

A KDE-Spatial Analysis of Jeepney Demand and Service Imbalance in Cagayan de Oro City

Kein Jake Culanggo, Qylle Christian Quiño, and GenheyLou Felisilda
University of Science and Technology of Southern Philippines

Abstract. This study addresses inefficiencies in Public Utility Vehicle (PUV) operations in Cagayan de Oro City by applying Kernel Density Estimation (KDE) to analyze spatial and temporal variations in passenger demand. The researchers developed a GIS-based system that generates heatmaps across selected time intervals to visualize high-demand and low-demand zones. Traditional jeepneys are the sole focus of this analysis. Passenger demand data were manually collected from key urban hotspots, and temporal demand shifts were modeled using discrete time slices. To identify transport imbalances, the study introduces an optimization metric based on the ratio of available jeepneys to total passenger demand within each zone. Initial testing with synthetic data demonstrated the system's potential to extract actionable insights and guide route planning. While integration of real-world data is forthcoming, the study offers a foundational, data-driven framework for improving transport planning and mobility strategies in similarly structured urban areas.

Keywords: Public Utility Vehicles (PUVs), Jeepney Transport, Kernel Density Estimation (KDE), Spatial Analysis, Temporal Demand Patterns, Transport Imbalance, Passenger Demand Mapping

1 Introduction

1.1 Aims. This study aims to optimize Public Utility Vehicle (PUV) deployment and route planning in Cagayan de Oro City by analyzing spatial and temporal patterns of passenger demand. Using Kernel Density Estimation (KDE), the study will generate heatmaps to identify high-density zones and peak ridership times based on geospatial data from malls, schools, terminals, markets, and other key locations. Additionally, an imbalance scale visualization will be provided to identify areas that are overserved, underserved, and optimal for PUV deployment, complementing the KDE heatmap analysis to maximize passenger ridership during fluctuating demand across different time slots and locations.

1.2 Background of the Study. Public Utility Vehicles (PUVs) play a vital role in urban mobility in Cagayan de Oro City, Philippines, supporting daily commutes and economic activities. However, PUV operations are marked by inefficiencies, such as traffic congestion, uneven route utilization, and long passenger wait times. These challenges are exacerbated by limitations in the city's Local Public Transport Route Plan (LPTRP), including the absence of designated loading and unloading zones and operators' reluctance to service low-demand areas [1].

Rapid urbanization and population growth have heightened the need for more adaptive and efficient transportation systems. Without optimization, issues such as overcrowding in high-demand zones, neglect of underserved areas, and operational imbalances persist. These inefficiencies result in commuter dissatisfaction, decreased economic productivity, and financial burdens on operators.

In a manner similar to the study of Nogueira et al. [2] conducted in Pasig City, which identifies the socioeconomic factors influencing public transport accessibility and walkability, our study on Public Utility Vehicle (PUV) demand in Cagayan de Oro City aims to explore the underlying factors contributing to underserved or overserved areas by jeepneys. The Pasig study found that barangays with fewer financial resources, higher population densities, and a larger share of low-income households and informal settlements tended to have lower walkability and poorer access to public transport. This suggests that socio-economic conditions play a crucial role in shaping transport accessibility.

To address these challenges, this study utilizes Kernel Density Estimation (KDE), a spatial analysis method widely employed in urban planning and crime mapping [3], to analyze passenger demand and identify high-ridership zones. A relevant example of KDE application in transportation is Li et al.'s [4] study in Wuhan, China, which aligned transport services with passenger density at bus stops. Building on this approach, the research will provide an imbalance scale visualization to identify areas that are overserved, underserved, and optimal for PUV deployment, complementing the KDE heatmap analysis to maximize passenger ridership during fluctuating demand across different time slots and locations.

By focusing on both spatial and temporal passenger data, this study aims to address critical gaps in transport service delivery, providing a foundation for enhancing urban mobility through data-driven solutions specifically tailored to Cagayan de Oro City.

1.3 Context. Public Utility Vehicles (PUVs) are vital to daily mobility in Cagayan de Oro City, yet operational inefficiencies persist, including congestion, long wait times, and uneven route utilization. These challenges are exacerbated by limitations in the city's Local Public Transport Route Plan (LPTRP), such as the absence of designated loading zones and gaps in routes with low ridership [1]. As urban growth increases demand on the PUV system, traditional zoning methods fail to address emerging needs effectively. Spatial analysis techniques, particularly Kernel Density Estimation (KDE), have been successfully applied in urban planning and crime mapping to identify hotspots and optimize resource allocation [3]. A similar approach was used by Li et al. [4] in Wuhan, where KDE was applied to align bus routes with areas of high passenger demand. These examples demonstrate the potential of geospatial methods to provide valuable insights into urban transport systems, particularly in understanding the spatial distribution of demand.

1.4 Research Problem. Cagayan de Oro lacks a data-driven framework to identify when and where Public Utility Vehicle (PUV) demand peaks, resulting in overcrowding in some areas, underutilization in others, and inefficient vehicle deployment. Existing planning tools do not capture the spatiotemporal dynamics of passenger movement. Without applying spatial analysis methods such as Kernel Density Estimation (KDE), transport agencies are unable to allocate services based on actual commuter demand patterns. This study addresses the need for a responsive, evidence-based system to guide route planning and PUV distribution, considering both temporal and spatial variations in demand.

1.4.1 Research Questions. To address inefficiencies in the deployment of Public Utility Vehicles (PUVs) in Cagayan de Oro City, this study applies spatial analysis techniques to assess the variations in passenger demand across both space and time. By leveraging Kernel Density Estimation (KDE) and imbalance scale visualizations, the research aims to identify areas that are overserved, underserved, or optimally served by the existing jeepney system. The following research questions guide the inquiry:

- How does passenger demand for jeepneys in Cagayan de Oro City vary spatially and temporally across key urban locations such as malls, schools, terminals, and markets?
- Which areas in Cagayan de Oro City are spatially overserved, underserved, or optimally served by jeepneys based on KDE-generated heatmaps and imbalance scale visualizations?
- What additional spatial, temporal, and operational factors must be integrated to more accurately capture jeepney passenger demand patterns in Cagayan de Oro and support more responsive deployment strategies?

1.5 Expected Outcomes. This study will apply Kernel Density Estimation (KDE) to analyze spatial and temporal variations in Public Utility Vehicle (PUV) passenger demand within Cagayan de Oro City. The expected outcomes are as follows:

1.3.1 Passenger Demand KDE Heatmap. The study will produce a KDE-based heatmap illustrating the spatial distribution of passenger demand. This visualization will identify areas of high and low concentration, serving as the foundation for evaluating the adequacy of service coverage.

1.3.2 Imbalance Calculation Based on Jeepney-to-Passenger Ratio. An imbalance ratio will be computed to quantify the service adequacy of jeepney deployment relative to passenger demand. This metric integrates several operational variables, including the total number of passengers observed, the frequency of jeepney arrivals, vehicle capacity, and the duration of the observation window. The resulting ratio provides a numerical indicator of service alignment: positive values suggest potential overservice, negative values indicate underservice, and values near zero reflect a balanced condition. This ratio serves as the core metric for evaluating how well jeepney supply matches localized demand across different locations and time periods.

1.5.4 Actionable Recommendations Based on Imbalance Analysis. Building on the spatial and temporal analyses, the study will provide recommendations for jeepney allocation strategies. These recommendations will be informed by the observed imbalance patterns and aim to improve service efficiency in both high-demand and low-demand zones.

These outcomes will serve as a foundation for data-informed decision-making in public transport planning. By revealing where and when mismatches between passenger demand and jeepney availability occur, the study will support the development of more efficient, responsive deployment strategies. Ultimately, the findings aim to contribute to a more equitable and optimized PUV system in Cagayan de Oro City, guided by spatial and temporal evidence rather than static route assumptions.

1.5 Significance. This study addresses inefficiencies in Public Utility Vehicle (PUV) operations in Cagayan de Oro City through the use of Kernel Density Estimation (KDE) to analyze passenger demand across spatial and temporal dimensions. Rather than relying on fixed-route assumptions, the study highlights demand imbalances by identifying areas that are underserved, overserved, or optimally served based on observed ridership data. A central contribution of this work is the development of an interactive GIS-based system that visualizes demand hotspots and imbalance zones, with features that support customization, scenario simulation, and editable density inputs.

The methodology and system prototype developed in this study offer a scalable and adaptable framework for transport planning. It equips local governments, transport agencies, and operators with a practical tool for evidence-based decision-making, enabling more responsive and equitable public transport strategies that align with dynamic passenger needs.

2 Literature Review

2.1 The Utility of Kernel Density Estimation (KDE) in Visualizing Urban Spatial Patterns

Kernel Density Estimation (KDE) is a non-parametric spatial analysis technique that transforms discrete data points into continuous density surfaces, enabling the visualization of concentration patterns across a geographic area. In urban studies, KDE has become a standard tool for detecting spatial intensities or "hotspots" critical to infrastructure planning, land-use optimization, and service distribution.

Mora García et al. [5] emphasized the sensitivity of KDE outputs to parameter selection, particularly bandwidth, which governs the level of smoothing in spatial surfaces. Their analysis, which evaluated multiple kernel types and smoothing parameters, confirmed that bandwidth selection critically influences the interpretation of urban density patterns, even more so than the choice of kernel function. This reinforces the need for calibrated KDE parameters when analyzing transport demand patterns to avoid misrepresentation of true density clusters.

Cai et al. [6], in the context of road networks, used KDE to reveal spatial concentrations of road infrastructure and its influence on landscape fragmentation. Their findings demonstrated that KDE can effectively uncover dense spatial formations within transportation systems and highlight their broader urban implications. Though environmental in focus, the study showcases KDE's utility in understanding spatial imbalances and guiding development interventions.

Yu et al. [7] introduced a network-constrained variant of KDE to delineate urban centers and commercial districts, adjusting for transport network accessibility. By integrating decay functions based on distance along roads, their method yielded more representative density surfaces for urban mobility studies. This approach is particularly relevant to transport planning in cities with complex road networks, such as Cagayan de Oro, where network topology significantly impacts travel behavior and ridership concentration.

These studies collectively highlight KDE's value in urban spatial analysis, especially in identifying uneven distributions of human activity and infrastructure. For this study, KDE provides a foundational method for visualizing the spatiotemporal variability in Public Utility Vehicle (PUV) demand, capturing both areas of persistent oversupply and underserved demand zones.

2.2 Applications of KDE in Public Transport and Passenger Demand Studies

KDE has also seen increasing adoption in public transport studies as a means to visualize spatial demand and service coverage. By converting discrete boarding points or passenger events into continuous surfaces, KDE allows researchers to identify patterns of ridership concentration and evaluate the distributional equity of transit services.

In a key example, Li et al. [4] assessed the alignment between passenger demand and bus service supply in Wuhan using KDE and a Coupling Degree Index (CDI). Their findings revealed notable mismatches in some zones, emphasizing KDE's strength in diagnosing areas with poor demand-supply correspondence. Although their study primarily addressed static demand snapshots, it laid the groundwork for ratio-based evaluations of service imbalance, which this study extends through a jeepney-to-passenger ratio heatmap.

Building on spatial realism, the study by Ebrahimpoor et al. [8] introduced Network KDE in their analysis of Tehran's transportation system. By embedding KDE within road network constraints and employing adaptive bandwidths, their model generated hotspot maps aligned with actual travel corridors. While their focus was on emergency planning and critical location identification, the methodology has broader implications for urban mobility analysis, particularly for aligning service distribution with road accessibility and movement flows.

A study referenced as [3], while outside the domain of transportation, demonstrated the integration of temporal decay functions into KDE to enhance predictive accuracy. Their method incorporated time-sensitive adjustments to density calculations, offering a promising direction for capturing temporal fluctuations in commuter behavior. This informs our approach by reinforcing the importance of time-segmented KDE analyses in identifying when demand intensifies or declines.

Additionally, while this study does not employ clustering, insights from Babapourdijojin et al. [9] are relevant. Their research combined K-means clustering with spatial analysis to examine commuting patterns in Italian cities, uncovering systemic issues such as accessibility gaps and modal mismatches. Though their methods differ, the study supports the broader strategy of segmenting demand to identify underserved locations, which aligns with our study's goal of refining jeepney allocation.

Altogether, these works inform our methodological design by validating KDE as a robust tool for visualizing and diagnosing public transport demand imbalances. Our project builds upon these foundations by implementing KDE not only to generate spatial density maps but also to quantify supply-demand mismatches, monitor temporal demand fluctuations, and guide corrective actions in jeepney deployment.

3 Methodology

This study employs Kernel Density Estimation (KDE) to model spatial and temporal passenger demand for Public Utility Vehicles (PUVs) in Cagayan de Oro City. The core methodology is implemented in Python, where KDE is used to convert discrete passenger locations into continuous demand surfaces. These spatial outputs are then integrated into an interactive frontend application developed using JavaScript, HTML, and CSS, allowing for real-time visualization and user interaction with the demand heatmaps.

The dataset used in this study, though currently synthetic, is structured to reflect realistic scenarios and will serve as the template for future real-world data collection. The primary data source is the `cdo_passenger_data.csv` file, which contains the following features:

- **category**: The functional type of the location (e.g., school, mall, office), representing the dominant commuter demographic.
- **orig_name**: The original or raw name of the hotspot location as captured from the source data.
- **name**: A standardized or cleaned identifier assigned to each passenger hotspot for consistent referencing.
- **longitude** and **latitude**: The geographic coordinates used both for spatial mapping and as input for KDE computation.
- **alley_capacity**: The maximum number of jeepneys that can simultaneously serve the location, reflecting physical infrastructure constraints.
- **jeep_count**: The number of jeepneys observed or estimated to pass through the location during the given time slot.
- **average_time_arrival_mins**: The average interval in minutes between jeepney arrivals, used as a proxy for congestion or service frequency.
- **total_passenger**: The number of passengers recorded or estimated at the location, used as the primary weight in KDE calculations.
- **time_slot**: A categorical variable indicating the time of day (e.g., morning, noon, evening), allowing temporal demand analysis.

The KDE model uses spatial coordinates and total passenger counts to generate density heatmaps that highlight demand intensity across various zones. The inclusion of time slots enables the modeling of temporal shifts in ridership patterns. GeoJSON files containing manually constructed jeepney routes, based on references such as Commute Tour Philippines, are overlaid onto these maps to provide contextual information about current PUV coverage. While these route layers are not part of the KDE computation, they enhance the interpretability of spatial demand imbalances.

The methodology proceeds through four main stages: **data acquisition**, **data preprocessing**, **KDE-based analysis**, and **interactive visualization**. Priority hotspots are manually observed for one full day to ensure reliable inputs, while synthetic data supplements less significant zones. KDE surfaces are validated visually and analytically against known spatial patterns to assess their plausibility and utility for guiding jeepney allocation strategies.

Table 1. Data Sources and Features for KDE Analysis

Dataset Source	Data	Description	Key Features Needed
CPDO (City Planning and Development Office)	Passenger Demand Locations	Geospatial data on key locations where high passenger demand is expected (e.g., malls, markets, universities, hospitals).	• Latitude & longitude of major landmarks (from SHP) • Zone classification (commercial, residential, mixed-use, etc.) (from DBF)
Research Team	Temporal Passenger Demand Data	Field-collected data on passenger movement in key urban areas. Includes passenger wait times, boardings, and alightings for different PUV types.	• Timestamped ridership counts • Passenger Demand Count
Commute Tour Philippines	Jeepney Route Data	Provides spatial data for jeepney routes, used to overlay demand data and evaluate route efficiency.	• Route Data (ID, Route Name, Geometry)

3.1 Problem Breakdown. Public Utility Vehicles (PUVs) remain the dominant mode of transport in Cagayan de Oro City; however, operational inefficiencies persist due to the absence of demand-driven route planning. The current mismatch between passenger demand and service coverage results in several issues: underutilized routes in low-demand areas, overcrowding in high-demand zones, prolonged passenger wait times, and financial strain on transport operators.

Although the Local Public Transport Route Plan (LPTRP) was introduced to address these inefficiencies, its implementation has encountered structural and behavioral barriers. These include the absence of designated loading and unloading zones, as well as operator hesitancy to serve routes with low ridership potential [1]. Consequently, significant portions of the city remain underserved, while others face service saturation, exacerbating traffic congestion and commuter dissatisfaction.

To respond to these challenges, this study proposes a structured, data-informed approach to characterize and analyze spatial and temporal ridership demand. Through the application of Kernel Density Estimation (KDE), the study identifies concentrations of passenger activity and reveals gaps in service coverage. These outputs serve as a foundation for more efficient PUV allocation, adaptive route planning, and responsive scheduling based on localized demand patterns.

3.2 Sub-Problems. This study focuses on addressing the following data-related gaps in public transport planning:

- **Lack of temporal and spatial insight into commuter demand shifts.** Current transport planning in Cagayan de Oro City fails to effectively track and visualize fluctuations in passenger demand across both time and space. Without dynamic tools that integrate spatial and temporal data, transport planners lack the

necessary insights to understand how commuter demand changes throughout the day. This results in inefficiencies such as overcrowding during peak hours, underutilization of services during low-demand periods, and overall misallocation of resources, leading to delays, increased waiting times, and suboptimal vehicle deployment.

- **Inability to visualize overserved and underserved areas based on demand.** Existing public transport frameworks in Cagayan de Oro City do not provide clear, demand-weighted geospatial visualizations to highlight areas that are either oversaturated or underserved by jeepney services. Without heatmaps or spatial tools that visualize demand patterns, it becomes challenging to identify regions with high ridership potential that lack sufficient coverage, as well as areas with an excess of jeepney services. This limits the effectiveness of current transport planning and prevents efficient allocation of vehicles to areas where they are most needed.
- **No data-driven decision-making framework for route optimization.** There is a lack of an integrated, data-driven system that combines both spatial and temporal passenger demand data with route network constraints for proactive decision-making. As a result, transport agencies are unable to make evidence-based decisions for jeepney deployment or route adjustments, relying instead on static and outdated information. This deficiency results in route inefficiencies, as the current planning process does not account for real-time demand fluctuations or the spatial dynamics of passenger movement, ultimately leading to poor service alignment and commuter dissatisfaction.

3.3 Solutions.

3.3.2 Implementation of Kernel Density Estimation (KDE) with Temporal Data. To address the lack of spatial and temporal insight into passenger demand, this study will apply Kernel Density Estimation (KDE) to generate heatmaps of passenger demand at specific locations across Cagayan de Oro City. The KDE will be applied to data segmented by predefined time slots, specifically 8 AM, 10 AM, 12 PM, 3 PM, and 5 PM, capturing the fluctuations in demand throughout the day. This time-based segmentation will allow for the visualization of demand shifts, providing a clearer picture of how passenger demand varies during different times of the day.

3.3.2.1 Kernel Density Estimation for Demand Hotspots.

KDE will be applied to the preprocessed passenger demand dataset, where each data point represents a public utility vehicle stop with an associated ridership count. To ensure the density estimation reflects actual ridership levels, this study will utilize Weighted Kernel Density Estimation (WKDE). In WKDE, stops with higher passenger counts will contribute more significantly to the density map. The WKDE function will be defined as follows:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n w_i K\left(\frac{x-x_i}{h}\right)$$

where:

- $\hat{f}(x)$ is the estimated density at location x ,
- n is the number of data points,
- h is the bandwidth (smoothing parameter),
- K is the kernel function (e.g., Gaussian).

3.2.2.2 Bandwidth Selection & Optimization. The bandwidth (h) is a crucial parameter that controls the level of smoothing. A smaller bandwidth results in more localized estimates, while a larger bandwidth gives smoother, broader estimates. Bandwidth selection can be done using methods like Silverman's method for automatic bandwidth selection or cross-validation to minimize the estimation error.

To balance bias and variance, this study employs Silverman's Rule of Thumb, a widely used automatic bandwidth selection method that adapts to data distribution. The optimal bandwidth h is computed as follows:

$$h = 0.9 \times \min\left(\hat{\sigma}, \frac{IQR}{1.34}\right) \times n^{-\frac{1}{5}}$$

where:

- h = bandwidth (smoothing parameter)
- $\hat{\sigma}$ = standard deviation of the dataset
- **IQR** = interquartile range (Q3 - Q1)
- n = number of data points

This formula optimizes bandwidth by selecting the smaller of the standard deviation or interquartile range, preventing extreme values. While Silverman's Rule is generally effective, it assumes a near-normal data distribution, which may not apply to skewed or multimodal datasets. Alternative methods like cross-validation offer more flexibility but come with higher computational costs. For public transport ridership data, Silverman's Rule strikes a balance between accuracy and efficiency.

3.3.4 Data-driven Spatial Analysis for Passenger Maximization and Service Coverage Improvement Building on the visualization and clustering outputs, this study conducts a detailed spatial analysis to derive actionable insights from the identified patterns and imbalances. By examining the spatial distribution of passenger demand and jeepney coverage, we aim to provide evidence-based recommendations for enhancing jeepney route planning and formulation of transportation policies. These recommendations will assist in optimizing service coverage, addressing underserved areas, and mitigating inefficiencies in the current PUV system.

3.3.4.1 Imbalance Ratio for Optimizing Jeepney Count Relative to Passenger Demand

A crucial component of this optimization process is identifying areas where jeepney services are either oversaturated or underserved. To evaluate this, an imbalance ratio is applied to measure the relationship between the number of passengers and the frequency of jeepney arrivals at a specific location within a defined time interval. This ratio accounts for the total passenger demand, the number of jeepneys serving the location, their average arrival intervals, and the standard vehicle capacity. By applying this metric, it becomes possible to detect spatial and temporal mismatches between supply and demand, thereby guiding necessary adjustments in jeepney deployment to ensure a more balanced and efficient service distribution.

Let the variables be defined as follows:

- **D** : Total passenger demand
- **J** : Number of jeepney arrivals (Jeep_Count)
- **A** : Alley capacity (number of jeeps that can load simultaneously)
- **T** : Time slot duration (in minutes)

- t : Dwell time per jeepney (in minutes)
- C : Passenger capacity per jeepney
- λ : Load factor (proportion of passengers carried per jeepney)
- β : Baseline adjustment factor

Step 1: Max Boarding Capacity

$$B_{max} = A \times \left(\frac{T}{t} \right)$$

Step 2: Effective Jeep Arrivals

$$J_{eff} = \min(J, B_{max})$$

Step 3: Passenger Supply

$$S = J_{eff} \times C \times \lambda$$

Step 4: Adjusted Supply

$$S_{adj} = S \times B$$

Step 5: Imbalance Ratio

$$Imbalance\ Ratio = \frac{D - S_{adj}}{D + S_{adj} + 1}$$

3.3.4.2 Time-Based Demand Shifts

Temporal segmentation was applied exclusively to the imbalance ratio analysis to capture fluctuations in the balance between supply and demand throughout the day. This approach highlighted time-sensitive routing opportunities, such as implementing one-way routes during peak hours or increasing jeepney frequency in consistently underserved areas.

3.3.4.3 Adjustment of Jeepney Deployment Based on Time-Specific Demand

Based on the findings from spatial analysis, the study provides actionable insights for improving jeepney deployment. Underserved areas will be identified for potential route expansions, while overserved zones will highlight the need for vehicle redistribution. The analysis of temporal demand fluctuations will inform scheduling adjustments, ensuring more efficient and dynamic jeepney operations that match actual passenger demand patterns.

3.4 Implementation

3.4.1 Initial Approach and Iteration. This report of our initial progress documents the progress of data preparation for the transport demand analysis system in Cagayan de Oro (CDO). The project's initial phase involved obtaining geospatial data, processing it into a structured format, and ensuring its compatibility for subsequent analysis. This report outlines the data acquisition process, the methods used for processing, and the next steps in the project implementation.

3.4.1.1 Data Acquisition. The project commenced with the collection of spatial data from the City Planning and Development Office (CPDO). The CPDO provided shapefiles containing geospatial records of key establishments in CDO, specifically:

- Shopping malls
- Universities
- Public markets
- Transportation terminals

These shapefiles included both polygon and point geometries, which represented the locations of these establishments. The objective was to consolidate these records and convert them into a structured format suitable for clustering and visualization.

3.4.1.2. Data Processing and Transformation. All data processing and transformation steps were executed in Jupyter Notebook to facilitate interactive data manipulation and visualization.

3.4.1.3 Data Transformation. To consolidate the spatial datasets provided by CPDO, we defined the directory containing the shapefiles and listed the specific files corresponding to key establishments in Cagayan de Oro, including universities, shopping malls, public markets, and transport terminals. By organizing the file paths systematically, we ensured efficient access to multiple datasets, allowing for streamlined data integration and processing.

3.4.1.4 Standardizing Spatial Data. The geospatial data was standardized by iterating through the provided shapefiles and converting them into a structured format. Each file was processed as follows:

- **Reading and Reprojecting Data:** The shapefiles were loaded using GeoPandas and converted to the WGS84 coordinate system (EPSG:4326) to ensure consistency in longitude and latitude values.
- **Extracting Spatial Information:** The centroid of each feature was calculated to obtain the longitude and latitude coordinates, allowing for point-based spatial representation.
- **Categorization and Data Cleaning:** Each record was assigned a category based on the filename, distinguishing between universities, malls, public markets, and transport terminals. Only essential columns (name, category, longitude, latitude) were retained to maintain relevance for analysis.

Consolidation of Data

The processed datasets were appended to a list for subsequent merging, ensuring systematic integration into a single structure. This step allowed the shapefile data to be transformed into a format that is easy to analyze for spatial clustering and demand estimation.

3.4.1.5 Merging and Exporting Data. After processing individual shapefiles, the extracted spatial data was merged into a unified dataset. Using Pandas, all dataframes were consolidated to create a dataset that integrates the locations of universities, malls, public markets, and transport terminals in Cagayan de Oro. The dataset was then exported as `cdo_passenger_data.csv`, ensuring accessibility for further analysis, including spatial clustering and temporal demand estimation.

Saving the data in CSV format enables seamless integration with analytical tools and programming environments, which will facilitate future refinement and expansion.

3.4.1.6 Synthetic Data Generation. To enhance the dataset, synthetic data was generated and added to `cdo_passenger_data.csv`. This step was necessary to simulate passenger demand patterns until real-world data becomes available. Random values were assigned to represent passenger counts at different times of the day, including peak and off-peak hours. Additional attributes were introduced, such as the number of passengers boarding and alighting from both traditional and electric jeepneys, as well as total waiting time.

Incorporating these synthetic values made the dataset more representative of actual transport activity, enabling preliminary analysis and model testing. Real-world passenger data will eventually replace these synthetic values to improve demand estimation accuracy.

3.4.1.7 Transition to Implementation. With data processing and transformation steps completed in Jupyter Notebook, we now have a structured dataset that integrates key spatial information and simulated passenger demand. The `cdo_passenger_data.csv` file is ready for further analysis, including spatial clustering and demand estimation. The next phase will involve transitioning to VS Code for system implementation, where we will develop the necessary scripts and models for analyzing the dataset and generating actionable insights for transport planning in Cagayan de Oro.

3.4.1.8 System Implementation in VS Code. With the data processing and transformation steps completed, we now shift our focus to system implementation using Node.js and JavaScript within a VS Code development environment. This section covers the setup of necessary dependencies, server configuration, and clustering logic for analyzing jeepney passenger demand in Cagayan de Oro.

File: package.json

Basic Project Information. The project configuration begins with basic metadata. The project name, `cdo-jeepney-analysis`, clearly identifies it as a specialized application for analyzing jeepney operations in Cagayan de Oro. The version is set to `1.0.0`, representing the initial stable release. The description specifies that it is an interactive web application for analyzing jeepney passenger demand in CDO. The main entry point is designated as `server.js`, establishing the primary execution file.

Script Commands. The scripts section defines two essential execution modes. The "start" command runs `server.js` directly using Node.js, suitable for production environments. The "dev" command uses nodemon to run `server.js`, providing automatic server reloading when code changes are detected, which is especially useful during development.

Core Dependencies. The dependencies section outlines the essential packages required for the application's operation. These include:

- `@turf/turf` for advanced geospatial analysis.
- `cors` for enabling cross-origin resource sharing.
- `csv-parse` and `papaparse` for robust CSV file handling.
- `dbscan` for implementing density-based spatial clustering algorithms.
- `ml-kmeans` for performing K-means clustering.

- Express for the web application framework.
- leaflet.markercluster for map clustering functionality.
- dotenv for managing environment variables.
- proj4 for coordinate system transformations.

Development Dependencies. For development, nodemon is included as the sole development dependency. It streamlines the development process by automatically restarting the server when code changes are detected, improving the workflow.

File: server.js

Core Server Setup. This section sets up the core server using Express.js to handle backend operations. The cors module allows cross-origin requests, ensuring accessibility from various sources. The path and fs modules help locate and read the cdo_passenger_data.csv file, which contains key passenger data. Papaparse is used for efficient CSV parsing, while proj4 ensures accurate coordinate transformations.

Coordinate System Configuration. This section defines the necessary coordinate systems for mapping passenger demand in Cagayan de Oro. The utmProjection variable sets up a Universal Transverse Mercator (UTM) projection for Zone 51N, suitable for Mindanao. The wgs84Projection defines the widely-used WGS84 geographic coordinate system.

Data Processing Functions. Functions are defined for processing and validating coordinate data. The convertCoordinates function transforms UTM coordinates into WGS84, ensuring spatial accuracy.

Data Aggregation System. The loadCSV function is responsible for reading and processing passenger demand data from a CSV file. It uses asynchronous handling to ensure smooth execution and data parsing.

API Endpoints. The API provides two key endpoints:

- /api/data: This endpoint loads passenger demand data from a CSV file and returns it as a JSON response.
- / (Main Page): This endpoint serves the index.html file from the public directory.

Server Management. The startServer function ensures that the server starts on an available port, automatically incrementing the port number if necessary.

File: clustering.js

This section handles clustering analysis using K-means and DBSCAN algorithms. The performKMeans function applies the K-means algorithm to group geographic points, while performDBSCAN uses density-based spatial clustering.

File: app.js

Map Initialization and Base Configuration. This section initializes the map, centered on Cagayan de Oro at coordinates (8.4815°N, 124.6472°E) with a zoom level of 13, using OpenStreetMap tiles for the base layer.

Global State Management. The Global State Management system defines variables for managing heatmap visualizations, route overlays, and passenger demand data, optimizing memory usage and ensuring smooth interactive map functionality.

Time Control System. A time control system is implemented using a slider interface, allowing users to select the time range for visualizing demand patterns.

Demand Analysis Functions. The `getHourlyDemand` function processes demand data, ensuring data integrity and validation for spatial analysis.

Heatmap Visualization System. The `updateHeatmap` function visualizes demand data dynamically, integrating real-time updates and error handling for improved visualization.

Road Network Integration. The `loadRoadNetwork` function retrieves and processes road network data from OpenStreetMap, facilitating visualization and routing.

Animation Control System. The `togglePlay` function controls the playback of animations, ensuring proper resource cleanup and functionality.

Disclaimer: Project Development History and Justification of Methodological Shift

The development of this project initially progressed with a strong sense of direction and confidence. In the early stages, preliminary reports were submitted that reflected a solid understanding of the goal and the tools available, which led the team to believe that the project was on a successful trajectory. Early implementations focused on clustering methods such as K-Means and DBSCAN. While these models were technically functional and produced outputs based on spatial demand clustering, the team quickly recognized a fundamental issue: although these models fell under the same broader category of spatial analytics, their purpose did not align well with the objectives of Kernel Density Estimation (KDE). Specifically, K-Means and DBSCAN focus on discrete cluster identification, whereas KDE aims to generate continuous spatial density surfaces. This conceptual mismatch led to difficulties in synthesizing their outputs meaningfully within a unified analytical framework.

Another significant obstacle arose in the implementation of KDE using JavaScript. The process of bandwidth selection, which is crucial to the quality and interpretability of KDE outputs, proved especially difficult in this environment due to the limited statistical tooling available. Although a working KDE visualization was achieved, the absence of rigorous bandwidth optimization severely limited its analytical reliability.

Parallel to these challenges, the team also sought to integrate jeepney route information into the spatial framework. These routes were manually constructed in QGIS and exported as GeoJSON files. While the resulting layers were successfully rendered as geographic features, it became evident that the format posed limitations in terms of embedding and querying essential attributes, such as jeepney identifiers and capacities. This hindered further plans to use route data as a dynamic input to supply modeling.

In revisiting the KDE methodology, the team realized that due to the unsupervised and continuous nature of the model, traditional validation metrics were inapplicable. Instead, the recommended approach was to compare multiple bandwidth parameters and assess their relative performance in capturing meaningful patterns. However, this comparative evaluation has not yet been executed due to the earlier constraints. Additionally, the proposed use of Moran's I and its associated significance metrics (e.g., z-scores or pseudo p-values) to assess spatial autocorrelation has not yet been implemented. Given these cumulative limitations and the fragmented progress across different implementation layers, the team has decided to restructure the workflow.

3.4.2 Restructured Workflow Approach.

Urban mobility in mid-sized Philippine cities faces increasingly complex challenges arising from population growth, infrastructure constraints, and fluctuating commuter demand. To support data-informed transportation planning, this project introduces an interactive dashboard for passenger density analysis in Cagayan de Oro. Developed using the Streamlit framework, the platform enables real-time visualization, imbalance diagnosis, and infrastructure-aware service modeling, with the goal of improving jeepney route efficiency and passenger experience.

3.4.2.1 System Architecture

Core Technologies. The system is built on **Python** and integrates several key libraries. Data handling is managed using **pandas** and **numpy**, while **folium** and **streamlit_folium** enable geographic visualization. Geospatial projection transformations are performed using **pyproj**, and clustering functionality is made available through **scikit-learn**. These components are orchestrated through Streamlit, which allows for a responsive and interactive front-end experience suitable for both technical and planning stakeholders.

Coordinate Reference Transformation. The original dataset utilizes **EPSG:3125**, the Philippine Reference System, which is incompatible with standard web mapping libraries. To address this, all coordinates are programmatically transformed into the WGS84 system (EPSG:4326) using **pyproj.Transformer**. This step ensures consistency across spatial visualizations and accurate alignment with base map layers.

```
transformer = Transformer.from_crs("EPSG:3125", "EPSG:4326")
lat, lon = transformer.transform(df['longitude'], df['latitude'])
df['lat_wgs84'], df['lon_wgs84'] = lat, lon
```

3.4.2.2 Data Processing and Temporal Management

Data Validation and Cleaning. Robust preprocessing ensures the reliability of the underlying data. Records with null values in spatial coordinates or passenger counts are systematically removed. Additionally, schema conformity is enforced by checking the presence of essential columns, including **time_slot**, **jeep_count**, and **alley_capacity**. This step is critical for maintaining consistency across downstream calculations.

Time Slot Structuring. Temporal analysis is enabled through predefined time segments representing distinct commuting phases. These include:

- 8:00:00 AM to 10:00:00 AM
- 10:00:00 AM to 12:00:00 PM
- 12:00:00 PM to 3:00:00 PM
- 3:00:00 PM to 5:00:00 PM
- 5:00:00 PM to 7:00:00 PM

Users interactively select these time slots via a graphical slider, which dynamically filters the dataset and triggers downstream recalculations in real time.

3.4.2.3 Imbalance and Capacity Modeling

Imbalance Ratio Computation. The core analytic engine of the dashboard is the imbalance ratio metric, designed to quantify service mismatches between passenger demand and jeepney supply. Three distinct computational approaches are supported, with the “Improved Ratio” set as the default.

Improved Ratio:

$$\text{imbalance_ratio} = \frac{\text{total_passenger} - \text{adjusted_supply}}{\text{total_passenger} + \text{adjusted_supply} + 1}$$

Simple Ratio:

$$\text{imbalance_ratio} = \frac{\text{total_passenger}}{\text{adjusted_supply} - 1}$$

Demand-Supply Gap:

$$\text{imbalance_ratio} = \frac{\text{total_passenger} - \text{adjusted_supply}}{\text{max_demand}}$$

Each method is parameterized by load factor assumptions, jeepney dwell time, total capacity, and temporal span. This enables flexible adaptation to both empirical observations and planning scenarios.

Infrastructure-Constrained Supply Estimation. Passenger supply is not computed solely from jeepney counts, but adjusted using the alley capacity, dwell time, and effective jeepney arrivals. The supply formula integrates physical infrastructure limits with behavioral assumptions regarding occupancy rates.

$$\begin{aligned} \text{df['max_boarding_capacity']} &= \text{df['alley_capacity']} * (\text{time_slot_duration} / \text{dwell_time_per_jeep}) \\ \text{df['effective_jeep_arrivals']} &= \text{np.minimum}(\text{df['jeep_count']}, \text{df['max_boarding_capacity']}) \end{aligned}$$

```
df['passenger_supply'] = df['effective_jeep_arrivals'] * passenger_capacity * load_factor
```

This model ensures that service capacity assessments are not decoupled from spatial constraints, making the tool particularly suitable for infrastructure-aware analysis.

3.4.2.4 Visualization Design

Heatmap and Marker Display. Spatial insights are rendered through a dual-layer visualization system. A Folium-based heatmap depicts kernel-weighted passenger densities, while CircleMarkers reflect location-specific metrics such as imbalance ratio and estimated jeepney throughput. Users can toggle these layers independently and customize visual parameters such as blur, radius, and gradient mapping.

Statistical Dashboard and Classification. A sidebar-driven dashboard summarizes key metrics, including total passenger count, average per location, and current service classification distribution. The imbalance ratio classification follows this schema:

- **Significantly Overserved:** < -0.99
- **Slightly Overserved:** -0.99 to -0.1
- **Well Balanced:** -0.1 to $+0.1$
- **Slightly Underserved:** $+0.1$ to $+0.3$
- **Significantly Underserved:** $> +0.3$

This classification provides urban planners with actionable insights on route optimization and deployment timing.

3.4.2.5 Interface and Export Features

User Controls. Users are offered an array of customization tools, including time slot sliders, map tile selection (e.g., OpenStreetMap, CartoDB), heatmap parameters, and imbalance method switching. These features support exploratory analysis and fine-tuned reporting.

Data Export and Tabular Review. All filtered and computed data are available for export in CSV format, appended with time stamps to aid in archiving. Additionally, users can view raw data tables, limit column visibility, and apply real-time filters for targeted analysis.

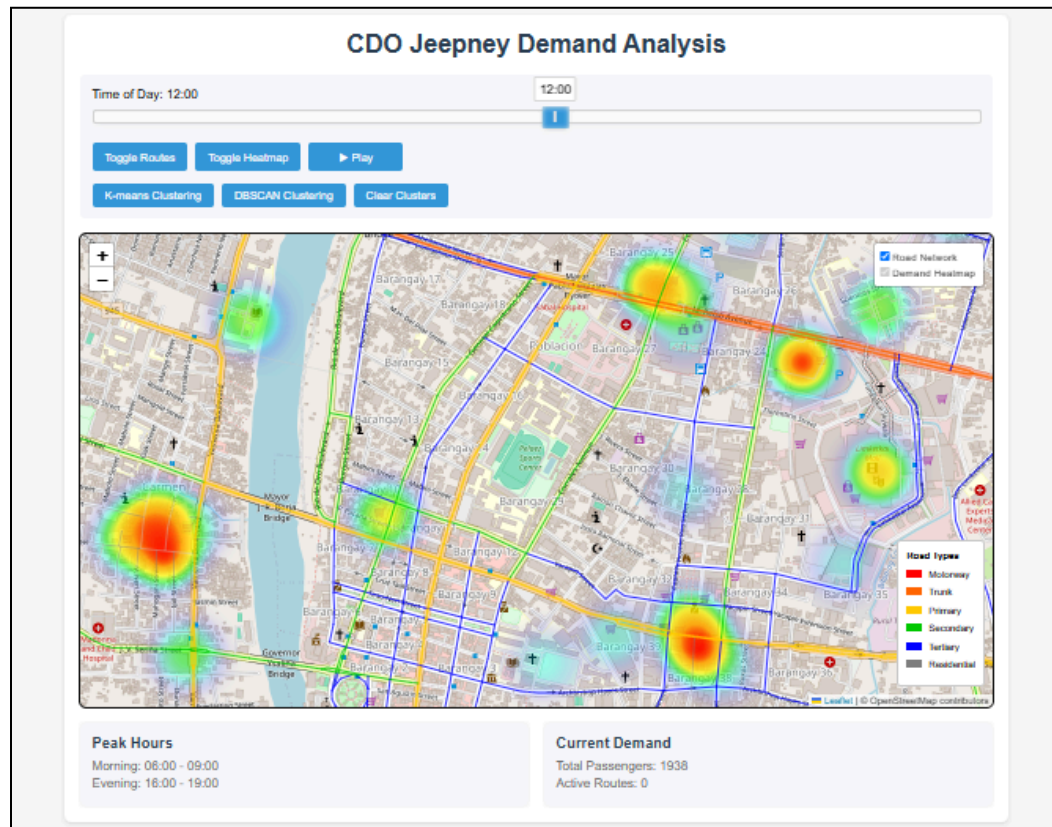
3.4.2.6 Quality Control and Performance.

The implementation employs caching (**@st.cache_data**) to reduce redundant computation and maintain responsiveness. Error handling routines detect missing files, null values, and column misconfigurations. All critical numeric operations are guarded against division-by-zero errors and clipped to prevent infinite values.

4 Results and Discussion

4.1 Outcomes. This study has yielded several significant results, demonstrating progress towards the development of a comprehensive transport demand analysis system for Public Utility Vehicles (PUVs) in Cagayan de Oro City.

4.2.1 KDE Heatmap Implementation.



The Kernel Density Estimation (KDE) heatmap has been successfully integrated into the system, providing a clear visualization of passenger demand patterns across the city. The heatmap enables transport planners to identify high-density and low-density areas effectively. A limitation remains with the time slider functionality, as the dataset currently contains data for only five specific intervals—6 AM, 8 AM, 12 PM, 5 PM, and 8 PM. Selection of unavailable times results in a blank heatmap, necessitating either the expansion of the dataset to cover all 24 hours or modification of the time slider to reflect the available intervals.

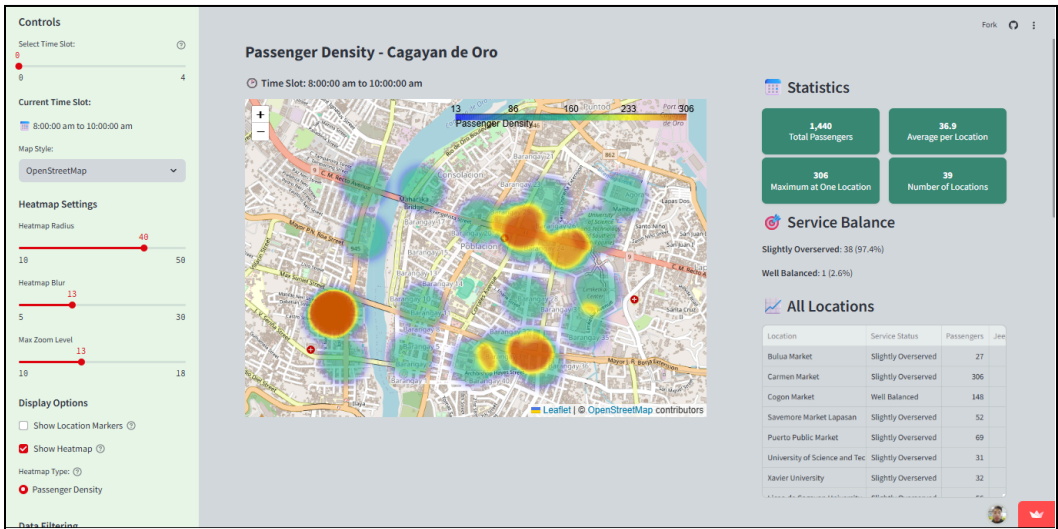
4.1.2 System Functionalities.



The system has achieved stable operation on localhost, resolving previous technical challenges caused by erroneous EPSG coordinate pairings. Road networks have been successfully overlaid on the map to represent the city's transport infrastructure, laying the foundation for route-based analysis. While map functionalities, including KDE heatmaps, K-Means clustering, and DBSCAN clustering, have been partially implemented, further debugging is required to ensure full functionality.

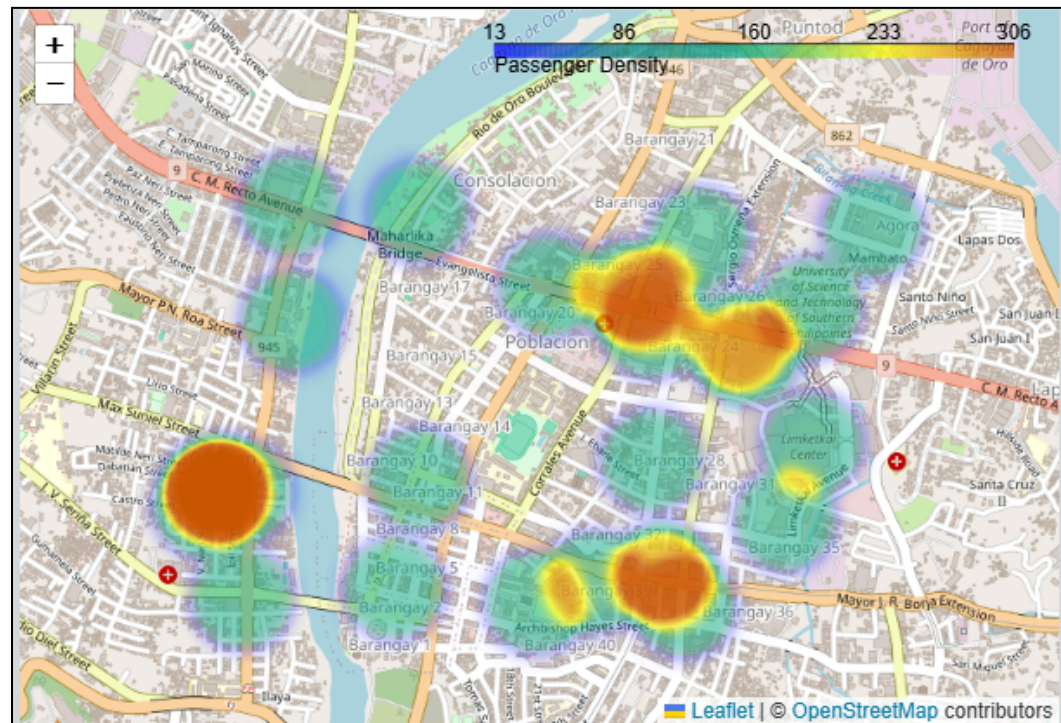
4.1.3 Proposed Interactive System Features. A key output of the system is its interactive functionality. This includes the generation of customizable maps, allowing users to toggle heatmap types, modify display parameters such as time and vehicle type, and interact with density points to access or edit attributes like passenger demand. While the interactive feature is not yet fully operational, progress is underway to integrate clickable and editable data points into the system.

4.2 Restructured Workflow Outcomes. This system presents a web-based geospatial analysis tool developed using Streamlit and deployed on Render. The platform is designed to identify spatial and temporal patterns in Public Utility Vehicle (PUV) demand and service imbalance within Cagayan de Oro City. By integrating Kernel Density Estimation (KDE), time-based imbalance metrics, and interactive mapping, the tool offers a comprehensive environment for visualizing jeepney passenger distribution and corresponding service adequacy across urban zones.



4.2.1 Spatial Demand Identification via KDE Heatmaps. The core visualization tool of this study is a spatially-grounded KDE heatmap built using the Folium library, overlaid on an interactive web-based map. By default, OpenStreetMap serves as the base layer, although

alternatives such as CartoDB Dark Matter and CartoDB Positron are available for varied visual preferences.

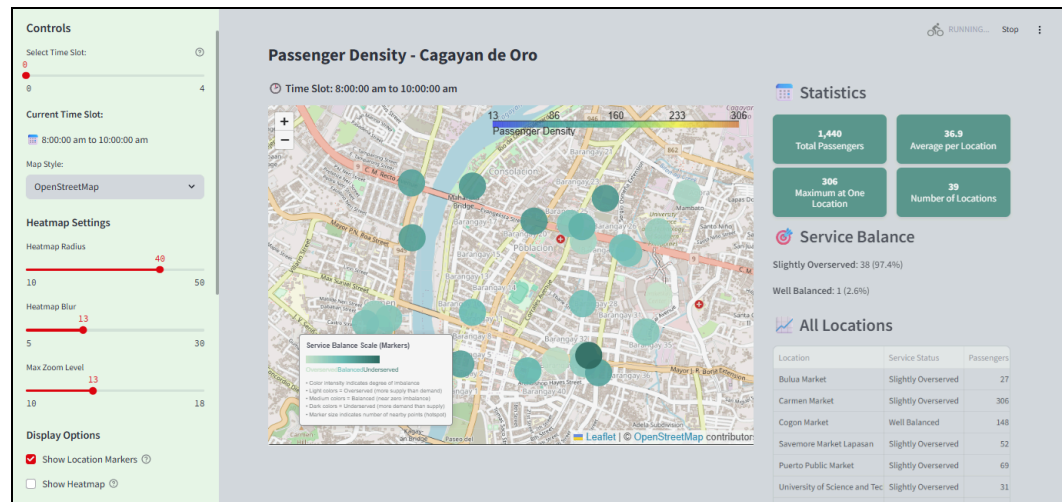


The KDE heatmaps effectively highlight zones of high passenger concentration derived from geocoded data collected around malls, terminals, schools, and other urban activity centers.

Users are afforded additional control through adjustable parameters—heatmap radius, blur intensity, and map zoom level. These controls, though represented in arbitrary slider units due to Folium's interface constraints, provide meaningful interpretive flexibility. They support manual bandwidth approximation, allowing users to refine the heatmap's density rendering and focus on areas of interest with different spatial granularities.

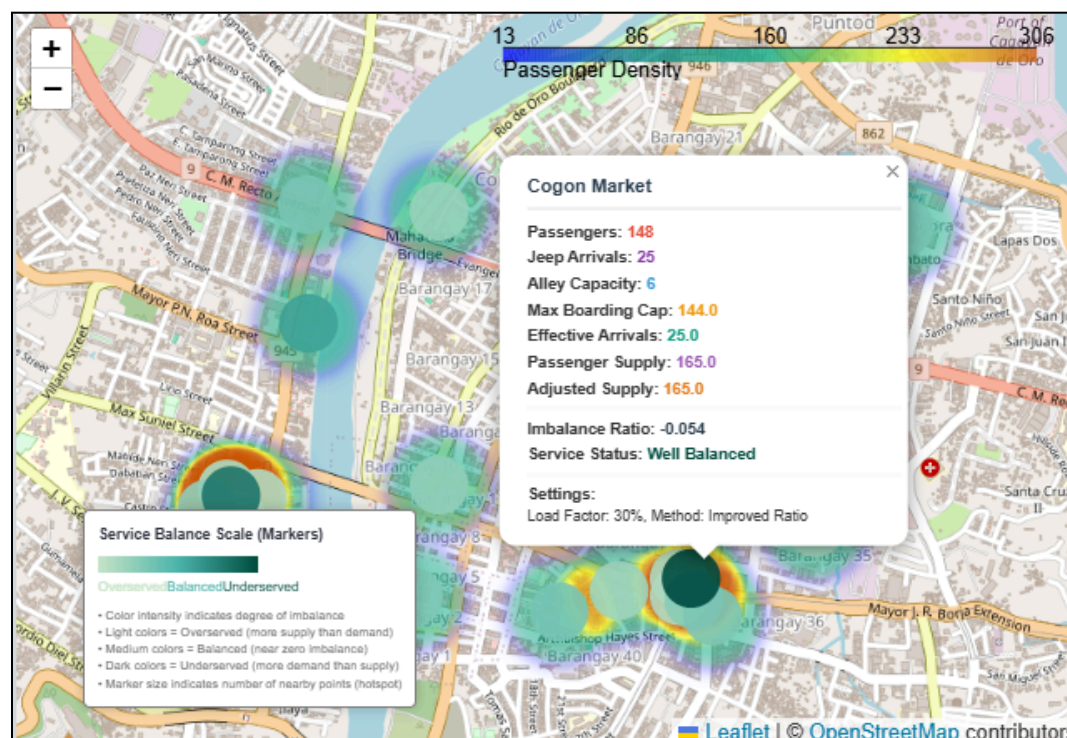
This spatial visualization successfully fulfills the objective of identifying zones of high passenger demand, setting the foundation for deeper analysis into service coverage and imbalance.

4.2.2 Temporal Service Imbalance Mapping. Beyond spatial analysis, the system introduces a dynamic temporal component through an **imbalance ratio map**, which evaluates the relationship between jeepney availability and passenger volume across defined time slots. This is implemented using a slider tool that enables users to explore service patterns at different periods of the day. Unlike the KDE heatmap, which is strictly spatial, the imbalance analysis incorporates time-based shifts in both supply and demand, providing insight into when and where service mismatches are most pronounced.



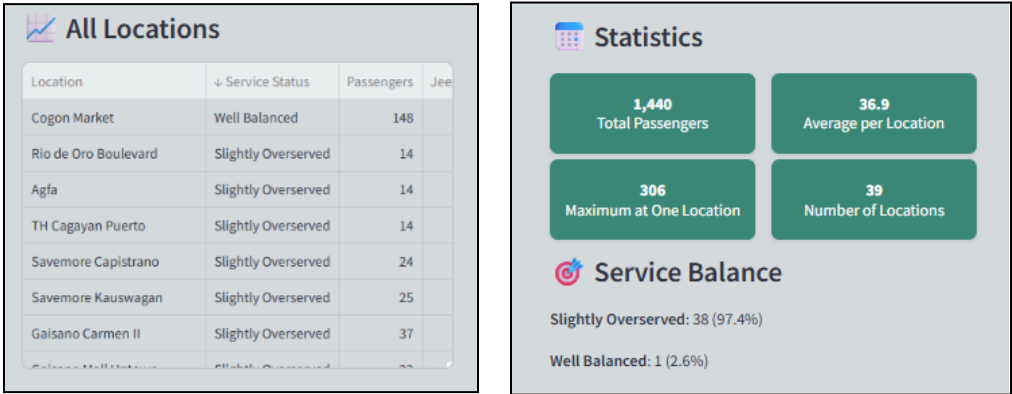
The imbalance is visualized through a color-coded mapping of location markers. These markers are styled in a gradient of blue-green, where darker green denotes areas that are likely underserved and lighter shades indicate potential oversupply. This allows immediate visual interpretation of the spatial-temporal service status at each location.

4.2.3 Data Point Markers and Interactive Feedback. To avoid visual clutter, **location markers** are disabled by default, but can be toggled on via a checkbox interface. When enabled, each marker becomes an interactive point of information. Users can click on any data point to access its full metadata, including timestamp, passenger volume, estimated jeepney count, and service classification (underserved, optimal, or overserved). This feature not only improves user engagement but also deepens the interpretability of the spatial and temporal datasets.



The graduated color scheme of the markers adds a further interpretive layer, reinforcing the classification of service status at a glance.

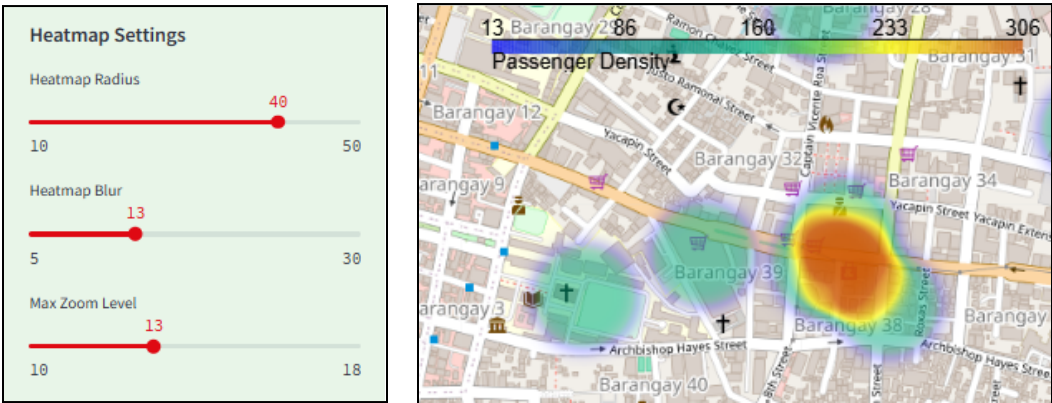
4.2.4 Tabular Visualization and Service Status Filtering. Complementing the map-based outputs, a comprehensive **data table** lists all collected and processed points, including their computed service status. Although the entries are not ranked by default, users can sort them interactively by clicking the column headers—particularly the service status column. This makes it possible to immediately identify the most underserved or overserved locations, aiding qualitative interpretation and potential decision-making.



Additionally, a **summary panel** displays the total count of points falling into each service classification category. These summary metrics are time-dependent and update in real time based on the selected time slot. This reinforces the non-static nature of the data and highlights how demand and service levels evolve throughout the day.

4.2.5 Synthesis and Implications. The integrated system developed in this study demonstrates a robust, multi-layered framework for identifying jeepney passenger demand and service imbalance in Cagayan de Oro City. Through a combination of KDE-based spatial analysis and temporally responsive imbalance mapping, the platform moves beyond simple data visualization toward actionable spatial-temporal insight.

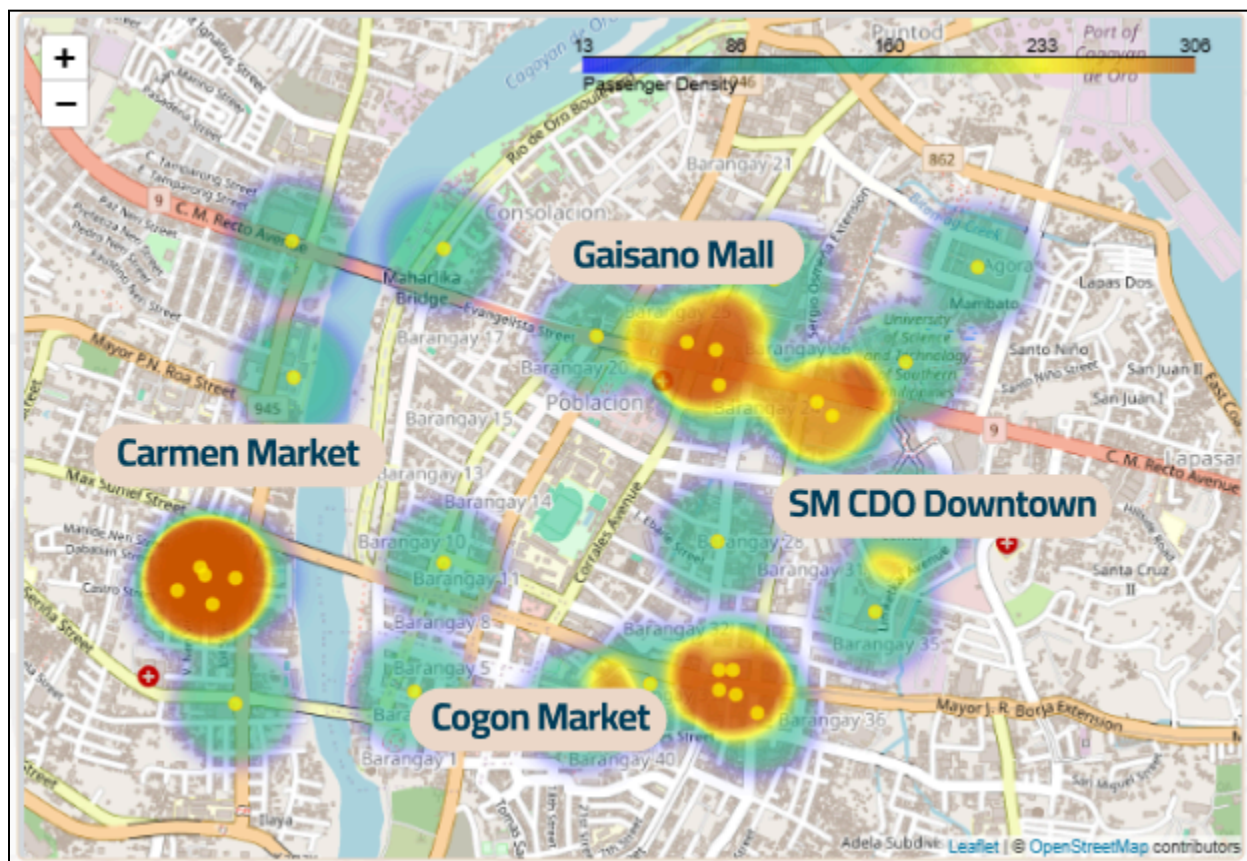
Although the bandwidth parameters for the KDE heatmap are not analytically calibrated but rather visually controlled, they still serve as an effective heuristic tool for pattern discovery. Meanwhile, the time-slot slider and dynamic imbalance metrics introduce a level of temporal resolution that would be difficult to achieve using static maps alone.



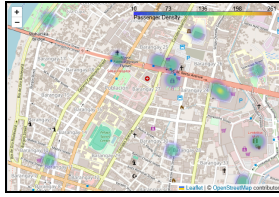
Taken together, the thematic tools presented provide a strong foundation for identifying mobility gaps in the city's traditional jeepney network. While the study does not directly prescribe optimization strategies, the system's outputs serve as a valuable diagnostic resource for policymakers, urban planners, and transport stakeholders.

5 Conclusion

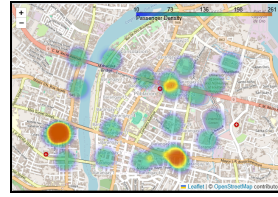
5.1 Synthesis of Analyses. The tool developed for this project, deployed via Streamlit and Render, demonstrates the potential of a geospatial approach to understanding jeepney service distribution in Cagayan de Oro City. Through the application of Kernel Density Estimation (KDE), a set of four spatially consistent high-demand hotspots emerged: Carmen, Cogon, Corrales, and Lapasan. These zones exhibited intense kernel overlaps due to both geographic proximity and high passenger volume, with Carmen peaking at 350 passengers and Cogon at 217 during the midday period. The spatial clustering observed—especially around major commercial and institutional hubs—confirms that even limited data coverage can yield valuable insights when urban features are selected carefully.



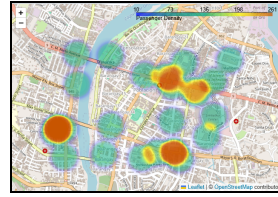
Manual bandwidth selection using Folium's HeatMap plugin, with a radius of 40 at zoom level 13, produced a balanced visual output. While not mathematically optimized, this setting offered a coherent compromise between detail and generalization. The result reflects realistic clustering without intensity distortion, making the map interpretable to both technical and non-technical users. Nonetheless, a sense of unease remains regarding the absence of a more systematic approach, particularly as existing statistical methods for bandwidth selection are often ill-suited to highly localized spatial contexts like this one.



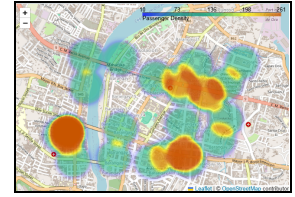
(a) radius = 20



(b) radius = 30

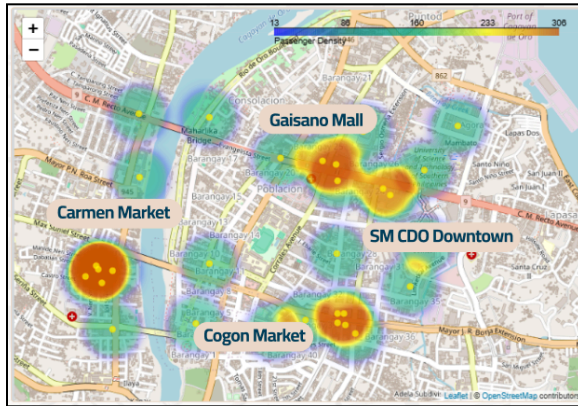


(c) radius = 40

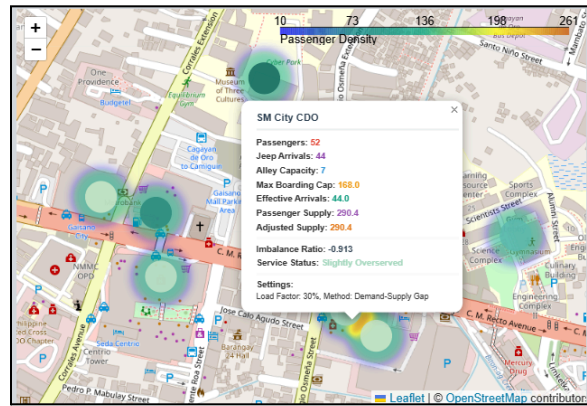


(d) radius = 50

Temporal imbalance analysis initially pointed to Cogon as a dynamic transport node, transitioning from balanced to underserved and back again across the observed time intervals. This inference, however, was primarily drawn under the improved ratio method, which upon closer evaluation may not reliably represent real-world demand-supply relationships. In contrast, the demand-supply gap method—introduced later—proved more consistent with patterns observed during manual data collection. While a full generalization remains difficult due to the limited and highly selective dataset, one recurring insight was particularly noteworthy: among the four expected red hotspots—Carmen, Cogon, Corrales, and Lapasan—SM Downtown appeared to diverge, exhibiting signs of underservice. This was unexpected, as spatially it aligns with high-demand clusters. A tentative explanation is its low jeepney accessibility due to traffic bottlenecks, which may discourage efficient boarding and alighting. However, this remains a hypothesis that warrants deeper spatial and operational investigation.



(a) The Four Consistent Red Hotspots;



(b) Actual Service Status of SM City CDO

Importantly, the imbalance ratio metric helped reveal systemic tendencies. That most zones trended toward overservice raises concern. If left unchecked, this surplus could lead to inefficiencies in route operations and unnecessary contributions to urban congestion. Even if not definitive, these findings offer a glimpse into latent structural imbalances that could inform future policy or planning adjustments.

5.2 Limitations and Recommendations. Several limitations temper the tool's current effectiveness. First, the KDE heatmap's spatial profile is heavily influenced by the clustering of select urban features—namely malls, universities, and markets—rather than a citywide distribution of data points. This narrow sampling introduces bias in identifying hotspots and constrains the ability to capture underserved zones that lie outside popular commercial centers.

Second, while the chosen heatmap radius and blur parameters produced visually effective results, they were tuned through trial-and-error rather than through a validated quantitative method. Although this may be defensible in applied geospatial work, it introduces subjectivity and limits reproducibility. Future versions of the system would benefit from integrating adaptive or data-driven bandwidth selection techniques tailored to urban spatial phenomena.

The imbalance ratio metric also rests on simplifying assumptions. Variables such as alley capacity, real-time passenger turnover, and jeepney availability were fixed or generalized, which limits the metric's interpretability under real-world variability. Additionally, the manual data collection process, constrained by limited manpower, may have introduced inaccuracies in counts and time stamping. These limitations could be addressed through automated or sensor-based data pipelines, which would enable higher-frequency, more spatially extensive input.

Despite these flaws, the tool marks a significant foundational step in making jeepney service patterns more interpretable and analyzable. It highlights the value of spatial-temporal diagnostics in transport planning and offers a functional framework that can scale with richer data and more advanced algorithms. With improved data access, broader urban coverage, and methodological refinement, this system can evolve into a robust decision-support platform for optimizing jeepney deployment and urban mobility more generally.

References

1. City Government of Cagayan de Oro: Council to Address Route CDO Plan Woes. [Online]. Available: <https://www.cagayandeorogov.ph/index.php/news-and-article/item/1859-council-to-address-route-cdo-plan-woes.html>. Last accessed: 2025/02/22.
2. Nogueira, A. C. C. S., & Diaz, C. E. D. (2023). Evaluating public transport accessibility and walkability in Pasig City (Philippines) and their correlation to socioeconomic factors at the Barangay level. *Asian Transport Studies*, 9, 100117. <https://doi.org/10.1016/j.eastsj.2023.100117>
3. Y. Hu, F. Wang, C. Guin, and H. Zhu: "A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation." *Applied Geography*, 99(5), 89–97 (2018). [Online]. Available: <https://doi.org/10.1016/j.apgeog.2018.08.001>.
4. B. Li, Z. Huang, J. Xia, W. Li, and Y. Zhang: "Coupling Degree between the Demand and Supply of Bus Services at Stops: A Density-Based Approach." In: *ISPRS International Journal of Geo-Information*, 10(3), 173 (2021). [Online]. Available: <https://doi.org/10.3390/ijgi10030173>.
5. Garcia, R. T. M. (2011). THE KERNEL DENSITY ESTIMATION FOR THE VISUALIZATION OF SPATIAL PATTERNS IN URBAN STUDIES. *International Multidisciplinary Scientific GeoConference SGEM* . . . <https://doi.org/10.5593/sgem2015/b21/s8.111>
6. Cai, X., Wu, Z., & Cheng, J. (2012). Using kernel density estimation to assess the spatial pattern of road density and its impact on landscape fragmentation. *International Journal of Geographical Information Science*, 27(2), 222–230. <https://doi.org/10.1080/13658816.2012.663918>
7. Yu, W., Ai, T., & Shao, S. (2015). The analysis and delimitation of Central Business District using network kernel density estimation. *Journal of Transport Geography*, 45, 32–47. <https://doi.org/10.1016/j.jtrangeo.2015.04.008>
8. Ghanbari, M., Karimi, M., Claramunt, C., & Lagesse, C. (2023). Identifying critical locations in urban transportation networks using network kernel density estimation (Net-KDE): A spatial analysis. *Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology*. Published June 14, 2023.
9. Babapour dijodin, M., Corazza, M. V., & Gentile, G. (2024). Systematic analysis of commuting behavior in Italy using K-Means clustering and spatial analysis: Towards Inclusive and Sustainable Urban Transport Solutions. *Future Transportation*, 4(4), 1430–1456. <https://doi.org/10.3390/futuretransp4040069>
10. Wang, Bin & Wang, Xiaofeng. (2007). Bandwidth Selection for Weighted Kernel Density Estimation