GRF) model
OOB prediction error (MSE)	0.097627	0.105
OOB R -squared %	47.27795	43.046
Mean squared error (Not OOB)	0.06	0.047
R -squared (Not OOB) %	67.682	74.683
AIC (Not OOB)	−559.293	−618.62
Source(s): Authors’ own work		

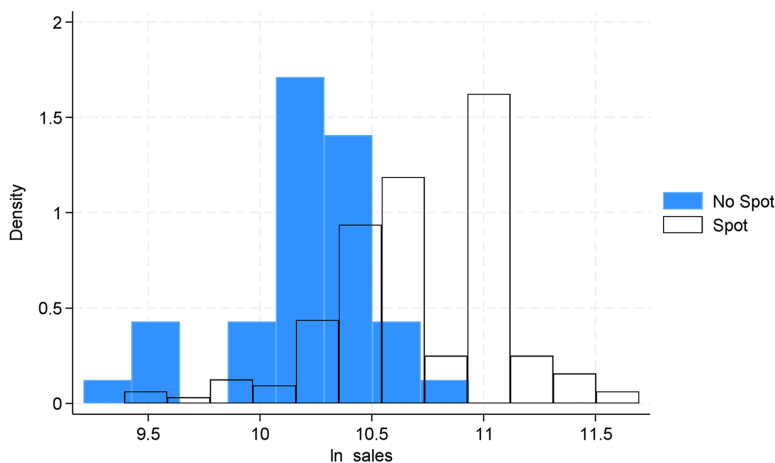
From a policy perspective, an important factor to highlight relates to formal banking services, not necessarily in the form of credits – which has been often underscored in the Bangladeshi context. Uddin (2021), for instance, highlights the importance of loans for improving the livelihoods of street vendors in Bangladesh. The results of the present study, however, identify that having a bank account is more important than having access to finance – a factor that was never statistically significant or selected as important in any of the models. The results in the LASSO model identified that having a bank account is associated with 7.6% higher sales on average. This finding thus suggests that this instrument and its possible policy implications on street vendors' income may be worth exploring further.

Our final discussion points focus on the importance of having a designated vending spot in street markets, given the robustness of this item across different models presented in this study. Figure 5 shows the difference in the distribution of the log of sales of street vendors, with a difference of 0.50 in logs, equivalent to approximately 17,500 BDT (equivalent to approximately 200 USD). This is a substantial difference in income for street vendors in Bangladesh, who historically have had to work in difficult circumstances, even when compared with street vendors in neighboring countries (Bhowmik, 2005).

Naturally, despite the improved modeling provided by the machine learning algorithms employed to produce robust and generalizable results, our study does not establish causality. Nonetheless, it offers an important starting point to reconsider how public spaces are managed, particularly in relation to the critical role of vending spots for street vendors.

The link between socio-spatial exclusion of street vendors and their ability to secure vending spots has been hinted at in previous studies (e.g. Bhowmik, 2005, in the case of Cambodia). Boonjubun (2022) provides a detailed account of the relationship between having a fixed spot and the competitiveness of street vendors. Fixed-stall vendors do not have to compete daily for a vending spot, which is crucial for maintaining consistent customer outreach. Additionally, these vendors save time and effort by not having to move their equipment and set up daily, making their activities more efficient. The absence of fixed spots leads to intense competition and a higher probability of eviction by authorities, resulting in uncertainty and instability.

This issue is particularly crucial in the context of Bangladesh. Securing a fixed spot for street vendors would require that some basic elements for operation are in place. As



Source(s): Authors' own calculations

Figure 5. Histogram for log of monthly sales

highlighted earlier, street vendors in Bangladesh are not only seen as a public nuisance and lack basic recognition, but they are also subject to direct harassment and often must bribe officials to be able to operate (Bhowmik, 2005; Uddin, 2021). This last item therefore suggests the possibility of some level of causal circularity, whereby poorer vendors are not able to afford bribes, losing the right to operate in the public space with a bit more certainty and reach to potential customers, making them poorer.

5. Conclusion

Utilizing machine learning techniques, specifically LASSO, RF and GRF, this study identifies the most critical factors influencing street vendors in Bangladesh. The study covers the New Market Area and Mirpur-1, with a sample of 243 vendors selected through simple random sampling. This approach allows for a robust comparison between individual socioeconomic factors and infrastructure-related factors, which have traditionally been explored separately or in qualitative settings (in the case of the latter factors).

The results indicate that having a designated vending spot is the most significant factor in predicting sales performance, far surpassing other variables in the preferred RF model. Vendors with designated spots make around 27% higher income than those without, emphasizing the need for secure and stable vending locations to improve income and livelihoods.

Other key urban infrastructure variables associated with better sales performance include the width of the sidewalk and the level of odors. This suggests that improved infrastructure conditions, such as wider sidewalks and lower levels of odors, are associated with better sales outcomes for street vendors – possibly more than many individual characteristics like age or even having a bank account.

Important individual characteristics – were in the preferred model – came second to infrastructure characteristics: identified salient features having a bank account, working experience and the type of goods sold, such as apparel. The initial economic situation of the vendor, measured by the log of initial capital, is also significant, especially when accounting for spatial heterogeneity in the GRF model. Full-time work is another characteristic that appears important – yet only in the less preferred LASSO model.

Policy-wise, the findings suggest that formal banking services, such as having a bank account, are more crucial for improving street vendors' income than access to credit – which is often highlighted in some studies in the Bangladeshi context. Yet the central contribution of the study is that while these items are important, the performance of street vending entrepreneurship is more strongly associated with overall public spaces and their infrastructure.

Thus, this study highlights the link between socio-spatial exclusion and vending spots, supporting previous research that vendors without fixed spots face intense competition, uncertainty and instability (Bhowmik, 2005; Boonjubun, 2022). Fixed-stall vendors benefit from not having to move their equipment daily, saving time and effort, which enhances their efficiency and competitiveness.

The situation of street vendors in Bangladesh reflects a common perception in many developing countries, where street vendors – at best – are seen as “invisible entrepreneurs” (Sekhani *et al.*, 2019; Uddin, 2021) or – at worst – as public nuisances subject to harassment and bribery by authorities. This study underscores the importance of secure vending spots and improved infrastructure to support these vendors.

It is important to acknowledge that while the study employs advanced machine learning models to improve the robustness and accuracy of predictions, it does not establish causality. However, the findings provide a quantified picture of the most significant factors associated with street vending “success,” primarily related to public urban infrastructure and access. These results call for a reevaluation of how public spaces are managed to better support street vendors, recognizing their contribution to the urban economy.

Notes

1. We opt to use this term as the income of street vendors may extend beyond their entrepreneurial activities on the streets, for instance, via welfare transfers. We investigate how the latter activities enhance their livelihoods and therefore focus on sales performance to be more precise.
2. In this case, the minimization is done on the basis of gradient descent. In an OLS, this is done on a least squares minimization algorithm.
3. As expected, the RF and GRF models outperform the OLS models on all goodness-of-fit measures for the training data (non-OOB in Table 3). When comparing the OLS specification with the best metrics (Model 3 in Table 2) against the metrics in Table 3 for the random forest algorithms, we observe that the RMSE of RF ($0.244 \approx \sqrt{0.06}$) and GRF ($0.217 \approx \sqrt{0.047}$) are lower than those of OLS (0.285). Additionally, the AIC values for RF (−559.293) and GRF (−618.62) are significantly lower than for OLS Model 3 (122.793), indicating a better fit. The R^2 values for RF (67.682%) and GRF (74.683%) are also higher than for OLS Model 3 (64.5%). Comparing Lasso with RF and GRF is more complex because Lasso's RMSE should ideally be evaluated on unseen data. Given the small sample size in our dataset and our primary use of Lasso as a variable selection tool, we only have the RMSE for the training data (0.322). This RMSE is outperformed by both RF and GRF, which have lower RMSE values on the training data (RF: 0.244 and GRF: 0.217).

References

- Adair, P. (2020), "The non-observed economy vs the shadow economy in the European Union: concepts, measurements methods and estimates revisited", in *Research Handbook on Development and the Informal Economy*, Edward Elgar Publishing, pp. 45–78.
- Adhikari, D.B. (2011), "Income generation in informal sector: a case study of the street vendors of Kathmandu Metropolitan City", *Economic Journal of Development Issues*, Vols 13&14 Nos 1&2, pp. 1–14, doi: [10.3126/ejdi.v13i0.7193](https://doi.org/10.3126/ejdi.v13i0.7193).
- Aliaga Linares, L. (2018), "The paradoxes of informalizing street trade in the Latin American city", *International Journal of Sociology and Social Policy*, Vol. 38 Nos 7–8, pp. 651–672, doi: [10.1108/IJSSP-09-2017-0119](https://doi.org/10.1108/IJSSP-09-2017-0119).
- Altman, N. and Krzywinski, M. (2017), "Ensemble methods: bagging and random forests", *Nature Methods*, Vol. 14 No. 10, pp. 933–935, doi: [10.1038/nmeth.4438](https://doi.org/10.1038/nmeth.4438).
- Amegah, A.K. and Jaakkola, J.J. (2014), "Work as a street vendor, associated traffic-related air pollution exposures and risk of adverse pregnancy outcomes in Accra, Ghana", *International Journal of Hygiene and Environmental Health*, Vol. 217 Nos 2–3, pp. 354–362, doi: [10.1016/j.ijheh.2013.07.010](https://doi.org/10.1016/j.ijheh.2013.07.010).
- Barbareco, G.Q., Reis, A.V.P., Lopes, G.D.R., Boaventura, L.P., Castro, A.F., Vilanova, T.C.F., Da Cunha Júnior, E.C., Pires, K.C., Pôrto Filho, R. and Pereira, B.B. (2019), "Effects of environmental noise pollution on perceived stress and cortisol levels in street vendors", *Journal of Toxicology and Environmental Health, Part A*, Vol. 82 No. 5, pp. 331–337, doi: [10.1080/15287394.2019.1595239](https://doi.org/10.1080/15287394.2019.1595239).
- Basu, S. and Nagendra, H. (2020), "The street as a workspace: assessing street vendors' rights to trees in Hyderabad, India", *Landscape and Urban Planning*, Vol. 199, 103818, doi: [10.1016/j.landurbplan.2020.103818](https://doi.org/10.1016/j.landurbplan.2020.103818).
- Bhowmik, S.K. (2005), "Street vendors in Asia: a review", *Economic and Political Weekly*, Vol. 40, pp. 2256–2264.
- Bhowmik, S. (2012), *Street Vendors in the Global Urban Economy*, Taylor & Francis, New Delhi.
- Boonjubun, C. (2022), "Struggles of street food vendors in Helsinki: A study of food truck and fixed stall vending using municipal land", *Local Environment*, Vol. 27 No. 7, pp. 801–823, doi: [10.1080/13549839.2022.2077711](https://doi.org/10.1080/13549839.2022.2077711).
- Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5–32, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- Bromley, R. (2000), "Street vending and public policy: a global review", *International Journal of Sociology and Social Policy*, Vol. 20 Nos 1/2, pp. 1–28, doi: [10.1108/01443330010789052](https://doi.org/10.1108/01443330010789052).

- Brownlee, J. (2020), "Statistical methods for machine learning", eBook, available at: https://machinelearningmastery.com/statistics_for_machine_learning/
- Caldarola, B. (2022), "Structural change(s) in Ghana: a comparison between the trade, formal and informal sectors", (SSRN Scholarly Paper 4306229), doi: [10.2139/ssrn.4306229](https://doi.org/10.2139/ssrn.4306229).
- Chen, M. and Carré, F. (2020), *The Informal Economy Revisited: Examining the Past, Envisioning the Future*, Taylor & Francis, London, p. 326.
- Daniels, P.W. (2004), "Urban challenges: the formal and informal economies in mega-cities", *Cities*, Vol. 21 No. 6, pp. 501-511, doi: [10.1016/j.cities.2004.08.002](https://doi.org/10.1016/j.cities.2004.08.002).
- De Soto, H. (1989), *The Other Path*, Harper & Row, New York, p. 17133.
- Elgin, C. (2020), *The Informal Economy: Measures, Causes, and Consequences*, Routledge, London.
- Faruque, Q., Haque, Q.F., Shekhar, H.U. and Begum, S. (2010), "Institutionalization of healthy street food system in Bangladesh: a pilot study with three wards of Dhaka City Corporation as a model", *National Food Policy Capacity Strengthening Programme (NFPCSP)*, Vol. 7, pp. 1-84.
- Georganos, S. and Kalogirou, S. (2022), "A forest of forests: a spatially weighted and computationally efficient formulation of geographical random forests", *ISPRS International Journal of Geo-Information*, Vol. 11 No. 9, 471, doi: [10.3390/ijgi11090471](https://doi.org/10.3390/ijgi11090471).
- Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E. and Kalogirou, S. (2019), "Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling", *Geocarto International*, Vol. 36 No. 2, pp. 121-136, doi: [10.1080/10106049.2019.1595177](https://doi.org/10.1080/10106049.2019.1595177).
- Ghatak, I. and Chatterjee, S. (2018), "Urban street vending practices: an investigation of ethnic food safety knowledge, attitudes, and risks among untrained Chinese vendors in Chinatown, Kolkata", *Journal of Ethnic Foods*, Vol. 5 No. 4, pp. 272-285.
- Gollin, D., Jedwab, R. and Vollrath, D. (2016), "Urbanization with and without industrialization", *Journal of Economic Growth*, Vol. 21 No. 1, pp. 35-70, doi: [10.1007/s10887-015-9121-4](https://doi.org/10.1007/s10887-015-9121-4).
- Grekousis, G., Feng, Z., Marakakis, I., Lu, Y. and Wang, R. (2022), "Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: a geographical random forest approach", *Health and Place*, Vol. 74, 102744, doi: [10.1016/j.healthplace.2022.102744](https://doi.org/10.1016/j.healthplace.2022.102744).
- Handoyo, E. and Wijayanti, T. (2021), "The factors affecting the welfare of street vendors in Indonesia", *Planning Malaysia: Journal of the Malaysian Institute of Planners*, Vol. 19 No. 4, pp. 231-243, doi: [10.21837/pm.v19i18.1048](https://doi.org/10.21837/pm.v19i18.1048).
- Hussain, S., Yasmin, S. and Islam, M.S. (2015), "Assessment of the socioeconomic aspects of street vendors in Dhaka city: evidence from Bangladesh", *Asian Social Science*, Vol. 11 No. 26, pp. 2256-2264, doi: [10.5539/ass.v11n26p1](https://doi.org/10.5539/ass.v11n26p1).
- Hussain, R. (2019), "City, Informality and poverty: the polarization of the street vendors in Dhaka City, Bangladesh", *Italian Sociological Review*, Vol. 9 No. 3, pp. 413-430.
- Israt, A.S. and Adam, M. (2017), "An investigation of physical attributes relevant to the informal street hawking in the urban spaces of Dhaka city", *Journal of Design and Built Environment*, Vol. 17 No. 1, pp. 75-86, doi: [10.22452/jdbe.vol17no1.5](https://doi.org/10.22452/jdbe.vol17no1.5).
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021), *An Introduction to Statistical Learning with Applications in R*, 2nd ed., Springer, available at: <https://www.statlearning.com>
- Kambara, C. and Bairagya, I. (2021), "Earnings and investment differentials between migrants and natives: a study of female street vendors in Bengaluru city", *Environment and Urbanization ASIA*, Vol. 12 No. 1, pp. 56-72, doi: [10.1177/0975425321990318](https://doi.org/10.1177/0975425321990318).
- Kesumasari, D. (2020), "Understanding physical settings of street vendors in Surakarta, Indonesia", *IOP Conference Series: Earth and Environmental Science*, Vol. 490 No. 1, 012002, doi: [10.1088/1755-1315/490/1/012002](https://doi.org/10.1088/1755-1315/490/1/012002).

- Lemessa, S.D., Watabaji, M.D., Yismaw, M.A. and Tadesse, A.B. (2021), "Evening street vending and the tragic living conditions of vendors: the case of eastern Ethiopia region", *Cities*, Vol. 108, 102947, doi: [10.1016/j.cities.2020.102947](https://doi.org/10.1016/j.cities.2020.102947).
- Loayza, N. (2018), "Informality: why is it so widespread and how can it Be reduced?", (SSRN Scholarly Paper 3360124), available at: <https://papers.ssrn.com/abstract=3360124>
- Loayza, N.V. and Rigolini, J. (2011), "Informal employment: safety net or growth engine?", *World Development*, Vol. 39 No. 9, pp. 1503-1515, doi: [10.1016/j.worlddev.2011.02.003](https://doi.org/10.1016/j.worlddev.2011.02.003).
- Luo, Y., Yan, J. and McClure, S. (2021), "Distribution of the environmental and socioeconomic risk factors on COVID-19 death rate across continental USA: a spatial nonlinear analysis", *Environmental Science and Pollution Research*, Vol. 28 No. 6, pp. 6587-6599, doi: [10.1007/s11356-020-10962-2](https://doi.org/10.1007/s11356-020-10962-2).
- Martínez, L., Short, J.R. and Estrada, D. (2017), "The urban informal economy: street vendors in Cali", *Colombia. Cities*, Vol. 66, pp. 34-43, doi: [10.1016/j.cities.2017.03.010](https://doi.org/10.1016/j.cities.2017.03.010).
- Michalak, L. (2020), "Street vendors in Tunisia: encountering the informal economy", in *Research Handbook on Development and the Informal Economy*, Edward Elgar Publishing, pp. 240-262.
- Moniruzzaman, A.K.M., Amin, M.R. and Alam, M.J. (2018), "Street vending as an alternative to self-employment and a way to income generation in Bangladesh", *Journal of Economics and Sustainable Development*, Vol. 9 No. 8, pp. 172-178.
- Nilakusmawati, D.P.E., Susilawati, M. and Wall, G. (2019), "A welfare model of street vendors: cases from Denpasar, Bali, Indonesia", *International Journal of Human Resource Studies*, Vol. 9 No. 3, pp. 1-20, doi: [10.5296/ijhrs.v9i3.14973](https://doi.org/10.5296/ijhrs.v9i3.14973).
- Pourhoseingholi, M.A., Baghestani, A.R. and Vahedi, M. (2012), "How to control confounding effects by statistical analysis", *Gastroenterology and Hepatology from Bed to Bench*, Vol. 5 No. 2, pp. 79-83.
- Recchi, S. (2021), "Informal street vending: a comparative literature review", *International Journal of Sociology and Social Policy*, Vol. 41 Nos 7/8, pp. 805-825, doi: [10.1108/ijssp-07-2020-0285](https://doi.org/10.1108/ijssp-07-2020-0285).
- Roy, A. (2005), "Urban informality: toward an epistemology of planning", *Journal of the American Planning Association*, Vol. 71 No. 2, pp. 147-158, doi: [10.1080/01944360508976689](https://doi.org/10.1080/01944360508976689).
- Roy, D. (2017), "IS – LM model revisited from the perspective of the underground economy", *Journal of Money Laundering Control*, Vol. 20, pp. 311-319, July 3, doi: [10.1108/jmlc-08-2016-0039](https://doi.org/10.1108/jmlc-08-2016-0039).
- Schneider, F. and Williams, C.C. (2013), "The shadow economy", *Social Science Research Network*, doi: [10.2139/ssrn.3915632](https://doi.org/10.2139/ssrn.3915632).
- Sekhani, R., Mohan, D. and Medipally, S. (2019), "Street vending in urban 'informal' markets: reflections from case-studies of street vendors in Delhi (India) and Phnom Penh City (Cambodia)", *Cities*, Vol. 89, pp. 120-129, doi: [10.1016/j.cities.2019.01.010](https://doi.org/10.1016/j.cities.2019.01.010).
- Teltscher, S. (1994), "Small trade and the world economy: informal vendors in Quito, Ecuador", *Economic Geography*, Vol. 70 No. 2, pp. 167-187, doi: [10.2307/143653](https://doi.org/10.2307/143653).
- Tibshirani, R. (1997), "The lasso method for variable selection in the cox model", *Statistics in Medicine*, Vol. 16 No. 4, pp. 385-395, doi: [10.1002/\(sici\)1097-0258\(19970228\)16:4<385::aid-sim380>3.0.co;2-3](https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3).
- Tuffour, J.K., Oppong, M., Nyanyofio, J.G.T., Abukari, M.F., Addo, M.N. and Brako, D. (2022), "Micro-entrepreneurship: examining the factors affecting the success of women street food vendors", *Global Business Review*, Vol. 0 No. 0, doi: [10.1177/09721509211072380](https://doi.org/10.1177/09721509211072380).
- Uddin, M.A. (2021), "Street vending: exploring an excluded economic sector in Chittagong city, Bangladesh", *Development in Practice*, Vol. 31 No. 3, pp. 380-392, doi: [10.1080/09614524.2020.1845613](https://doi.org/10.1080/09614524.2020.1845613).
- Williams, C.C. (2008), "Beyond necessity-driven versus opportunity-driven entrepreneurship: a study of informal entrepreneurs in England, Russia and Ukraine", *The International Journal of Entrepreneurship and Innovation*, Vol. 9 No. 3, pp. 157-165, doi: [10.5367/000000008785096647](https://doi.org/10.5367/000000008785096647).
- Williams, C. (2023), *A Modern Guide to the Informal Economy*, Edward Elgar Publishing, Cheltenham.

Further reading

Irakunda, D. and Van Bergeijk, P.A. (2020), "Financial inclusion of urban street vendors in Kigali", *Journal of African Business*, Vol. 21 No. 4, pp. 529-543, doi: [10.1080/15228916.2019.1695182](https://doi.org/10.1080/15228916.2019.1695182).

Corresponding author

Beatriz Calzada Olvera can be contacted at: calzadaolvera@ihs.nl