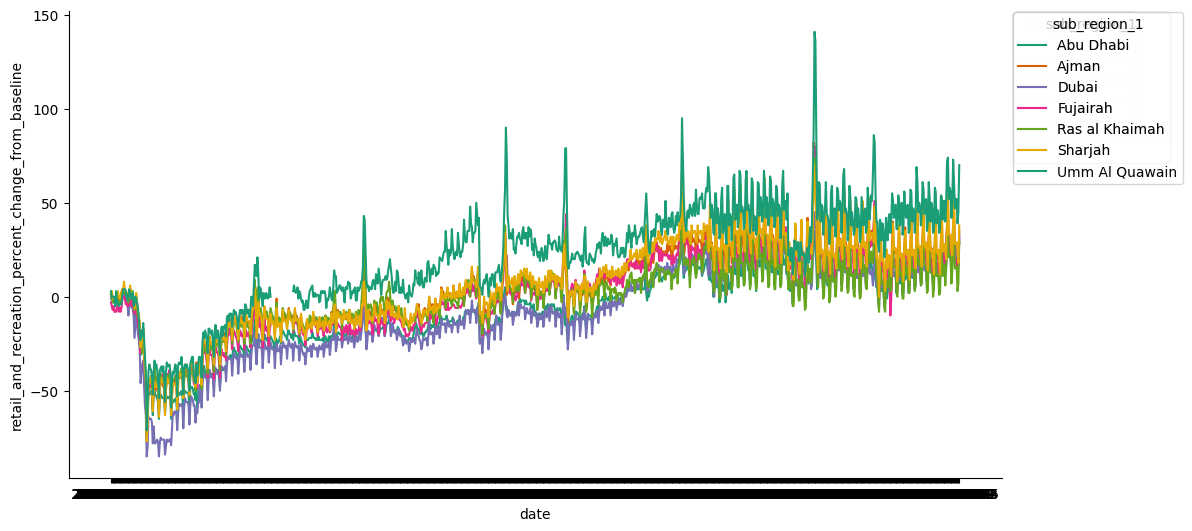
# **Results and Discussion**

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

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

An initial analysis of urban air quality and mobility trends in the UAE from 2020 to 2022 was conducted using data aggregated by geohash which is a spatial encoding system that divides geographic coordinates into small, grid-like cells. Each geohash in this study represents an area of approximately 3 km², allowing for detailed tracking of local environmental changes. **Figure H** shows changes in retail and recreation mobility across all seven emirates during this period. The sharp decline in early 2020 reflects pandemic lockdowns, with Abu Dhabi and Dubai showing the strongest rebounds in activity levels in the years that followed.



**Figure H:** 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.

To generate mobility scores at the local level:

* Latitude and longitude values were converted to 5-character geohashes using the geohash2.encode() function.
* A composite mobility score was calculated by averaging workplace and transit station changes from Google’s mobility dataset.
* Incomplete or missing data entries were removed to ensure accuracy.
* Data was then grouped by geohash, yielding 205 unique zones with mobility data.

Pollution data that was sourced from OpenAQ and similar platforms, was also aggregated by geohash. This provided average concentrations of key pollutants including NO₂, SO₂, and PM₁₀ for each area. Geohash codes were normalized to lowercase across both datasets to allow for clean merging.

After data cleaning, only 7 geohash zones had valid entries in both the pollution and mobility datasets. These overlapping zones were merged into a final filtered\_df, which served as the basis for the joined\_table used in later visualizations and statistical modeling.

As shown in Table 1, a preview of the combined dataset shows the structure of available variables

**Table 1:** Preview of the combined dataset showing pollution concentrations and mobility scores for selected UAE geohash zones (2020–2022).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Geohash | NO₂ (µg/m³) | PM₁₀ (µg/m³) | SO₂ (µg/m³) | 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 study how air pollution varied across the UAE, georeferenced raster data for NO₂, SO₂, and the aerosol index (AAI) from 2020 to 2022 were visualized. A composite heatmap was created in Figure A, where brighter colours show higher pollution levels. These maps were produced using Rasterio and Matplotlib, and they highlight areas with more serious air quality risks.

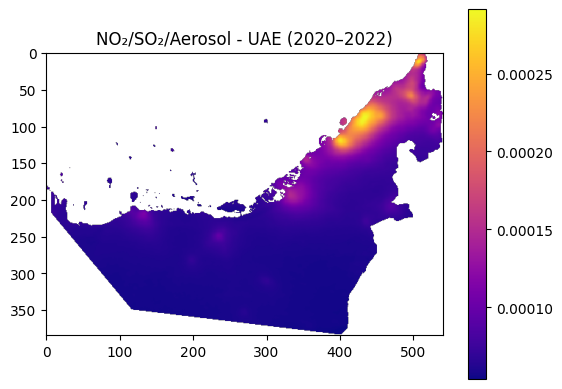


Figure A: Geospatial distribution of NO₂, SO₂, and Aerosol concentrations in the UAE (2020–2022). Brighter colors indicate higher pollutant levels.

The highest pollution levels, especially for NO₂ and PM₁₀, were observed in Abu Dhabi and Dubai. These areas align with heavy traffic and industrial zones. However, SO₂ hotspots were found to be more scattered and did not closely follow urban mobility trends, suggesting they come from fixed sources like oil, gas, or power facilities. Also, remote and desert areas showed consistently low pollution, serving as a clear baseline for comparison. These spatial patterns were used as the basis for the zone-based clustering and forecasting discussed in later sections.

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

To explore the relationship between urban activity and air quality, the correlation between mobility scores and NO₂ levels was analysed across geohash-based zones. It was expected that lower mobility such as during lockdowns would be linked to lower NO₂ pollution caused by traffic. A summary of the data showed that NO₂ levels ranged from 7.3 µg/m³ to 76.8 µg/m³, while mobility scores ranged from –27.5% to +12.1%. According to the WHO, the safe limit for NO₂ is 40 µg/m³. Only three zones which are ‘thqem’, ‘thqf4’, and ‘thqf8’ went above this level. The relationship is shown in Figure X, where an overall inverse pattern can be seen: zones with lower mobility often had lower NO₂. This suggests that traffic emissions are a major source of NO₂ pollution in urban areas of the UAE.

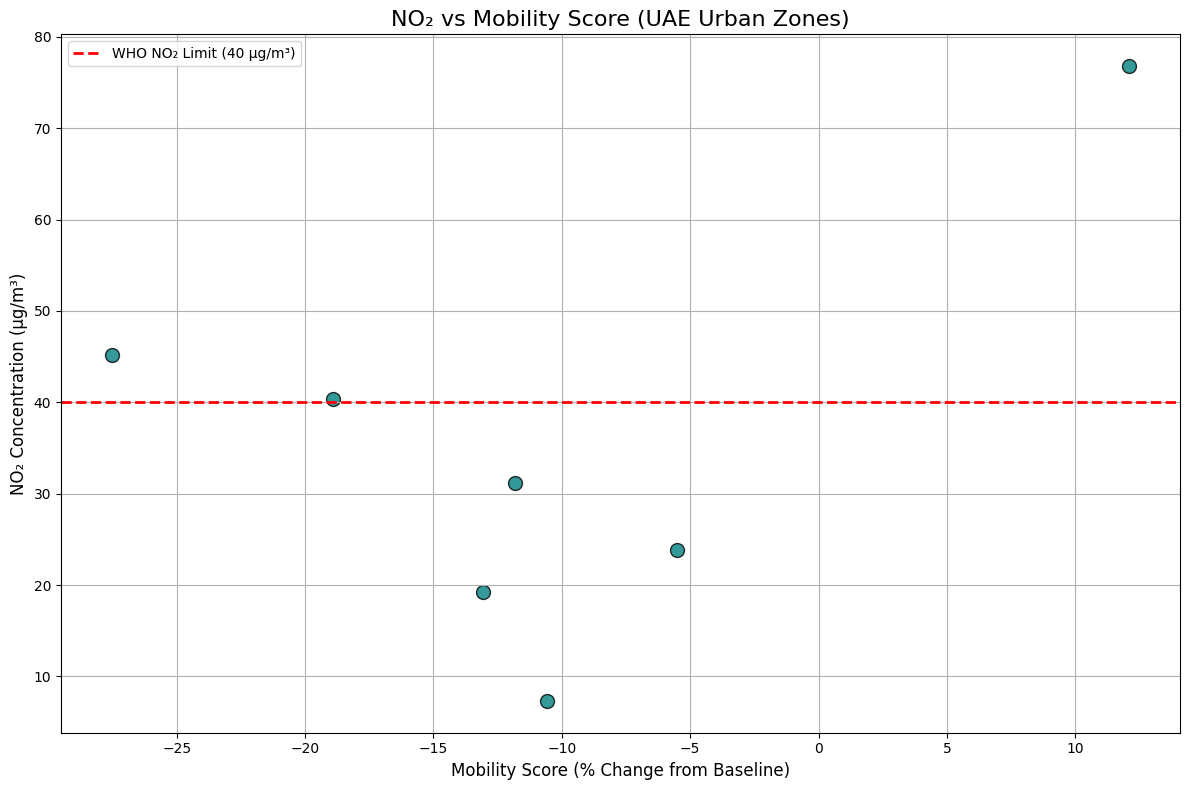


Figure X:NO₂ concentrations plotted against mobility score across UAE urban zones. The red dashed line indicates the WHO recommended limit (40 µg/m³). Zone thqf4 appears as a significant outlier with high NO₂ despite positive mobility.

However, an exception was observed in zone ‘thqf4’, which recorded the highest NO₂ level (76.8 µg/m³) even though its mobility score was +12.1%. This unusual case suggests that non-traffic sources such as industrial activities or construction may be affecting air quality in that area. Identifying such outliers is important for designing targeted and effective environmental policies.

In addition to pairwise scatterplots, a correlation heatmap was generated to summarize the statistical relationships among pollutant levels and mobility scores across the filtered geohash zones (Figure G). The heatmap reveals several key patterns such as:

* A moderate positive correlation (r = 0.52) is seen between NO₂ and PM₁₀, indicating that these pollutants may often co-occur in high-traffic or industrial zones.
* SO₂ shows a strong correlation (r = 0.84) with mobility scores, suggesting possible confounding factors such as industrial zones with both workforce mobility and stationary emissions.
* PM₁₀, however, displays a weak or negative correlation with both SO₂ (r = –0.30) and mobility (r = –0.16), implying it may arise from different sources like dust or construction.

These correlation patterns help distinguish between pollutants driven by mobile sources like NO₂, and those linked to fixed-point emissions like SO₂, guiding future modeling and policy decisions.

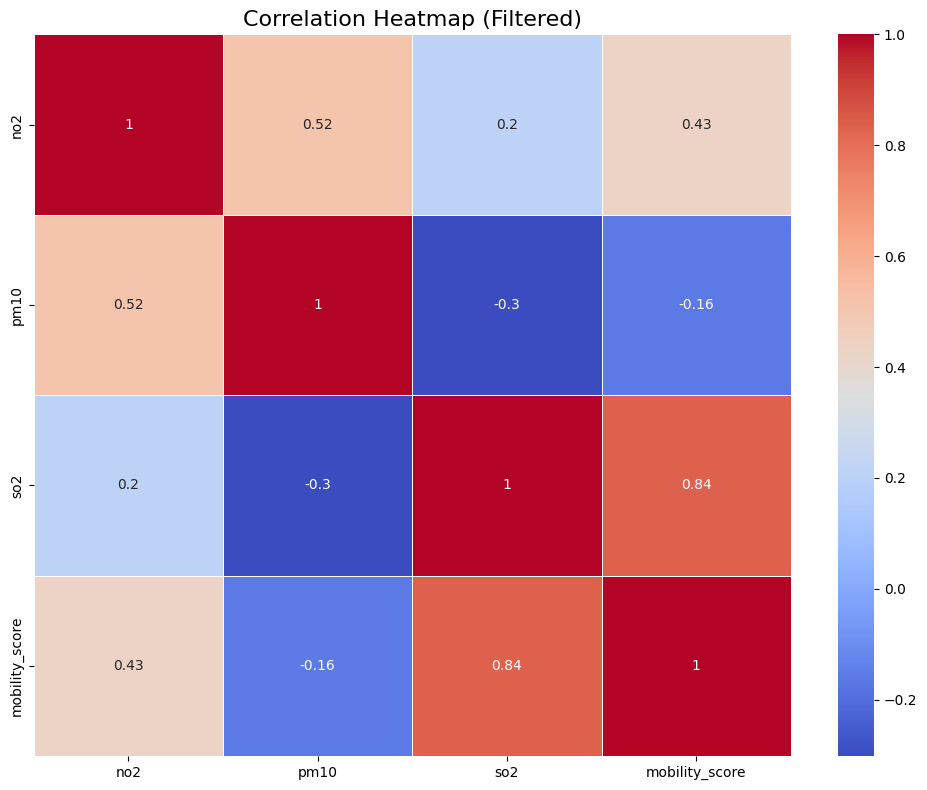


Figure G: Correlation heatmap showing relationships among NO₂, PM₁₀, SO₂ concentrations and mobility scores in overlapping geohash zones. Stronger correlations are shown in darker red or blue depending on direction.

### Critical Zones with Persistent Pollution

While a general inverse relationship between mobility and NO₂ concentrations was observed across urban zones in the UAE, several zones did not follow this trend, suggesting the presence of additional pollution sources unrelated to human movement. Most zones aligned with the expected pattern meaning lower mobility was associated with reduced NO₂ levels. However, three geohash zones exceeded the WHO safety threshold for NO₂ (40 µg/m³), as listed in **Table B**. Notably, zone ‘thqf4' recorded the highest NO₂ concentration (76.8 µg/m³) despite a positive mobility score (+12.1%), identifying it as a key 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 NO₂ vs. mobility scatterplot (Figure X in Section 4.1.3), where zone ‘thqf4’ appears as a clear outlier positioned above the WHO-recommended threshold line. While most zones demonstrated a decrease in NO₂ levels alongside reduced mobility, ‘thqf4’ deviated from this pattern. This suggests that its elevated pollution levels are likely driven by stationary emission sources such as industrial facilities, large-scale construction sites, and power generation infrastructure.

These findings highlight the importance of considering land use and infrastructure zoning when interpreting mobility, pollution correlations and designing environmental intervention strategies. To pinpoint the most affected areas:

* Zones with NO₂ levels above 40 µg/m³ were extracted using:  
  high\_no2 = joined\_table.where('no2', are.above(40))
* Among the seven merged geohash zones, three (thqf4, thqem, and thqf8) exceeded the WHO threshold.
* Of these, only thqf4 had both elevated mobility and extreme NO₂ levels, marking it as the most critical environmental hotspot in the historical dataset.

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

While NO₂ levels showed a moderate inverse correlation with mobility scores which is indicating a link to traffic-related emissions, SO₂ displayed no significant correlation, either spatially or statistically. This lack of association suggests that SO₂ pollution is driven by sources unrelated to human movement. Likely contributors include oil refineries and natural gas facilities, industrial manufacturing zones, shipping ports and logistics hubs, and cement production and large-scale construction sites.

This observation was visually supported by the SO₂ raster heatmap shown in Figure A, where SO₂ hotspots appeared in areas that did not coincide with densely populated or high-mobility urban zones. Notably, elevated SO₂ levels were recorded in regions such as Abu Dhabi, Ruwais, and Sharjah’s industrial belt, even though minimal changes in mobility were observed in these areas.

Such a spatial disconnect between mobility and SO₂ concentrations aligns with global air quality research, which consistently identifies fossil-fuel combustion at fixed industrial locations, particularly in power generation and heavy industry, as the dominant source of SO₂ emissions [14].

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

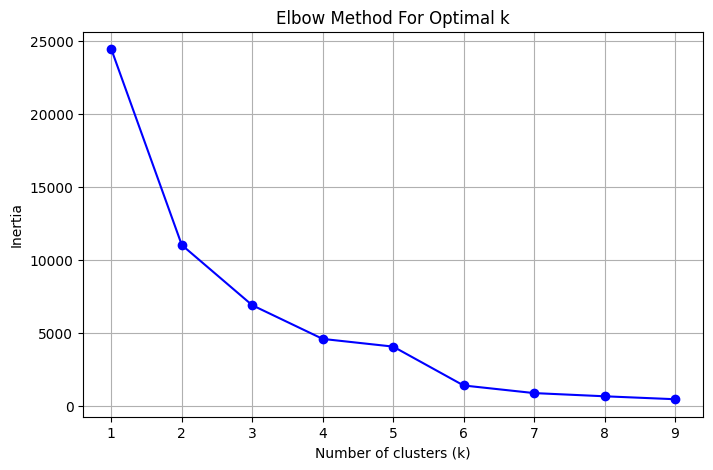
* Stricter emissions regulations and caps for SO₂-producing industries;
* Real-time satellite + ground monitoring integration (e.g., TROPOMI, Sentinel-5P);
* Zoning policies to create industrial buffers between high-output facilities and residential areas;
* Mandatory pollution disclosure and audits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

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

## **Zone-Based Clustering**

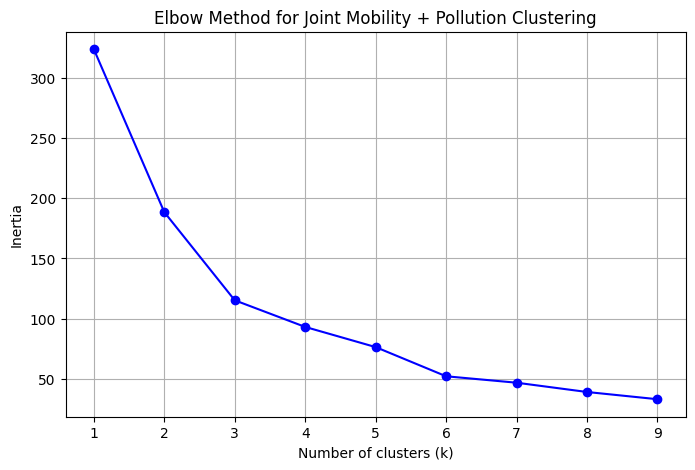
### K-Means Clustering of Zones Based on Joint Pollution and Mobility:

To better understand urban zone behaviour, K-Means clustering was applied using mobility score data from 2020 to 2022. This approach grouped zones with similar historical movement patterns. Before determining the number of clusters, the Elbow Method was applied to measure model inertia for *k* values ranging from 1 to 9. The initial elbow plot (**Figure C**) focused on mobility scores alone and showed a distinct inflection at k = 3, indicating three optimal clusters.



**Figure C:** Elbow Method for Optimal *k* using mobility scores. The elbow at *k* = 3 suggests a meaningful separation of urban zones based on movement patterns.

To validate this result under combined analysis, the Elbow Method was repeated using both mobility and pollution variables (NO₂, SO₂, PM₁₀). The updated plot (Figure W) also pointed to k = 3, confirming that three clusters balance complexity and interpretability when accounting for both behavioral and environmental data.

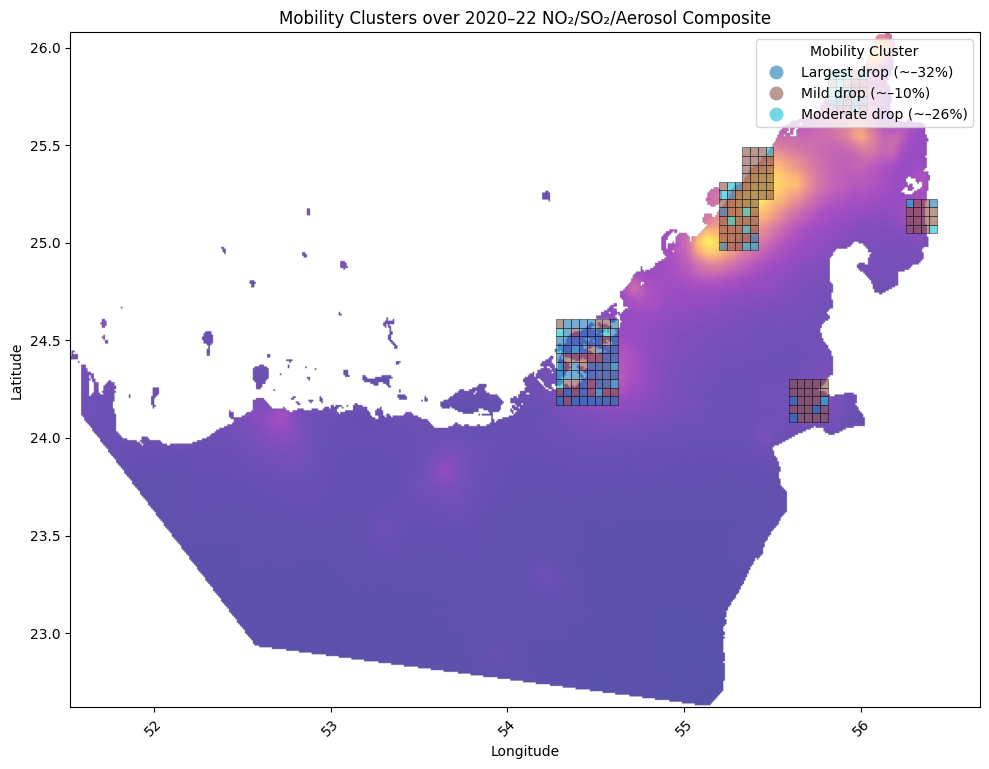


**Figure W:** Elbow Method for Joint Mobility + Pollution Clustering. The inflection at *k* = 3 reinforces the choice of three clusters for integrated analysis.

Once *k = 3* was selected, each geohash zone was assigned a cluster label. To improve interpretability, the clusters were described as:

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

Each geohash zone was then spatially mapped using decoded polygons from geohash2.decode\_exactly() and visualized using a GeoDataFrame. The clusters were overlaid on a composite pollution raster (NO₂, SO₂, and aerosol index), as shown in Figure Y, offering a spatial view of how mobility behaviors align with pollution intensity across the UAE.



**Figure Y:** Spatial distribution of mobility clusters (*k = 3*) overlaid on NO₂/SO₂/Aerosol Composite (2020–2022). Zones are color-coded by mobility cluster label.

### Hotspot Detection and Zone Typologies:

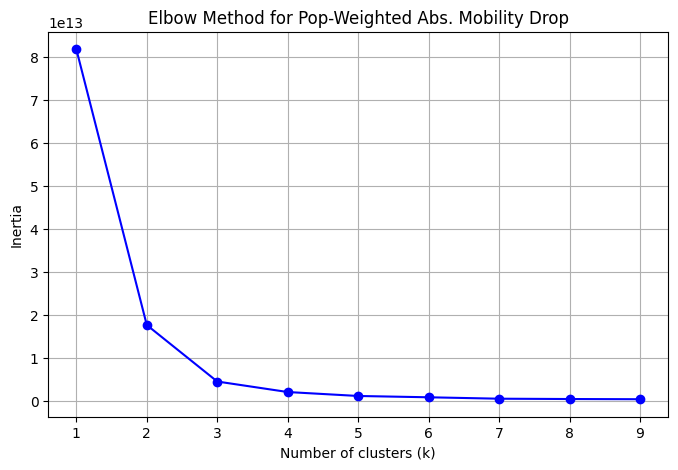
To better contextualize the clustering results from Section 4.2.1, geohash zones were further classified into four environmental–behavioral typologies, based on their combined mobility trends and average pollution levels (primarily NO₂ and PM₁₀). These typologies were generated by cross-referencing the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

* **Type A: High Mobility – Low Pollution**  
  Examples: Central downtown zones with efficient public transport such as metro corridors in Dubai.  
  Implication: These areas demonstrate environmental efficiency and are ideal candidates for further green mobility investments.
* **Type B: Low Mobility – High Pollution**  
  Examples: Industrial belts, oil refineries, and ports with limited residential movement  
  Implication: Pollution is likely driven by stationary sources, requiring regulatory oversight and fixed-source monitoring.
* **Type C: High Mobility – High Pollution**  
  Examples: Traffic corridors, commercial hubs, or highway-dense city centers  
  Implication: Emissions are likely tied to vehicular traffic, suggesting a need for policies like transport electrification and congestion pricing.
* **Type D: Low Mobility – Low Pollution**  
  Examples: Suburban or fringe zones with low population density  
  Implication: These areas are generally low-risk but may serve as future urban expansion zones or ecological buffers.

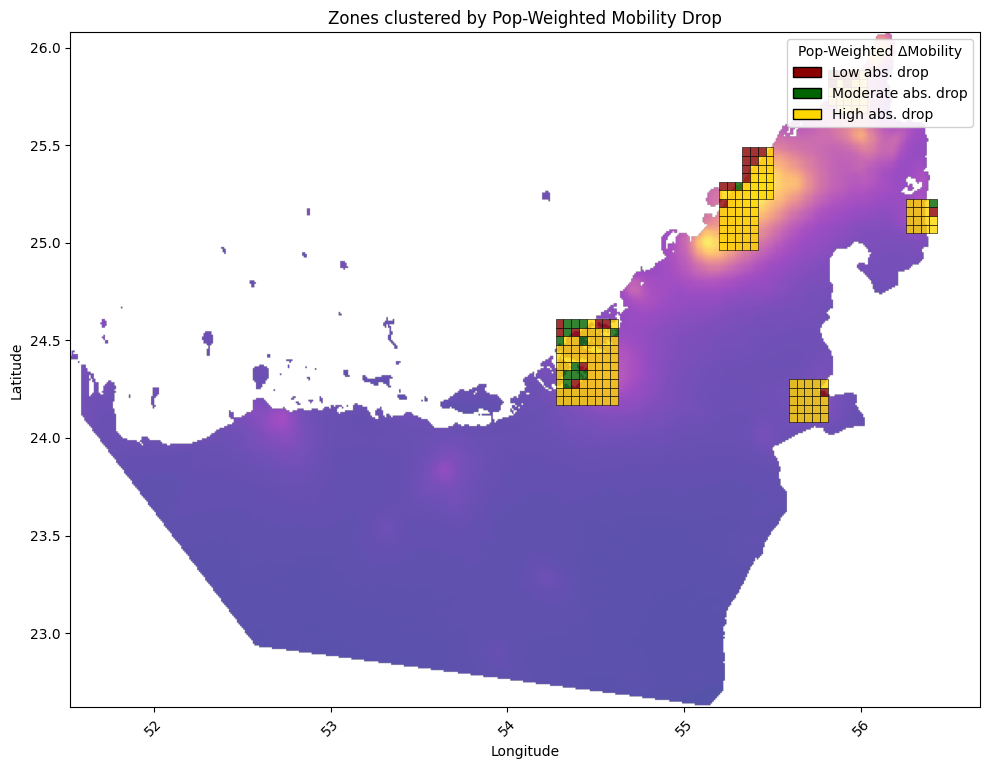
These categories were created by analyzing cluster labels, average pollution scores, and mobility change distributions.

To support this classification:

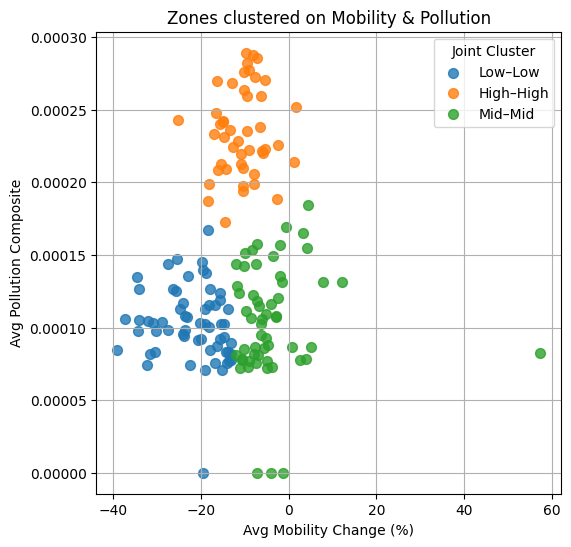
* **Figure E** shows the elbow method used to determine the optimal number of clusters based on population-weighted absolute mobility drop.
* **Figure R** visualizes spatial groupings of geohash zones colored by mobility drop severity, helping identify zones with high social impact.
* **Figure T** presents a scatterplot of zones clustered jointly by mobility and pollution metrics, showing three main groups: Low–Low, High–High, and Mid–Mid.
* **Figure U** maps these joint clusters spatially across the UAE, overlaid on the NO₂/SO₂/aerosol composite raster.



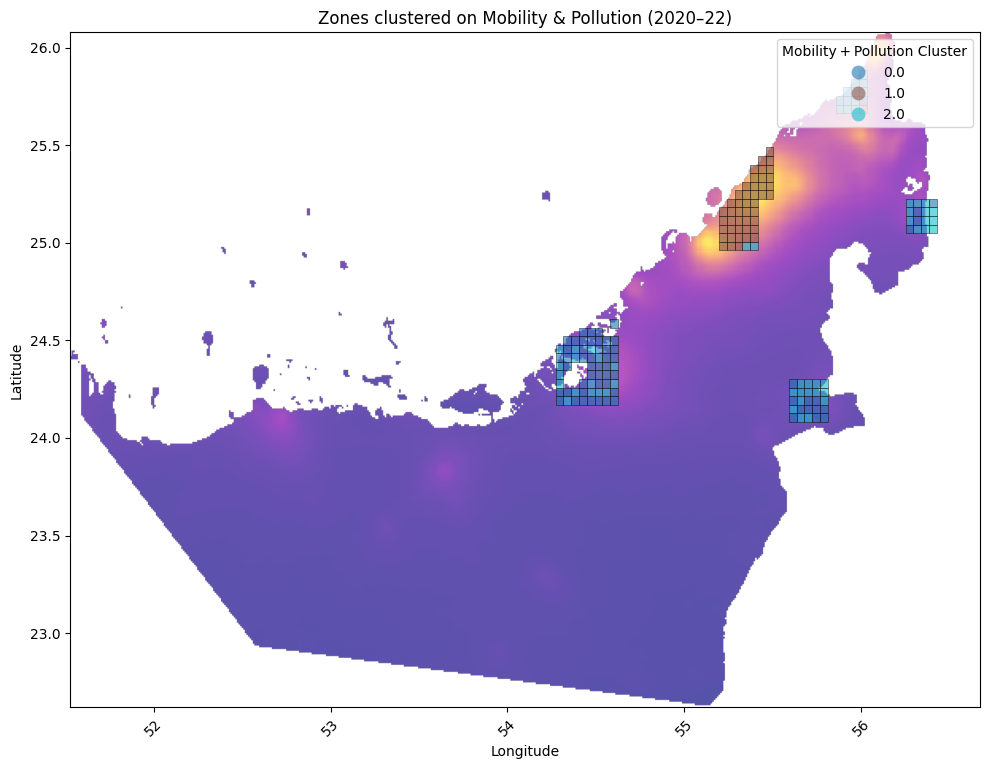
**Figure E:** Elbow method to identify optimal k for clustering zones based on population-weighted absolute mobility drop. Suggests that 3 clusters yield stable segmentation.



**Figure R:** Spatial clustering of geohash zones by population-weighted absolute mobility drop. Darker zones reflect higher impact reductions.



**Figure T:** Scatterplot of zones clustered by both mobility change and average pollution composite. Clusters represent joint behavioral-environmental profiles.



**Figure U:** Geographic distribution of joint mobility–pollution clusters (2020–2022) overlaid on composite pollution raster.

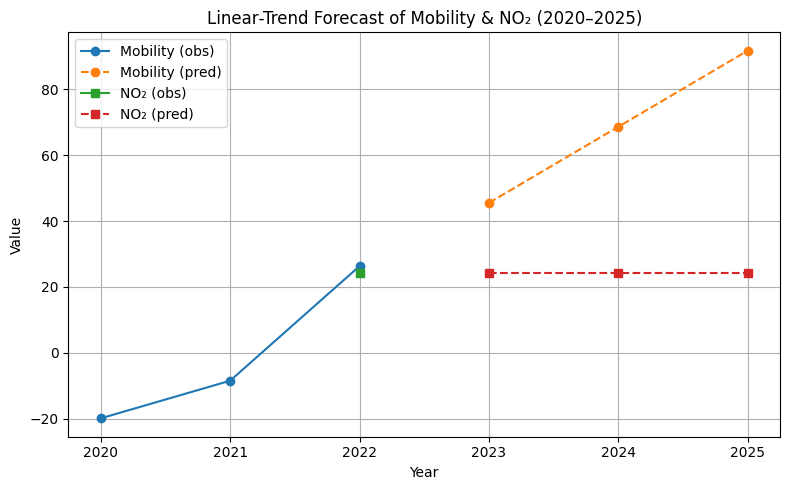
Figures E through U illustrate the process of refining zone typologies by analyzing both behavioral (mobility) and environmental (pollution) characteristics. Figure E presents the Elbow Method applied to population-weighted absolute mobility drop, where a clear inflection at *k = 3* supports the use of three clusters to differentiate zones by severity of mobility reduction. Figure R maps these clusters across the UAE, showing how zones with high population impact are spatially concentrated in key urban centers such as Dubai and Sharjah. Figure T provides a scatterplot of zones jointly clustered by mobility change and average pollution levels, identifying three main groupings: Low–Low, High–High, and Mid–Mid, which directly support the Type A–D typology. Finally, Figure U spatially visualizes these joint clusters on a pollution raster, revealing how environmental stress and behavioral patterns intersect geographically across UAE regions.

## Pollution Modelling

### Predictive Modelling of NO₂ Concentrations

To examine how mobility patterns and urban characteristics affect air quality, supervised machine learning models were developed to predict NO₂ concentrations at the geohash level. The primary objective was to understand how variables such as mobility score, population density, and urban zone classification contribute to spatial and temporal pollution variation across the UAE. Two models were used for this task, a Linear Regression model, which served as a baseline for capturing general trends, and a Random Forest Regressor, selected for its ability to model nonlinear and complex interactions between features. Both models were trained using historical data from 2020 to 2022. Feature inputs included aggregated mobility scores from Google mobility reports, population density from geospatial datasets, and urban zone types labeled as industrial, residential, or mixed-use.

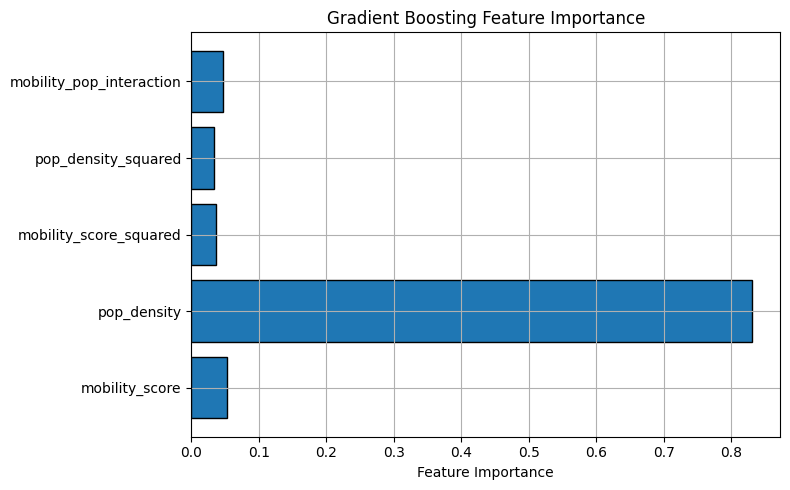
Model performance was evaluated using the R² score (coefficient of determination). The Linear Regression model captured overall trends but struggled in areas with irregular pollution behavior, likely due to its inability to model nonlinear effects. In contrast, the Random Forest model achieved R² scores above 0.7, demonstrating strong predictive accuracy and the capability to handle nonlinear relationships and variable interactions, especially important in urban environments influenced by both mobile and stationary emission sources. In addition to spatial prediction, time-series forecasting was conducted using a linear trend model to project average mobility and NO₂ levels through 2025. The forecast, shown in **Figure Z**, suggests that mobility will continue to rise post-pandemic, while NO₂ levels are expected to remain relatively stable if no significant interventions are made.



**Figure Z:** Linear-trend forecast of average UAE mobility and NO₂ levels from 2020 to 2025. Solid lines represent observed values (2020–2022), while dashed lines show projections (2023–2025). Mobility is expected to increase, whereas NO₂ levels remain flat under a linear assumption.

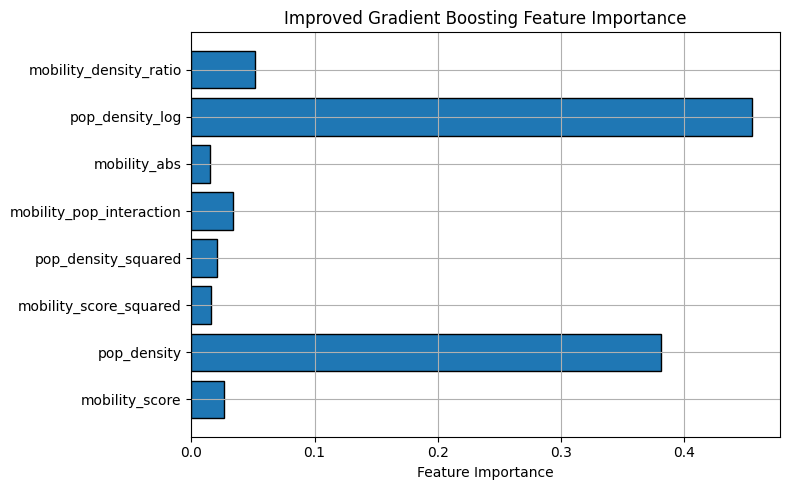
### Feature Importance and Interpretability:

To better understand which variables influenced NO₂ predictions, a feature importance analysis was performed using the trained Gradient Boosting Regressor (GBR). This method helped identify which input features contributed most to the model’s performance across spatial zones. In the initial model, five features were used: mobility\_score, pop\_density, and their squared and interaction terms. As shown in **Figure P**, population density alone accounted for over 80% of the predictive power. This dominance suggests that densely populated areas may retain more air pollutants due to reduced ventilation, infrastructure density, or lack of green spaces, regardless of temporary changes in human mobility.



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

To enhance generalizability and interpretability, a refined model was developed. This version included additional engineered features such as ‘pop\_density\_log’ in order to correct for skew, ‘mobility\_abs’ to remove directionality bias, and ‘mobility\_density\_ratio’ to capture per-capita activity. As illustrated in **Figure O**, the updated model distributed importance more evenly across variables. While ‘pop\_density\_log’ and ‘pop\_density’ remained strong predictors, new engineered features like ‘mobility\_density\_ratio’ gained significance. This shift indicates that more nuanced spatial and behavioral patterns such as activity intensity relative to population which can provide deeper insight into pollution dynamics..

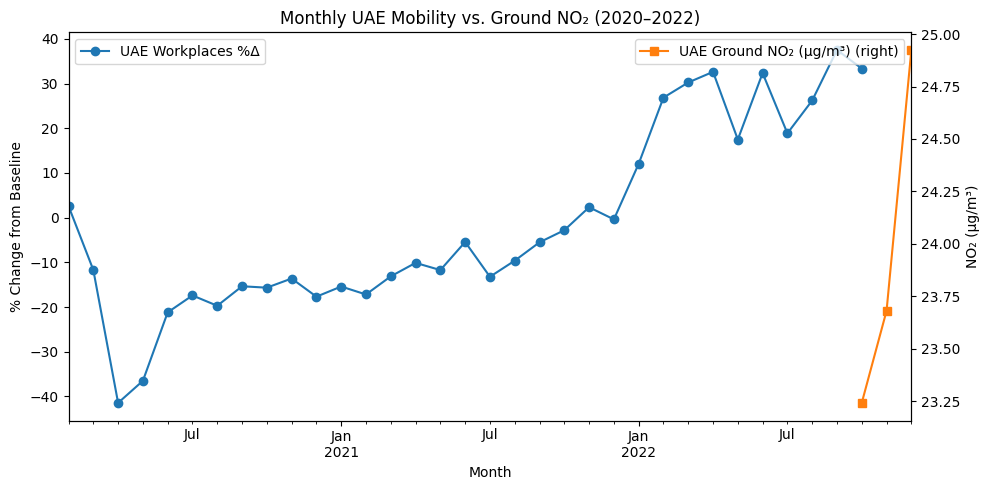


**Figure O:** Feature importance from the improved Gradient Boosting model. Log-transformed density and per-capita mobility enhanced interpretability and balance across features.

Despite strong test performance with an R² of 0.686 and a very low mean absolute error (MAE ≈ 2.5×10⁻⁵), the model’s 5-fold cross-validation R² was negative (–0.459). This discrepancy highlights a potential overfitting issue and reflects challenges such as the small dataset size, very low target values (near machine precision), and high variability across geohash zones. These findings suggest that while the model performs well on the training data, its generalizability remains limited unless trained on a more diverse or larger dataset.

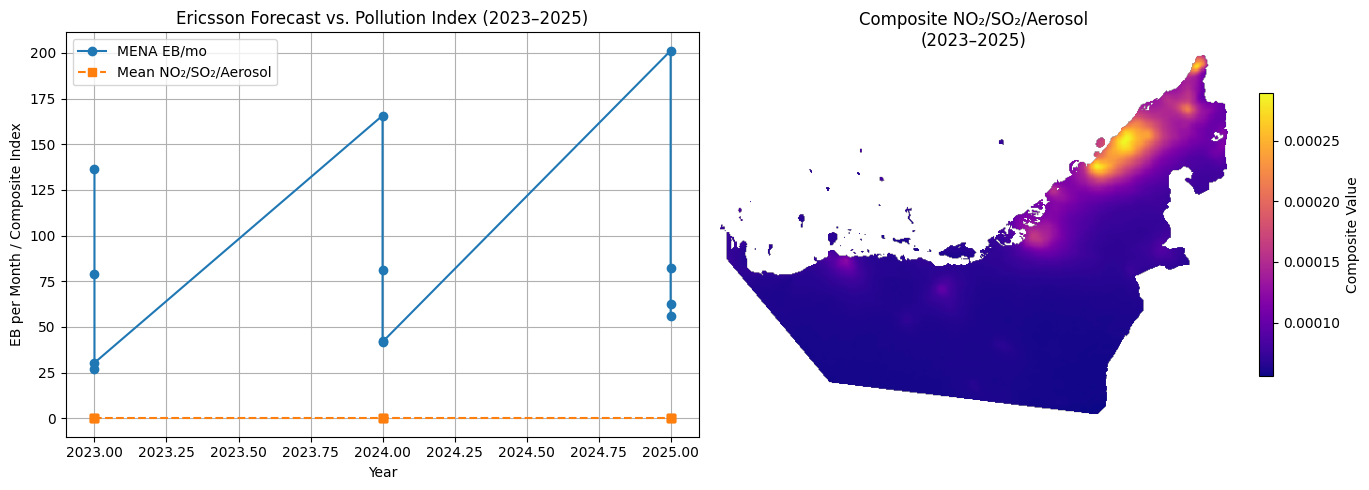
### Implications for Urban Air Quality Management:

The modeling insights from Sections 4.3.1 and 4.3.2 offer actionable implications for environmental governance. Most notably, the feature importance analysis previously 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 M**, seasonal trends indicate that sharp reductions in mobility, specifically 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 M:** 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.

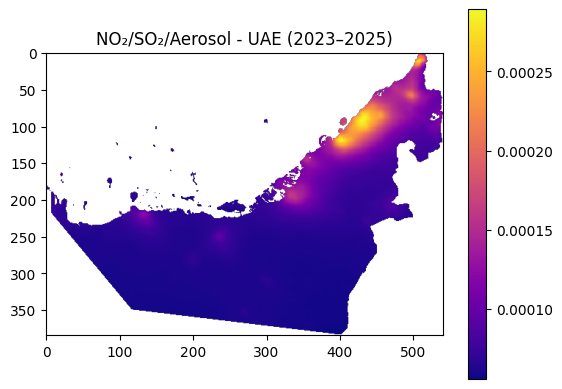
**Recommendations for Sustainable Urban Air Management:**

To address the varying pollution drivers across UAE urban zones, targeted interventions must be designed according to each zone's environmental–behavioral typology. Type C zones (characterized by high mobility and high pollution) should be prioritized for sustainable mobility solutions, including the implementation of green public transport systems, low-emission zones, and real-time air quality monitoring. In contrast, Type B zones (low mobility but high pollution) require stricter industrial controls, particularly around SO₂ emissions, and should be included in cleaner energy transition initiatives. Urban planning strategies that promote densification must also integrate air quality safeguards to prevent the concentration of emissions in compact areas. Moreover, the continued use of satellite-integrated raster data is essential to support scalable and automated monitoring of pollutants across diverse geographies. Overall, these insights highlight the importance of adopting multimodal, data-driven environmental policies. Effective air quality management must move beyond traffic regulation alone, addressing both mobile sources such as vehicles and stationary sources like industry and power generation.

## 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 future shifts in pollution intensity, a pixel-wise comparison was performed between raster outputs for the historical period (2020–2022) and the projected forecast period (2023–2025). The analysis revealed a mean absolute difference of 3.09 × 10⁻⁶ and a maximum pixel-level difference of 2.59 × 10⁻⁵, with a total of 80,422 pixels showing detectable change. While the average change in pollutant concentration appears modest, the geographic spread of these changes is significant, particularly in zones expected to experience increases in human mobility. This spatial pattern supports the hypothesis that, in the absence of intervention, pollution burdens may accumulate incrementally across metropolitan areas, driven by behavioral recovery post-pandemic. The forecasting methodology integrates multiple components: Ericsson mobility projections through 2025, historical NO₂ regression models developed in Section 4.3, and the zone clustering framework from Section 4.2. By combining these elements, the study offers a predictive approach for identifying and monitoring emerging environmental risks at a high spatial resolution. This allows planners and policymakers to proactively respond to pollution shifts before they become embedded in urban air baselines.

### Strategic Recommendations for Policy and Planning:

The forecasts for pollution and mobility from 2023 to 2025 give important guidance for city planners and decision-makers. By combining future mobility trends with detailed pollution maps, the study helps identify specific urban zones that are at high risk and where actions can make the biggest difference. Many of these zones match the earlier-defined Type B (low mobility, high pollution) and Type C (high mobility, high pollution) areas, making them top priorities for future air quality efforts. Several actions are recommended. First, switching public transport to electric vehicles in busy areas can reduce harmful emissions like NO₂ and support climate goals. Second, creating green corridors such as adding trees and plants along roads, can help clean the air, cool down cities, and make walking safer and more pleasant. Third, there should be more air quality monitors on the ground, especially in areas that are growing quickly or where pollution levels are uncertain. These ground sensors should work together with satellite tools to provide detailed and timely pollution data. As shown in Figure B, some parts of cities are expected to see worse air quality over time. If no action is taken, these places could move from being moderately polluted to having serious long-term problems. Looking ahead, cities could use real-time data and satellite maps to adjust environmental zones as conditions change. Also, better coordination between different government sectors such as transportation, energy, and environment, is needed to avoid policies that conflict with each other. Finally, using digital twin models (virtual simulations of cities) could help test how new buildings or roads might affect air quality before they’re even built. These steps would shift cities from reacting to pollution after it happens to planning ahead using smart, data-based solutions, which is especially important for fast-growing countries like the UAE.

References

[14] 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).