Earthquake Vulnerability and Small-Scale Production Ecosystems in Istanbul *(Extended Abstract)*

Mehmet Emin Sarıhan[1]

[1]Dept of Urban and Regional Planning, Faculty of Architecture, Istanbul Technical University, Taşkışla Campus, 34367 Şişli/İstanbul

Abstract

Within the metropolitan landscape of Istanbul, specialized clusters of small-scale production areas remain tightly integrated into residential neighborhoods. As a sprawling metropolis, Istanbul is characterized by specialized zones of small-scale production deeply embedded within its residential fabric. This complex urban spatial structure faces profound threats from seismic risk and compounding fragilities, compromising the resilience of both the city's inhabitants and its economic cores. This study *seeks for traces* of the "invisible" spatial intersection of small-scale production ecosystems and seismic vulnerability in Istanbul, a dimension often overlooked in traditional land-use planning. The aim is to show the relationship between small scale production density, socio-economic status, and predicted building damage to identify latent earthquake risk zones *based on neighborhood level*. Utilizing a "Geospatial Big Data" approach, the methodology *employed to visualize in GIS around* 48,000 Point of Interest (POI) data entries *extracted by Python* from Google Maps to serve as a proxy for *small-scale production locations'* intensity. The study employs K-Means Cluster Analysis on Istanbul's neighborhoods to identify non-linear spatial typologies. *Most of the* results reveal a significant "spatial mismatch": the highest structural vulnerability (Cluster 5) is found in residential zones with low *production* activity, driven by building fragility rather than function. However, a critical "mixed-use" typology (Cluster 3) was identified where high production intensity co-locates with high damage risk *around urban residential areas,* creating *risks* for secondary chain reaction hazards. The study concludes that urban risk mitigation must shift from broad policies to typology-specific interventions, specifically targeting these hidden, hybrid production behind residential clusters.

Keywords: ***Seismic Vulnerability, Urban Resilience, Small-Scale Production, K-Means Clustering, Geospatial Big Data, Istanbul.***

Materials and Methods

A seismic clustering model is an effective approach for organizing extensive datasets in data collecting (Shafapourtehrany et. al., 2022). It was firstly suggested by MacQueen in 1967, it is regarded as one of the simplest unsupervised learning techniques and was used to address a prominent clustering problem (Karmenova et. al, 2020). K-means is a partitioned clustering method in which each cluster is associated with a centroid, and each data point is assigned to the cluster with the nearest centroid. The procedures used in this approach, as per Han et al. (2022). In data-poor environments, K-Means Clustering is applied to stratify neighborhoods based on variable similarity, serving as a mechanism to establish ground-truth datasets and map spatial vulnerability (Shafapourtehrany et. al., 2022). Consequently, the utilization of geospatial big data, specifically from Google Maps and POIs, makes it possible to scientifically reveal the intersection of concealed structural weaknesses, industrial risks, and socio-economic risk strata (co-location effects) within the dense urban fabric of Istanbul (Uckardesler, 2025; Peng et. al., 2026).

This study employs a sequential, a quantitative framework designed to interrogate the relationship between functional urban usage (specifically small-scale industrial density) and physical seismic vulnerability. Through K-Means Clustering this study employing the statistical analysis of Exploratory Spatial Data Analysis (ESDA) by K-Means Clustering to uncover subtle, non-linear spatial typologies and co-location patterns.

Data Sources:

1. **Small Production Data (Independent Variable):** Point data extracted from GoogleMaps data by using Python (geopandas library) and approximately 48.000 points of representing small-scale manufacturing and industrial units visualised. (aggregated to total number of buildings in a neighborhood via ArcGIS, 2025).
2. **Seismic Vulnerability Data (Dependent Variable):** Predicted "Heavily Damaged" building counts based on neighborhoods prepared for a M-w >7.5 earthquake scenario (Source: IMM/Kandilli Observatory Disaster Lots Estimates Booklets, 2020).
3. **Socio-Economic Status (SES):** A categorical index (Scale 1–8) representing the development level of the neighborhood by “Mahallem Istanbul Project, 2021”.

4. **Population Density:** Persons per hectare, calculated via spatial join of demographic data to neighborhood geometries from “TUIK, 2024”.

The study integrates neighborhood-level (Neighborhoods=957 in total) datasets-based Istanbul Neighborhoods. To quantify functional land use, small-scale industrial point data were spatially joined to administrative boundaries using GIS and normalized against the total building stock to derive comparable production density ratios.

Data processing and spatial analysis were conducted using Python, utilizing the GeoPandas library for spatial joins and coordinate transformations. The data extraction is held by **“geopandas” Phyton** tool for extraction of POI data from Google Maps via keywords. This process consists 2 phases from extraction of Turkish keywords (around 35.000 POI data) and English keywords (15.000 POI data) with withdrawal of overlapping data.

Small-Scale Production Data Extraction Words

Main Sector Group	Covered Subsectors (Raw Values in the Data Set)
1. Automobile and Vehicle Light-Industry	auto_body_shop, auto_customization, auto_detailing, auto_electrical_repair, auto_glass_service, auto_manufacturers, auto_parts_supply, auto_restoration, automobile_leasing, automobile_registration, automotive, automotive_repair, bike_repair, boat_parts, boat_rental, boat_service, car_stereo_store, car_window_tinting, engine_repair, mobility_equipment, motorcycle_manufacturer, motorcycle_repair, tire_dealer, truck_repair
2. Construction and Hardware Industry	aluminum_supplier, awning_supplier, building_supply, carpenter, commercial_industrial, construction_services, countertop_installation, glass_manufacturer, granite_supplier, home_developer, industrial_cleaning, industrial_company, industrial_equipment, iron_and_steel, lumber_store, metal_fabricator, metal_plating, metal_supplier, metals, paper_mill, plastic_fabrication, steel_fabricators, water_heater_repair, water_treatment, windows_installation, wood_and_pulp, furniture_manufacturers
3. Textiles and Apparel	bags_luggage_company, clothing_company, fabric_wholesaler, leather_goods, linen, mattress_manufacturing, sewing_and_alterations, shoe_repair
4. Logistics and Infrastructures	automotive_storage_facility, junk_removal_and_hauling, junkyard, packing_supply, storage_facility, warehouses, automation_services
5. Manufacturing and Services	appliance_manufacturer, appliance_repair, appliance_store, business_equipment, business_manufacturing, carpet_cleaning, caterer, cleaning_products, computer_hardware, computer_wholesaler, food_beverage_distribution, restaurant_equipment, sign_making

In the Table 1, “geopandas” Phyton keywords for the extraction keywords are listed.

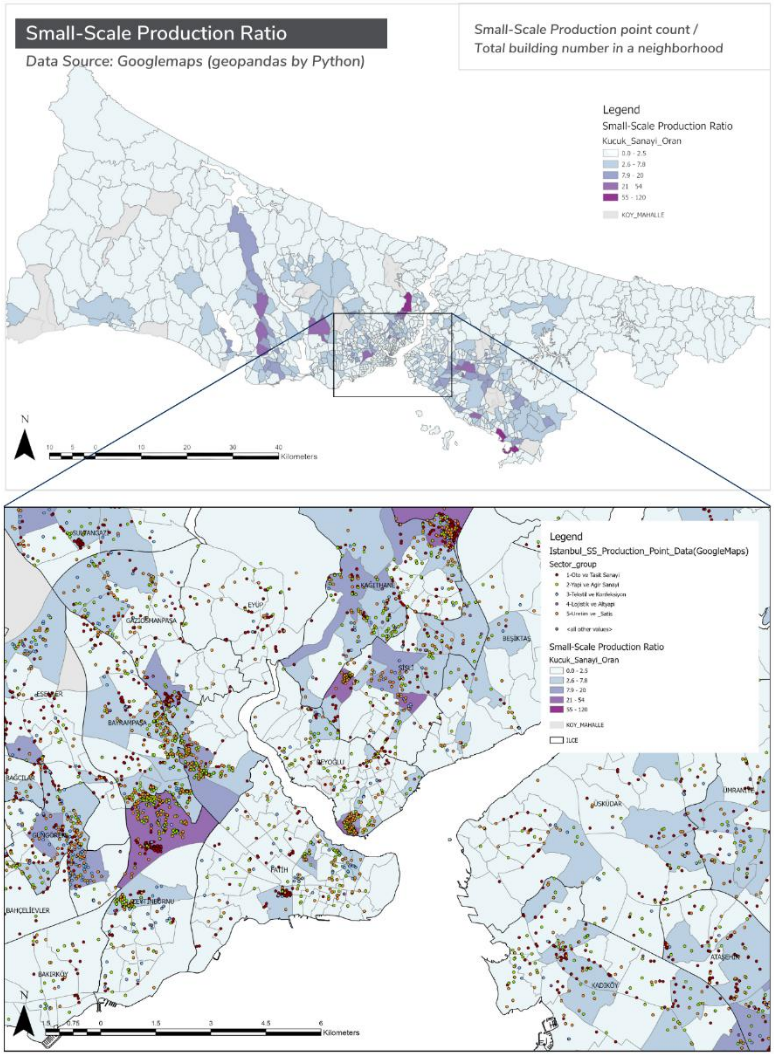

Figure 2. Istanbul Spatial Small Scale Production Point Data & Ratio by Neighborhoods

Statistical Method: K Means Cluster Analysis

The study explores through to **K-Means Cluster Analysis**, a multivariate statistical technique used to partition the dataset into distinct, non-overlapping subgroups (clusters) based on feature similarity.

Data Pre-processing: Standardization (Z-Scores)

Given the scalar disparity between variables (e.g., population in thousands vs. small scale production density in decimals), all datasets were transformed into **Z-scores** prior to clustering. This ensured that no single variable dominated the Euclidean distance calculations due to magnitude, allowing for an unbiased structural grouping.

The Clustering of Neighborhoods:

The K-Means algorithm iteratively classified neighborhoods into **five distinct clusters** based on the multidimensional proximity of their Z-score values. This method allowed for the identification of **"Spatial Co-location Patterns"** in which areas where specific variables (e.g., high risk and low Socio-Economic Scores) intersect, even if they do not correlate globally.

4. Results and Discussion

The K-Means Clustering analysis of the resulting clusters will be revealed in table and mapping visualizations format for comparisons (in full paper format).

The primary problem addressed in this research is the inadequacy of traditional, static urban inventories to capture the dynamic and "invisible" risk factors in rapidly urbanizing metropolises. Specifically, small-scale production facilities (KSS) often infiltrate residential fabrics ("merdiven altı"), creating unregistered pockets of systemic vulnerability that official zoning maps fail to record. The purpose of this paper was to operationalize digital "proxy data" (Google Maps POIs) to expose these hidden intersections between economic activity and physical fragility.
With the K-Means Cluster Analysis, it reveals a spatial divergence. The results indicate that high-risk zones (Cluster 5) form a distinct neighborhood typology characterized by moderate density and low industrial activity. Conversely, areas with high industrial concentration (Cluster 3) exhibit lower-than-average predicted damage. This suggests that urban risk mitigation strategies should prioritize specific neighborhood typologies rather than focusing solely on land-use functions.

Conclusion

The findings conceptually reframe Istanbul's risk landscape as one defined by "Spatial Mismatch" rather than direct correlation. The study proves that the zones with the highest probability of collapse (Cluster 5) are residential and economically stagnant. In contrast, the zones with the highest potential for secondary cascading disasters (Cluster 3) are economically vibrant yet structurally compromised. The "Hidden" Small-Production City: The identification of Cluster 3 demonstrates that industrial risk is not confined to designated "Organized Industrial Zones" (Cluster 2) but is embedded within the living tissue of low-income neighborhoods.

To improve this work, future research should integrate temporal mobility data to accurately calculate human exposure levels (casualties) distinguishing between working hours and sleeping hours. Additionally, incorporating specific secondary hazard modeling (e.g., fire spread simulation or chemical dispersion models) into the identified "Cluster 3" zones would transition this study from a risk inventory to an actionable disaster management plan. Finally, validating the "digital" POI findings with localized field surveys in selected high-risk neighborhoods would bridge the gap between digital proxies and ground truth.