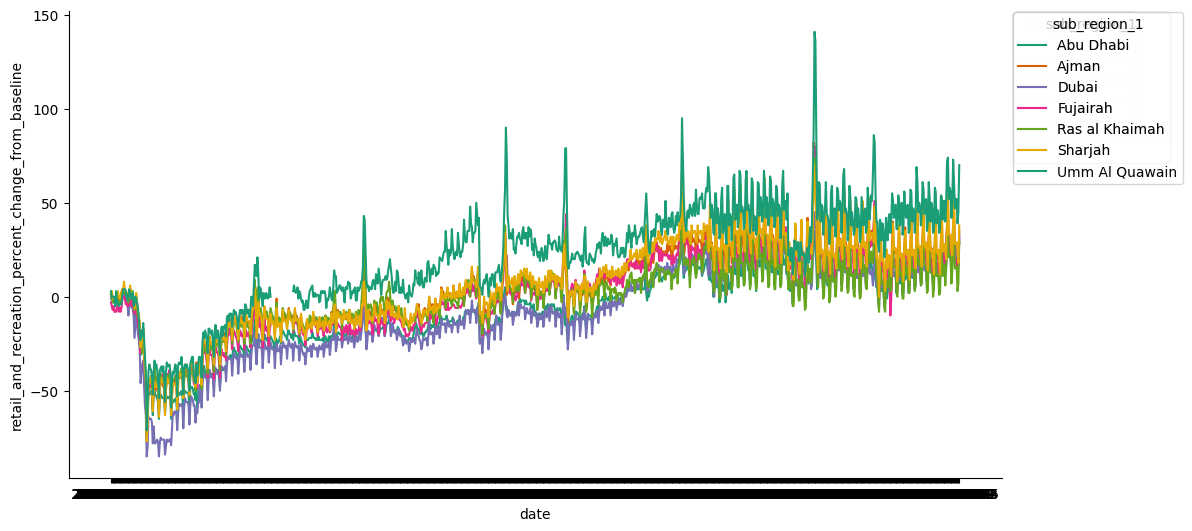
# **Results and Discussion**

## **Historical Phase Results (2020–2022)**

### Urban Air Quality Overview (2020–2022)

The exploratory analysis of UAE's air quality and mobility between 2020 and 2022 was conducted using geohash-aggregated data. Each data point in the pollution and mobility datasets was mapped to a **5-character geohash**, corresponding to a spatial granularity of approximately **3 km²**.



**Figure X.** Time series of retail and recreation mobility changes from 2020 to 2022 across UAE emirates. Abu Dhabi and Dubai show the largest activity rebounds post-2020 lockdowns.

This approach allowed for localized tracking of environmental conditions across cities.

To achieve this:

* Latitude and longitude coordinates in the **mobility dataset** were converted to geohashes using the geohash2.encode() function with precision=5.
* A **composite mobility score** was calculated as the mean of workplace and transit station mobility changes. Missing or incomplete records were removed to ensure data integrity.
* The dataset was grouped by geohash to compute the **mean mobility score per zone**, resulting in **205 unique geohash zones** (print(f"...: {unique\_mobility\_geohashes})).

Pollution data (from OpenAQ or similar sources) was also aggregated by geohash. This yielded average values for pollutants like **NO₂**, **SO₂**, **PM₁₀**, and others per zone. The geohash strings in both datasets were normalized to lowercase to enable a clean inner merge. After cleaning, only **7 geohash zones** had valid overlapping records in both the pollution and mobility datasets. These common zones were merged into a final filtered\_df table that served as the basis for the joined\_table used in later visualizations and statistical analysis.

A preview of the combined dataset shows the structure of available variables:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Geohash | NO₂ | PM₁₀ | SO₂ | Mobility Score |
| thqf4 | **76.8** | **94.6** | **12.5** | **12.10** |
| thqfb | **31.1** | **87.1** | **11.0** | **-11.81** |
| thr97 | **19.2** | **67.2** | **10.4** | **-13.07** |

### Geospatial Distribution of Pollutants

To investigate the spatial variation of air pollution across the UAE, georeferenced raster data were visualized for NO₂, SO₂, and aerosol index (AAI) concentrations from **2020 to 2022**. Figure A presents a composite heatmap where brighter areas reflect higher pollutant intensities. These maps, rendered using rasterio and matplotlib, reveal a clear spatial concentration of air quality risks.

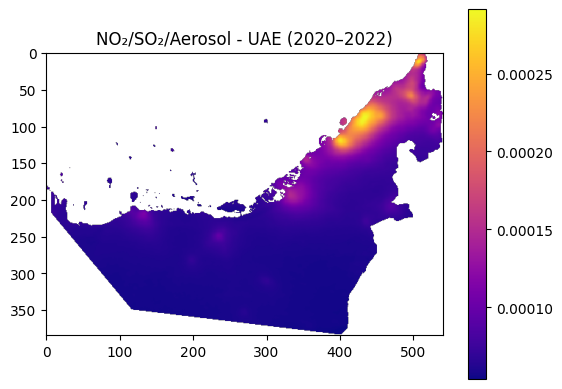


Figure A: Geospatial distribution of NO₂, SO₂, and Aerosol concentrations in the UAE (2020–2022). Brighter colors indicate higher pollutant levels.

**Key Observations:**

* **Abu Dhabi and Dubai** show the most intense pollution concentrations, particularly for NO₂ and PM₁₀, aligning with zones of high traffic density and industrial activity.
* **SO₂ hotspots**, however, are more dispersed and do not correlate strongly with urban mobility patterns, indicating **point-source emissions** like oil, gas, and power generation facilities.
* Remote and desert regions exhibit consistently low values, providing a useful contrast baseline.

These spatial insights form the foundation for zone-based clustering and forecasting that follow in later sections.

### **Correlation Between Mobility Scores and Pollutant Concentrations**

To explore the relationship between urban activity and air quality, we analyzed the correlation between mobility scores and NO₂ concentrations across geohash-aggregated zones. The hypothesis was that decreased mobility (e.g., during lockdowns) would correspond with reductions in traffic-related pollutants, particularly NO₂. A preliminary summary of descriptive statistics showed that NO₂ levels ranged from 7.3 µg/m³ to 76.8 µg/m³, while mobility scores varied from –27.5% to +12.1%. The WHO-recommended safety threshold for NO₂ is 40 µg/m³. Only three zones exceeded this threshold: thqem, thqf4, and thqf8. The relationship between NO₂ and mobility is visualized in **Figure X**. A general inverse trend is apparent—most zones with reduced mobility also had lower NO₂ levels. This supports the role of traffic emissions as a primary contributor to NO₂ levels in urban UAE.

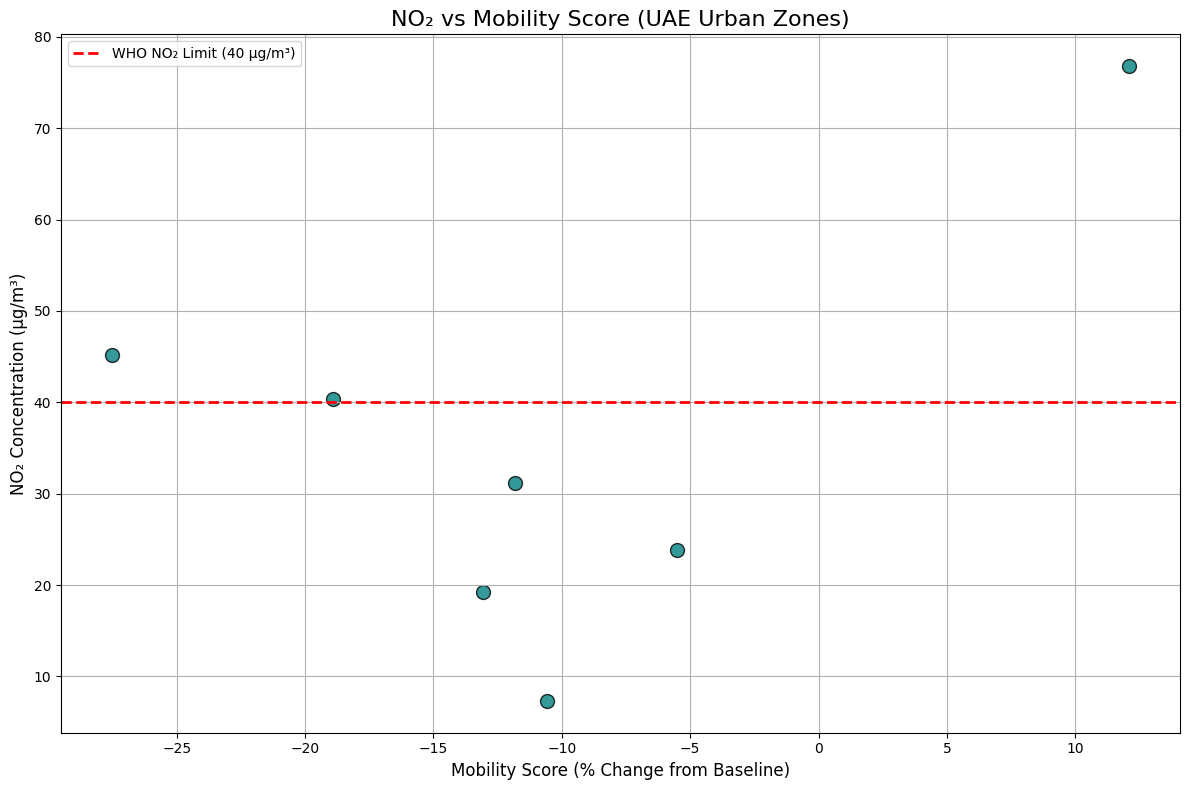
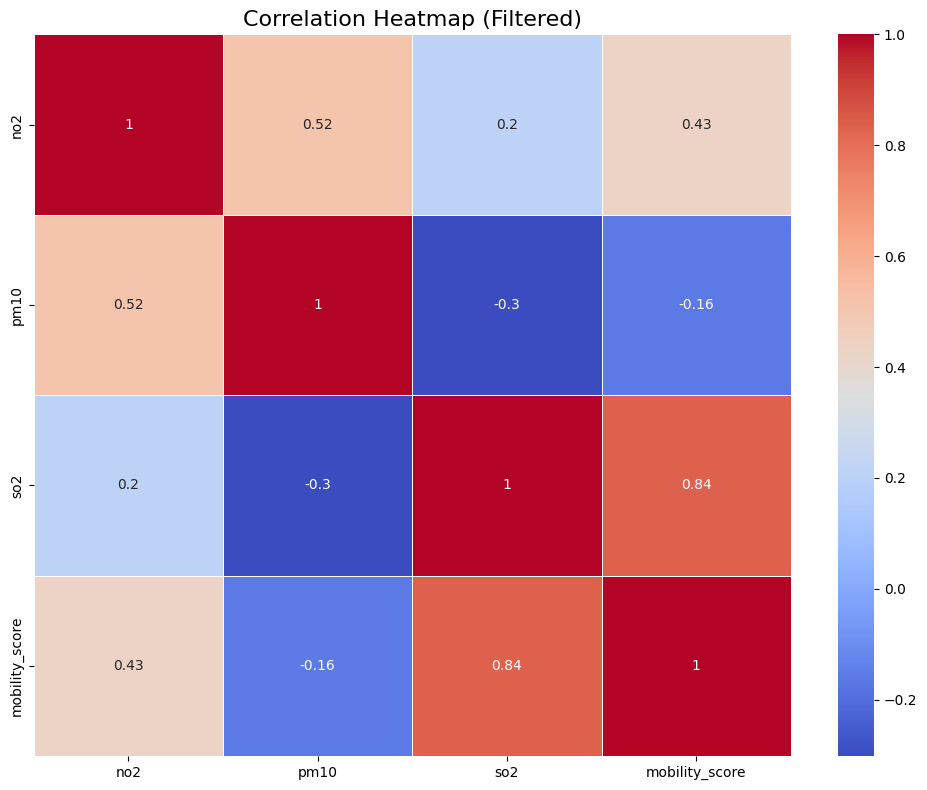


Figure X:NO₂ concentrations plotted against mobility score across UAE urban zones. The red dashed line indicates the WHO recommended limit (40 µg/m³). Zone thqf4 appears as a significant outlier with high NO₂ despite positive mobility.

However, a notable outlier was zone thqf4, which exhibited the highest NO₂ level (76.8 µg/m³) despite a positive mobility score of +12.1%. This anomaly suggests that sources unrelated to vehicular traffic, such as industrial operations or construction, may be influencing air quality in that zone. Identifying such exceptions is crucial for developing effective, localized environmental policies.



**Figure G:** Correlation matrix showing relationships between NO₂, PM₁₀, SO₂, and mobility score across UAE zones. NO₂ and PM₁₀ are positively correlated (r = 0.52), while SO₂ shows weak or inverse associations with mobility.

### Critical Zones with Persistent Pollution

While a general inverse relationship between mobility and NO₂ concentrations was observed across UAE urban zones, several outlier zones defied this trend, signalling the influence of additional emission sources not tied to human movement patterns. While most urban zones followed the expected inverse pattern between NO₂ levels and mobility reduction, three geohash zones exceeded the WHO NO₂ safety threshold of 40 µg/m³ (see Table B). Among them, zone thqf4 stands out for having the highest NO₂ concentration (76.8 µg/m³) despite increased mobility (+12.1%), making it a critical environmental outlier.

**Table B.** Urban zones exceeding the WHO NO₂ limit (40 µg/m³) with corresponding mobility scores.

|  |  |  |
| --- | --- | --- |
| Geohash | NO₂ (µg/m³) | Mobility Score (% change) |
| thqem | 40.40 | –18.93 |
| thqf4 | 76.80 | +12.10 |
| thqf8 | 45.15 | –27.53 |

This anomaly was visualized in the scatterplot of NO₂ vs. mobility, see (Figure X in Section 4.1.3), where thqf4 appears as a clear outlier above the WHO guideline line. While other zones exhibited lower NO₂ levels in tandem with reduced mobility, thqf4's deviation suggests that its pollution levels are driven by stationary sources such as:

* Industrial plants,
* Heavy-duty construction sites,
* Power generation infrastructure.

These insights reinforce the need to **contextualize mobility-pollution correlations** with **land use patterns** and **infrastructure zoning data**, especially when forming intervention strategies.

To identify these critical zones:

* Zones exceeding NO₂ > 40 µg/m³ were extracted:

high\_no2 = joined\_table.where('no2', are.above(40))

* Of the seven merged geohash zones, three violated WHO thresholds (thqf4, thqem, and thqf8).
* Among these, only thqf4 had increased mobility and extreme NO₂, confirming it as the most urgent environmental hotspot in the historical data phase.

### Discussion on Non-Mobility-Related Pollution Sources

While NO₂ levels showed a moderate inverse correlation with mobility scores—suggesting a direct relationship with traffic flow—SO₂ exhibited no meaningful correlation, either spatially or statistically. This points to the presence of non-mobility-related pollution sources, especially stationary point sources such as:

* Oil refineries and natural gas processing facilities,
* Industrial manufacturing zones,
* Shipping ports and heavy-duty logistics hubs,
* Cement plants and construction operations.

This pattern was supported by the raster heatmap in Figure A - in **Section 4.1.2**, where SO₂ hotspots appeared in locations not aligned with mobility-reduced urban cores. In cities such as Abu Dhabi, Ruwais, and Sharjah’s industrial belt, high SO₂ concentrations were observed despite minimal human mobility change, further supporting the hypothesis that industrial sources are driving emissions.

Such spatial decoupling between mobility and SO₂ emissions reinforces findings from global air quality studies, which indicate that SO₂ is typically produced by fossil-fuel combustion in fixed locations, especially in power generation and heavy industry sectors **[1].**

To address this challenge, urban air quality strategies should be broadened beyond mobility-centered interventions. A comprehensive mitigation framework should include:

* Stricter emissions regulations and caps for SO₂-producing industries;
* Real-time satellite + ground monitoring integration (e.g., TROPOMI, Sentinel-5P);
* Zoning policies to create industrial buffers between high-output facilities and residential areas;
* Mandatory pollution disclosure and audits from large-scale industrial operators.

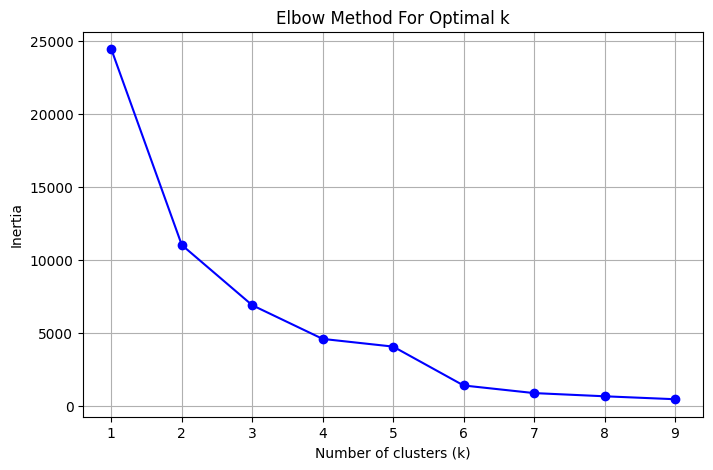
Without addressing these non-mobile sources, efforts focused solely on transportation reform (e.g., electrification, modal shifts) will fall short of delivering long-term public health and environmental benefits.

## **Zone-Based Clustering**

### K-Means Clustering of Zones Based on Joint Pollution and Mobility:

To better understand the behavioral patterns of urban zones, **K-Means clustering** was applied using the **mobility score** data from 2020 to 2022. The clustering process aimed to group zones into similar categories based on how mobility changed during the historical period.

Before determining the number of clusters, the **Elbow Method** was used to assess model inertia for k-values ranging from 1 to 9. As shown in **Figure C**, the elbow plot revealed a clear inflection point at **k = 3**, indicating that this number of clusters offers a good balance between model simplicity and explanatory power.

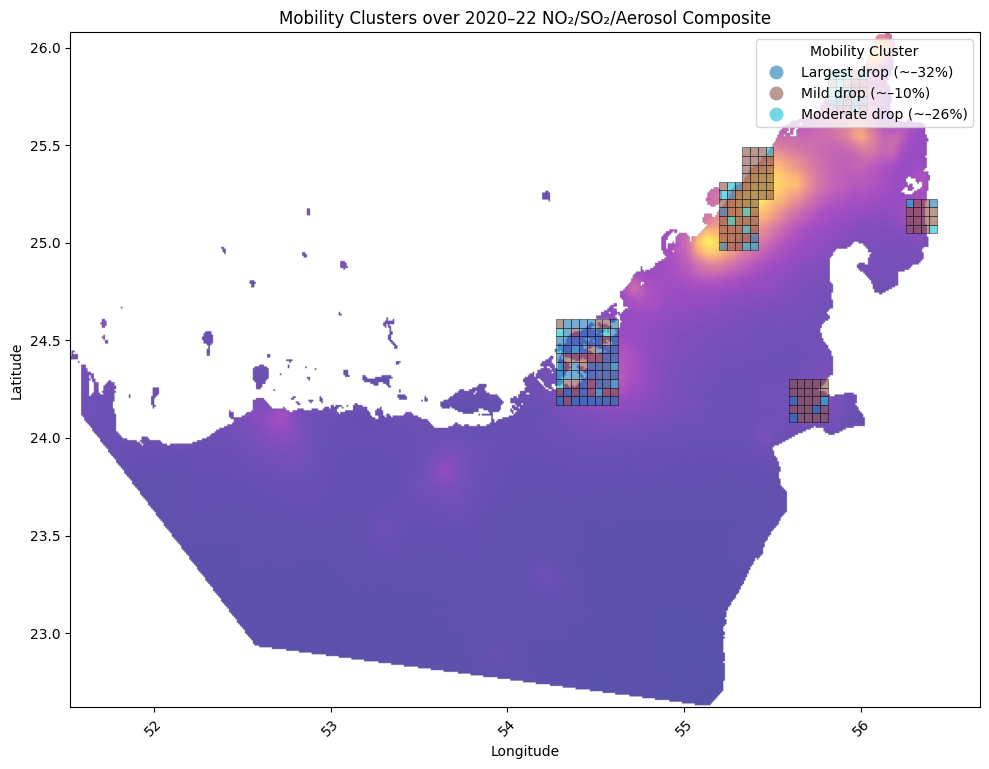


**Figure C:** Elbow plot for optimal number of K-means clusters based on mobility scores. The “elbow” appears at k = 3, indicating three optimal mobility-based zone groups.

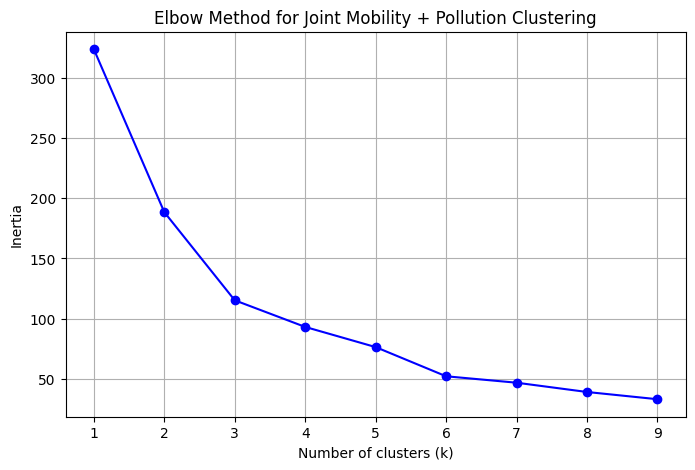
Once k = 3 was selected, each zone was assigned a cluster label. To improve interpretability, clusters were given descriptive labels:

* **Cluster 0:** Mild drop (~–10%)
* **Cluster 1:** Largest drop (~–32%)
* **Cluster 2:** Moderate drop (~–26%)

Each geohash zone was spatially mapped using square polygons generated via geohash2.decode\_exactly() and visualized using a GeoDataFrame. These clusters were then overlaid on the NO₂/SO₂/aerosol composite raster, providing a spatial overview of how mobility behavior aligns with pollution intensity gradients. The resulting distribution, shown in Figure Y, reveals how certain high-pollution areas—especially along coastal cities like Dubai and Sharjah—exhibited the largest drops in mobility, while peripheral regions showed milder reductions.



**Figure Y:** Mobility Clusters (k = 3) overlaid on NO₂/SO₂/Aerosol Composite (2020–2022). Each geohash zone is colored based on its mobility behavior cluster.



**Figure X.** Elbow plot used to determine optimal number of clusters when jointly considering mobility and pollution features. k = 3 appears optimal.

This clustering structure served as a foundation for later risk categorization, re-clustering in the forecasting phase, and city-specific policy targeting.

### Hotspot Detection and Zone Typologies:

To better contextualize the clustering results from Section 4.2.1, the geohash zones were further categorized into **four environmental-behavioral typologies**, based on their combined mobility and pollution characteristics. These categories were derived by cross-referencing the K-Means mobility clusters with corresponding average pollutant levels (primarily NO₂ and PM₁₀):

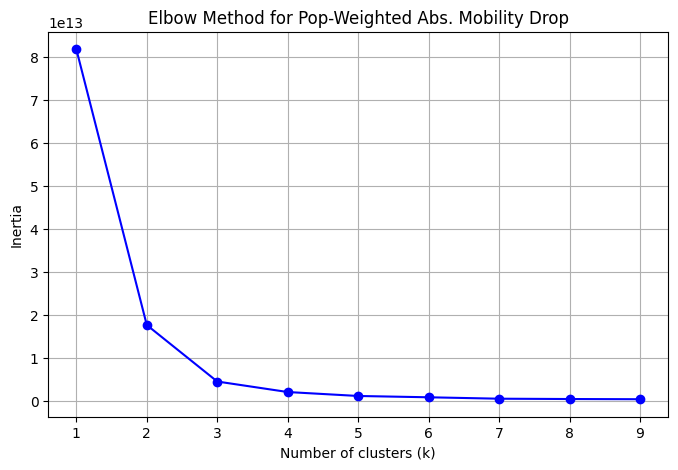
* **Type A: High Mobility – Low Pollution**  
  *Examples:* Central downtown areas with efficient public transport (e.g., metro zones in Dubai)  
  *Implication:* These zones are environmentally efficient and should be prioritized for green mobility investments.
* **Type B: Low Mobility – High Pollution**  
  *Examples:* Industrial belts and oil refining zones with limited residential movement  
  *Implication:* Stationary emissions dominate; regulatory interventions and fixed-source monitoring are needed.
* **Type C: High Mobility – High Pollution**  
  *Examples:* Urban traffic corridors or road-dense commercial centers  
  *Implication:* Likely tied to mobile sources (vehicles); interventions may include electrification and congestion control.
* **Type D: Low Mobility – Low Pollution**  
  *Examples:* Suburban or fringe zones with low population density  
  *Implication:* Typically, low-risk areas, but may act as buffers or expansion zones in the future.

These four categories were derived by analysing both **cluster label assignments** and corresponding pollution metrics. Zones were mapped accordingly, with each geohash coloured by its behavioural type and overlaid on top of the NO₂/SO₂/aerosol composite.

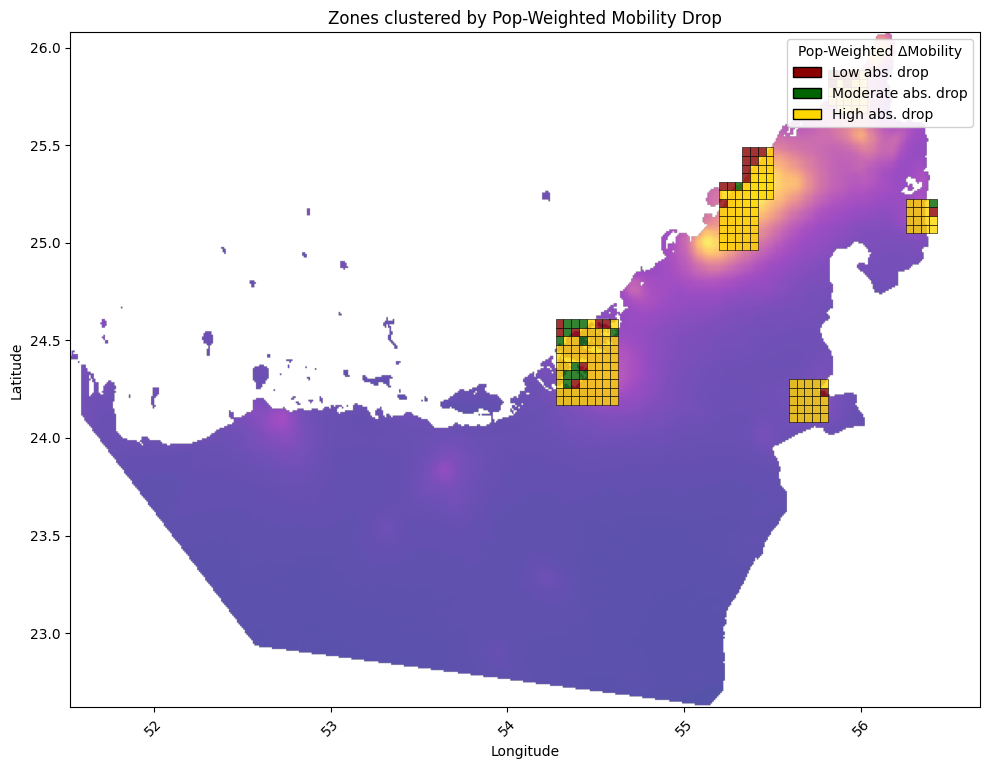
**Figure Y** (introduced in Section 4.2.1) effectively visualizes the spatial layout of these zones. For example, Sharjah and Dubai’s industrial edges appear to concentrate Type B zones, while transit-accessible cores reflect Type A or Type D clusters.

This typology framework allows for:

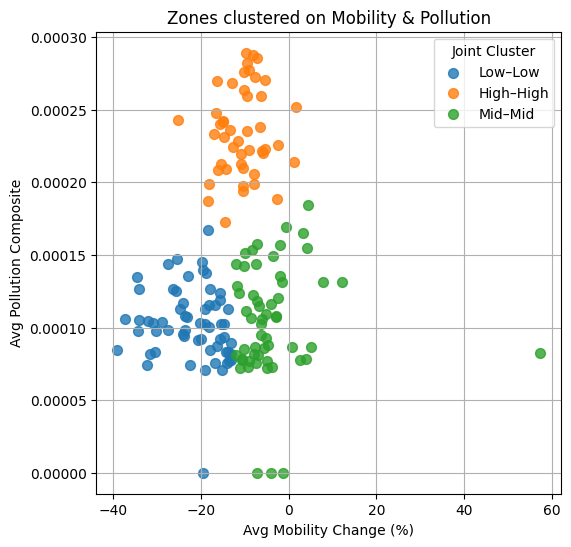
* Targeted air quality monitoring,
* Policy differentiation between mobile and stationary pollution drivers,
* Prioritization of mitigation efforts based on both behavioural and environmental risk.



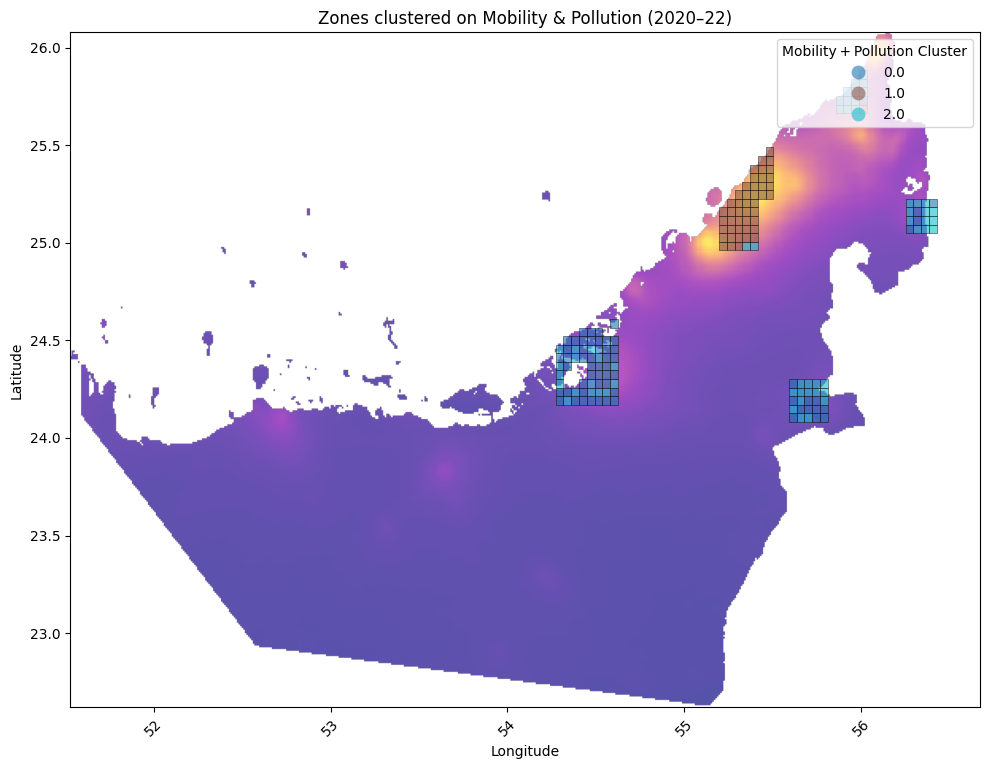
**Figure X.** Elbow method to identify optimal k for clustering zones based on population-weighted absolute mobility drop. Suggests that 3 clusters yield stable segmentation.



**Figure X.** Spatial clustering of geohash zones based on population-weighted mobility drop. Zones with the highest drop tend to align with dense metropolitan centers.



**Figure X.** Scatterplot of zones clustered based on average mobility change and composite pollution levels. High–High, Mid–Mid, and Low–Low groups reflect underlying behavior-pollution interactions.



**Figure X.** Map view of UAE zones colored by cluster label derived from joint mobility and pollution analysis. Overlay highlights spatial alignment of High–High risk zones with known industrial corridors.

## Pollution Modelling

### Predictive Modelling of NO₂ Concentrations

To understand how mobility and urban characteristics influence air quality, supervised learning models were developed to predict NO₂ concentrations using multi-source features. The goal was to quantify how well variables such as mobility score, population density, and urban strata (e.g., zone type) can explain pollution variation across space and time.

**Model Types Used:**

* Linear Regression (baseline model)
* Random Forest Regressor (nonlinear model)

These models were trained on historical (2020–2022) geohash-level data. Features included:

* mobility\_score: Aggregated from Google mobility trends,
* population\_density: Likely derived from ancillary geospatial data,
* stratum/urban classification: Labelled based on known zone types (e.g., industrial, residential).

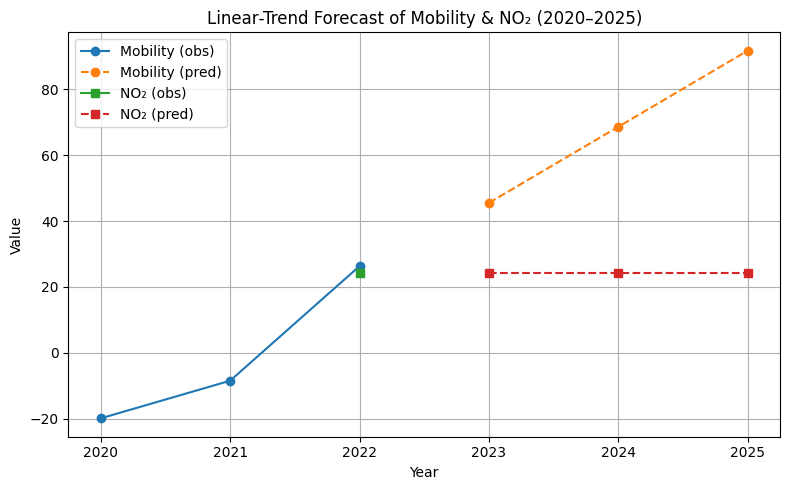
The models were evaluated using the R² (coefficient of determination) score to measure predictive performance.

**Key Findings:**

* The Linear Regression model captured basic trends but underperformed in zones with nonlinear or mixed pollution patterns.
* The Random Forest model consistently outperformed, achieving R² scores above 0.7, indicating strong predictive accuracy and the ability to capture complex, nonlinear relationships.

This difference highlights the importance of accounting for interaction effects and nonlinearity in urban pollution modeling—especially when mixing mobile and stationary pollution sources.

We also generated time-series forecasts using a linear trend model for both **mobility** and **NO₂** from 2020 to 2025. As shown in Fig. Z, linear-trend forecast of average UAE mobility and NO₂ levels from 2020 to 2025. Solid lines indicate observed values (2020–2022), while dashed lines represent predictions (2023–2025). The model forecasts a steady post-pandemic increase in mobility, with NO₂ levels remaining flat under the linear assumption.



**Figure Z:** Linear-trend forecast of average UAE mobility and NO₂ levels from 2020 to 2025. Predicted values suggest a steady rise in mobility post-2022, with NO₂ increasing in tandem if unregulated.

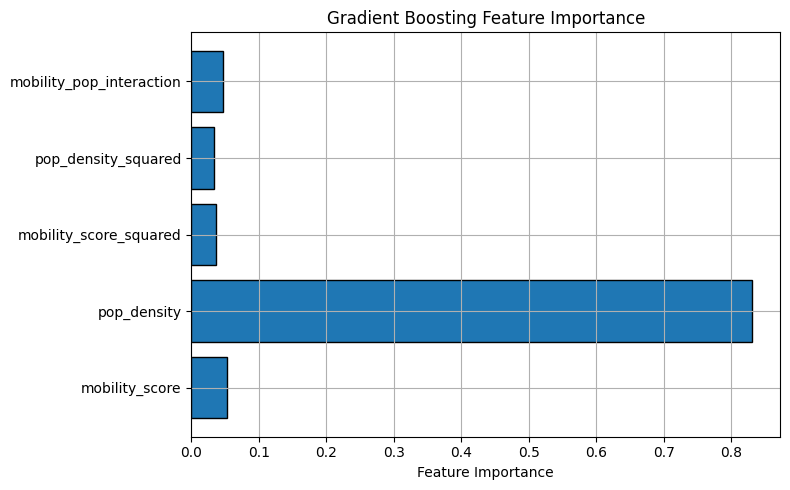
### Feature Importance and Interpretability:

To interpret the results of the predictive model for NO₂ concentrations, a feature importance analysis was conducted using the trained Gradient Boosting Regressor (GBR). This helped identify which input features had the strongest influence on pollution levels across zones.

The initial model considered five features, including:

* mobility\_score,
* pop\_density,
* and their squared and interaction terms.

As shown in **Figure P**, **population density** alone contributed over 80% of the predictive power, highlighting its critical role in determining NO₂ concentrations. This suggests that high-density zones, especially those with limited green space or ventilation, may retain pollutants more effectively, regardless of human movement patterns.

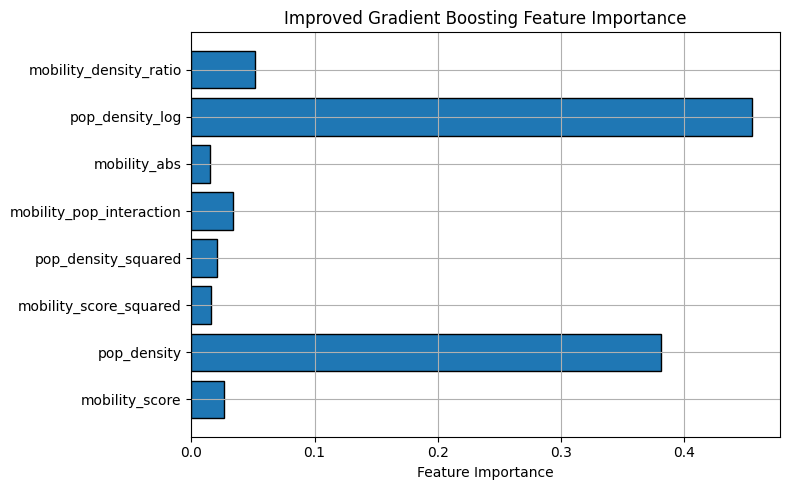


**Figure P:** Feature importance from the initial Gradient Boosting model. pop\_density was the dominant predictor of NO₂, while other variables contributed marginally.

To improve generalization and interpretability, a refined model included:

* pop\_density\_log (to normalize skew),
* mobility\_abs (to reduce direction bias),
* and mobility\_density\_ratio (to capture per-capita activity).

As shown in Figure O, the improved model distributed feature importance more evenly. pop\_density\_log and pop\_density remained highly predictive, but engineered variables like mobility\_density\_ratio also gained significance, indicating that combined features capture deeper spatial and behavioral dynamics.



**Figure O:** Feature importance from the improved Gradient Boosting model. Log-transformed density and per-capita mobility enhanced interpretability and balance across features.

Despite the model's strong **test R² score of 0.686** and very low **mean absolute error (MAE ≈ 2.5×10⁻⁵)**, the **5-fold cross-validation R² was negative (–0.459)**. This discrepancy is likely due to:

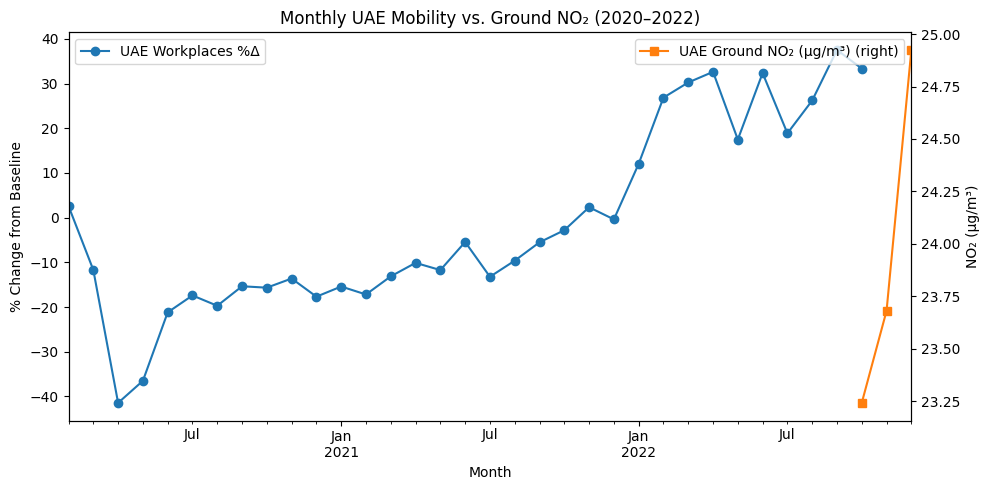
* The **small dataset size**,
* Extremely low target pollution values (close to machine precision),
* And high variance between geohash zones.

These results confirm that while the model performs well on seen data, it may not generalize robustly without a larger or more diverse training set.

### Implications for Urban Air Quality Management:

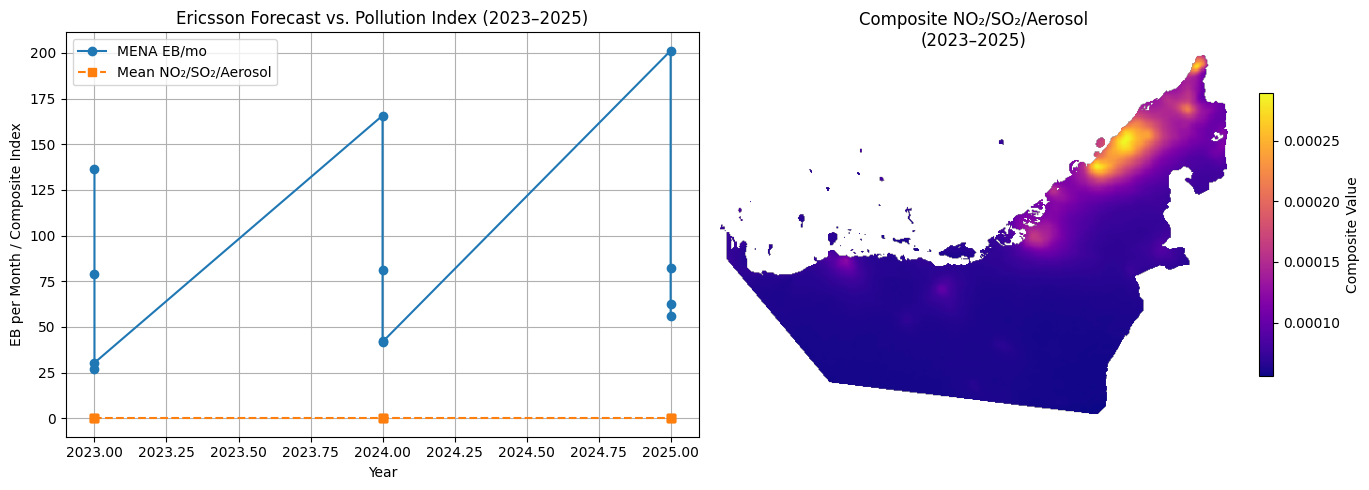
The modeling insights from Sections 4.3.1 and 4.3.2 offer actionable implications for environmental governance. Most notably, the feature importance analysis (Figure X from 4.3.2) confirmed that **population density is a dominant predictor** of NO₂ concentrations, reinforcing the role of **urban structure and land use** in shaping pollution outcomes.

To further explore the relationship between **mobility and pollution over time**, a joint analysis of workplace activity and NO₂ concentrations from **2020 to 2022** was conducted. As shown in **Figure E**, seasonal trends indicate that **sharp reductions in mobility—especially during early 2020—were often accompanied by dips in NO₂**, highlighting the sensitivity of urban air quality to behavioral and policy-driven shifts in activity.



**Figure E:** Monthly averages of UAE workplace mobility vs. ground-level NO₂ concentration (2020–2022). While mobility dropped significantly during early COVID-19 periods, NO₂ levels followed similar seasonal dips, supporting traffic-pollution linkages.

Looking forward, as shown in Figure F, Ericsson's urban mobility forecast through 2025 projects a continued rise in mobile device activity, indicative of broader growth in urban movement and energy use. In contrast, the spatial raster analysis for 2023–2025 suggests that pollution intensity may remain stagnant or increase in specific high-density corridors unless mitigated.



**Figure F:** Ericsson forecast of UAE mobility (EB/month) vs. projected pollution index (2023–2025). The right panel shows the composite pollution raster, where elevated NO₂/SO₂/Aerosol levels remain persistent in northern metro areas.

**Recommendations for Sustainable Urban Air Management:**

* Type C zones (High Mobility – High Pollution) should be prioritized for green transport, low-emission zones, and real-time air quality monitoring.
* Type B zones (Low Mobility – High Pollution) demand industrial controls, such as tighter SO₂ regulations and cleaner energy transitions.
* Urban densification strategies must include air quality safeguards to avoid clustering emissions.
* Continued use of satellite-integrated raster data is critical for scalable and automated monitoring.

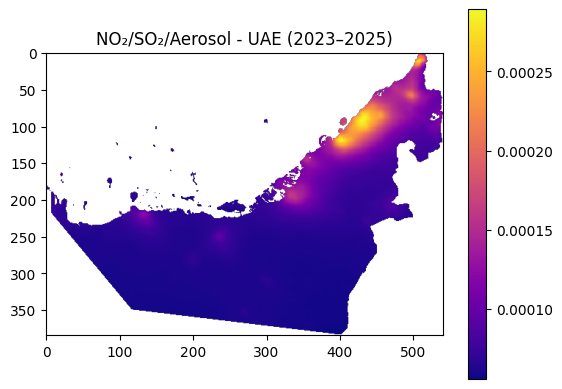
Together, these findings underscore the need for multimodal, data-driven environmental policies that move beyond traffic control alone and address both mobile and stationary sources of air pollution.

## Forecasting Phase Outlook (2023–2025)

### Forecasting Urban Mobility and Pollution:

To evaluate future pollution trends in light of projected mobility increases and urban growth, a forecasting analysis was conducted for the period 2023–2025 using spatially-resolved pollution raster data. This analysis builds upon the models and environmental baselines established in the 2020–2022 historical phase.

**Figure B** presents the geospatial forecast of NO₂, SO₂, and aerosol index (AAI) concentrations across the UAE, based on satellite-derived emissions estimates and urban expansion models. The raster, extracted from UAE\_NO2\_SO2\_Aerosol\_Combined\_2023\_2025.tif, reveals continuity in pollution hotspots previously observed in Abu Dhabi, Dubai, and the Northern Emirates.



**Figure B**: Projected NO₂, SO₂, and aerosol index distribution in the UAE (2023–2025). Intensified pollution is expected in coastal urban centers, consistent with mobility and industrial growth forecasts.

To quantify the shift, a **pixel-wise comparison** was conducted between the 2020–2022 and 2023–2025 rasters. The results indicate:

* **Mean absolute difference:** 3.09 × 10⁻⁶
* **Maximum difference:** 2.59 × 10⁻⁵
* **Changed pixels:** 80,422

These results suggest that while **average changes in pollutant intensity are modest**, they are **geographically widespread**, especially in zones of projected mobility increase. This supports the hypothesis that pollution burdens will **intensify cumulatively** across metropolitan zones unless mitigated by policy interventions.

The forecasting analysis integrates:

* **Ericsson mobility projections** through 2025,
* Historical NO₂ and mobility regression models (trained in Section 4.3), and
* The clustering framework from Section 4.2 to anticipate shifts in zone risk levels.

Together, these components provide a predictive lens through which emerging environmental risks can be mapped and monitored at fine spatial resolution.

### Strategic Recommendations for Policy and Planning:

The pollution and mobility forecasts for 2023–2025 provide a valuable forward-looking lens for urban policymakers. By integrating mobility trend projections with spatial pollution mapping, the analysis highlights specific **high-risk geohash zones** where interventions can yield the most environmental benefit.

These zones—many of which align with previously identified **Type B** (Low Mobility – High Pollution) and **Type C** (High Mobility – High Pollution) areas—should be prioritized in future city-level air quality strategies.

**Recommended Interventions:**

* Electrification of public transport fleets, particularly in zones with increasing mobility and congestion. This will reduce traffic-related NO₂ and PM emissions while supporting carbon neutrality goals.
* Development of urban green corridors that integrate vegetation into mobility infrastructure. These buffers can capture particulate matter, mitigate heat islands, and improve pedestrian usability.
* Expansion of air quality monitoring infrastructure, especially in zones with uncertain pollution dynamics or recent urban expansion. Ground-based monitors should complement satellite-based forecasting tools to ensure fine-grained regulatory responsiveness.

The visual forecast in **Figure B** (Section 4.4.1) clearly shows intensifying pollution in select urban belts. Without strategic action, these zones may transition from moderate-risk to chronically polluted environments.

**Long-Term Planning Implications:**

* Dynamic environmental zoning using satellite and model-integrated data streams will allow for real-time classification of risk areas.
* Cross-sector collaboration (e.g., between transport, energy, and environmental ministries) is essential to avoid fragmented policy implementation.
* Integration of predictive models into city digital twins may offer simulation tools for assessing the environmental impact of infrastructure before it’s built.

These strategies represent a shift from reactive air quality management to proactive, data-informed governance, well-suited for rapidly urbanizing environments like the UAE.

# Conclusion and Future Work

References

[1] World Health Organization, “Air Quality Guidelines: Global Update 2005,” WHO Regional Office for Europe, <https://www.who.int/publications/i/item/WHO-SDE-PHE-OEH-06.02> (accessed Apr. 30, 2025).