Spatial Analysis of the 2023 Earthquakes in Turkiye- An Executive Summary

Autor : Gozde Yazganoglu Delgado Doblas

Tutor: Diego José Bodas Sagi

September, 2023

CUNEF
UNIVERSIDAD

Thesis presented for the Degree of
Master of Data Science

1. Introduction

In February 2023, in Kahramanmaras, Turkiye experienced a catastrophic earthquake, a somber reminder of nature's volatile power. The repercussions of this seismic event reached as far as Syria, affecting 17 cities and almost 19 million individuals. The human cost was harrowing with around 57,000 casualties and approximately 130,000 seriously injured. Financially, the damages mounted to $34.2 billion, leading to a 4% dip in Turkey's economy. Many buildings got damaged, destroyed resulting with death and mobility of the people in the middle of hard winter.

As earthquake is an unknown, unpredicted phenomena, we are not able to avoid it to happen. One of the very limited we know is that, they are likely to happen on sensitive zones that are likely to move over broken faluts. However, we should know about the risks and in order to determine the risks, we should be able to get sufficient learnings from previous disastrous events. If buildings are sensitive, around the sensitive zones, or are there economic factors that leads more damage…these are some of the risk factors that should be considered for the existing buildings and the buildings that have demolished or severely damaged.

1.1. Objectives of the study

This study focuses on using the data collected from Kaggle (Dincer, 2023) that is sourced from EU Copernicus disaster information (European Commission - Copernicus Emergency Management Service, 2023) in a machine learning model to predict patterns of the damaged buildings just after the disaster. Of course, we do not hold more information about the land properties or building materials. However, sometimes, most of the time buildings collapse due to bad business decisions. In this study we aim to find answers to following questions.

- If there is a relationship with the geophysical fault information?
- How are clustered buildings and different level of damages?
- By using a machine learning algorithm can we find what makes an area more vulnerable?
- Which regions are affected worse compared to other regions?

For the objectives have been set, there are serious sub-objectives that could be established to achieve main goals. As all data science projects needs to start with, a proper data collection should be made. Here the data set is consisting of several dataset and in orders to make a proper analysis we need external information to have.

2. Studies of Machine Learning on Earthquake

Research in machine learning has demonstrated its potential in earthquake analysis. Notably, a 2020 study by the Earthquake Engineering Research Institute (EERI) utilized artificial neural network (ANN) modeling for hazard detection. Priambodo et al. (2020) executed a temporal-spatial analysis on Indonesia's seismic data using ANN, emphasizing the need for more geo-specific data for accurate predictions. Garg, Masih, and Sharma (2021) focused on bridge damage assessments post-earthquakes, underscoring the significance of distance from the earthquake center. Mangalathu et al. (2020) employed machine learning to assess building damage, achieving a 65% accuracy with their model, emphasizing building age and fault distance as key variables. Further studies have expanded the realm of earthquake data analysis, leveraging it for purposes such as insurance risk assessment.

3. Methodology

In this project the usual methodology of Data Science

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

3.1. Data Collection and Preparation

This study uses a compound dataset of several datasets mentioned below.

- EU Copernicus Disaster Data: That has buildings, roads, facilities as target and elevation, water sources, anciliary information as auxiliary information
- AFAD Earthquake Data information: Earthquakes of 6th of February is filetered to include in main dataset.
- TURKSTAT: Regional Socioeconomic information is used.
- ATAG Fault Data: Fault information is used.

Specifically, we have used distances to important points as other variable and as much as socioeconomic knowledge added to the data set on locality basis.

3.2. Spatial Analysis Methods

Spatial analysis methods are heavily used in this study to understand spatial relationships to see damage level. Some of the methods should be marked below.

- Point Pattern Analysis
- Ball Tree Algorithm
- Spatial Lags, Weights and Autocorrelation

These techniques are used to find patterns, calculate distances, find spatial variables and measure the effects of location in general in machine learning models.

3.3. Machine Learning Algorithms

In this study supervised machine learning models were used to understand which variables have more affects on damage level of the building dataset and unsupervised model is used to cluster regions according to spatial and socioeconomic variables. Among the models tried, as the best supervised model is Random Forest and as a clustering model K-means is the best model.

4. Analysis

As mentioned in the methodology a explanatory per-analysis made all data set and findings were combined with spatial analysis results. Later, multiclass clasification experiment and clustering experiment was conducted.

4.1. EDA (Explanatory Data Analysis)

Analysis suggests that Kirikhan and Kahramanmaras have received most number of damaged and destroyed buildings. Damages buildings are observed mostly in the residential buildings. However in different localities such GAZIANTEP, MALATYA, KAHRAMANMARAS have observed also sever damages on roads and airfields.

(damage_gra 4: destroyed, damage_gra = 3: damaged, damage_gra =2, possibly damaged, damage_gra = 1: no visual damage, damage_gra = 0: no damage information)

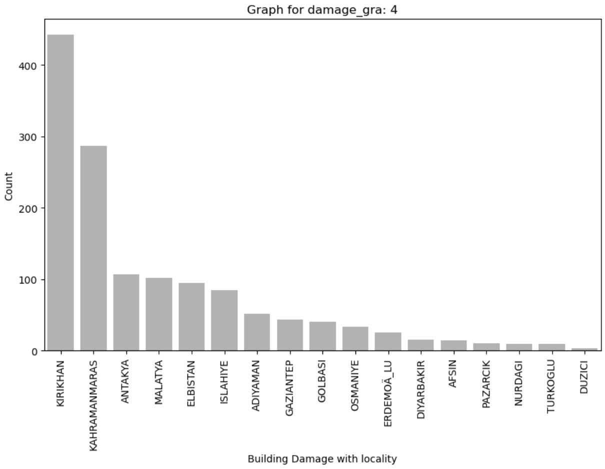

Figure 2 Destroyed Buildings in Cities

Source: Yazganoglu, G (2023)

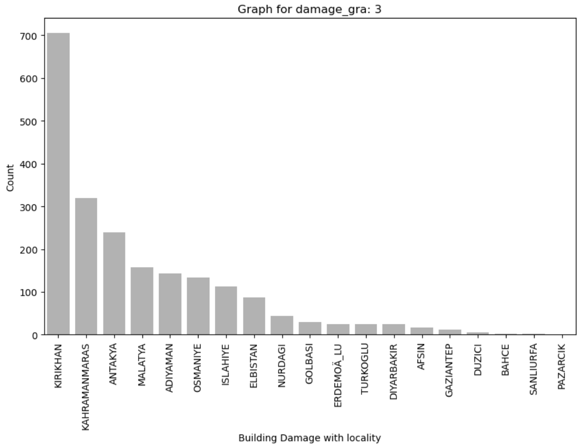

Figure 1 Damaged Buildings in Cities

Source: Yazganoglu, G (2023)

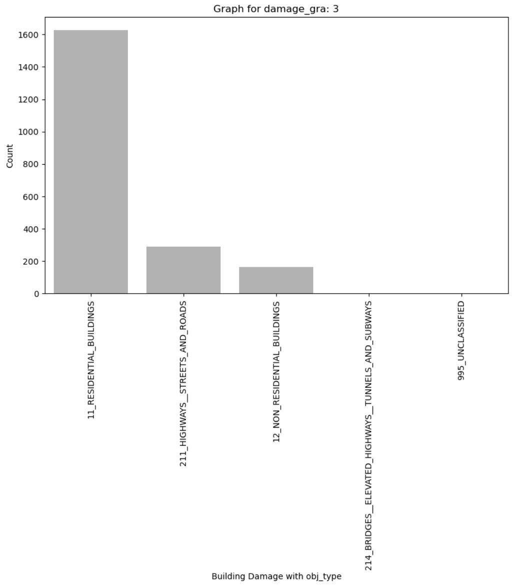

Figure 3 Damaged buildings according to object type

Source: Yazganoglu G (2023)

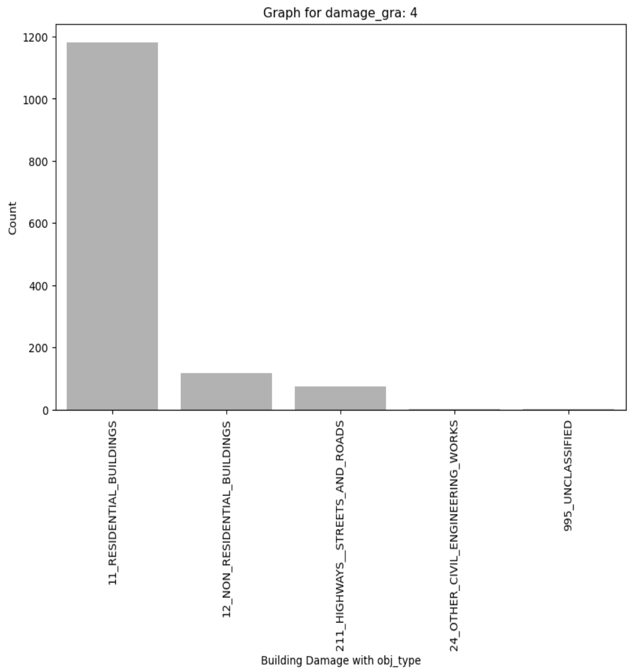

Figure 4 Destroyed Buildings according to obj_type

Source: Yazganoglu, G (2023)

4.1.2. ESDA (Explanatory Spatial Data Analysis)

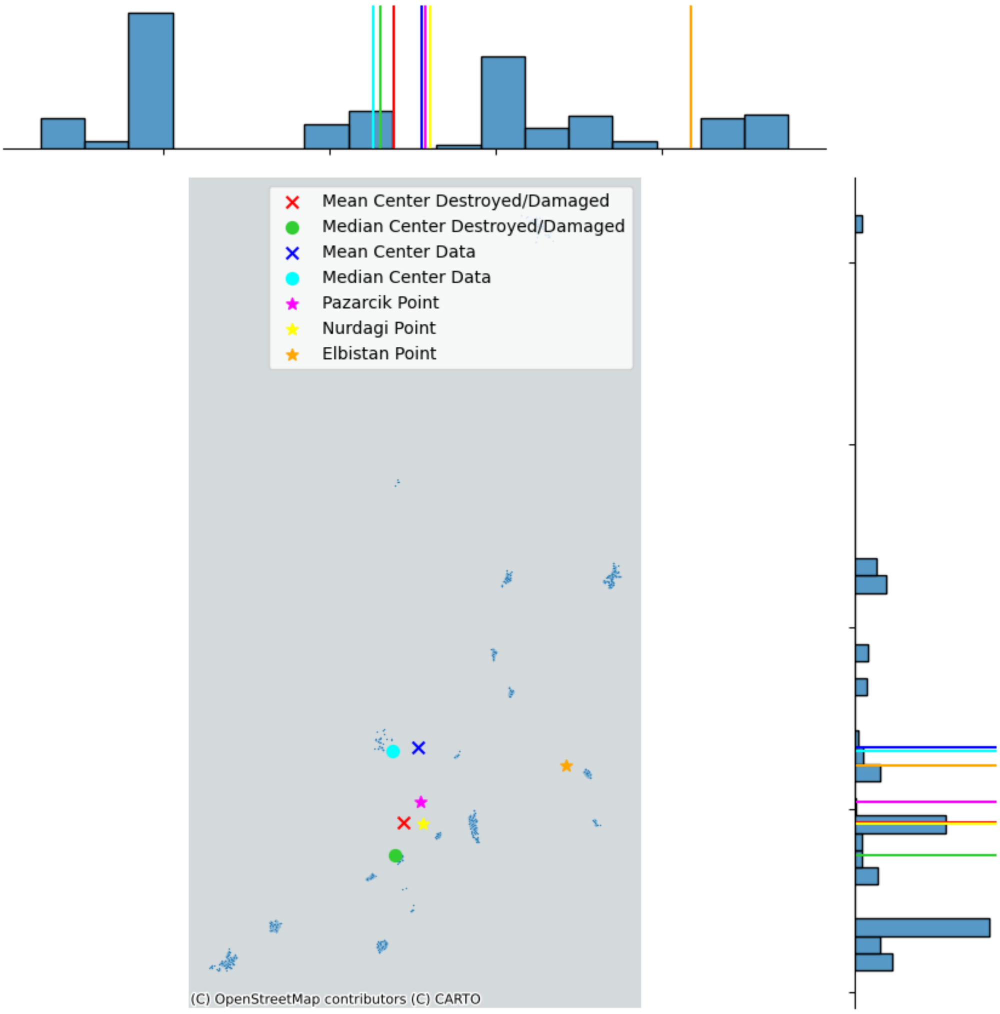

Figure 5 Point Pattern Analysis

Source: Yazganoglu, G (2023)

Point pattern analysis showcased that different localities are clustered in different regions and this is not random at all. When we filter only damaged and destroyed data we end up with same result, however, their mean/median locations are in different places. Mean/median points are centered more close to the Pazarcik and Nurdagi earthquakes. Elbistan Earthquake point doesn't have the same impact and not surrounded by density of the damaged/destroyed observations.

Column	**Moran's I Value**	**Interpretation**
percentage	0.658767	Moderate positive spatial autocorrelation. Similar values tend to cluster together.
damage_gra	0.532654	Positive spatial autocorrelation but not as strong as others

Table 1Moran's values for spatial variables and interpretation

Source: Yazganoglu, G (2023)

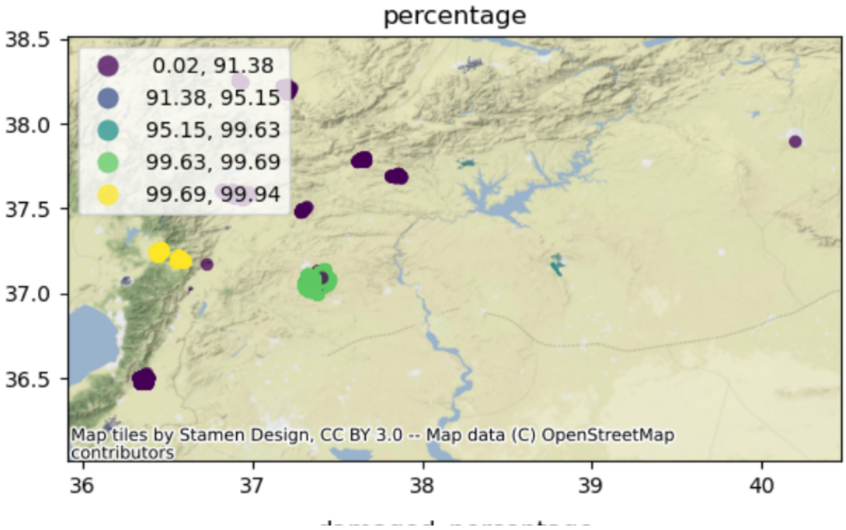

Figure 6: Spatial Mapping for the percentage variable

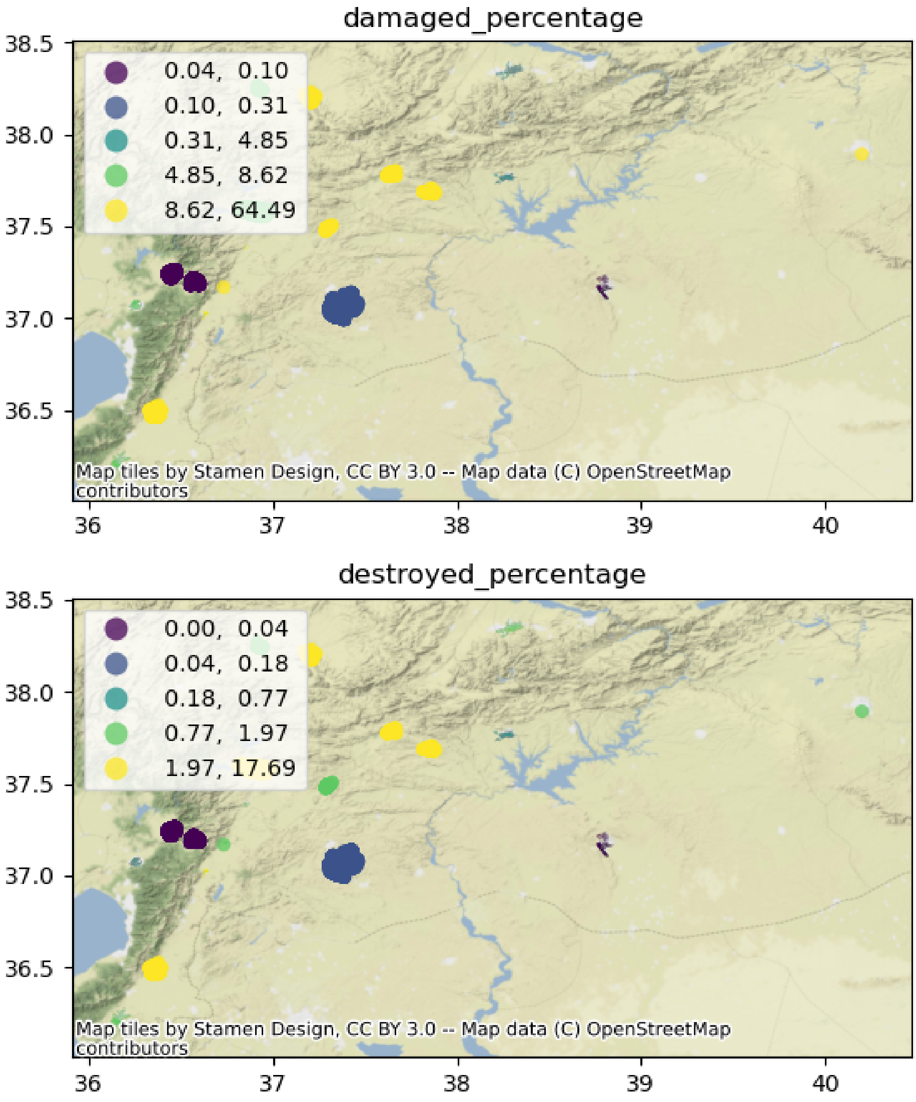

Figure 7 Spatial Mapping of Damaged and Destroyed Percentage in the Locality

When we look at spatial autocorrelations to the variables created as percentage, and target value damage level, we see that damage level and over all percentage of the damage level we observed moderate spatial relations which shows that similar damage levels are tend to be observed in similar locations.

4.2. Machine Learning Models

Multiclass Classification experiment showcased that the best model is Random Forest that determine damage level. According to the best model, spatial lags, geographic variables and building type is found as important variables

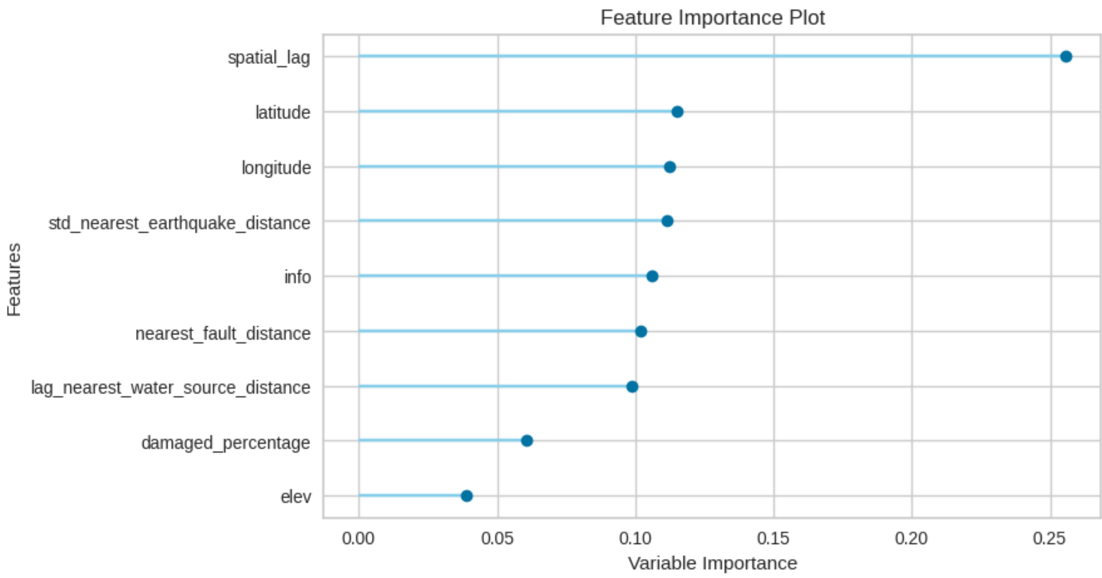

Figure 8: Feature Importance

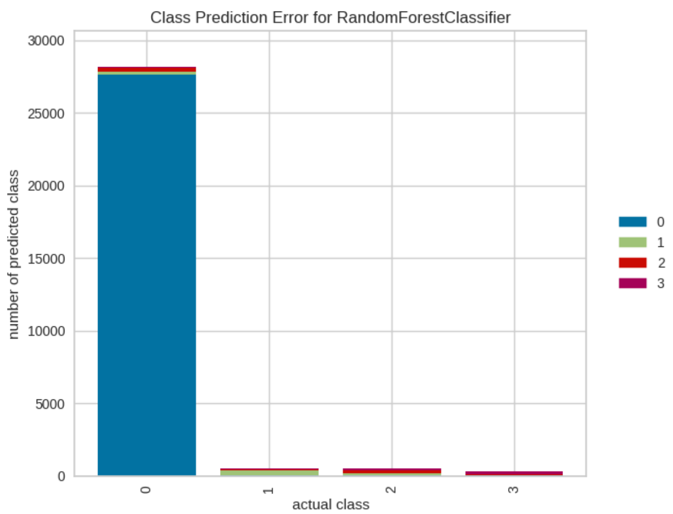

Figure 9: Class Prediction Error

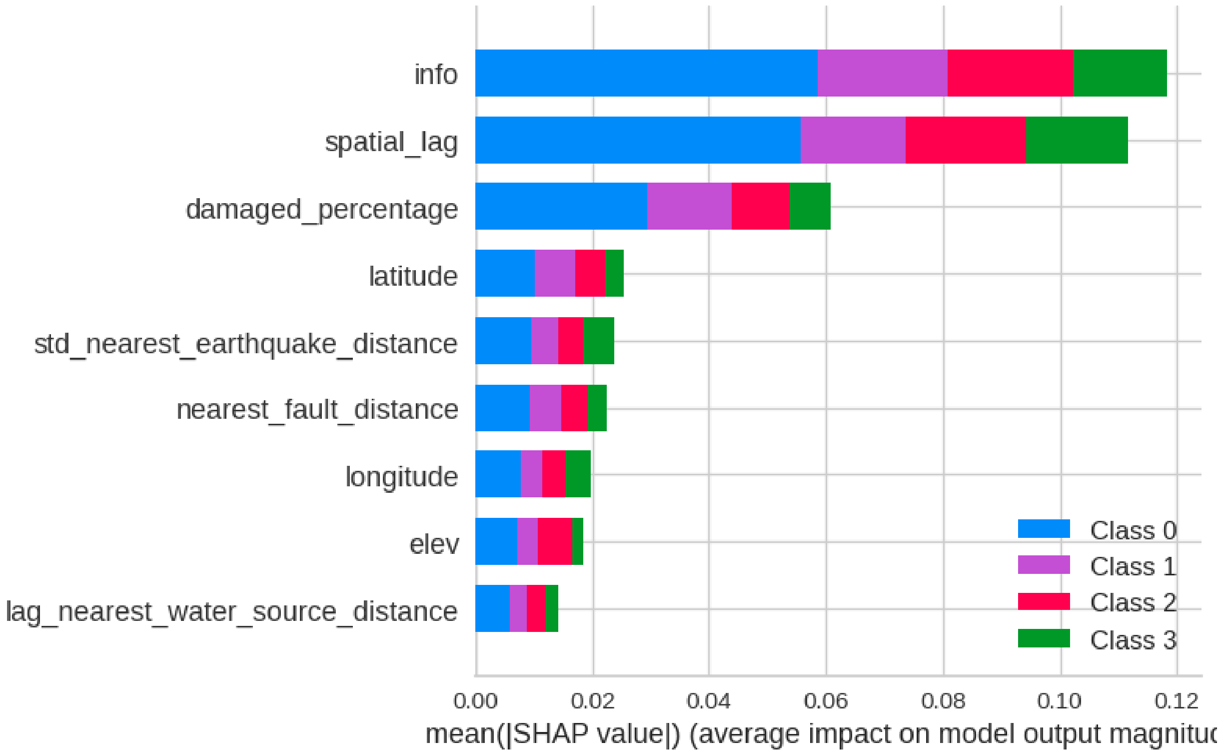

Figure 10: Mean SHAP Value Impact

Interpreting SHAP values in a multiclass classification model adds an extra layer of complexity compared to binary classification or regression models, as we need to consider the contribution of features towards each class prediction.

As figure reflects, building 'info' and 'spatial_lag' s are the important values to determine building damage level in overall. We also observe fault distance and earthquake distance are also following important values. These 2 variables also is the most used ones for all classes.

4.2.2. Unsupervised Machine Learning

The optimum number of clusters have been found as 4. Below is the summary of the clusters and their features.

Cluster 0 is primarily located in 'ANTAKYA' and 'KIRIKHAN' and contains a moderate number of observations. It experienced moderate damages, particularly affecting residential buildings. Despite having a low population, it surprisingly boasts the third-highest income levels but reports the lowest figures in total and secondhand sales. Damages are also observed in railways, electricity lines, and bridges, requiring extra attention in rebuilding.

Cluster 1 is concentrated in 'GAZIANTEP' with a significant presence in 'KAHRAMANMARAS' and 'MALATYA.' It has the fourth largest number of observations and faced severe damages, mainly impacting residential buildings. Despite low population density, it ranks second in income levels and third in both total and secondhand sales metrics. No damages were observed in a weighted damage model.

Cluster 2 is highly localized in 'SANLIURFA' and has the third-highest number of observations. Interestingly, it reported no complete destructions and had minimal damage. Residential buildings, highways, streets, and roads were the most affected. It is characterized by a densely populated locality, the lowest income levels, but paradoxically, it has the highest numbers in total and secondhand sales.

Cluster 3 spreads across multiple localities, with a significant presence in 'ADIYAMAN,' 'DUZICI,' and 'BAHCE.' It has the highest number of observations and experienced moderate damages, particularly affecting highways, streets, and roads. This locality is densely populated, boasts the highest income levels, and ranks second in both total and secondhand sales. The data suggests potential for post-reconstruction development in this cluster, particularly in the market sector.

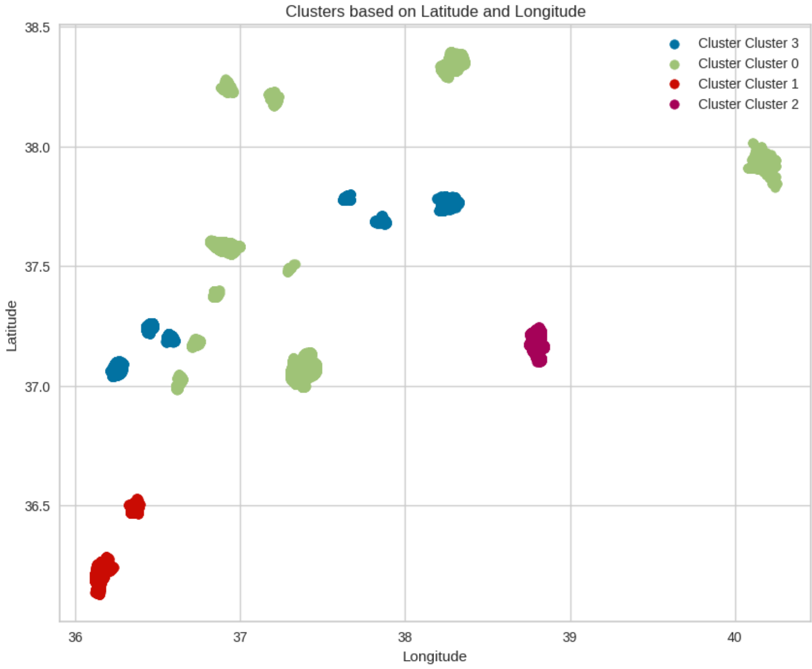

Figure 11 Clusters Reflected on the Map

Source: Yazganoglu, G (2023)

Conclusions:

Objective & Data Sources:

This study embarked upon understanding damages and losses from a major disaster event in 2023 using satellite datasets from the EU Copernicus database. These datasets detailed damages following earthquakes on 6th February 2023 across 19 locations, including buildings, roads, and facilities. Additional data layers like distances to water sources, faults, earthquake locations, and emergency campsites were incorporated, supplemented with socioeconomic data from TUIK.

Analysis:

Point-pattern and spatial autocorrelation analyses revealed geographic factors influencing damage levels. Pazarcik and Nurdagi earthquakes were found closer to the most affected sites than the Elbistan earthquake. Damage patterns weren't random, suggesting spatial clustering of damaged and non-damaged buildings. A Random Forest classifier was the best-fit model for multiclass supervised classification, emphasizing building type and spatial variables as significant predictors. K-Means clustering was preferable to DBSCAN for mapping problematic areas, indicating more influence from socioeconomic variables.

Findings:

Distances to fault lines and earthquake epicenters play a pivotal role in predicting damage. There's a tendency for damaged or destroyed buildings to cluster near disaster-prone regions. Machine learning can identify which regions are most vulnerable.

Limitations & Future Directions

The research lacked specific local information, making it challenging to derive detailed insights.

- Earthquake impacts were based largely on broad locality data due to the unavailability of detailed information.
- A more localized methodology, using tools like geocoding and specific spatial weights, could improve insights.
- Richer datasets can yield a more in-depth understanding of earthquake dynamics. The study highlights the importance of interdisciplinary collaboration, integrating expertise from seismologists, economists, architects, civil engineers, and data scientists.
- **Implications**:
- Urban planning and construction techniques need careful consideration to mitigate earthquake damage.
- Economic protections, like comprehensive earthquake insurance, should be explored.
- The housing market's close ties with the loan market highlight the importance of collateral evaluation. With homes as primary collateral for many loans, financial institutions and insurance companies should consider auditing buildings for earthquake safety.

In essence, this research underscores the importance of comprehensive earthquake preparedness, recovery strategies, and interdisciplinary collaboration in understanding and mitigating the impacts of such natural disasters.