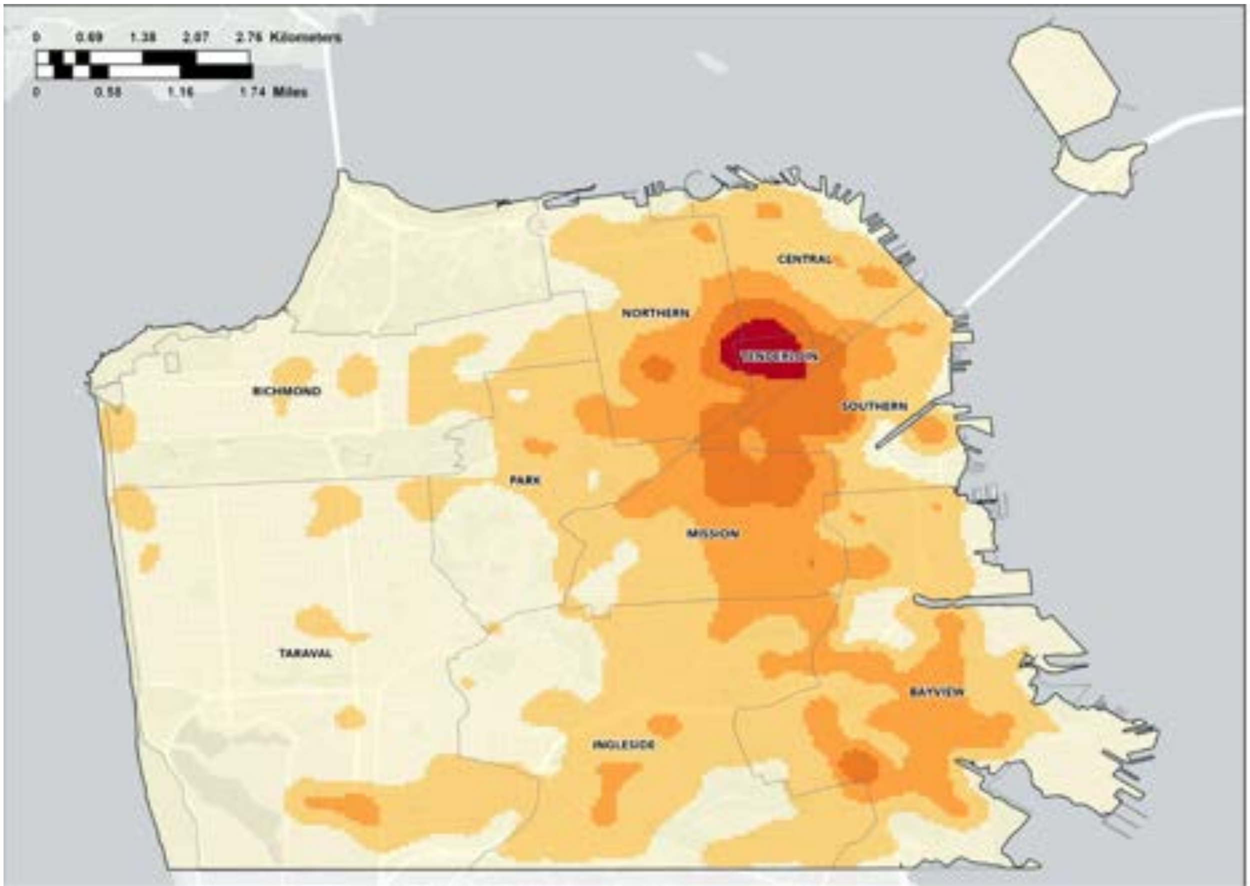
GIS RESEARCH REPORT · SAN FRANCISCO, CALIFORNIA



San Francisco Motor Vehicle Crime Spatial Analysis

Spatial distribution, administrative aggregation, density, hot spots, and road proximity · 2021–2025

Alex Tour-Sarkissian

Independent GIS Research · ArcGIS Pro 3.6

Project Overview

This personal GIS research project investigates the spatial distribution of crime incidents across San Francisco using ArcGIS Pro. The study applies spatial analysis techniques to identify crime patterns, detect statistically significant hotspots, and examine the relationship between crime and selected geographic features.

The project follows a complete GIS research workflow, including data collection, data preparation, spatial analysis, map production, statistical interpretation, and report writing. The results are intended to demonstrate practical GIS skills and provide insights that could support public safety planning and crime prevention strategies.

Table of Contents

List of Figures

List of Tables

Abstract

This independent GIS research project examines the spatial distribution of 15,149 mapped motor vehicle crime incidents in San Francisco from 2021 through 2025. The completed workflow uses point distribution mapping, Spatial Join aggregation by police district and neighborhood, Kernel Density estimation, Getis-Ord Gi* Hot Spot Analysis, and Near analysis relative to major roads. All distance- and density-based analysis was conducted in NAD 1983 UTM Zone 10N (EPSG:26910). Kernel Density was generated using a 50-meter cell size, 500-meter search radius, planar method, and square-kilometer area units. Results show that Bayview recorded the highest police-district incident total (2,599), while the strongest continuous density and statistically significant hot-spot patterns occur in central and northeastern San Francisco, especially around Tenderloin and adjacent portions of Central, Southern, and Mission. Road-proximity analysis shows that 58.7% of mapped incidents occurred within 250 meters of the nearest major road. These findings identify spatial concentration and association rather than causal relationships. The final deliverable is a six-map analytical series designed to demonstrate an end-to-end GIS research workflow, spatial-statistical interpretation, cartography, and technical reporting.

1. Introduction

1.1 Background and Context

Motor vehicle crime is an important urban crime category that can vary considerably across geographic locations. The occurrence of these incidents may be influenced by population density, land use, transportation infrastructure, commercial activity, parking areas, and accessibility. Understanding where motor vehicle crimes occur and whether they form identifiable spatial patterns can provide useful information for urban safety planning and crime prevention.

Geographic Information Systems (GIS) provide effective tools for examining the spatial characteristics of crime. By representing crime incidents as geographic features, GIS can be used to visualize incident locations, compare crime levels across administrative areas, identify areas of high concentration, detect statistically significant spatial clusters, and examine relationships between crime incidents and geographic features such as major roads.

This study focuses on motor vehicle crime incidents recorded in San Francisco between 2021 and 2025. The analysis uses ArcGIS Pro to investigate the spatial distribution and concentration of incidents through multiple complementary spatial analysis techniques.

1.2 Problem Statement

Although crime records provide information about individual incidents, a list of incident locations alone does not clearly reveal broader spatial patterns. Motor vehicle crime may be unevenly distributed across the city, with certain police districts, neighborhoods, or locations experiencing higher concentrations than others.

There is therefore a need to examine the spatial characteristics of motor vehicle crime using GIS-based analytical methods. In particular, understanding the distribution, density, statistically significant hot spots, and proximity of incidents to major roads can provide a more comprehensive view of the geographic pattern of motor vehicle crime in San Francisco.

1.3 Aim and Objectives

Aim

The main aim of this study is to analyze the spatial distribution and geographic patterns of motor vehicle crime in San Francisco from 2021 to 2025 using GIS and spatial analysis techniques in ArcGIS Pro.

Objectives

The specific objectives are to:

1. Map the spatial distribution of motor vehicle crime incidents across San Francisco.
2. Examine the distribution of motor vehicle crime by police district.
3. Analyze motor vehicle crime patterns by neighborhood.
4. Identify areas of high crime density using Kernel Density Estimation.
5. Detect statistically significant hot and cold spots using Getis-Ord Gi* Hot Spot Analysis.
6. Examine the spatial proximity of motor vehicle crime incidents to major roads.
7. Compare the findings from different spatial analysis methods to identify consistent patterns of crime concentration.
8. Demonstrate a complete GIS workflow for urban crime spatial analysis.

1.4 Research Questions

The study is guided by the following research questions:

1. **How are motor vehicle crime incidents spatially distributed across San Francisco between 2021 and 2025?**
2. **Which police districts and neighborhoods contain the highest concentrations of motor vehicle crime incidents?**
3. **Where are the major areas of motor vehicle crime density identified through Kernel Density Estimation?**
4. **Where do statistically significant hot and cold spots of motor vehicle crime occur?**
5. **What is the spatial relationship between motor vehicle crime incidents and major roads?**
6. **Do different spatial analysis methods identify consistent areas of motor vehicle crime concentration?**

1.5 Scope of the Study

This study focuses on motor vehicle crime incidents recorded within San Francisco during the period **2021–2025**. The analysis is conducted using **ArcGIS Pro** and focuses primarily on the spatial characteristics of reported incidents.

The study includes six main analytical components: incident distribution, administrative aggregation by police district and neighborhood, Kernel Density Estimation, Getis-Ord Gi* Hot Spot Analysis, and proximity analysis relative to major roads.

The study is limited to the available crime and geographic datasets and therefore represents patterns within the recorded data rather than all unreported or unrecorded incidents. The analysis focuses on identifying and interpreting spatial patterns and does not attempt to establish direct causal relationships between motor vehicle crime and individual geographic factors.

1.6 Significance of the Study

The study demonstrates how GIS can be applied to analyze urban crime from multiple spatial perspectives. Combining descriptive mapping, administrative aggregation, density estimation, statistical hot spot analysis, and road proximity analysis provides a more comprehensive understanding of where motor vehicle crime is concentrated.

The findings may provide useful spatial insights for public safety planning, crime prevention, resource allocation, and further urban crime research. From a GIS perspective, the project also demonstrates practical skills in spatial data preparation, geoprocessing, spatial statistics, cartographic visualization, and interpretation of geographic patterns.

1.7 Report Structure

This report is organized into seven main sections. **Chapter 1** introduces the study, including its background, problem statement, aim, objectives, research questions, scope, and significance. **Chapter 2** describes the study area and its geographic and administrative characteristics. **Chapter 3** presents the datasets, data preparation procedures, GIS workflow, and spatial analysis methods used in the study. **Chapter 4** presents and summarizes the results of the spatial analyses. **Chapter 5** discusses and interprets the major findings and their implications. **Chapter 6** provides the overall conclusions and recommendations for future analysis. **Chapter 7** presents the references and data sources, followed by **Appendix A**, which summarizes the analytical parameters used in the GIS analysis.

2. Study Area

2.1 Geographic Context of San Francisco

San Francisco occupies a compact urban peninsula on the northern end of the San Francisco Peninsula, bounded by the Pacific Ocean to the west and San Francisco Bay to the north and east. The city combines a dense central business and residential core, extensive eastern waterfront and industrial areas, lower-density western residential districts, steep topography, and a highly connected street network. These geographic characteristics make San Francisco a useful study area for examining how urban incident patterns vary across administrative units, continuous space, and transportation corridors.

The six final maps use a consistent citywide extent so that patterns can be compared across analytical methods. The San Francisco city boundary is used as the principal study-area reference, while police districts, neighborhood polygons, and major roads provide additional spatial frameworks for aggregation and proximity analysis.

2.2 Administrative and Police District Structure

The police-district analysis uses ten San Francisco Police Department districts represented in the project boundary layer: Bayview, Central, Ingleside, Mission, Northern, Park, Richmond, Southern, Taraval, and Tenderloin. These districts provide operationally meaningful administrative units for comparing mapped incident totals. DataSF identifies the current police-district layer as SFPD districts in effect since July 19, 2015.

Police-district totals are useful for broad comparison, but they should not be interpreted as standardized crime rates. Districts vary in geographic size, land use, street density, vehicle activity, residential population, and other characteristics. For that reason, district aggregation is interpreted alongside neighborhood aggregation, Kernel Density, and Getis-Ord Gi* results rather than as a standalone measure of risk.

2.3 Road Network and Urban Characteristics

San Francisco contains a dense street system ranging from local residential streets to arterials, major streets, freeway segments, and ramps. The project road data were prepared into a major-road reference layer for proximity analysis. Rather than treating every street equally, the final analysis focuses on the prepared major-road network as a transportation framework against which incident distances can be measured.

The road network is relevant to this study because motor vehicle crime occurs in a transportation environment shaped by vehicle movement, parking, curb access, commercial activity, and street connectivity. The proximity analysis does not test those mechanisms directly; it evaluates only the geographic distance between mapped incidents and the nearest major road.

2.4 Crime and Motor Vehicle Crime Context

The source incident dataset is derived from San Francisco Police Department incident reports published through DataSF. The project subset contains 15,149 mapped motor vehicle crime records with incident dates from 2021 through 2025. The working spreadsheet includes fields for incident date and year, incident type and description, resolution, intersection, Centerline Network Number (CNN), police district, and analysis neighborhood.

DataSF notes that published incident locations are mapped to nearby intersections to protect privacy rather than representing exact event coordinates. DataSF also notes that the intersection-mapping process changed on April 24, 2024, which may introduce slight reporting irregularities in analyses spanning that date. This limitation is relevant to the interpretation of fine-scale point, density, and proximity patterns and is addressed again in Chapter 5.

2.5 Study Area Extent and Spatial Units

The study area is limited to the mapped San Francisco municipal extent used in the project. Multiple spatial units are intentionally used because each supports a different analytical question. Individual incident points preserve event locations; police districts provide broad administrative aggregation; neighborhood polygons provide finer local aggregation; a fishnet grid supports Getis-Ord Gi* analysis; a raster grid supports Kernel Density; and road lines provide the reference geometry for Near analysis.

Using multiple spatial representations is central to the design of the study. A high count within a large polygon, a high density value in a raster cell, and a statistically significant Gi* hot spot are related concepts but are not equivalent. The report therefore treats the six map outputs as complementary evidence rather than interchangeable representations.

3. Data and Methods

3.1 Data Sources and Data Description

The analysis combines a core motor vehicle crime incident dataset with municipal boundary, police-district, neighborhood, and road reference layers. Derived analytical layers were created for police-district aggregation, neighborhood aggregation, Kernel Density, crime-analysis fishnet cells, Getis-Ord Gi* output, road buffers, and road-distance classes. Table 1 summarizes the principal datasets used in the completed six-map workflow.

Table 1. Data Sources and Dataset Description

Dataset	Geometry	Role in Analysis	Primary Source
Motor vehicle crime incidents, 2021-2025	Point	Core event dataset for all analyses	San Francisco Police Department / DataSF
San Francisco city boundary	Polygon	Study extent, mask, clipping, and reference	San Francisco Open Data / DataSF
Police districts	Polygon	Administrative aggregation and map context	Current Police Districts / DataSF
Neighborhood boundaries	Polygon	Fine-scale aggregation and local pattern comparison	Project neighborhood boundary layer / DataSF-derived geography
Major roads	Line	Near analysis and supporting buffers	San Francisco street centerline / project road source
Crime analysis fishnet	Polygon grid	Spatial units for Getis-Ord Gi* analysis	Derived in ArcGIS Pro
Kernel Density raster	Raster	Continuous density representation	Derived in ArcGIS Pro

3.2 Motor Vehicle Crime Incident Data (2021-2025)

The working incident dataset contains 15,149 records covering the five-year period 2021-2025. The project filtered the broader police incident source to a motor vehicle crime subset and retained the geographic and descriptive fields needed for analysis. Key fields in the working spreadsheet include incident date and year, incident type, incident description, resolution, intersection, CNN, source police district, and source analysis neighborhood.

The five-year study period was selected to provide a sufficiently large sample for comparing spatial patterns while still representing a recent period. Combining multiple years reduces the chance that the final maps are dominated by short-term fluctuations in a single year. At the same time, the analysis should be interpreted as an aggregate 2021-2025 spatial pattern rather than a temporal trend analysis; changes from year to year were not modeled in the final six-map series.

3.3 Administrative Boundaries and Police District Data

Police-district polygons were used as target features in a Spatial Join so that each mapped incident intersecting a district could contribute to that district's Join_Count value. This approach intentionally uses spatial location rather than relying only on the source police-district attribute stored in the incident table. The final choropleth therefore represents the project's spatial assignment of mapped incident points to district polygons.

A quality-control comparison found that the ten district values displayed on the final map sum to 15,143, six fewer than the overall 15,149 incident total. The report treats this difference as an unresolved QA item rather than assuming a cause. It may reflect boundary placement, unmatched geometry, or another spatial-join condition and should be checked in the final geodatabase if exact district reconciliation is required.

3.4 Neighborhood Boundary Data

Neighborhood polygons were used for a second Spatial Join to provide a finer geographic view of mapped incident counts. The neighborhood map contains substantially more spatial units than the ten police districts, allowing localized differences to become visible that would otherwise be absorbed into larger administrative areas.

The source incident spreadsheet also includes an analysis-neighborhood attribute. For descriptive context, Table 3 later in this report summarizes that source field. The final neighborhood choropleth, however, is interpreted primarily from the spatial-join output shown in Figure 3. This distinction is important because source-record neighborhood coding and polygon-based spatial assignment can differ at boundaries or where different neighborhood definitions are used.

3.5 Major Road Network Data

The project road source was prepared into a major-road feature layer used as the reference geometry for Near analysis. The road layer was also used to create 100 m, 250 m, and 500 m buffers for cartographic context. The final map emphasizes the relationship between incident points and major transportation corridors without implying that other streets are unimportant to motor vehicle crime.

3.6 Data Preparation and Quality Control

Data preparation followed a structured ArcGIS Pro workflow. Authoritative base layers were separated from derived analytical outputs, and new analysis products were stored in a dedicated research geodatabase. A core motor vehicle crime feature class was retained as the source layer, while tools that modify input features, such as Near, were run on derived analysis copies. This approach preserved the original analytical input and reduced the risk of accidental field changes to the core data.

- Reviewed field names, null values, and geographic attributes before analysis.
- Standardized distance- and density-based analysis in a projected coordinate system with meter units.
- Created a dissolved San Francisco boundary for masking and map extent control.
- Used consistent final map layouts, legends, north arrows, scale bars, and map details to support cross-map comparison.

- Checked analytical outputs against the overall incident total and identified unreconciled district and neighborhood-field differences as QA notes rather than silently forcing totals to match.

3.7 Coordinate Reference System and GIS Environment

All distance- and density-based analysis was conducted in **NAD 1983 UTM Zone 10N (EPSG:26910)**. This projected coordinate system uses meters and is appropriate for local-scale analysis in the San Francisco region. A projected coordinate system was necessary for the 50 m Kernel Density cell size, 500 m search radius, 100/250/500 m road buffers, and Near-distance categories.

ArcGIS Pro 3.6 was the principal analysis and cartographic environment, with QGIS included in the broader project toolset. ArcGIS Pro geoprocessing tools used in the completed workflow include Spatial Join, Kernel Density, Hot Spot Analysis (Getis-Ord Gi*), Near, Buffer, field calculation, and Summary Statistics.

3.8 Spatial Analysis Framework

The analytical framework progresses from descriptive mapping to increasingly transformed spatial representations. First, incident points are mapped directly. Second, incidents are aggregated to police districts and neighborhoods. Third, Kernel Density produces a continuous surface independent of administrative boundaries. Fourth, Getis-Ord Gi* evaluates statistically significant clustering using fishnet cells. Fifth, Near analysis measures the distance between each incident and the nearest major road. This sequence allows each method to answer a distinct research question while remaining grounded in the same core incident dataset.

3.9 Motor Vehicle Crime Spatial Distribution

3.9.1 Incident Point Mapping

The baseline map displays all 15,149 mapped motor vehicle crime incidents as point features. Police-district boundaries and the San Francisco boundary provide orientation while the incident points remain the primary visual variable. This representation preserves individual event locations and establishes the spatial baseline for all subsequent analyses.

3.9.2 Temporal Study Period: 2021-2025

The point distribution combines incidents recorded from 2021 through 2025 into one citywide map. The analysis does not weight incidents by year and does not model temporal change; every record is treated as one mapped event in the five-year spatial distribution.

3.9.3 Spatial Distribution Patterns

Visual inspection of the point map indicates that incidents are distributed throughout the city but are more densely packed in central, northeastern, and southeastern areas. Western districts contain many incidents but generally show a more dispersed pattern. These visual differences motivate the formal aggregation, density, and hot-spot analyses that follow.

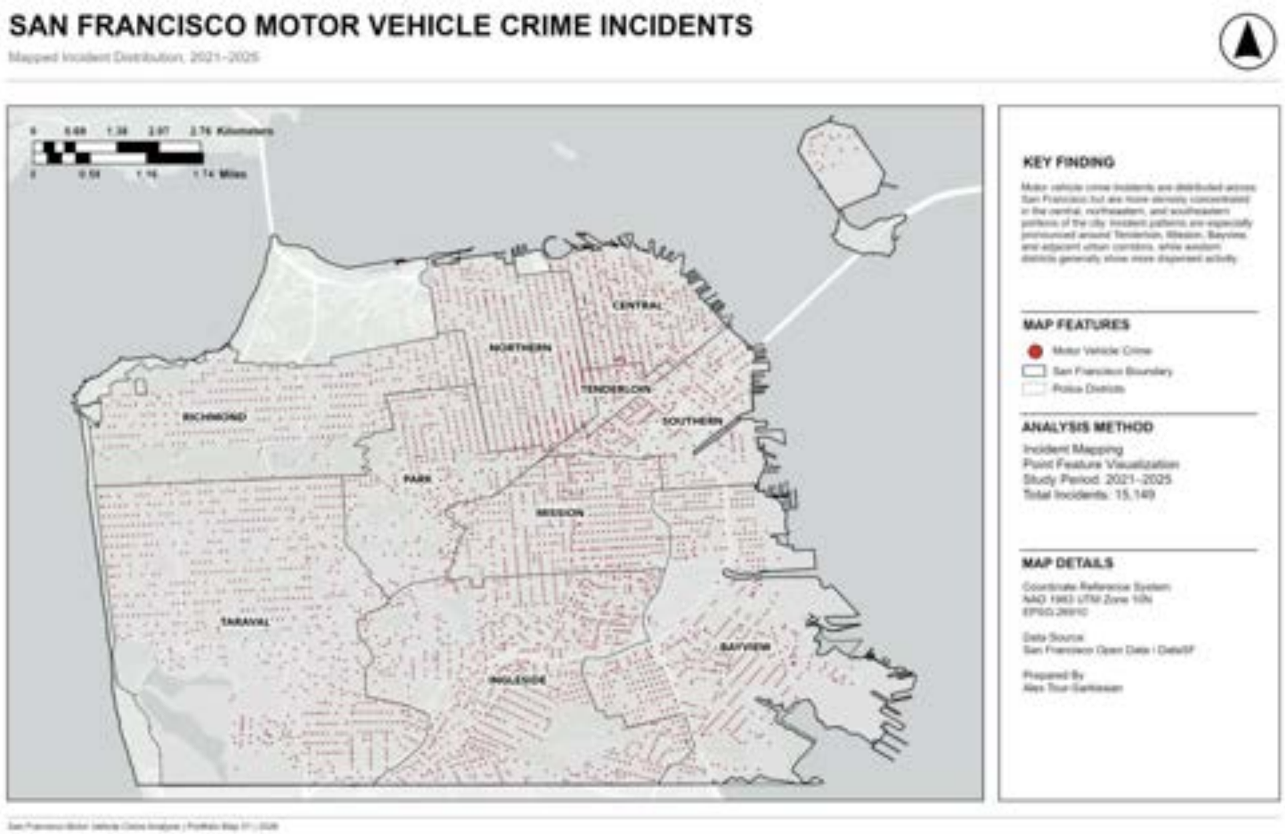


Figure 1. Motor Vehicle Crime Incident Distribution, San Francisco, 2021-2025.

3.10 Administrative Aggregation Analysis

3.10.1 Crime by Police District

Spatial Join was used with police districts as target polygons and mapped crime incidents as join features. The resulting Join_Count values were symbolized using five Natural Breaks (Jenks) classes from Very Low to Very High. District labels show both the police-district name and final mapped incident count.

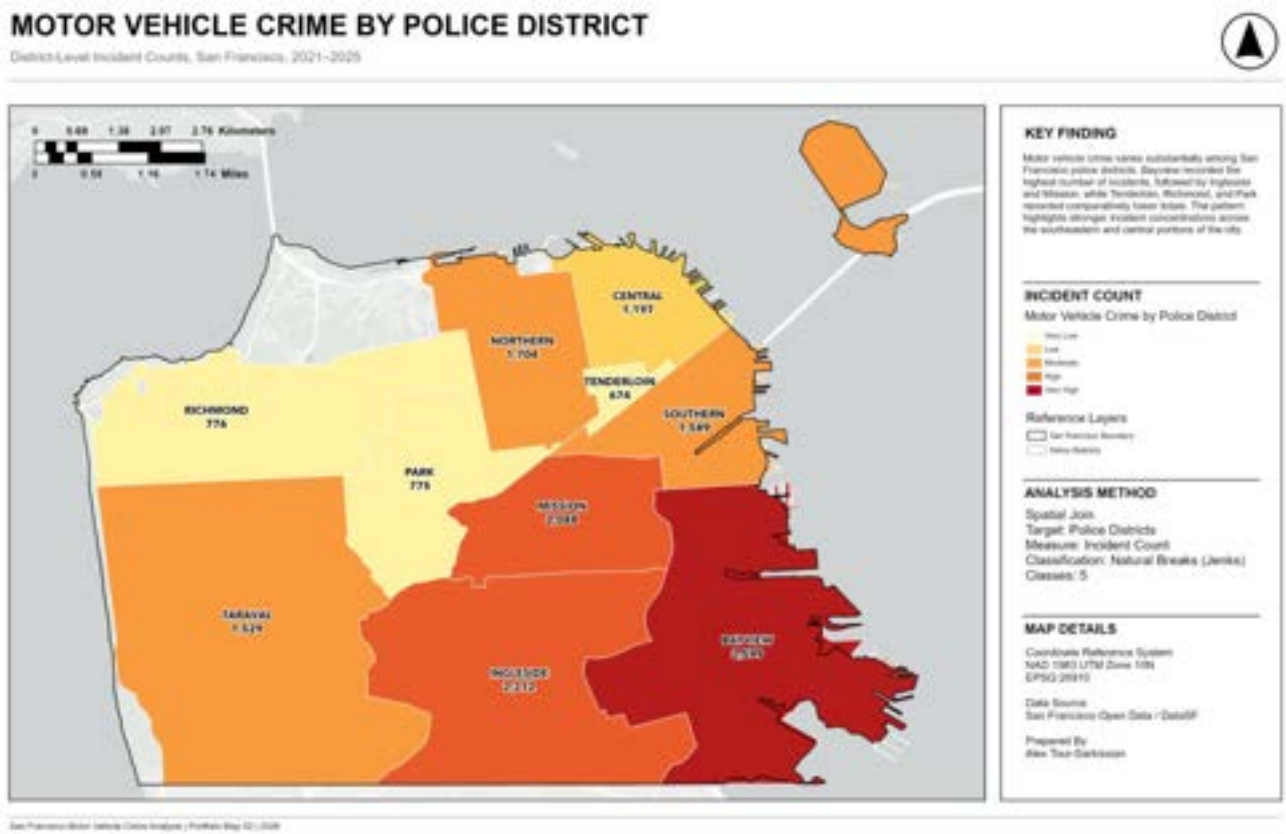


Figure 2. Motor Vehicle Crime by Police District, using Spatial Join and five-class Natural Breaks (Jenks) symbology.

3.10.2 Crime by Neighborhood

A second Spatial Join aggregated incident points to the project neighborhood polygons. The final choropleth uses five Natural Breaks (Jenks) classes, allowing higher-count neighborhood areas to be distinguished from the broader lower-count pattern. Because of the number and size of neighborhood polygons, the final portfolio map intentionally limits or omits neighborhood labels to preserve readability.

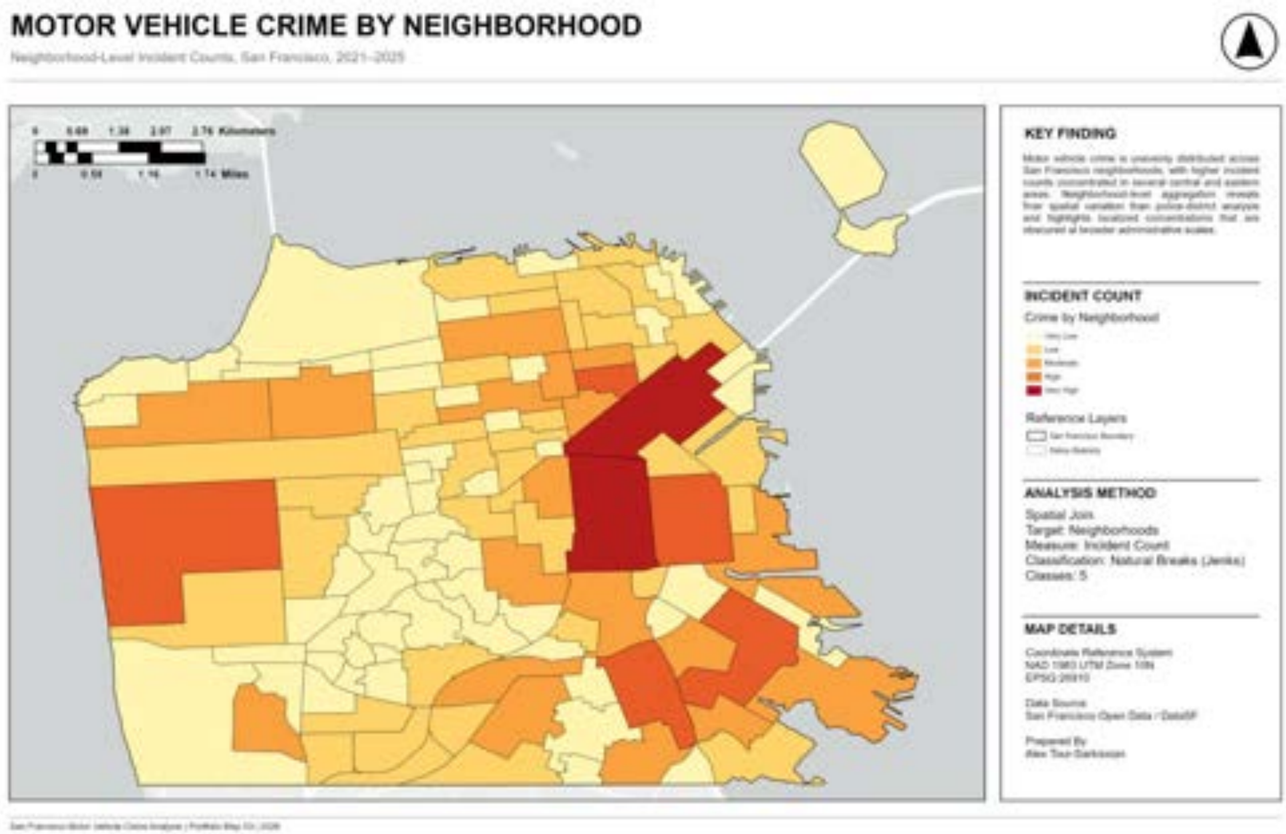


Figure 3. Motor Vehicle Crime by Neighborhood, showing neighborhood-level incident-count classes.

3.10.3 Comparison of Administrative Crime Patterns

The district and neighborhood maps are designed to be interpreted together. Police districts provide a broad operational view, whereas neighborhoods reveal smaller pockets of variation inside those districts. Neither map is population-normalized, so the values represent mapped incident counts rather than rates per resident, vehicle, or area.

3.11 Crime Density Analysis

3.11.1 Kernel Density Estimation

Kernel Density Estimation was applied to the crime incident points to calculate a magnitude-per-unit-area surface. Each incident was counted equally because no population weighting field was used. The final raster was created with a 50 m cell size, 500 m search radius, square-kilometer area units, density output values, and planar method. The San Francisco boundary was used to control processing extent and mask the final density surface.

3.11.2 Identification of High-Density Crime Areas

The resulting density raster was classified into five categories - Very Low, Low, Moderate, High, and Very High - using a sequential light-yellow to dark-red palette. Increasing color intensity therefore represents increasing modeled incident density.

3.11.3 Interpretation of Crime Density Surfaces

Kernel Density is independent of police-district and neighborhood boundaries. It smooths nearby incidents into a continuous surface and is therefore useful for identifying broad concentration zones that cross administrative boundaries. It does not test statistical significance and should not be interpreted as a probability or causal surface.

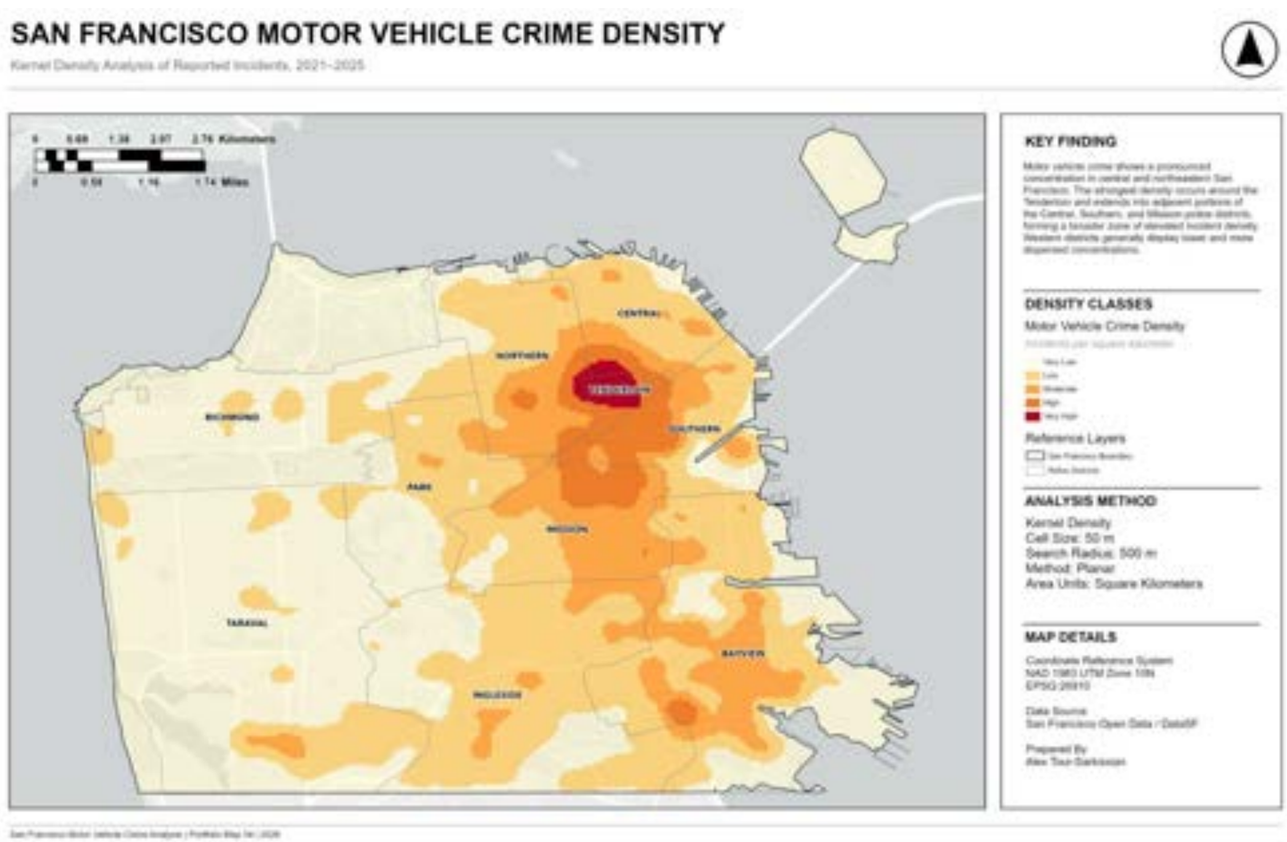


Figure 4. Kernel Density of motor vehicle crime incidents using a 50 m cell size and 500 m planar search radius.

3.12 Hot Spot Analysis

3.12.1 Getis-Ord Gi* Method

Hot Spot Analysis (Getis-Ord Gi*) was applied to a crime-analysis fishnet so that each grid cell could be evaluated in the context of neighboring cells. The method identifies statistically significant spatial clusters of high values (hot spots) and low values (cold spots). The ArcGIS Pro output includes z-scores, p-values, and a Gi_Bin confidence-level field for each analyzed feature.

3.12.2 Spatial Clustering of High Crime Values

A high value alone is not sufficient for a statistically significant hot spot. A cell must have a high local value and be surrounded by other high values strongly enough that the local clustering is unlikely to result from random spatial variation. The same logic applies to statistically significant cold spots of low values.

3.12.3 Statistically Significant Hot and Cold Spots

The final map displays cold spots at 99%, 95%, and 90% confidence; non-significant cells; and hot spots at 90%, 95%, and 99% confidence. A blue-to-red diverging palette differentiates low-value and high-value statistical clusters and clearly distinguishes the Gi* map from the warm sequential count and density maps.

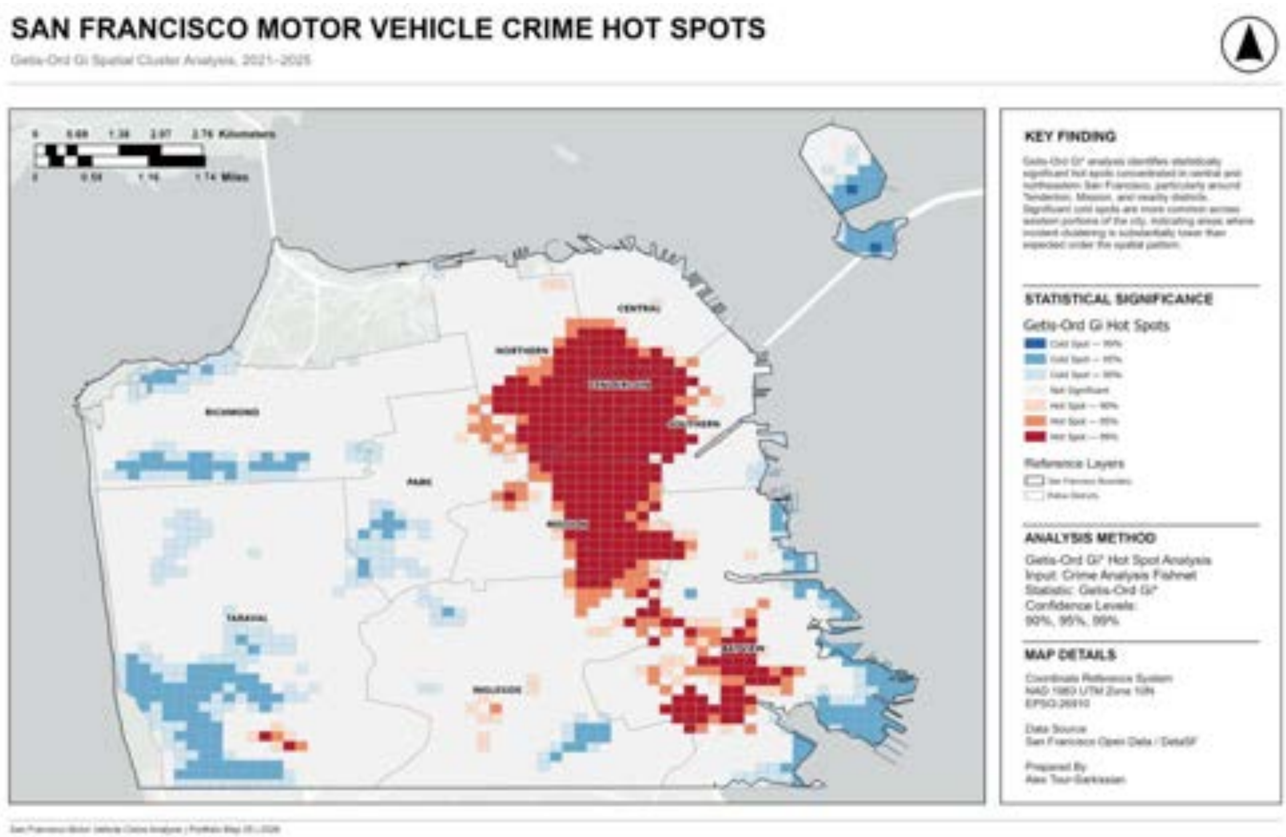


Figure 5. Getis-Ord Gi Hot Spot Analysis with 90%, 95%, and 99% confidence classes.*

3.13 Road Proximity Analysis

3.13.1 Major Road Network

The prepared major-road layer represents the transportation corridors used as reference features for proximity analysis. The final cartographic treatment emphasizes major roads with a dark line symbol, while semi-transparent 100 m, 250 m, and 500 m buffers provide visual context.

3.13.2 Crime Incidents in Proximity to Major Roads

The Near tool calculated the distance from every incident point to the nearest major-road feature. Because Near modifies its input, the tool was run on a derived analysis copy rather than the core crime source layer. The NEAR_DIST field was then classified into mutually exclusive distance categories: 0-100 m, 101-250 m, 251-500 m, and greater than 500 m.

3.13.3 Spatial Relationship Between Crime and Roads

Summary Statistics was used to count incidents in each distance category and calculate the percentage of the overall mapped total. The road-proximity map includes a supporting bar chart so the spatial pattern and quantitative distribution can be interpreted together.

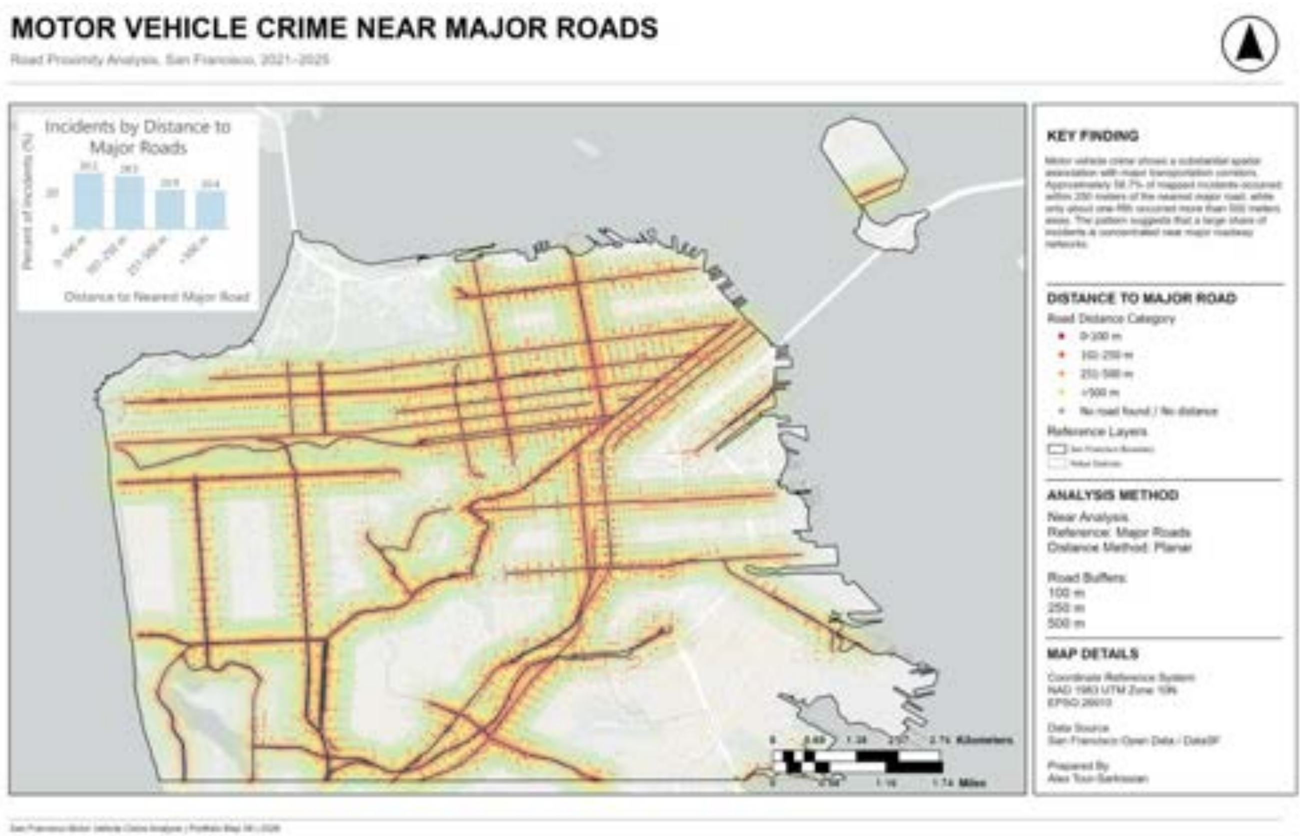


Figure 6. Motor Vehicle Crime Near Major Roads, including road-distance categories, road buffers, and percentage summary chart.

3.14 Analytical Parameters and GIS Workflow

3.14.1 Spatial Analysis Parameters

Table 4 consolidates the principal parameters used in the completed analyses. Maintaining a parameter record supports reproducibility, technical review, and future refinement of the project.

Table 4. Spatial Analysis Methods and Parameters

Analysis	Input	Key Parameters / Output
Incident Distribution	15,149 crime points	Point visualization; police districts and city boundary as context
Police District Spatial Join	Crime points + police districts	Join_Count; 5 Natural Breaks (Jenks) classes
Neighborhood Spatial Join	Crime points + neighborhood polygons	Join_Count; 5 Natural Breaks (Jenks) classes
Kernel Density	Crime points	50 m cell; 500 m radius; planar; square kilometers; densities
Getis-Ord Gi*	Crime analysis fishnet	Gi_Bin confidence classes: 90%, 95%, 99% hot/cold
Near Analysis	Crime points + major roads	Planar distance; NEAR_DIST; 4 distance categories
Road Buffers	Major roads	100 m, 250 m, 500 m
Summary Statistics	Road-distance categories	Frequency and percent of total incidents

3.14.2 Distance and Search Parameters

All reported distance thresholds are in meters. Kernel Density used a 500 m search radius and 50 m output cell size. Road-proximity classes were defined as 0-100 m, 101-250 m, 251-500 m, and greater than 500 m, with supporting buffers at 100 m, 250 m, and 500 m.

3.14.3 Statistical Significance Parameters

The Getis-Ord Gi* output was interpreted using 90%, 95%, and 99% confidence classes. Positive statistically significant z-scores represent clusters of high values, while negative statistically significant z-scores represent clusters of low values. Non-significant cells were retained to show where the analysis did not support a hot-spot or cold-spot interpretation.

3.14.4 GIS Processing Workflow

- Acquire and organize the incident and geographic reference data.
- Filter and prepare the 2021-2025 motor vehicle crime subset.
- Standardize distance- and density-based analysis in NAD 1983 UTM Zone 10N.
- Create police-district and neighborhood Spatial Join outputs.
- Generate the Kernel Density raster and classify the density surface.
- Create the crime-analysis fishnet and run Getis-Ord Gi* Hot Spot Analysis.
- Run Near against major roads, classify NEAR_DIST, and calculate road-distance percentages.
- Apply consistent portfolio cartography, review outputs, and export the six final map layouts.

4. Results

4.1 Overall Spatial Distribution of Motor Vehicle Crime

Figure 1 shows that motor vehicle crime incidents are distributed throughout San Francisco but are not spatially uniform. The densest point patterns are visible in central and northeastern parts of the city, with additional concentrations in southeastern San Francisco. Western areas contain substantial numbers of incidents but generally display greater spacing between points and less visually continuous concentration.

The map establishes a broad east-west contrast that is revisited in every subsequent analysis. Central districts around Tenderloin, Mission, Southern, and neighboring areas repeatedly emerge as important parts of the citywide pattern, while the western side tends to be characterized by lower or more dispersed intensity.

4.2 Crime Patterns by Police District

Figure 2 shows substantial variation in mapped incident totals among police districts. Bayview records the highest displayed total at 2,599 incidents, followed by Ingleside at 2,212 and Mission at 2,088. Northern records 1,704 incidents, Southern 1,589, Taraval 1,529, and Central 1,197. Richmond, Park, and Tenderloin show the lowest displayed district totals in the final map.

Table 2. Motor Vehicle Crime Records by Police District

Police District	Mapped Incident Count
Bayview	2,599
Ingleside	2,212
Mission	2,088
Northern	1,704
Southern	1,589
Taraval	1,529
Central	1,197
Richmond	776
Park	775
Tenderloin	674

Note: The ten final map labels sum to 15,143 incidents, six fewer than the project total of 15,149. The cause of this difference was not established in the supplied final materials and remains a QA item.

The district choropleth shows the highest class concentrated primarily in Bayview, with Ingleside and Mission also in the upper range. The administrative pattern is useful for comparing broad operational areas, but it does not mirror the Kernel Density or Getis-Ord Gi* maps exactly because those methods measure different spatial properties.

4.3 Crime Patterns by Neighborhood

Figure 3 shows more localized spatial variation than the police-district map. High and very-high incident classes appear in several central and eastern neighborhood polygons, while many western and interior polygons fall into lower classes. The finer geography exposes within-district variation that is hidden when all incidents are summarized to only ten police districts.

For descriptive context, the incident source spreadsheet contains an analysis-neighborhood attribute with 41 named categories. The source-field counts below are not presented as exact replacements for the spatial-join choropleth because the final map uses the project neighborhood polygons and boundary-based spatial assignment. They are included to document the incident dataset and provide a comparable neighborhood-level reference.

Table 3. Motor Vehicle Crime Records by Neighborhood (source analysis_n field)

Analysis Neighborhood	Count	Analysis Neighborhood	Count
Bayview Hunters Point	1,809	Pacific Heights	266
Mission	1,778	Haight Ashbury	251
South of Market	982	Russian Hill	246
Tenderloin	855	Visitacion Valley	243
Bernal Heights	603	Mission Bay	220
Sunset/Parkside	577	Inner Sunset	209
Excelsior	529	Lone Mountain/USF	197
Financial District/South Beach	519	North Beach	171
Western Addition	482	Glen Park	152
West of Twin Peaks	449	Inner Richmond	142
Potrero Hill	446	Golden Gate Park	134
Nob Hill	428	Presidio Heights	120
Castro/Upper Market	386	Chinatown	96
Hayes Valley	377	Twin Peaks	87
Portola	365	Japantown	69
Outer Mission	343	Treasure Island	57
Outer Richmond	341	McLaren Park	25
Noe Valley	294	Seacliff	25
Oceanview/Merced/Ingleside	290	Lincoln Park	7
Marina	287	Presidio	5
Lakeshore	282		

Note: The source analysis_n field contains 15,144 nonblank records across 41 categories; five records have no analysis-neighborhood value. The final Map 03 choropleth was created from the project neighborhood polygon layer and should be interpreted from its spatial-join output rather than this source-field table alone.

4.4 Crime Density Patterns

Figure 4 identifies the strongest continuous motor vehicle crime density in central and northeastern San Francisco. The most prominent very-high-density core is centered around Tenderloin and extends into adjacent portions of Central, Southern, and Mission. Moderate-to-high density also extends southward and into several eastern and southeastern locations.

Western districts such as Richmond and Taraval generally exhibit lower and more dispersed density. This result reinforces the broad pattern visible in Figure 1 while removing the visual clutter of thousands of overlapping points. Because density is modeled continuously, the pattern also crosses police-district and neighborhood boundaries rather than being constrained by them.

4.5 Statistically Significant Hot Spots

Figure 5 identifies statistically significant hot-spot clustering across a large central corridor, including Tenderloin and Mission and adjacent areas. Additional significant hot-spot cells occur in parts of Bayview. Significant cold spots appear more frequently in western San Francisco and along portions of the outer study area.

The spatial overlap between the central Kernel Density maximum and the central Gi* hot-spot cluster is analytically important. Kernel Density indicates strong concentration, while Gi* provides evidence that neighboring high values form statistically significant clusters. The methods therefore support related but distinct interpretations of the same underlying incident pattern.

4.6 Motor Vehicle Crime and Major Road Proximity

Figure 6 shows a strong visual alignment between many incident locations and the prepared major-road network. The quantitative summary confirms that 4,582 incidents (30.2%) occurred within 100 m of the nearest major road and another 4,313 incidents (28.5%) occurred between 101 and 250 m. Together, these two classes represent 8,895 incidents, or approximately 58.7% of the mapped total, within 250 m of a major road.

Road Proximity Summary

Distance to Nearest Major Road	Incidents	Share of Total
0-100 m	4,582	30.2%
101-250 m	4,313	28.5%
251-500 m	3,171	20.9%
>500 m	3,083	20.4%
Total	15,149	100.0%

The remaining 3,171 incidents (20.9%) occurred between 251 and 500 m, while 3,083 incidents (20.4%) occurred more than 500 m away. These percentages describe a spatial association only. They do not demonstrate that major roads cause motor vehicle crime or establish which roadway-related conditions, if any, contribute to incident locations.

4.7 Integrated Spatial Findings

When interpreted collectively, the six maps identify a broadly consistent citywide pattern. The point map suggests greater concentration in central, northeastern, and southeastern San Francisco. Administrative aggregation confirms substantial differences among districts and neighborhoods. Kernel Density identifies a continuous central concentration around Tenderloin and neighboring districts, while Getis-Ord Gi* confirms statistically significant hot-spot clustering across much of the same corridor. The road-proximity analysis adds a transportation dimension by showing that a majority of mapped incidents fall within 250 m of a major road.

The analyses do not produce identical boundaries or rankings because each method answers a different question. The strongest interpretation comes from areas that remain important across multiple methods rather than from assuming that any one map is definitive.

4.8 Summary of Key Results

Table 5. Summary of Key Spatial Findings

Analysis	Key Spatial Finding
Incident Distribution	15,149 mapped incidents; denser central/northeastern and southeastern point patterns
Police Districts	Bayview highest displayed district total (2,599), followed by Ingleside (2,212) and Mission (2,088)
Neighborhoods	Fine-scale polygons reveal localized high-count areas within broader district patterns
Kernel Density	Strongest continuous density centered on Tenderloin and adjacent Central, Southern, and Mission areas
Getis-Ord Gi*	Statistically significant central/northeastern hot spots; cold spots more common in western San Francisco
Major Roads	58.7% of mapped incidents within 250 m of nearest major road; 20.4% more than 500 m away

5. Discussion

5.1 Interpretation of Spatial Distribution Patterns

The baseline point map demonstrates that the incident pattern is geographically uneven, but visual point density alone is insufficient for formal interpretation. Areas with many points may appear dominant because of overlap, street geometry, or the compactness of the local urban fabric. The subsequent analyses therefore transform the same incident dataset in different ways to separate administrative totals, continuous concentration, statistical clustering, and proximity.

The persistence of a central and northeastern concentration across the point, density, and hot-spot maps suggests that this part of San Francisco is a robust feature of the five-year spatial pattern rather than an artifact of a single representation. At the same time, the high Bayview district total and additional Bayview hot-spot cells show that the southeastern part of the city is also analytically important.

5.2 Administrative Differences in Crime Concentration

Police-district and neighborhood maps demonstrate the effect of aggregation scale. Police districts are useful for broad operational comparison, but large polygons can hide local variation. Neighborhood polygons reveal smaller high-count areas that may be obscured inside a district-level total. This scale dependency is a practical example of the modifiable areal unit problem: results can change when the same events are summarized to different boundary systems.

Neither administrative map is population-normalized. A district or neighborhood with more incidents is not necessarily more risky for an individual vehicle owner because the analysis does not control for residents, visitors, parked vehicles, street mileage, parking supply, land use, or exposure time. The maps should therefore be interpreted as incident-count distributions, not standardized crime rates.

5.3 Density and Clustering of Motor Vehicle Crime

Kernel Density provides a boundary-independent view of concentration and is especially useful in San Francisco because important spatial patterns cross administrative lines. The 500 m search radius creates a smoothed local surface that emphasizes neighborhood-scale concentrations rather than individual street corners. The 50 m cell size provides sufficient spatial detail for cartographic interpretation while maintaining a continuous surface.

The central density core around Tenderloin and adjacent districts is consistent with the dense point pattern in Figure 1. The smoother surface also reveals extensions of moderate and high density that are difficult to see in the raw point map. Because Kernel Density is sensitive to the selected search radius and cell size, the result should be understood as one scale-specific representation of concentration rather than a fixed boundary of crime activity.

5.4 Hot Spot Patterns and Spatial Significance

Getis-Ord Gi* adds inferential context by testing whether neighboring high or low values form statistically significant clusters. The 90%, 95%, and 99% confidence classes provide a hierarchy of statistical evidence. In practical terms, the method distinguishes areas that are merely high from areas where high values are embedded within a broader cluster of high neighboring values.

The overlap between high Kernel Density and high-confidence Gi* hot spots in central San Francisco strengthens the interpretation that the central corridor is an important concentration area in the 2021-2025 dataset. Cold-spot patterns in western areas provide the complementary finding that some parts of the city contain spatially clustered low values relative to the citywide distribution.

5.5 Relationship Between Crime and Major Roads

The road-proximity analysis shows that mapped motor vehicle crime incidents are frequently located close to major transportation corridors. Approximately 58.7% of incidents occur within 250 m of the nearest major road, and 30.2% occur within 100 m. This result is consistent with the visual alignment between incident points and the major-road network in Figure 6.

The finding should not be interpreted causally. Major roads are correlated with many other urban characteristics, including vehicle volumes, parking demand, commercial land uses, transit activity, intersection density, and accessibility. The current analysis does not isolate any of those factors. A future explanatory model would need to include exposure measures and additional covariates before drawing conclusions about why incidents are concentrated near particular corridors.

5.6 Comparison of Analytical Findings

Comparison of the six outputs demonstrates why a multi-method workflow is more informative than a single map. Point mapping preserves event locations but becomes visually dense. Administrative aggregation simplifies comparison but depends on polygon boundaries. Kernel Density models continuous concentration without testing significance. Getis-Ord Gi* tests local clustering but requires a defined neighborhood structure and spatial units. Near analysis measures a different relationship entirely: distance to transportation infrastructure.

The strongest evidence is therefore relational rather than map-specific. Central and northeastern San Francisco appear important in the point, density, and hot-spot analyses. Bayview is prominent in the police-district map and also contains significant hot-spot cells. Western San Francisco is comparatively lower in the density map and contains more cold-spot cells. The road-proximity analysis cuts across these geographic patterns and shows that major-road accessibility is a common spatial characteristic for a large share of incidents citywide.

5.7 Implications for Urban Crime Analysis

The analysis demonstrates how GIS can support a structured public-safety research workflow. District and neighborhood summaries can help identify administrative areas for follow-up review. Density and hot-spot maps can define smaller geographic focus areas for field investigation, parking-management review, environmental assessment, or targeted data collection. Road-proximity results can identify transportation corridors for additional analysis of parking, land use, lighting, vehicle exposure, or traffic activity.

The project is best understood as decision-support analysis rather than a predictive or causal model. Its value lies in organizing public incident data into multiple spatial perspectives and identifying where additional evidence would be most useful.

5.8 Limitations of the Spatial Analysis

- Published incident locations are privacy-protected and mapped to nearby intersections rather than exact event coordinates; DataSF also notes an intersection-mapping change beginning April 24, 2024.
- The analysis combines 2021-2025 incidents and therefore does not measure year-to-year change or temporal seasonality.
- Police-district and neighborhood results are raw mapped counts rather than population-, vehicle-, parking-, or exposure-normalized rates.
- The final police-district map totals sum to 15,143, six fewer than the 15,149 project incident total; the cause remains a QA item in the supplied materials.
- The neighborhood source-field counts and final polygon-based spatial-join map may differ because different neighborhood definitions or boundary assignments can be involved.
- Kernel Density results depend on the selected 50 m cell size and 500 m search radius.
- Getis-Ord Gi* results depend on the fishnet and spatial-neighborhood configuration; statistical significance does not explain the causal processes creating the clusters.
- Road proximity measures association only. It does not control for street density, traffic volume, parking supply, land use, accessibility, or other confounding variables.
- The completed six-map series does not include the originally considered transit-station proximity or citywide crime-category analyses.

6. Conclusion

6.1 Summary of the Study

This study analyzed the spatial distribution of 15,149 mapped motor vehicle crime incidents in San Francisco from 2021 through 2025. A six-part ArcGIS Pro workflow was developed to move from direct incident mapping to administrative aggregation, continuous density estimation, statistically significant hot-spot analysis, and proximity to major roads. Using the same core incident dataset across multiple methods allowed the study to compare how spatial conclusions change with geographic representation and analytical technique.

6.2 Key Findings

- Motor vehicle crime incidents are distributed citywide but are more densely concentrated in central, northeastern, and southeastern San Francisco.
- Bayview has the highest displayed police-district count at 2,599 incidents, followed by Ingleside (2,212) and Mission (2,088).
- Neighborhood aggregation reveals finer local variation that is hidden at the police-district scale.
- Kernel Density identifies the strongest continuous concentration around Tenderloin and adjacent portions of Central, Southern, and Mission.

- Getis-Ord Gi* identifies statistically significant hot spots across much of the central corridor and portions of Bayview, while cold spots are more common in western San Francisco.
- Approximately 58.7% of mapped incidents occurred within 250 m of the nearest major road.

6.3 Answers to the Research Questions

Answers to the Research Questions

Research Question	Answer from Completed Analysis
RQ1 - Spatial distribution	Incidents occur throughout San Francisco but form visibly denser patterns in central, northeastern, and southeastern areas.
RQ2 - Districts and neighborhoods	Bayview has the highest displayed district total, while neighborhood aggregation shows additional fine-scale variation across central and eastern parts of the city.
RQ3 - Kernel Density	The strongest modeled density is centered around Tenderloin and extends into adjacent Central, Southern, and Mission areas.
RQ4 - Hot and cold spots	Getis-Ord Gi* identifies statistically significant central/northeastern hot spots and more extensive cold-spot patterns in western San Francisco.
RQ5 - Major roads	A substantial spatial association exists: 58.7% of mapped incidents occur within 250 m of the nearest major road.
RQ6 - Consistency across methods	Yes, several methods identify central and eastern San Francisco as important, but each method defines and measures the pattern differently.

6.4 Contribution of the GIS Analysis

The project contributes a reproducible portfolio-scale example of an end-to-end GIS research workflow. It demonstrates data preparation, geodatabase organization, coordinate-system management, Spatial Join, raster density analysis, local spatial statistics, proximity analysis, summary statistics, charting, cartographic design, and QA/QC. More importantly, it demonstrates analytical discipline by separating raw counts, density, statistical significance, and proximity rather than presenting them as equivalent measures.

6.5 Recommendations for Future Analysis

- Reconcile the six-record difference between the total incident count and the final police-district spatial-join totals.
- Preserve and analyze annual incident counts separately to evaluate temporal change between 2021 and 2025.
- Normalize administrative counts using relevant exposure measures such as registered vehicles, parking supply, street mileage, daytime population, or other appropriate denominators.
- Test sensitivity of Kernel Density results to alternative search radii and cell sizes.
- Document the fishnet size and spatial-neighborhood conceptualization used for Getis-Ord Gi* and test sensitivity to alternative configurations.
- Extend the corridor analysis using road functional class, traffic volume, parking concentration, commercial land use, or intersection density.
- If desired, complete the originally considered transit-station proximity analysis as a separate extension rather than merging it into the completed road-proximity result.
- Develop a reproducible ArcGIS ModelBuilder or Python workflow so core analysis outputs can be regenerated from documented parameters.

6.6 Final Conclusion

The completed analysis shows that San Francisco motor vehicle crime from 2021-2025 is spatially uneven and that the most informative interpretation emerges from combining multiple GIS methods. Administrative counts identify broad geographic differences, Kernel Density reveals continuous concentration, Getis-Ord Gi* distinguishes statistically significant clustering, and road proximity quantifies the relationship between incidents and major transportation corridors. Together, the six maps form a coherent analytical narrative that is stronger than any individual map and provide a documented foundation for future GIS-based urban crime research.

7. References and Data Sources

7.1 Academic and Technical References

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. Geographical Analysis, 24(3), 189-206.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. Geographical Analysis, 27(4), 286-306.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall.

7.2 GIS and Spatial Analysis References

Esri. (2026). Kernel Density (Spatial Analyst) - ArcGIS Pro documentation. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/kernel-density.htm

Esri. (2026). How Kernel Density works - ArcGIS Pro documentation. https://pro.arcgis.com/en/pro-app/3.6/tool-reference/spatial-analyst/how-kernel-density-works.htm

Esri. (2026). Hot Spot Analysis (Getis-Ord Gi*) (Spatial Statistics) - ArcGIS Pro documentation. https://pro.arcgis.com/en/pro-app/3.6/tool-reference/spatial-statistics/hot-spot-analysis.htm

Esri. (2026). Near (Analysis) - ArcGIS Pro documentation. https://pro.arcgis.com/en/pro-app/3.6/tool-reference/analysis/near.htm

7.3 Crime Data Sources

City and County of San Francisco, Police Department / DataSF. Police Department Incident Reports: 2018 to Present. https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783

DataSF dataset metadata notes that incident locations are mapped to nearby intersections to preserve anonymity and that intersection mapping changed on April 24, 2024; these conditions are relevant to multi-year fine-scale spatial analysis.

7.4 Geographic and Road Network Data Sources

City and County of San Francisco / DataSF. Current Police Districts / Map of Current Police Districts. https://data.sfgov.org/-/Map-of-Current-Police-Districts/wkhw-cjsf

City and County of San Francisco / DataSF. Analysis Neighborhoods. https://data.sfgov.org/Geographic-Locations-and-Boundaries/Analysis-Neighborhoods/j2bu-swwd

City and County of San Francisco / DataSF. Streets - Active and Retired. https://data.sfgov.org/Geographic-Locations-and-Boundaries/Streets-Active-and-Retired/3psu-pn9h

Project-derived geographic layers include the San Francisco city boundary, prepared major-road layer, crime-analysis fishnet, spatial-join outputs, Kernel Density raster, road buffers, and road-distance analysis feature classes.

Appendix A. Analytical Parameters Summary

A.1 Dataset Summary

Appendix Table A1. Dataset and Output Summary

Dataset / Output	Role
CRIME_MotorVehicle_2021_2025	Core incident feature class for 2021-2025 motor vehicle crime
BASE_SF_City_Boundary	Study boundary, mask, and map extent control
BASE_Police_Districts	Police-district aggregation and context
BASE_Neighborhoods	Neighborhood aggregation
BASE_Major_Roads	Road-proximity reference feature
ANA_Crime_By_Police_District	Spatial-join district output
ANA_Crime_By_Neighborhood	Spatial-join neighborhood output
ANA_Crime_Fishnet	Grid used for statistical analysis
ANA_HotSpots_GiStar	Getis-Ord Gi* output
ANA_KernelDensity_MotorVehicle	Kernel Density raster
ANA_Crime_RoadDistance	Incident features with NEAR_DIST and road-distance categories
SUM_Road_Proximity	Road-distance frequency and percentage summary

A.2 Coordinate Reference System

Analysis coordinate system: NAD 1983 UTM Zone 10N (EPSG:26910). Linear unit: meter. The projected coordinate system was used for Kernel Density, Near analysis, and road buffers. Cartographic basemaps were used only for visual context.

A.3 Spatial Analysis Parameters

Appendix Table A2. Core Spatial Analysis Parameters

Analysis	Parameter	Value
Police District Spatial Join	Join operation	JOIN_ONE_TO_ONE / spatial count
Police District Spatial Join	Match option	INTERSECT
Neighborhood Spatial Join	Join operation	JOIN_ONE_TO_ONE / spatial count
Neighborhood Spatial Join	Match option	INTERSECT
Administrative choropleths	Classification	Natural Breaks (Jenks), 5 classes
Point distribution	Study period	2021-2025
Point distribution	Mapped incident total	15,149

A.4 Kernel Density Parameters

Appendix Table A3. Kernel Density Parameters

Parameter	Value
Input features	Motor vehicle crime incidents, 2021-2025
Population field	None
Output cell size	50 m
Search radius	500 m
Area units	Square kilometers
Output cell values	Densities
Method	Planar
Mask	San Francisco boundary
Classification	5 classes, Natural Breaks (Jenks) for final cartography

A.5 Getis-Ord Gi* Parameters

Appendix Table A4. Getis-Ord Gi* Parameters

Parameter	Value
Input framework	Crime analysis fishnet
Statistic	Getis-Ord Gi*
Output field used for symbology	Gi_Bin
Cold-spot classes	90%, 95%, 99% confidence
Hot-spot classes	90%, 95%, 99% confidence
Non-significant class	Retained in final map

A.6 Road Proximity Parameters

Appendix Table A5. Road Proximity Parameters

Parameter	Value
Input incidents	ANA_Crime_RoadDistance derived analysis copy
Near features	BASE_Major_Roads
Method	Planar
Search radius	Blank / nearest road located without a fixed search limit
Distance field	NEAR_DIST
Distance categories	0-100 m; 101-250 m; 251-500 m; >500 m
Road buffers	100 m; 250 m; 500 m
Summary output	SUM_Road_Proximity
Within 250 m	8,895 incidents / 58.7%

A.7 GIS Software and Processing Tools

Appendix Table A6. Software and Processing Tools

Software / Tool	Use in Project
ArcGIS Pro 3.6	Primary GIS analysis, geoprocessing, spatial statistics, cartography, layout export
QGIS	Supporting GIS environment listed in project documentation
Spatial Join	District and neighborhood incident aggregation
Kernel Density	Continuous crime-density surface
Hot Spot Analysis (Getis-Ord Gi*)	Statistical hot-spot and cold-spot analysis
Near	Nearest-major-road distance calculation
Buffer	100 m, 250 m, and 500 m road-buffer visualization
Summary Statistics	Road-distance frequencies and percentages
Field Calculator	Road-distance text and numeric class fields