# **The Influence of Urban Form and Socio-Demographics on Robotaxi Demand: A San Francisco Case Study**

## **1. Introduction**

Autonomous vehicle (AV) technology continues to evolve, with autonomous ride-hailing or robotaxi services emerging as a potentially transformative force in urban mobility. Transportation planners are closely monitoring this development. Advanced robotaxi services, initially with human safety operators, have gained traction in cities worldwide, including Phoenix, San Francisco, Guangzhou, and Wuhan. Notably, companies like Cruise and Waymo have begun testing and commercializing fully autonomous (safety-operator-free) robotaxis in cities such as San Francisco.

Since their commercial introduction in the early 2020s, adoption has surged dramatically. In San Francisco alone, monthly paid trips exploded from approximately 12,600 in August 2023 to over 400,000 by December 2024 – a staggering 30-fold increase in less than 1.5 years (Figure 1). By November 2024, Waymo had captured a significant 22% market share of the city's total ride-hailing industry, directly competing with established players like Lyft and Uber. This rapid and substantial growth presents both opportunities and challenges for urban mobility. While current data (Aug 2023 - Dec 2024) suggests Robotaxis have primarily drawn riders from existing TNCs (Uber/Lyft) rather than impacting stable public transit and traditional taxi ridership yet, the sheer scale and pace of adoption necessitate an urgent and deep understanding of the underlying demand drivers. Effective integration requires proactive planning to ensure Robotaxis contribute positively to urban goals of sustainability, equity, and efficiency, rather than potentially exacerbating congestion or creating new disparities.

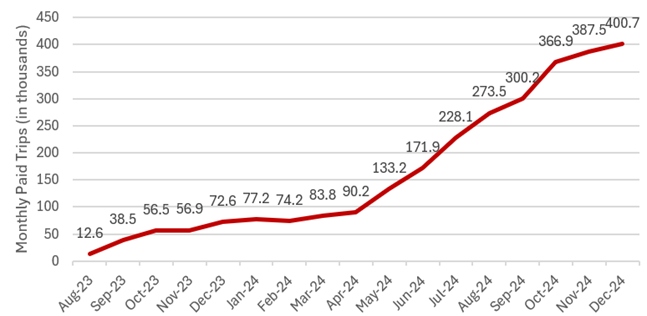


Figure 1. Figure 1 Waymo Monthly Paid Trips in San Francisco

### **1.1 Research Gaps and Objectives:**

### 

Recency and Scale: Many studies used data from before robotaxis became this popular. Large-scale commercial deployments and dramatic ridership increases observed in cities like San Francisco. Research using current, extensive empirical data (like the Aug 2023-Dec 2024 period) is needed to capture these new dynamics.

Comprehensive Factor Integration: While studies often consider either built environment or socio-demographics, few have comprehensively integrated both macro-scale urban form characteristics (the 5Ds: Density, Diversity, Design, Destination Accessibility, Distance to Transit) and micro-scale streetscape quality features (like those derivable from AI-based analysis of street view imagery) alongside detailed socio-demographic variables to explain Robotaxi demand.

To examine how built environment and socio-demographic factors influence Robotaxi demand and inform sustainable mobility strategies. As Robotaxis grows globally, understanding their relationship with urban form is critical to balance innovation, sustainability, and equity.

To identify and quantify the key drivers of Robotaxi demand at the census tract level in San Francisco, focusing on the interplay between the urban environment, population characteristics, and service usage patterns.

· Identify significant macro-scale built environment factors (Density, Diversity, Design, Destination Accessibility, Distance to Transit - the 5Ds) associated with Robotaxi trip generation.

· Assess the influence of micro-scale streetscape quality features (derived using AI-based image analysis) on Robotaxi demand.

· Determine the impact of key socio-demographic and economic factors (e.g., income, density, vulnerability) on Robotaxi adoption.

· Develop and compare predictive models (OLS, Random Forest, and Neural network) to understand the complex, potentially non-linear relationships between the identified factors and Robotaxi demand.

· Generate actionable insights and recommendations for promoting sustainable and equitable mobility futures incorporating Robotaxis.

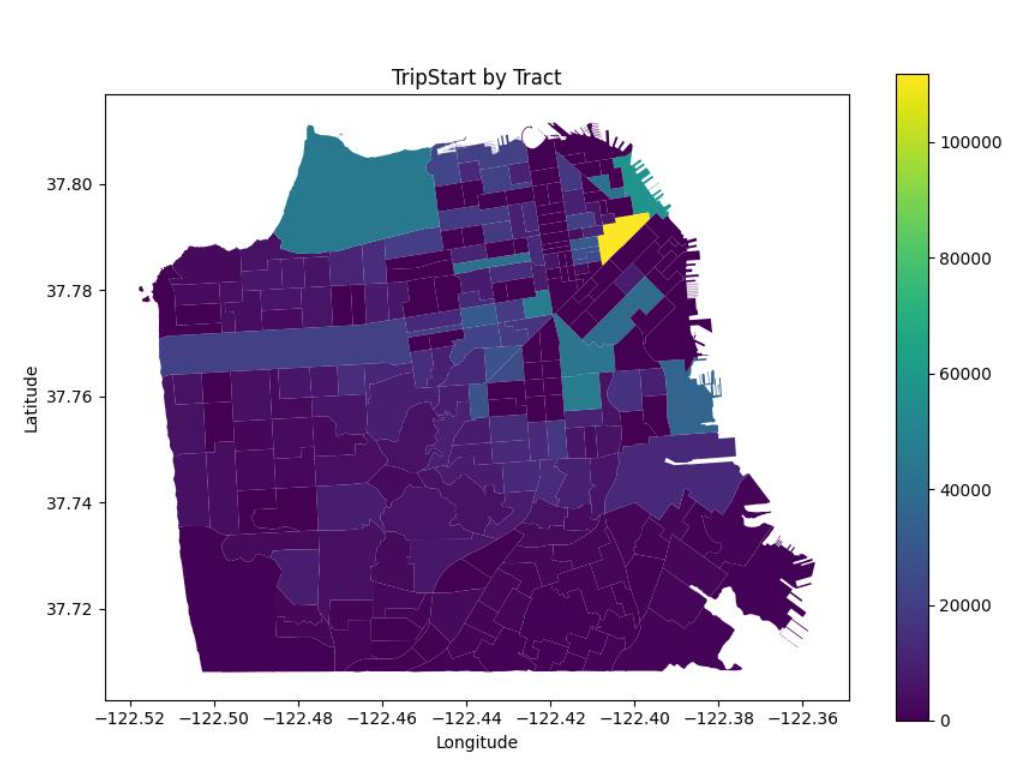
## **2. Data and Methodology**

### **2.1 Study Area**

The study focuses on San Francisco County, California. Its status as an early and extensive adopter of robotaxi services, combined with its diverse and complex urban fabric, makes it an ideal location for investigating the interplay between AVs and the built environment. The unit of analysis is the census tract level.

### **2.2 Data Collection**

* Dependent Variable: Trip start density  
  Monthly Robotaxi trip origin counts per census tract. (Source: Aggregated, anonymized data from service providers, potentially via public utility commission filings or data-sharing agreements).

  
*Figure 2. Number of trip start in census tract level*

* Independent Variables: Built Environment and Socio-Demographics A comprehensive set of independent variables was compiled, categorized as follows: (Refer to Appendix A for detailed variable descriptions)
  1. **Macro-scale Built Environment Features (5D):** Capturing broader urban form and function based on the "5D" framework.
     + ***Density:*** Population density, building density, intersection density, road network density, traffic signal density, etc.
     + ***Diversity****:* Land use mix (e.g., proportions of Commercial, Industrial, Residential land uses; Gini\_Simpson index), commercial POI density.
     + ***Design:*** Tree density , open space area proportion, mean street slope, mean elevation.
     + ***Destination Accessibility:*** (Partially captured through POI densities).
     + ***Distance to Transit:*** Bus stop density, bus line density, metro/subway station density, parking meter density.
  2. **Micro-scale Streetscape Quality Features (SVI):** Quantifying perceptual attributes of the street environment using Street View Indices derived from street-level imagery.
     + ***Openness/Enclosure:*** Sky view index, enclosure index.
     + ***Greenness:*** Green view index / vegetation percentage.
     + ***Walkability:*** Spatial walkability / sidewalk percentage.
     + ***Other:*** Street furniture percentage, street obstacles percentage.
  3. **Socio-Demographic Features:**
     + ***Economic:*** Median household income, median property value, median rent, unemployment rate.
     + ***Demographic:*** Population density, proportions of White, Black, Asian populations, ethnic diversity.
     + ***Social:*** Homeless density, crime densities.

### **2.3 Analytical Framework**

The analysis proceeded through the following stages:

1. **Variable Filtering:** To address potential multicollinearity among independent variables, Variance Inflation Factor (VIF) scores were calculated. Variables exhibiting high VIF values (e.g., > 10) were systematically removed to ensure model stability and interpretability.
2. **Model Development:**
   * *Ordinary Least Squares (OLS) Regression:* An OLS model was initially fitted to serve as a baseline, examining linear relationships between the filtered independent variables and robotaxi crash density.
   * *Random Forest (RF) Regression:* Recognizing the potential for non-linear relationships and complex interactions, a Random Forest model was employed as the primary analytical tool. RF, an ensemble method, is robust to outliers and capable of capturing intricate patterns while providing measures of feature importance.
   * Neural network (NN):
3. **Model Evaluation and Interpretation:**
   * *Model Performance:* Models were evaluated using standard metrics including R-squared (R2), Mean Squared Error (MSE), and Mean Absolute Error (MAE).
   * *Feature Importance:* RF provides an inherent measure of feature importance based on contribution to model accuracy.
   * *SHAP (SHapley Additive exPlanations):* To interpret the RF model's predictions, the SHAP framework was utilized. SHAP values quantify the marginal contribution of each feature to individual predictions, enabling both global interpretation (e.g., overall feature importance, summary plots) and local interpretation (e.g., dependence plots showing how changes in a feature's value impact predictions).

## **3. Results**

### **3.1 Variable Filtering Results**

Following VIF analysis, 18 independent variables were retained for modeling. The descriptive statistics and final VIF values for these variables are presented below, indicating that multicollinearity was adequately addressed (all VIF < 10).

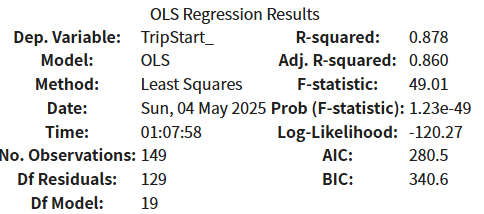
*Table 1: Descriptive Table for final independent variables*

|  | **Variable Name** | **count** | **mean** | **std** | **min** | **max** | **VIF** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Macro-scale Built Environment | |  |  |  |  |  |  |
|  | Intersection Density | 156 | 476.60 | 241.40 | 39.90 | 1567.20 | 6.10 |
|  | Road Density | 156 | 31.00 | 8.20 | 7.40 | 57.10 | 3.50 |
|  | Gini Simpson Index | 156 | 0.30 | 0.20 | 0.00 | 0.70 | 1.90 |
|  | POI Shanno Entropy | 156 | 2.90 | 0.50 | 1.00 | 4.10 | 1.40 |
|  | Parking Meter Density | 156 | 1132.50 | 2110.40 | 0.00 | 9342.30 | 5.00 |
|  | Transit Stop Density | 156 | 95.40 | 58.10 | 9.90 | 329.30 | 2.90 |
| Micro-scale Streetscape | |  |  |  |  |  |  |
|  | Visual Enclosure | 156 | 0.20 | 0.10 | 0.00 | 0.80 | 4.70 |
|  | Visual Walkability | 156 | 0.10 | 0.00 | 0.00 | 0.10 | 2.00 |
|  | Visual Vegetation | 156 | 0.10 | 0.00 | 0.00 | 0.20 | 2.50 |
|  | Visual Furniture | 156 | 0.00 | 0.00 | 0.00 | 0.00 | 1.90 |
| Socio-Demographic | |  |  |  |  |  |  |
|  | Population Density | 156 | 35865.30 | 30496.70 | 8.60 | 301895.50 | 3.60 |
|  | Whites % | 156 | 42.30 | 20.50 | 0.00 | 79.60 | 9.10 |
|  | Black % | 156 | 5.30 | 7.50 | 0.00 | 50.90 | 2.70 |
|  | Asian % | 156 | 33.90 | 19.80 | 0.00 | 100.00 | 7.90 |
|  | Ethnic Diversity | 156 | 0.60 | 0.10 | 0.00 | 0.80 | 2.30 |
|  | Felony Density | 156 | 453.00 | 699.40 | 2.30 | 6175.70 | 6.80 |
|  | Homeless Density | 156 | 326.70 | 724.00 | 2.70 | 5535.10 | 2.40 |

**3.2 Model Performance**

* OLS Model: The baseline OLS regression model achieved an R-squared of 0.878 and an adjusted R-squared of 0.86. The F-statistic was highly significant (p < 0.001), indicating that the selected variables collectively explain a substantial portion of the variance in robotaxi crash density.

Although the OLS model shows a high R-squared, it's potentially less suitable for this analysis. OLS assumes linear relationships, likely oversimplifying the complex, non-linear effects of the built environment on crashes which Random Forest can better capture. Furthermore, the table indicates Covariance Type: nonrobust, meaning the model's significance tests (p-values) could be unreliable if the underlying assumption of constant error variance is violated.

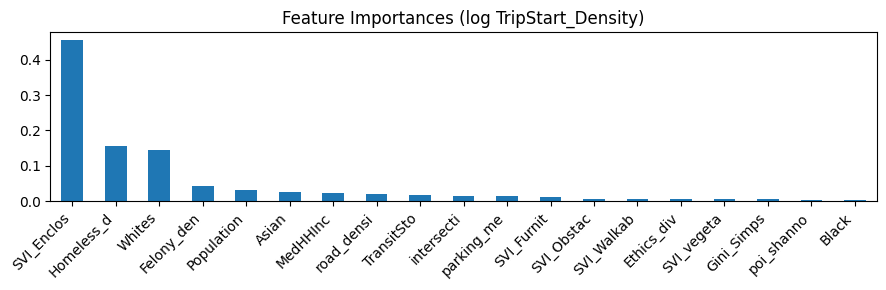


*Figure 3. OLS Regression Results*

* **Random Forest Model:** The RF model demonstrated comparable explanatory power, achieving an R-squared of 0.825 on standardized data, with an MSE of 0.3656 and MAE of 0.4437. Given its ability to handle non-linearities, subsequent interpretation focuses primarily on the RF model.

### **3.3 Feature Importance**

The feature importance scores derived from the RF model (see figure below) highlight the relative influence of different variables. SVI Enclosure (SVI\_Enclos) emerged as the most influential predictor, followed by homeless density and white population. Other variables with notable importance include felony density (Felony\_den), intersection density (intersecti), and sidewalk percentage (SVI\_sidewa).



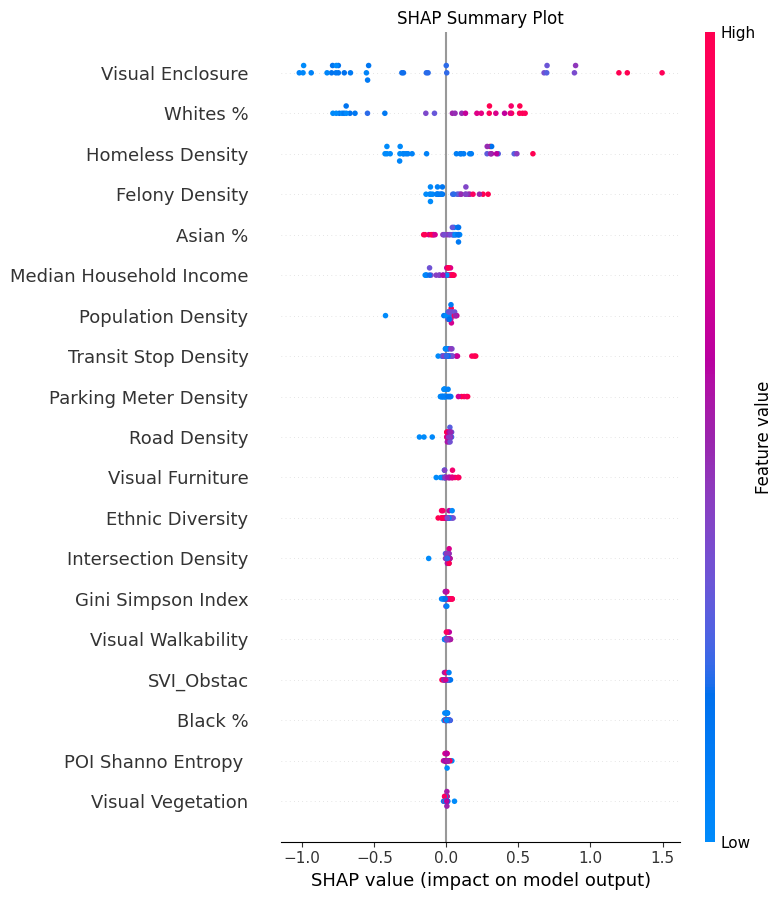
*Figure 4. RF Feature Importance*

### **3.4 SHAP Analysis: Interpreting Built Environment Impacts**

SHAP analysis provided deeper insights into how each feature influences the RF model's predictions of robotaxi crash density.

**3.4.1 SHAP Summary Plot**

Figure 5 presents the SHAP summary plot, which provides a global overview of feature importance and impact direction on the Random Forest model's prediction of Robotaxi trip start density. Features are ranked vertically based on their overall importance, calculated as the mean absolute SHAP value across all observations (census tracts). Each point on the plot represents a single census tract; its horizontal position indicates the SHAP value (the impact of that feature on the model's output for that specific tract), while its color represents the feature's value (red for high values, blue for low values). Visual Enclosure emerges as the most influential predictor, with its high values (red dots) consistently pushing the predicted trip demand higher (large positive SHAP values). Following Visual Enclosure, Homeless Density, Whites %, Felony Density, and Asian % are the next most important features, generally showing that higher values tend to increase the predicted trip demand, although the distribution of SHAP values indicates complex interactions. Other notable positive contributors to predicted demand include Population Density and Transit Stop Density. Conversely, features like the Gini Simpson Index (land use diversity) exhibit a negative association, where higher values tend to decrease predicted trip demand. Features lower down the plot have progressively less impact on the model's predictions, with SHAP values clustering closer to zero.



*Figure 5. Results of SHAP feature importance analysis.*

**3.4.2 Non-linear Relationships (from SHAP Dependence Plots):** Examining SHAP dependence plots reveals specific patterns:

* 1. **Macro-scale Built Environment (5D) Impacts** (Figure 5)
     + *Density*

Intersection Density (X1) exhibits a strong positive relationship, where predicted demand increases sharply at lower densities and continues to rise, though more slowly, at higher densities. Similarly, Road Density (X2) shows a generally positive trend, with demand increasing with density but flattening out at very high levels. Parking Meter Density (X5) displays a more complex pattern; its impact is minimal at low densities but becomes strongly positive, indicating higher predicted trip demand, after exceeding a threshold (around 2000-4000 meters/area), likely reflecting concentrated commercial activity.

* + - *Diversity*

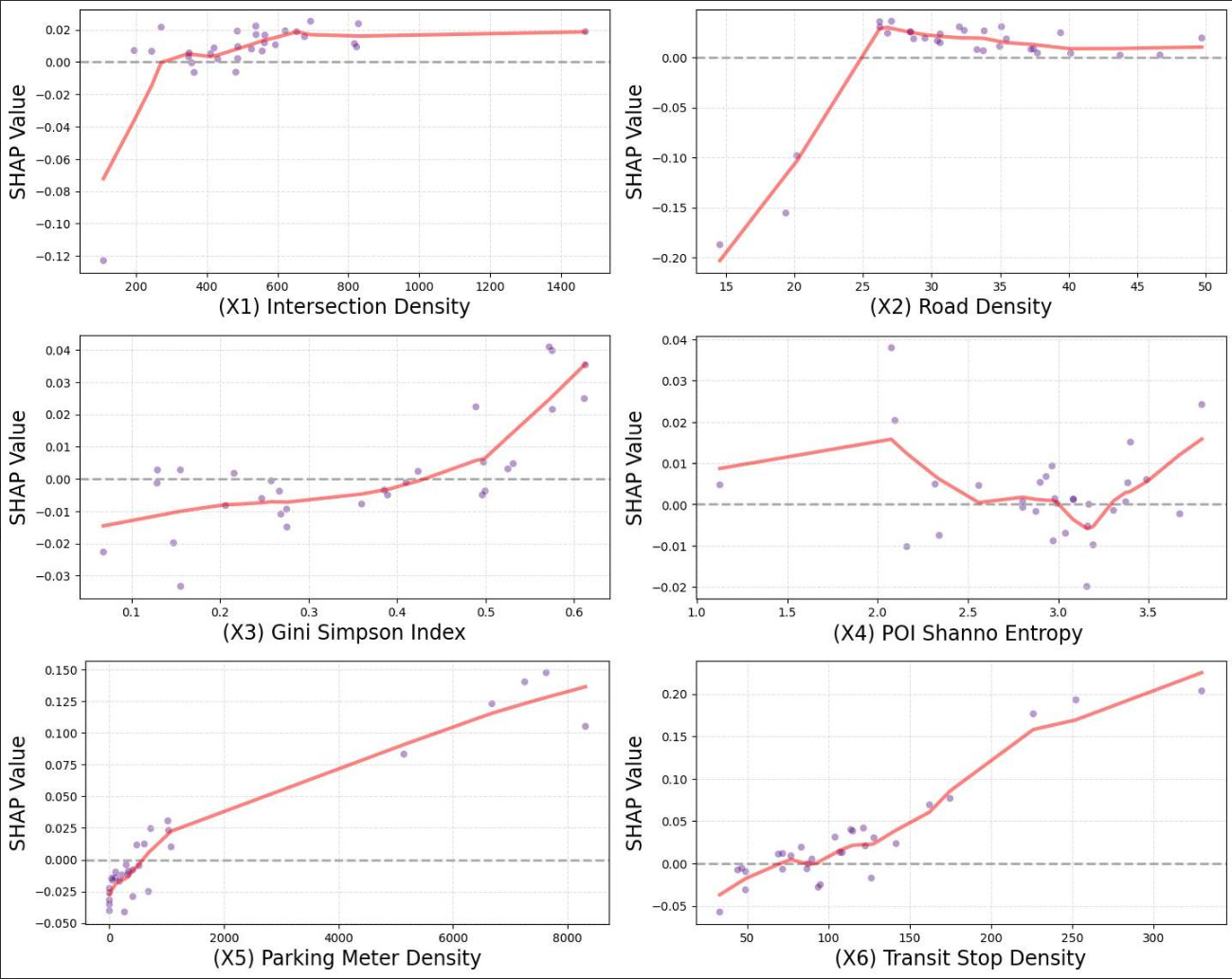
The Gini-Simpson Index (X3), representing land use mix, demonstrates a consistent negative non-linear relationship; as diversity increases, the predicted trip density decreases, suggesting that more mixed-use environments are associated with lower Robotaxi demand. Conversely, POI Shannon Entropy (X4), indicating the variety of points of interest, has a relatively flat relationship with predicted demand, showing only a slight increase at the highest entropy levels.

* + - *Design*

Intersection Density (X5) shows a strong positive association, where predicted risk increases sharply at lower densities and continues to rise, albeit more slowly, at higher densities. Open Space (X6) exhibits a slightly positive, near-linear trend, suggesting marginally higher predicted risk with more open space, which might be counter-intuitive and warrants further investigation. Elevation (X7) displays a distinct non-linear pattern, with predicted risk being lowest at very low and very high elevations, peaking at moderate elevation levels (around 50-100 units). Lastly, Street Slope (X8) demonstrates a relationship that is relatively flat but shows a slight overall positive trend, suggesting that steeper slopes are associated with a marginally higher predicted crash density.

* + - *Accessibility/ Distance to Transit:*

Transit Stop Density (X6) displays a clear non-linear positive association. The impact on predicted trip demand is negligible at low densities but increases steadily and significantly beyond a threshold of approximately 100-150 stops per area unit, suggesting Robotaxis are frequently used in areas well-served by public transport, potentially for first/last-mile connections or as an alternative.



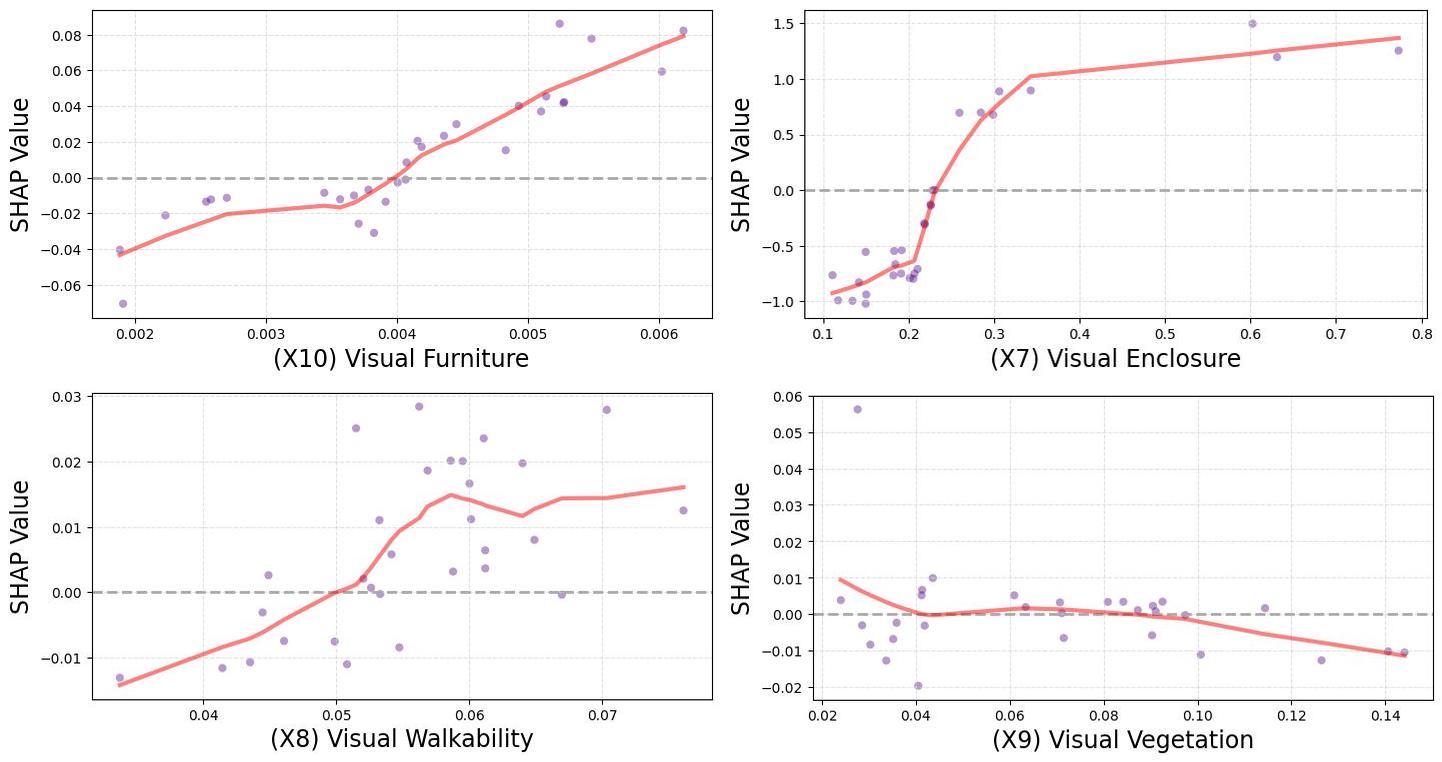
*Figure 6. Non-linear relationships between trip density and macro-scale built environment variables.*

* 1. **Micro-scale Streetscape Quality (SVI) Impacts** (Figure 6)
     + *Enclosure/Constraint*

As identified in the feature importance analysis, Visual Enclosure (X7) demonstrates the strongest impact among SVI features, exhibiting a distinct positive sigmoidal relationship. At low enclosure levels (below approximately 0.2), its influence on predicted trip demand is minimal. However, the predicted demand increases sharply and substantially as enclosure rises between approximately 0.2 and 0.4, after which the positive impact plateaus at a high level. This suggests a critical threshold where the 'street canyon' effect, often associated with dense urban cores, becomes strongly linked with higher Robotaxi trip generation.

* + - *Pedestrian Environment & Greenness:*

Features related to the pedestrian environment and greenness show more varied and generally weaker non-linear effects. Visual Walkability (X8), representing visible sidewalk space, displays a slightly positive overall trend, indicating marginally higher predicted trip demand in areas with more walking space, potentially reflecting correlation with high-activity areas. Visual Furniture (X10) shows a complex pattern, being negatively associated with demand at very low levels but becoming positively associated at moderate to high levels. Conversely, Visual Vegetation (X9) exhibits a generally negative association; predicted trip demand tends to decrease as visible vegetation increases, particularly beyond 0.04, suggesting lower demand in greener, potentially less urbanized settings.



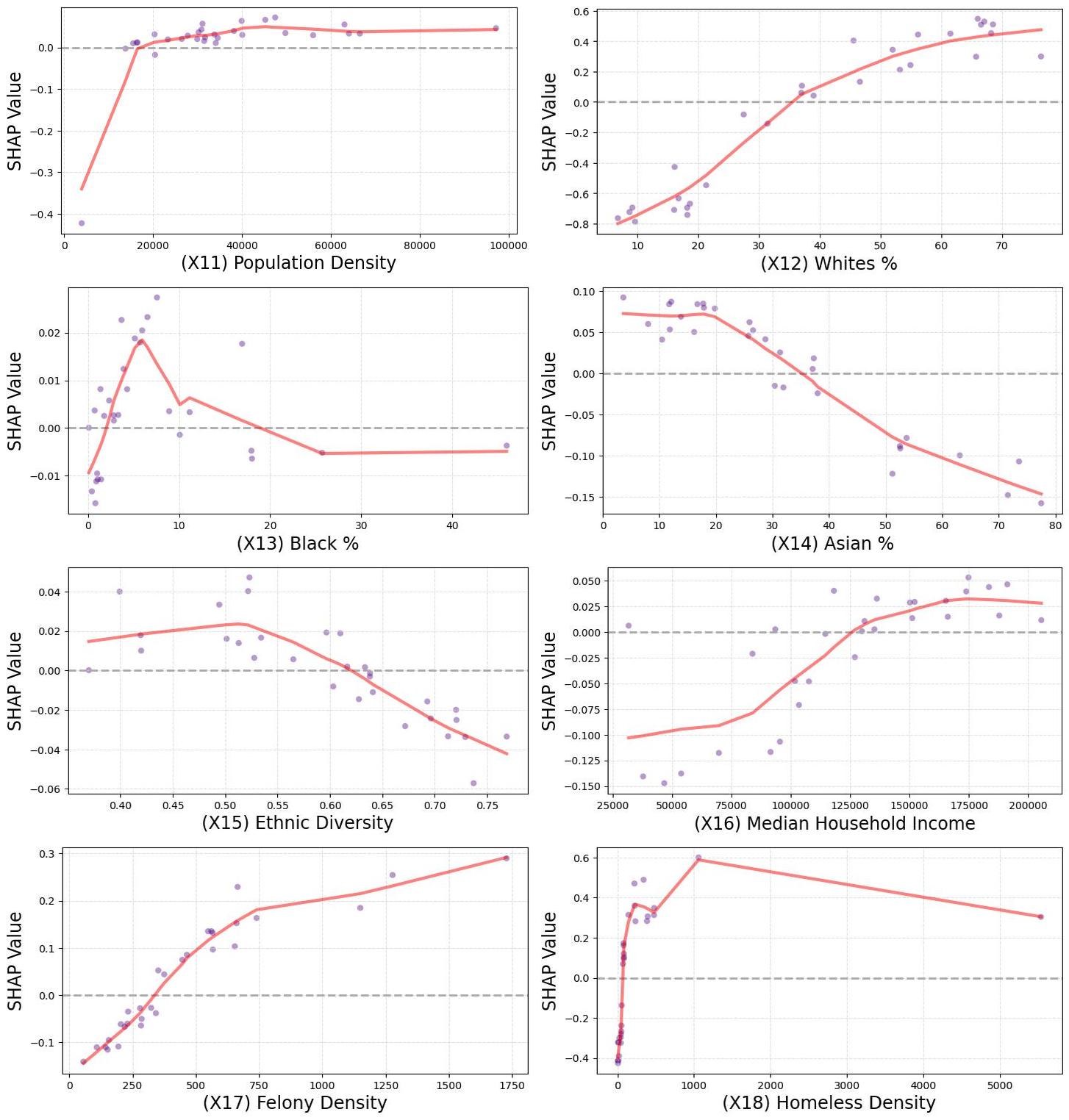
*Figure 7. Non-linear relationships between trip density and streetscape variables*

* 1. **Socio-Demographic Impacts**
     + *Demographic*

The impacts of demographic variables on predicted Robotaxi trip density reveal complex and non-linear patterns (Figure 8, plots X11-X15). Population Density (X11) exhibits a strong positive association, where predicted trip demand increases sharply at lower densities before the rate of increase slows down considerably at higher densities, indicating diminishing marginal returns but confirming its role as a major driver of trip generation. Ethnic composition variables display more intricate relationships. Whites % (X12) shows a non-linear, somewhat S-shaped positive curve initially, flattening out and potentially peaking at mid-to-high percentages (around 60-70%), suggesting the highest positive impact on demand in these tracts. Asian % (X14) shows a generally negative relationship. It has a slightly positive impact at lower percentages (below ~25-30%), but this trend reverses sharply, becoming strongly negative as the percentage increases further. In contrast, Black % (X13) demonstrates a weaker, more complex pattern, peaking with a small positive impact at very low percentages (around 5-10%) and then declining, suggesting a slight negative association at higher percentages. Finally, Ethnic Diversity (X15) exhibits a relatively consistent negative trend, indicating that census tracts with higher ethnic diversity are associated with slightly lower predicted Robotaxi trip demand. These varied relationships underscore the nuanced ways demographic factors correlate with Robotaxi usage patterns, likely intertwined with other socio-economic and geographic characteristics.

* + - *Economic/Social*

Median Household Income (X16) shows a very weak, mostly flat relationship with predicted trip density, with SHAP values remaining close to zero across the income spectrum, suggesting minimal direct influence within this model after accounting for other factors. In contrast, social indicators like Felony Density (X17) and Homeless Density (X18) both show strong positive non-linear trends. Predicted trip demand increases substantially as these densities rise from low levels, but the rate of increase slows considerably at higher densities, eventually flattening out. This suggests that while areas with higher crime or homelessness density are associated with increased Robotaxi demand, potentially due to their correlation with central, high-activity locations, the marginal impact diminishes at very high levels.



*Figure 8. Non-linear relationships between trip density and socio-demographic variables.*

## **4. Discussion**

This study investigated the influence of built environment factors on robotaxi crash frequency in San Francisco using DMV collision data and a range of environmental and socio-demographic variables. By employing Random Forest modeling and SHAP interpretation, we identified key attributes associated with predicted crash risk.

### **4.1 Summary of Key Findings**

* **Density and Centrality Indicators Strongly Drive Predicted Trip Demand**

Multiple density metrics (Population, Intersection, Road, Parking Meter, Transit Stop densities) and social indicators often associated with central, high-activity locations (Felony Density, Homeless Density) exhibit strong positive, albeit often non-linear, associations with predicted Robotaxi trip start density. This aligns well with classic central place theory and accessibility principles, where areas with high concentrations of people, activities, and infrastructure naturally generate and attract more travel demand. The finding confirms that Robotaxi demand distribution largely follows traditional urban spatial patterns. However, the observed non-linearities, often showing diminishing marginal effects at very high densities, suggest that demand concentration does not increase indefinitely with density, adding a layer of complexity to simple density-demand assumptions.

* **Visual Enclosure is the Most Dominant Predictor of Trip Demand**

The micro-scale streetscape feature 'Visual Enclosure' emerged as the single most important predictor in the model. High enclosure, characteristic of dense urban cores ("street canyons"), is strongly linked to significantly higher predicted trip demand. This likely reflects its role as a strong proxy variable, effectively capturing the concentration of activities and trip origins/destinations typical of these high-density environments. This highlights that beyond macro-scale density, street-level morphology significantly reflects travel demand concentration.

* **Land Use Diversity Shows an Inverse Relationship with Predicted Demand**

In contrast to density measures, higher land use diversity (measured by the Gini-Simpson Index) consistently correlates with lower predicted Robotaxi trip start density. This finding resonates with theories on mixed-use development, which propose that functionally diverse neighborhoods can satisfy more needs locally, potentially reducing the generation of external trips, including those potentially served by Robotaxis. It suggests demand might be more dispersed in highly mixed-use areas compared to specialized zones.

* **Socio-Demographic Influences are Complex and Nuanced**

While Population Density is a major driver, other socio-demographic variables display intricate non-linear patterns. Ethnic composition variables (e.g., a peak positive impact at mid-to-high levels for Whites %, generally negative for Asian %) likely reflect underlying socio-economic factors, lifestyle preferences, or correlations with other built environment features rather than direct causality. Notably, Median Household Income showed minimal direct influence on predicted trip demand in this model after accounting for other factors. These complex relationships underscore the need for cautious interpretation, avoiding simplistic links between demographic labels and travel behavior and recognizing the interplay with unmeasured variables.

### **4.2 Policy and Practical Implications**

* **Manage Demand in High-Density/High-Enclosure Zones**

The strong positive correlation between predicted trip demand and various density metrics (Intersection, Road, Population, Transit Stop, Parking Meter densities), coupled with the dominant role of high Visual Enclosure, confirms that dense urban cores and "street canyons" are primary Robotaxi demand hotspots. Planners must proactively manage these areas to accommodate Robotaxi operations without hindering other modes or exacerbating congestion. This includes:

* Curbside Management: Designating sufficient and strategically located pick-up/drop-off (PUDO) zones specifically for Robotaxis and other TNCs, potentially using dynamic pricing or time restrictions.
* Traffic Flow Optimization: Adjusting signal timing and lane configurations in high-demand corridors to account for Robotaxi movements, especially during peak hours.
* Infrastructure Design: Considering street designs that balance vehicle movement with pedestrian safety and comfort, particularly in areas with high Visual Walkability which, despite potentially higher demand, need prioritized pedestrian infrastructure.
* **Leverage Land Use Planning for Demand Management**

The finding that higher land use diversity (Gini-Simpson Index) is associated with lower predicted Robotaxi demand supports planning strategies aimed at reducing vehicle dependency. Planners can:

* Promote Mixed-Use Development: Continue encouraging developments that mix residential, commercial, and recreational uses to shorten trip distances and potentially moderate the need for vehicular trips, including Robotaxis.
* Integrate Robotaxis with Local Accessibility: Ensure that land use plans consider how Robotaxis can complement local accessibility in diverse neighborhoods, potentially serving niche trips not well-covered by walking or transit, rather than primarily facilitating longer commutes from specialized zones.
* **Address Equity and Accessibility**

The complex relationships observed between predicted Robotaxi trip density and socio-demographic variables demand careful consideration through an equity lens. The findings indicate potential disparities in service provision or uptake across racial groups. Based on the analysis, current Robotaxi usage appears concentrated in areas with specific racial demographics (higher White, lower Asian presence) and does not show strong penetration in areas with higher Black populations, potentially indicating service inequality or differential access. Planners must:

* Monitor Service Distribution: Track Robotaxi service areas and usage patterns across different neighborhoods and demographic groups to ensure equitable access and avoid creating new mobility divides.
* Promote Equitable Service Design: Work with operators to ensure service design considers the needs of diverse communities. This could involve targeted outreach, multilingual support, accessible vehicle options, and ensuring service reliability in underserved areas. Policies may be needed to incentivize or mandate minimum service levels across the entire city.
* Consider Vulnerable Populations: While high Homeless Density and Felony Density correlate with high demand (likely due to location), planners need to consider the specific mobility needs and safety concerns of vulnerable populations in these areas, ensuring Robotaxi integration doesn't negatively impact them.
* **Integrate Robotaxis with Public Transit**

The positive association between Transit Stop Density and predicted Robotaxi demand suggests a potential for complementary roles. Planners should:

* Facilitate First/Last-Mile Connections: Design PUDO zones near major transit hubs (bus stops, metro stations) to encourage Robotaxis as feeders to the public transit network.
* Monitor Competition: Continuously monitor data to understand whether Robotaxis are complementing or competing with transit routes and adjust policies accordingly to protect transit viability.

## **Appendices**

**Appendix A: Variable Descriptions**

| Category 1 | Category 2 | Variables | Description | Name in CSV | Data Source | mean | std |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sociodemographics (SD) |  | Median household income | The total income of households | MedHHInc |  | 150413.6274 | 61399.4068 |
|  | White Americans proportion | White Americans proportion | Whites |  | 43.7559 | 21.5279 |
|  | African Americans proportion | African Americans proportion | Black |  | 4.7649 | 8.0432 |
|  | Asian Americans proportion | Asian Americans proportion | Asian |  | 33.9921 | 21.3277 |
|  | Native American proportion | Native American proportion | Native\_Ame |  | 0.5649 | 2.0972 |
|  | Ethnic diversity | Simpson’s diversity index of races | Ethics\_div |  | 0.5772 | 0.1381 |
|  | Median Rent | Median house rent per unit | MedRent |  | 2368.388 | 789.3619 |
|  | Population density | residential population / total area | Population |  | 13543.7982 | 12361.6683 |
|  | Property value | The property value was at the CT level | MedHVal |  | 1434334.441 | 377868.0486 |
|  | Homeless density | Total homeless incidents divided by total area | Homeless\_d |  | 479.9015 | 1205.658 |
|  | Felony density | Total felony incidents divided by total area | Felony\_den |  | 650.6372 | 1358.7998 |
|  | Misdemeanor density | Total misdemeanor incidents divided by total area | Misdemeano |  | 1043.1205 | 2711.4294 |
|  | Violation density | Total violation incidents divided by total area | Violation. |  | 481.167 | 1635.7148 |
|  | Homeless density | Total homeless incidents divided by total area | Homeless\_d |  | 479.9015 | 1205.658 |
|  | Unemployment rate | Rate of unemployed people | Unemployme |  | 5.4139 | 6.1345 |
| Macro-scale built environment (BE) | Density | Building density | Building area / total area | Building\_d |  | 0.3597 | 0.1091 |
| Intersection density | Number of intersections / total area | intersecti |  | 507.0996 | 317.7923 |
| Road network density | Length (mi) of roads / total area | road\_densi |  | 32.1464 | 10.4633 |
| Traffic signal density | Number traffic signal / total area | traffic\_si |  | 55.3139 | 73.521 |
| Diversity | Commercial POI density | Number of commercial-related points of interest (POIs) / total area | commerci\_1 |  | 4978.3824 | 6126.9465 |
| Public proportion | Proportion of land use designated as green area | Public |  | 10.1925 | 16.8866 |
| Commercial land use proportion | Proportion of land use designated as commercial area | Commercial |  | 4.7189 | 17.5151 |
| Industrial land use proportion | Proportion of land use designated as industrial area | Industrial |  | 1.7151 | 8.9932 |
| Residential land use proportion | Proportion of land use designated as residential area | Residentia |  | 64.1538 | 33.7559 |
| Mixed Use land use proportion | Proportion of land use designated as Mixed Use | Mixed.Use |  | 19.2197 | 26.9177 |
| Gini\_Simpson | 1−HHI (Gini-Simpson) | Gini\_Simps |  | 0.2853 | 0.1999 |
| Design | Tree density | Number of tree / total area | tree\_densi |  | 6114.1665 | 3371.0951 |
| Open space area | Open space area/ total area | open\_space |  | 6.7496 | 16.2824 |
| Street slope | The average slope degree within each CT | mean\_slope |  | 5.9028 | 3.3147 |
| Elevation | The average elevation degree within each CT | mean\_eleva |  | 57.3088 | 39.2553 |
| Distance to transit & Accessibility | Bus stop density | Number of bus stops / total area | Bus.Stop.D |  | 103.1959 | 89.8999 |
| Bus line density | Length (mi) of bus lines / total area | Bus.Line.D |  | 27.2088 | 36.8634 |
| Metro/subway station density | Number of metro/subway stops / total area | Metro.Stop |  | 17.4152 | 45.2954 |
| Metro/subway line density | metro/subway lines (mi) / total area | Metro.Line |  | 3.6615 | 13.2444 |
| Transit stop density | Number of bus and metro stops / total area | TransitSto |  | 120.6111 | 113.7594 |
| Parking Meter Density | Number of parking meters / total area | parking\_me |  | 1723.4671 | 2794.0202 |
| Eye-level street environment (SE) |  | Sky view index | The percentage of sky in the images | SVI\_sky | Google SVIs | 0.3507 | 0.0561 |
|  | Enclosure index | The percentage of pixels indicating walls, fences, banister, and rails in buildings | SVI\_Enclos | Google SVIs | 0.277 | 0.1552 |
|  | Green view index | The pixel percentage of vegetation pixels in images | SVI\_vegeta | Google SVIs | 0.0676 | 0.0359 |
|  | Spatial walkability | The pixel percentage of walking space (sidewalks, paths, stairs, and stairway) in roadways | SVI\_sidewa | Google SVIs | 0.023 | 0.0052 |
|  | Street furniture | The percentage of street furniture ( benches, streetlights, traffic light, pole) in the images | SVI\_Furnit | Google SVIs | 0.0039 | 0.0012 |
|  | Street obstacles | The percentage of pixels indicating walls, fences, banister, and rails in buildings | SVI\_Obstac | Google SVIs | 0.0044 | 0.0035 |
|  |  | Visual motorization index | The percentage of pixels indicates roads, vehicles and traffic lights and traffic signs | VMI | Google SVIs | 0.4262 | 0.0144 |