Interactive Exploration and Explanation of Spatio-temporal Anomalies with Graph-LLM Integration

Juanpablo Heredia*, Leighton Estrada-Rayme*, Jeremy Matos-Cangalaya† and Jorge Poco*
*School of Applied Mathematics, Fundação Getulio Vargas (FGV)
†Computer Science, Universidad de Ingeniería y Tecnología (UTEC)
{juanpablo.parillo, jorge.poco}@fgv.br, rayme.leighton.2025@fgv.edu.br, jeremy.matos@utec.edu.pe

***Abstract*—Understanding spatiotemporal anomalies is critical in domains such as urban safety, mobility, and environmental monitoring. These scenarios involve complex dynamics that are effectively modeled using graph-based representations, where the spatial structure is encoded through data connectivity, and each node corresponds to a time series. Anomaly detection in such data is crucial for identifying unusual or significant events, but it requires complex methods involving pattern recognition, prediction, and classification. Interpreting these anomalies remains challenging. To address this, we introduce an interactive system that combines spatiotemporal visualizations with Large Language Models (LLMs) to generate context-aware explanations by unifying temporal, spatial, and textual insights. We guide the LLM using a structured prompting strategy grounded in the data to reduce hallucinations and improve plausibility. As a demonstration of functionality, we analyze crime anomalies in São Paulo, uncovering links to events such as Carnival and religious holidays.**

I. Introduction

Spatiotemporal (ST) data are increasingly available across various domains, including urban monitoring, mobility, environmental sensing, and video analytics. These datasets support tasks like forecasting, classification, imputation, and anomaly detection [1]. Among these, anomaly detection is particularly valuable for uncovering unusual or critical events. However, most existing methods are black boxes, flagging outliers without providing insight into their causes or context [2], [3]. This limitation is especially problematic in areas like public safety or environmental analysis, where understanding anomalies is as important as detecting them. Although some systems offer trend visualizations, few integrate spatial and temporal reasoning with natural language explanations to support human decision-making [4], [5]. The resulting lack of interpretability undermines user trust and hinders the adoption of anomaly-detection models in real-world, complex scenarios.

To address this gap, we propose a hybrid framework that combines graph-based modeling, interactive visualization, and large language models (LLMs) for explaining anomaly contexts and generating hypotheses. The approach has three main components: (i) a graph-based representation that models ST data as a graph — nodes contain time series — capturing spatial structure and temporal variation; (ii) an interactive visual analytics interface in which users explore trends and anomalies within a coherent spatial and temporal context; and (iii) Possible explanations from an LLM prompted with localized ST context produce human-readable hypotheses and justify them via web search before displaying them. This graph abstraction mitigates sparsity by encoding neighborhood structure and relational context beyond simple spatial proximity.

Our setup builds on recent advances in prompting LLMs with structured data to reduce hallucinations and foster interpretable reasoning [6], [7]. The LLM serves not as the final decision-maker but as a generative companion that assists analysts in formulating and validating possible explanations.

Urban environments offer a particularly consequential setting for ST anomaly analysis. Crime data are especially challenging, extremely sparse, and shaped by complex social dynamics [8]. In this context, anomalies rarely represent mere statistical noise; they often signal substantive shifts — such as sudden surges in historically safe neighborhoods or unexpected lulls in known crime hotspots — making urban crime an ideal testbed for assessing the interpretability and real-world utility. Therefore, we apply our framework to weekly crime reports from São Paulo, one of the world's largest and most socially intricate cities. Two real-world scenarios (Carnival and major religious holidays) demonstrate how the system detects, contextualizes, and provides possible explanations using external evidence. We also held brief informal interviews with two external domain experts, who praised the system's novelty and usefulness, offering suggestions for improvement.

Our contributions are threefold: (1) a methodology for analyzing and explaining ST anomalies that integrates interactive contextualization, hypothesis generation, and validation; (2) an interactive visualization tool that supports exploration of ST data; and (3) case-based experiments showing that the framework yields interpretable, actionable insights for both experts and non-experts.

The code and other materials are accessible at https://visualdslab.com/papers/AnomalyExplain/.

II. Related Work

A. Spatiotemporal Anomaly Detection in Urban Contexts

Identifying anomalies requires first defining normal behavior. In urban settings, behaviors considered abnormal in one location or period may be typical in another [9]. ST patterns in cities — such as traffic flows or crowd dynamics — are shaped by both routine and irregular events. This variability demands context-aware approaches to anomaly detection [9], [10].

Anomaly detection is commonly tackled using machine learning or statistical methods, often through reconstruction-or forecasting-based strategies that identify deviations from expected patterns. Machine learning models typically achieve high predictive performance. At the same time, statistical approaches tend to offer greater interpretability—an essential feature in urban analytics, where understanding the underlying reasons behind detected anomalies is crucial [2].

Traditional techniques include matrix and tensor decompositions, as well as autoregressive models. Decomposition-based methods (e.g., PCA, NMF, tensor factorization) capture normal behavior via low-rank structures, treating deviations as anomalies [11], [12]. These are generally unsupervised and algebraic rather than purely statistical in nature.

In contrast, autoregressive models — such as STARMA [13] and STAR [14] — are grounded in statistical theory and model local ST dependencies for forecasting. Anomalies are detected when observed values deviate considerably from predictions.

B. LLMs and Urban Data Graphs

Recent studies have adapted LLMs to urban-data problems, where ST dependencies and scarce labeled samples present unique challenges [15], [16]. These models seek to generalize across diverse tasks, including traffic forecasting, crime prediction, and policy planning.

Graph-based methods provide a principled way to anchor LLMs in the relational urban context, enhancing relevance and reducing redundancy [17]. A common strategy is to linearize graph elements (nodes, edges, and paths) into text that an LLM can ingest. Fatemi et al. [6] show that LLM performance is highly sensitive to graph encoding choices, structural granularity, and task design. Complementing this, Perozzi et al. [18] introduce a GNN encoder that maps graph information directly into the LLM token space, enabling tighter alignment. Such techniques often pair with knowledge-graph retrieval via indexed subgraph reconstruction [7], [19]. In contrast, our system eschews global indexing, retrieving local ST context surrounding a user-selected node.

In the urban domain, UrbanGPT [15] and UrbanLLM [16] illustrate the promise of ST-aware LLM. UrbanGPT embeds ST dependencies and uses instruction tuning for zero-shot prediction, while UrbanLLM decomposes tasks and routes subtasks to specialized models. Inspired by these, we embed spatial and temporal features into a structured prompt that guides the LLM through step-by-step reasoning to generate context-aware hypotheses.

C. Urban Data Visualization and Interactive Analysis

Interactive visual analytics is key to deciphering complex ST patterns in urban data. Cities continually generate heterogeneous, high-velocity streams, and carefully designed visualizations help uncover correlations, trends, and anomalies across space and time. Deng et al. [20] emphasize that exploring urban causal relationships requires tools capable of handling noise, temporal variation, and spatial dependence.

Several systems target pattern discovery and decision support in ST contexts by integrating multi-level visualizations,

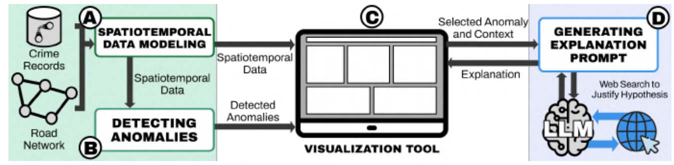

Fig. 1. Method pipeline: A) Model ST data using the crime records and urban road network. B) Use the ST data to calculate and flag anomalies. C) Receive constant data from A, B, and D to communicate via visualizations. D) Receive the ST context of the selected anomaly and return the LLM response.

coordinated views, and specialized techniques such as ST-Heatmaps [21]. A common focus is on human-computer interaction and interpretability [21], [22]. For example, *Crim-Analyzer* [22] pairs NMF-based hotspot detection with linked views to examine crime dynamics, while *JamVis* [21] aids congestion analysis of Waze alerts through clustering and visual summarization. Going beyond descriptive pattern identification, *Compass* [20] embeds causal inference by extending Granger causality to reveal dynamic relationships in urban time series, accompanied by tailored visual explanations.

Our work shares the aim of empowering domain experts through interactive exploration, but further augments these visual methods with LLM-guided reasoning, thereby fusing visual and linguistic interpretability within a single framework.

III. Methodology

Our workflow (Fig. 1) comprises four coupled components:

A. Spatiotemporal Data Modeling

Following Hassan et al. [23], we begin by selecting a study area in the target city and extracting its road network using `OSMnx` [24], resulting in an intersection-centric graph where nodes represent street intersections and edges represent street segments. Because our analysis focuses on street segments, we invert this representation: every original edge becomes a node, and two nodes are linked when their corresponding street segments intersect.

The graph is enriched with both temporal and spatial attributes from two sources: (i) available incident records — such as crime reports, traffic accidents, or 311 service calls — and (ii) amenity information available through `OSMnx`.

Incident data. Geocoded events are aggregated over a regular temporal interval Δt (e.g., daily, weekly, or monthly) chosen to balance resolution and sparsity. Each event is mapped to the nearest street-segment node, yielding a time series of incident counts per node. Public holidays and domain-specific events could be flagged to support richer contextual interpretation.

Amenity data. For each node, we compute the distance to the nearest facility of key categories (e.g., police stations, transit hubs, entertainment venues) within 200 m. These distances form a fixed-length static vector encoding local urban context.

B. Detecting Anomalies with STARMA

While any anomaly detection model could work, we prioritize interpretability and favor statistical approaches. Accordingly, we chose the Space-Time Autoregressive Moving-Average (STARMA) model, which extends ARMA/ARIMA to

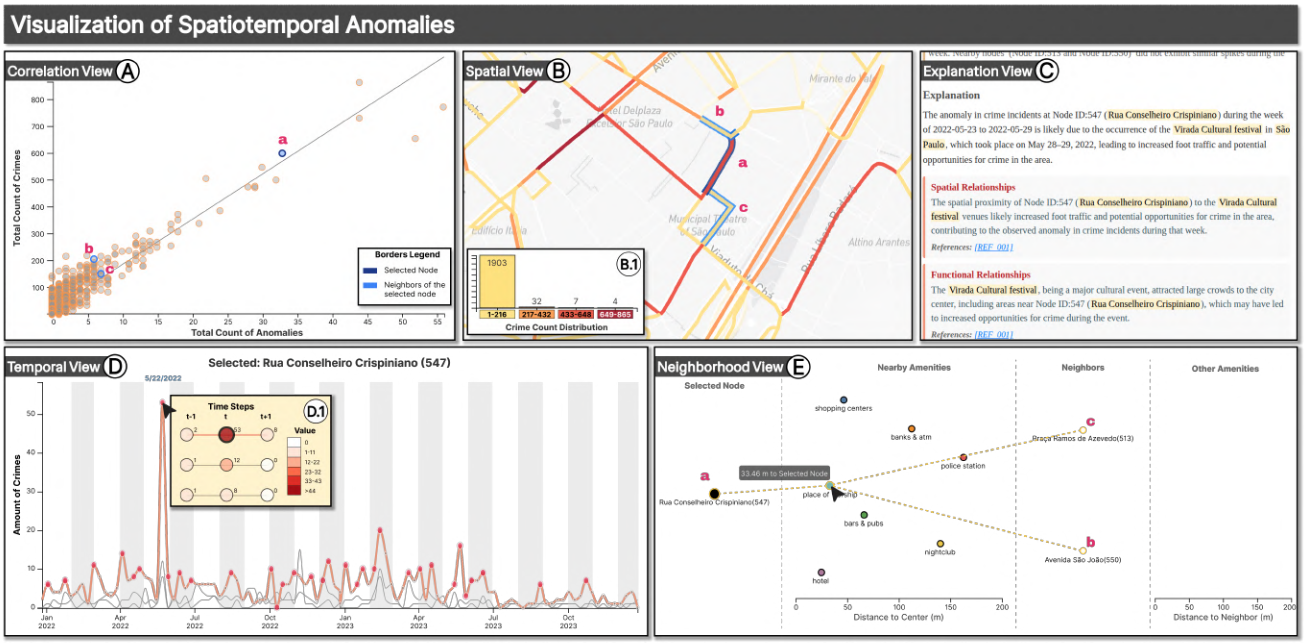

Fig. 2. The Visualization Tool: A) Scatter plot illustrating the correlation between crime counts and anomalies per node. B) Map layout of nodes; clicking selects a node and highlights its neighborhood. Node colors reflect crime count distribution (B.1). C) Explanation results from LLM. D) Time series of the selected node; hovering over detected anomalies shows a tooltip with local ST context (D.1). E) Subgraph representing the selected node neighborhood.

capture both temporal and spatial dependencies in multivariate time series [13]. For N spatial locations observed at discrete time points, the STARMA(p, q, r, s) formulation is

$$\mathbf{Y}_t = \underbrace{\sum_{i=1}^{p}\sum_{j=0}^{r} \boldsymbol{\phi}_{ij}\mathbf{W}^{(j)}\mathbf{Y}_{t-i}}_{\text{Autoregression (AR)}} + \underbrace{\sum_{i=1}^{q}\sum_{j=0}^{s} \boldsymbol{\theta}_{ij}\mathbf{W}^{(j)}\boldsymbol{\varepsilon}_{t-i}}_{\text{Moving Average (MA)}} + \boldsymbol{\varepsilon}_t \tag{1}$$

Where $\mathbf{Y}_t \in \mathbb{R}^N$ is the vector of observations at time t; p and q denote the temporal AR and MA orders, respectively; r and s specify the spatial AR and MA orders — that is, how many neighborhood "rings" are considered at each lag; $\mathbf{W}^{(j)} \in \mathbb{R}^{N\times N}$ is the j-th-order spatial weight matrix (here, derived from the graph adjacency); $\boldsymbol{\phi}_{ij}$ and $\boldsymbol{\theta}_{ij}$ are the AR and MA coefficient matrices for temporal lag i and spatial lag j; and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ represents white-noise innovations. In this process, the current observation $\mathbf{Y}_t$ depends on past values (AR term) and past disturbances (MA term), not only from the same location but also from neighboring areas.

Model fitting. Coefficients $\{\boldsymbol{\phi}_{ij}, \boldsymbol{\theta}_{ij}\}$ are estimated by maximum likelihood. At the same time, the hyperparameters (p, q, r, s) are selected via a grid search guided by the Akaike and Bayesian Information Criteria.

Anomaly detection. After fitting, the model yields one-step-ahead forecasts $\widehat{\mathbf{Y}}_t$. We flag an observation as anomalous when any component of the residual $\mathbf{r}_t = \mathbf{Y}_t - \widehat{\mathbf{Y}}_t$ exceeds ± 3 standard deviations of the in-sample residual distribution. Since STARMA embeds both spatial contiguity and temporal dynamics, this residual-based rule highlights localized deviations that purely temporal or purely spatial models might miss.

C. Visualization Tool

To support exploration, our system provides an interactive interface of five coordinated, task-specific views.

Correlation View. This scatterplot maps each street-segment node to a point whose coordinates encode the total recorded incidents (e.g., crimes, traffic accidents) and the corresponding anomaly count (Fig. 2-A). The view enables the rapid detection of outliers and clusters, serving as a gateway to deeper exploration. Users can click points or brush across multiple ones to drive the Spatial View. To minimize visual clutter, we display only nodes that have at least one recorded incident and utilize pan-and-zoom controls to facilitate navigation. Point outlines convey selection state — orange by default, **blue** for the active node, and **light blue** for its immediate neighbors.

Spatial View. This map-based pane depicts the study area's spatial domain (Fig. 2-B). Each street-segment node appears as a selectable object, and the view reflects any selections made in the *Correlation* or *Temporal Views* — whether via brushing or time-range filtering — thus analysts can focus on either the entire network or a chosen subset. Node fill color represents the total incident count, with progressively darker hues indicating higher values. Counts are divided into four equal-width quartile bands, shown in the legend (Fig. 2-B.1). Opacity encodes anomaly status: nodes without detected anomalies or outside of filters are desaturated and unselected. Clicking a node highlights it in **blue**, while its first-order neighbors in a **lighter-blue**. All selections are propagated back to the *Correlation View*, ensuring a consistent coloring across the interface. Fig. 2 shows this selection across views — selected node with **a**, and its neighbors with **b** and **c**.

Explanation View. This panel presents LLM-generated explanations, formatted for clarity and linked to the visual interface

Problem Statement	An anomaly has been detected in a week and street in particular (detailed after), a temporary spacio analysis is needed to recognize the possible causes and provide a plausible explanation. An anomaly is an unexpected value compared to spatial neighbors, temporal patterns, or both.
Task Structure	Complete these three sequential parts. Do not let later parts influence earlier analysis. **Part 1: Initial Hypotheses (Data-Only Analysis)** - Identify the anomaly's spatial extent and temporal duration based on the provided graph data. - Generate at least 2 distinct hypotheses explaining the indicated anomaly. - Base hypotheses exclusively on the graph data: amenities, crime patterns, holidays, and spatial relationships between the nodes. - Consider why this specific location/time was affected differently than its direct spatial neighbors. **Part 2: Web Validation** - Crucial Geographic Focus: Search for verifiable evidence exclusively from São Paulo city center, focusing on events, news, or public safety reports that occurred specifically in the vicinity described by the graph during the anomalous period. - Prioritize information related to the specific streets mentioned in the graph, or very immediate surrounding blocks. Avoid general city-wide news or events that are not spatially specific to this small area. - Focus on: local news outlets, official municipal event calendars, public safety reports, and government announcements relevant to the São Paulo city center. - Evaluate each hypothesis generated in Part 1 against the web evidence found. - Select the single best-supported hypothesis. - If no relevant web evidence is found specifically for the designated small geographic area and time, explicitly state this. - If the evidence found is not highly consistent with the precise spatial or temporal location (e.g., it's a city-wide event or in a different neighborhood), disregard it and continue searching for more specific evidence, or state that no relevant evidence was found. **Part 3: Final XML Report** Format your conclusion using the specified XML structure, including only web-verified sources in the References section.
Spatio-Temporal Context	**Textual Graph:** Contains nodes with street info, nearby amenities, crime time series, holiday markers, and an explicit anomaly indicators === Nodes === {Node ID}({node name}, {node type}) Nearby amenities (distances in meters): - {amenity}: {distance} - ... Crime Incidents (Weekly): - {start week date} to {end week date}: {crimes count} {**(Anomaly)**} - ... === Edges === {node name} ({node id}) -- {node name} ({node id}) ...
Extra Spatial Context	**Amenity Security Risk Reference:** - {amenity} : {short behavior description}. - Bars/Pubs: Alcohol-related violence, late hours - ...
Extra Temporal Context	**Holiday list per week** - {start week date} to {end week date} include {holidays on this week}
Output Structure	**Final considerations** - Always reference nodes by ID: (id: xxxx) - Use only real, accessible URLs in References - Base conclusions on verifiable evidence only - You must adhere strictly to the following output structure: <AnomalyExplanation> ... </AnomalyExplanation>

Fig. 3. Prompt structure: To the left of each prompt component is a general description, and to the right is an example without specific details.

(Fig. 2-C). Hovering over a node ID in the text highlights the corresponding elements across all views, connecting narrative and visualization. A drop-down menu allows analysts to switch between *GPT-4o Mini* and *Gemini 2.0 Flash*, which share a common prompt template (see Sec. III-D).

Temporal View. When a node is selected, its incident-count time series is plotted (Fig. 2-D); otherwise, the network-wide aggregate series is shown. Moving the cursor reveals the exact date and public holidays. Neighboring nodes' series appear as faint background traces to provide local context. Hovering over an anomaly marker (red dots) displays a tooltip with incident counts for the selected node and its neighbors in the previous, current, and subsequent intervals (Fig. 2-D.1). Two interactions are available: (i) clicking an anomaly marker requests an LLM explanation, retrieving the associated subgraph and a temporal window of -4 to $+2$ intervals; (ii) holding `Alt` while clicking activates a time-range selection that synchronizes the *Correlation View* and *Spatial View*.

Neighbourhood View. This panel organizes the spatial context of the selected node into four regions (Fig. 2-E): (1) the selected node; (2) amenities within a fixed radius of that node; (3) first-order neighboring nodes; and (4) amenities that are close to at least one neighbor but not to the selected node. This view complements the *Spatial* and *Temporal View* by showing connectivity and proximity as spatial evidence for or against LLM-generated hypotheses. Hovering over an amenity reveals its distance and linkage (see cursor in Fig. 2-E). It provides a clear and interactive representation of the spatial context.

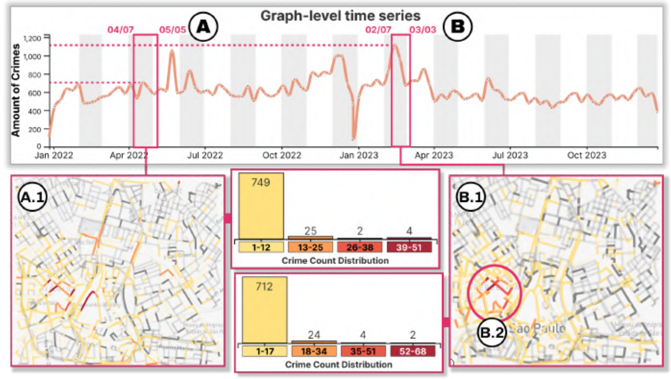

Fig. 4. Carnival periods: Weeks from 04-07-2022 to 05-05-2022 (A) and 02-07-2023 to 03-03-2023 (B) cover the full duration of Carnival for each year. A.1 and B.1 show the corresponding maps with their respective legends. B.2 indicates the part with the most crimes.

D. *Prompting for Anomaly Explanation*

Explanations are generated through a structured prompt (Fig. 3) that anchors the LLM in the local ST context. The prompt consists of (1) a concise problem statement; (2) a step-by-step task description; (3) the textualised ST context of the selected anomaly; (4) descriptions of nearby amenities (additional spatial context); (5) a list of holidays or notable events (additional temporal context); and (6) the required XML output schema (Appendix A). A preceding message defines the model's role and global constraints—such as language, tone, and persona. The task has three steps: (1) generate hypotheses from the provided information, (2) search the web for supporting public events and cite sources, and (3) select the best hypothesis and return it in the requested format. This structure reflects practices that enhance LLM controllability, reduce ambiguity, and improve multi-step reasoning [25], [26]. Outputs are constrained to machine-readable XML format, following best-practice guidelines [27]. Overall, this prompt design functions as a risk-reduction strategy that promotes evidence-grounded hypothetical explanations.

We evaluated two web-augmented LLM configurations: *GPT-4o Mini* with built-in web search, and *Gemini 2.0 Flash* with the Google Search tool. GPT-4o Mini consistently produced concise, well-cited hypotheses, while Gemini returned richer references but with less consistency.

IV. Usage Scenarios

To demonstrate our workflow, we analyze official incident data from a compact downtown district of São Paulo, mirroring the spatial extent used by Hassan et al. [23]. The setting is challenging due to high incident counts at some street segment levels and high social sensitivity. We use official cell phone theft records from the State of São Paulo, and aggregated them weekly to balance temporal detail against sparsity. We present two representative scenarios below.

A. *Carnival of 2023*

Across Brazil, Carnival brings a surge in street activity — characterized by tourists, parades, increased alcohol consumption, and crowding — that often correlates with a rise in

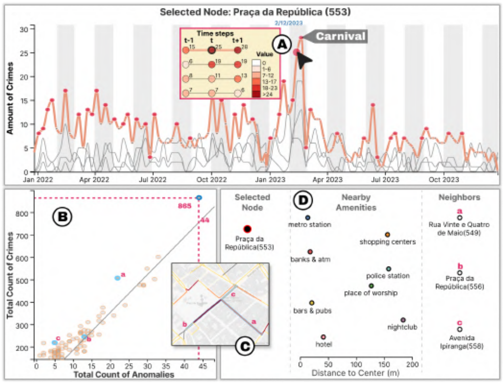

Fig. 5. Detail of more "dangerous" street segment. A) Local context of the higher peak during the Carnival. B) Correlation View highlighting the selected node and its neighbors, similar to Spatial View (C). D) Neighborhood View showing details of local space context.

crime. In 2022, a small resurgence of the pandemic delayed celebrations to 16–30 April, while in 2023, they occurred normally from 11–19 February. Fig. 4 (*Temporal View* A and B) marks the weeks surrounding both events. We focus on node 553, near *Praça da República*, which recorded the highest counts in both years (see the spatial context in Fig. 4-B.2). Its time series shows several anomalies, with a sharp peak during Carnival 2023 (*Temporal View* in Fig. 5). Node 553 also stands out in the *Correlation View*, while neighbors remain near the origin (Fig. 5-B). The Neighborhood View highlights a few direct connections but many nearby crime-linked amenities — such as metro stations, bars, and pubs — known to be associated with crime.

Peak weeks analysis. The peak of incidents spans two consecutive weeks: the week of Carnival and one preceding it. The first peak is predictably high, and the LLM correlates it well, citing risk-prone amenities and contextual evidence, such as:

- "*The Pagu Block paraded on Tuesday (21) during Carnival at Praça da República in São Paulo.*"
- "*Crowds with distracted people during their leisure time are a magnet for opportunistic actions... thefts and robberies increase during Carnival.*"

The week-before-carnival's anomaly, however, may seem unexpected — especially for non-local users. Here, the LLM's response proves especially useful, as it uncovers real-world context not directly present in the data. For example, the model found that pre-Carnival events contribute to the early spike: "*The Ilú Obá de Min block opens the São Paulo carnival today (17) with a street opera. The parade will take place at night in the center of the capital's streets.*"

Take-away. In cities like São Paulo, Carnival celebrations often begin before the official dates. This type of contextual knowledge may seem trivial to locals, but it is not for all users.

B. Religious Holidays

In our initial exploration, most detected anomalies were *positive*, i.e., sudden increases in incident counts — expected, as low-incident streets make any spike noticeable. However, a pronounced *decrease* can be equally anomalous, especially for locations with a historically high baseline.

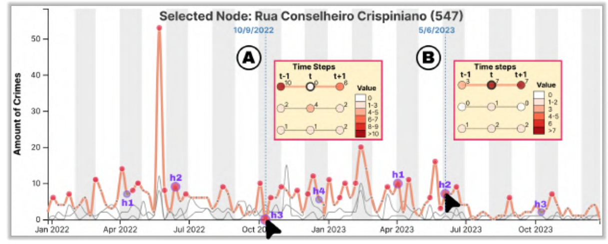

Fig. 6. Anomalies due to catholic holidays. A) Context of dip anomaly; and B) Light anomalous peak. Main catholic holidays: (h1) Holy Week, (h2) Corpus Christi, (h3) Our Lady of Aparecida, and (h4) Christmas.

To examine this case, we focus on the upper-right quadrant of the Correlation View and select node 547 (Rua Conselheiro Crispiniano; highlighted in Fig. 2). Its time series contains several anomalies, including a marked dip (Fig. 6). The decline coincides with the holiday of *Our Lady of Aparecida*. Because a Catholic church is situated nearby, it is plausible that religious observance reduced routine street activity and, thus, opportunities for crime. However, does every major Catholic holiday produce a similar effect?

Comparative holiday analysis. Fig. 6 also marks three other Catholic holidays (labels `h1`–`h4`). To minimize tourism-related confounds (prominent in `h1` and `h4`), we compare `h3` in 2022 with `h2` in 2023:

- **`h3`—Our Lady of Aparecida 2022** (Fig. 6-A): the node recorded *zero* incidents during the holiday week.
- **`h2`—Corpus Christi 2023** (Fig. 6-B): incidents spiked during the holiday week and remained elevated thereafter.

Node 547 is situated in a high-traffic corridor lined with amenities, making it inherently vulnerable to crime. Nevertheless, the LLM generated distinct but meaningful responses: `h3`: *"Likely due to religious observances associated with the Feast of Our Lady of Aparecida, leading to increased community presence and vigilance."* Consistent with *"Our Lady of Aparecida coincides with Children's Day in Brazil, leading to gatherings and a reflective, family-oriented atmosphere."* `h2`: *"Heightened social activities and gatherings during Corpus Christi created opportunities for crime, both on the node itself and in nearby Praça Ramos de Azevedo."* Supported by *"Guide to events and activities in São Paulo over the Corpus Christi long weekend."*

Take-away. Religious holidays do not uniformly suppress crime — even near churches. Whether incident counts rise or fall depends on the nature of the celebration: contemplative observances may dampen street activity, whereas festive, tourism-oriented events can generate the opposite effect.

V. Discussion, Limitations, and Future Work

We introduced an interactive system that explains spatiotemporal anomalies in urban data through LLM-generated narratives. Structured prompts—augmented with local spatial and temporal context—anchor the model's output in the underlying data. Our case studies demonstrated that the resulting LLM's responses can surface plausible, data-supported hypotheses linking anomalies to social dynamics such as holidays, mass events, and amenity density. By coupling visual cues with natural language insights, the system helps both specialists and lay users interpret complex patterns more easily.

Collaborators feedback. Two criminology experts in São Paulo praised the system's novelty and its LLM-based contextualization, which streamlines analysis and speeds insight generation. They suggested expanding data sources beyond news and public events to capture additional factors and upgrading the anomaly detector. Overall, they believe the tool can cut analysis time during pattern exploration and be deployed in security offices after targeted refinements.

Limitations. An evident limitation is the *output variability*. Despite our prompt engineering, LLM responses can still vary in detail or drift into hallucinations. External search tools (e.g., Google) did not consistently improve relevance, often yielding trivial, off-topic citations or omitting known solutions due to a lack of web references. Another limitation is the *lack of general user evaluation*. While the interface was designed with standard usability heuristics, we have not conducted formal studies to assess its impact on comprehension, trust, or decision-making. Another limitation is response correctness, as real explanations are not available for every anomaly point.

Future work. We plan to extend our framework to include *forecast explanations*, enabling the system to elucidate predicted trends (and their uncertainties) — supporting a more complete spatiotemporal analysis pipeline. Once this functionality is integrated, we will perform a *comprehensive user study* with domain experts and novices to assess interpretability gains, reliability, practical usefulness, result correctness, and potential design improvements. We also aim for *domain generalization*. Beyond crime data, the approach can also be applied to traffic, pollution, epidemiology, or even neural activity networks, underscoring its versatility.

VI. Conclusion

We present a system that combines coordinated visualizations with LLM-generated narratives to try to explain spatiotemporal anomalies. A structured prompting strategy constrains output, ensuring concise contextualizations. Applied to weekly incident data in São Paulo, the method produced interpretations aligned with known events, showing its potential to enrich data exploration and understanding.

Acknowledgment

This work was supported by FAPESP (#2021/07012-0, #2023/04868-7), FAPERJ (#E-26/210.585/2025), CNPq (#311144/2022-5), CAPES (#88887.184850/2025-00) Finance Code 001, and FGV. The opinions, hypotheses, conclusions, and recommendations expressed in this material are the responsibility of the authors and do not necessarily reflect the views of FAPESP, FAPERJ, CNPq, CAPES, and FGV.

References

[1] M. Jin, Q. Wen, Y. Liang, C. Zhang, S. Xue, X. Wang, J. Zhang, Y. Wang, H. Chen, X. Li *et al.*, "Large models for time series and spatio-temporal data: A survey and outlook," *arXiv preprint:2310.10196*, 2023.

[2] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint:1901.03407*, 2019.

[3] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, and K.-R. Müller, "A unifying review of deep and shallow anomaly detection," *Proceedings of the IEEE*, 2021.

[4] A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S. K. Rumi, and F. D. Salim, "Spatiotemporal data mining: a survey on challenges and open problems," *Artificial Intelligence Review*, 2022.

[5] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *ITU Journal: ICT Discoveries*, 2018.

[6] B. Fatemi, J. Halcrow, and B. Perozzi, "Talk like a graph: Encoding graphs for large language models," in *International Conference on Learning Representations*, 2024.

[7] X. He, Y. Tian, Y. Sun, N. Chawla, T. Laurent, Y. LeCun, X. Bresson, and B. Hooi, "G-retriever: Retrieval-augmented generation for textual graph understanding and question answering," *NeurIPS*, 2024.

[8] S. N. De Melo, D. V. Pereira, M. A. Andresen, and L. F. Matias, "Spatial/temporal variations of crime: A routine activity theory perspective," *International Journal of Offender Therapy and Comparative Criminology*, 2018.

[9] H. Zhang, Y. Zheng, and Y. Yu, "Detecting urban anomalies using multiple spatio-temporal data sources," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 2018.

[10] M. Zhang, T. Li, Y. Yu, Y. Li, P. Hui, and Y. Zheng, "Urban anomaly analytics: Description, detection, and prediction," *IEEE Transactions on Big Data*, 2020.

[11] L. Tišljarić, S. Fernandes, T. Carić, and J. Gama, "Spatiotemporal traffic anomaly detection on urban road network using tensor decomposition method," in *International Conference on Discovery Science*, 2020.

[12] S. E. Sofuoglu and S. Aviyente, "Gloss: Tensor-based anomaly detection in spatiotemporal urban traffic data," *Signal Processing*, 2022.

[13] S. Rathod, B. Gurung, K. Singh, and M. Ray, "An improved space-time autoregressive moving average (starma) model for modelling and forecasting of spatio-temporal time-series data," *Journal of the Indian Society of Agricultural Statistics*, 2018.

[14] L. Padilla, B. Lagos-Álvarez, J. Mateu, and E. Porcu, "Space-time autoregressive estimation and prediction with missing data based on kalman filtering," *Environmetrics*, 2020.

[15] Z. Li, L. Xia, J. Tang, Y. Xu, L. Shi, L. Xia, D. Yin, and C. Huang, "Urbangpt: Spatio-temporal large language models," in *Proceedings of the 30th ACM SIGKDD*, 2024.

[16] Y. Jiang, Q. Chao, Y. Chen, X. Li, S. Liu, and G. Cong, "UrbanLLM: Autonomous urban activity planning and management with large language models," in *EMNLP*, 2024.

[17] B. Peng, Y. Zhu, Y. Liu, X. Bo, H. Shi, C. Hong, Y. Zhang, and S. Tang, "Graph retrieval-augmented generation: A survey," *arXiv preprint:2408.08921*, 2024.

[18] B. Perozzi, B. Fatemi, D. Zelle, A. Tsitsulin, M. Kazemi, R. Al-Rfou, and J. Halcrow, "Let your graph do the talking: Encoding structured data for llms," *arXiv preprint:2402.05862*, 2024.

[19] Y. Hu, Z. Lei, Z. Zhang, B. Pan, C. Ling, and L. Zhao, "GRAG: Graph retrieval-augmented generation," in *NAACL*, 2025.

[20] Z. Deng, D. Weng, X. Xie, J. Bao, Y. Zheng, M. Xu, W. Chen, and Y. Wu, "Compass: Towards better causal analysis of urban time series," *IEEE TVCG*, 2021.

[21] E. Rodriguez, N. Ferreira, and J. Poco, "JamVis: exploration and visualization of traffic jams," *EPJST*, 2022.

[22] G. Garcıa, J. Silveira, J. Poco, A. Paiva, M. B. Nery, C. T. Silva, S. Adorno, and L. G. Nonato, "CrimAnalyzer: Understanding crime patterns in são paulo," *IEEE TVCG*, 2019.

[23] W. Hassan, M. M. Cabral, T. R. Ramos, A. C. Filho, and L. G. Nonato, "Modeling and predicting crimes in the city of são paulo using graph neural networks," in *BRACIS 2025*. Springer, 2024.

[24] G. Boeing, "Modeling and analyzing urban networks and amenities with osmnx," *Geographical Analysis*, 2025.

[25] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," in *NeurIPS*, 2022.

[26] A. Leidinger, R. van Rooij, and E. Shutova, "The language of prompting: What linguistic properties make a prompt successful?" in *EMNLP*, 2023.

[27] Anthropic, "Use xml tags in prompts," https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/use-xml-tags, 2024, accessed: 2025-08-26.

APPENDIX

A. Requested XML Structure for LLM Responses

```xml
<AnomalyExplanation>
    <Details>
        <Spatial>Describe affected nodes/areas.</Spatial>
        <Temporal>Start and end dates/times of anomaly.</Temporal>
        <Patterns>Key patterns observed in data.</Patterns>
    </Details>
    <Hypothesis>
        <Statement>State the chosen hypothesis clearly.</Statement>
        <Reasoning>
            <Relational>
                <Description>How spatial relationships support hypothesis.</Description>
                <Evidence>Comma separated list of references, e.g., REF_001,REF_002</Evidence>
            </Relational>
            <Functional>
                <Description>How functional relationships support hypothesis.</Description>
                <Evidence>Comma separated list of references.</Evidence>
            </Functional>
            <Temporal>
                <Description>How timing and duration support hypothesis.</Description>
                <Evidence>Comma separated list of references.</Evidence>
            </Temporal>
        </Reasoning>
    </Hypothesis>
    <Conclusion>Summary of findings and overall explanation.</Conclusion>
    <Keywords>Comma separated list of relevant terms.</Keywords>
    <References>
        <Item>
            <Id>Reference id, e.g., REF_001<Id>
            <Source>e.g., Local News Article Title</Source>
            <URL>https://example.com/news-article</URL>
            <Excerpt>Quote or summarize relevant part supporting claim.</Excerpt>
            <Date>Date of news/event</Date>
        </Item>
    </References>
</AnomalyExplanation>
```

Fig. 7. XML output structure.

B. Implementation and Execution Details

The entire processing pipeline was developed in Python 3.9 using open-source packages such as NumPy, Pandas, GeoPandas, and Scikit-learn. Anomaly detection was executed in a Google Colab environment without GPU support. The visualization system follows a client-server paradigm. The server handles data filtering and selection, communicates with OpenAI and Gemini APIs, and parses and processes their responses. The local server was implemented with Python's Flask package, with all APIs using the GET method and written from scratch. For the client (visualization system), we used JavaScript with the Svelte framework. All visualizations were created using the *D3.js* library. The raw data was dowloaded from https://www.ssp.sp.gov.br/estatistica/consultas, the official site of São Paulo State Public Security portal.

Execution Time: he anomaly detection procedure takes $\sim 4.5\ s$ per run ($\sim 4.0\ s$ for STARMA fitting and $\sim 0.5\ s$ for residual computation and thresholding). LLM-based explanations require $\sim 14\ s$ per query, depending on service latency and context length. STARMA hyperparameter tuning takes $\sim 1\ h$, but it is an offline, one-time step.

C. Spatiotemporal Context in the Prompt: Carnival of 2023

Spatio-Temporal Context

```
=== Nodes ===
Node ID:553 (Praça da República, secondary street)
Nearby amenities (distances in meters):
 - metro station: 13.63
 - shopping centers: 156.10
 - banks & atm: 17.98
 - police station: 158.04
 - place of worship: 127.09
 - bars & pubs: 20.86
 - nightclub: 183.87
 - hotel: 41.67
Crime Incidents (Weekly):
 - 2023-01-16 to 2023-01-22: 12
 - 2023-01-23 to 2023-01-29: 19
 - 2023-01-30 to 2023-02-05: 12
 - 2023-02-06 to 2023-02-12: 15
 - **2023-02-13 to 2023-02-19: 25 (Anomaly)**
 - 2023-02-20 to 2023-02-26: 28
Node ID:549 (Rua Vinte e Quatro de Maio, residential street)
Nearby amenities (distances in meters):
 - shopping centers: 6.24
 - banks & atm: 15.11
 - police station: 6.36
 - place of worship: 127.09
 - bars & pubs: 61.53
 - nightclub: 119.52
 - hotel: 30.99
Crime Incidents (Weekly):
 - 2023-01-16 to 2023-01-22: 6
 - 2023-01-23 to 2023-01-29: 7
 - 2023-01-30 to 2023-02-05: 14
 - 2023-02-06 to 2023-02-12: 6
 - 2023-02-13 to 2023-02-19: 19
 - 2023-02-20 to 2023-02-26: 19
Node ID:556 (Praça da República, secondary street)
Nearby amenities (distances in meters):
 - metro station: 21.20
 - shopping centers: 168.36
 - banks & atm: 50.64
 - bars & pubs: 24.20
 - hotel: 87.28
Crime Incidents (Weekly):
 - 2023-01-16 to 2023-01-22: 4
 - 2023-01-23 to 2023-01-29: 3
 - 2023-01-30 to 2023-02-05: 2
 - 2023-02-06 to 2023-02-12: 8
 - 2023-02-13 to 2023-02-19: 11
 - 2023-02-20 to 2023-02-26: 13
Node ID:558 (Avenida Ipiranga, secondary street)
Nearby amenities (distances in meters):
 - shopping centers: 174.16
 - banks & atm: 16.62
 - police station: 151.49
 - place of worship: 114.53
 - bars & pubs: 13.33
 - nightclub: 137.02
 - hotel: 18.04
Crime Incidents (Weekly):
 - 2023-01-16 to 2023-01-22: 1
 - 2023-01-23 to 2023-01-29: 0
 - 2023-01-30 to 2023-02-05: 1
 - 2023-02-06 to 2023-02-12: 7
 - 2023-02-13 to 2023-02-19: 7
 - 2023-02-20 to 2023-02-26: 6
=== Edges ===
"Praça da República" (id:553) -- "Rua Vinte e Quatro de Maio" (id:549)
"Praça da República" (id:553) -- "Praça da República" (id:556)
"Praça da República" (id:553) -- "Avenida Ipiranga" (id:558)
```

Extra Spatial Context

```
Amenity Security Risk Reference
 - Nightclubs: Alcohol/drugs, crowds, late hours
 - Metro Stations: High traffic, anonymity, theft opportunities
 - Bars/Pubs: Alcohol-related violence, late hours
 - Fuel Stations: Cash transactions, 24/7 isolated locations
 - Banks/ATMs: Cash targets, robbery risk
 - Places of Worship: Respected venues, occasional large gatherings
 - Shopping Centers: Controlled, surveilled environments
 - Hotels: Internal security, but tourists attractive targets
 - Police Stations: Crime deterrent, safety perception
```

Extra Temporal Context

```
Holiday list per week
 - 2023-01-16 to 2023-01-22
 - 2023-01-23 to 2023-01-29 include: São Paulo City Anniversary
 - 2023-01-30 to 2023-02-05
 - 2023-02-06 to 2023-02-12
 - 2023-02-13 to 2023-02-19
 - 2023-02-20 to 2023-02-26 include: Carnival Monday, Carnival Tuesday, Ash Wednesday
```

Fig. 8. Spatiotemporal context for the 2023 Carnival and Pre-Carnival celebrations.

D. Spatiotemporal Context in the Prompt: Religious Holidays

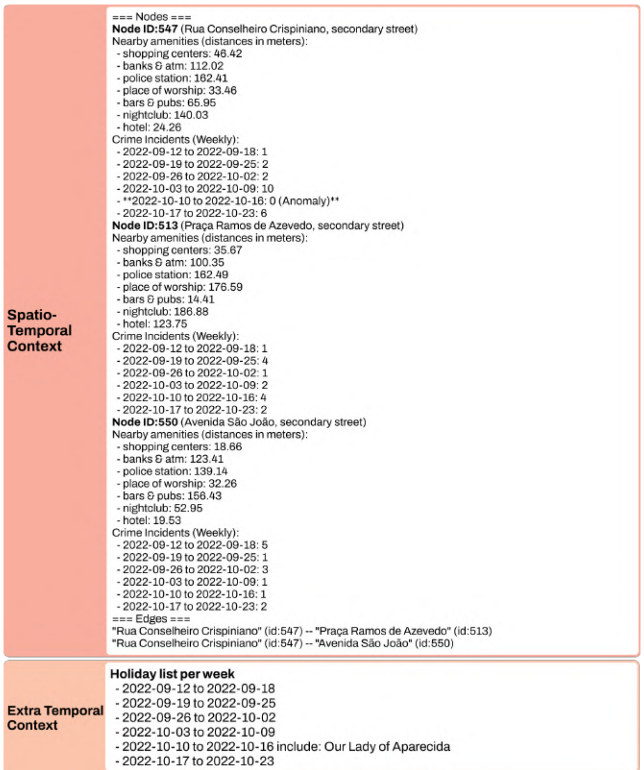

Fig. 9. Spatiotemporal context for the 2022 Our Lady of Aparecida celebration scenario. *Extra Spatial Context* omitted since it matches Fig. 8.

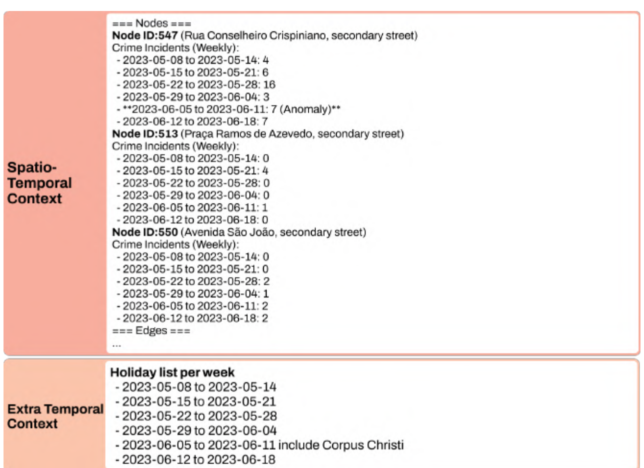

Fig. 10. Spatiotemporal context of the 2023 Corpus Christi celebration scenario. Node details and *Extra Spatial Context* are omitted since they match Fig. 9 and Fig. 8.

E. LLM Response: Carnival of 2023

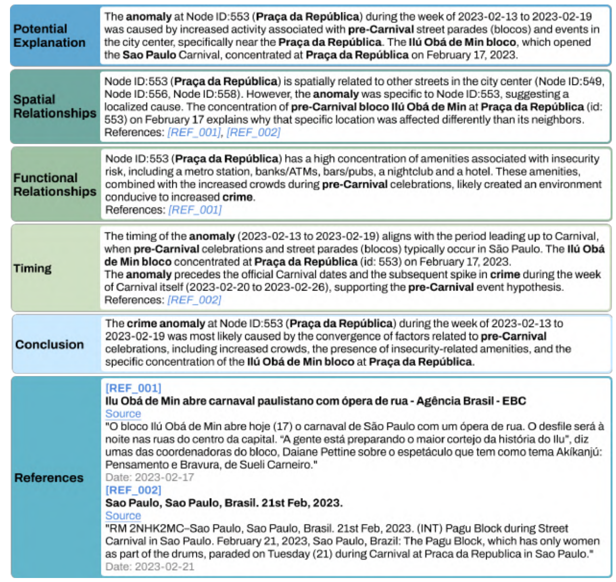

Fig. 11. Parsed LLM Response. Anomaly during *Pre-carnival* week. Using Google Gemini Flash 2.0.

F. LLM Response: Religious Holidays

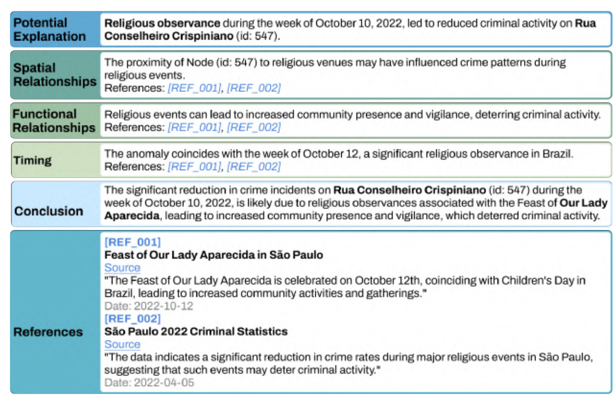

Fig. 12. Parsed LLM Response. Anomaly during *Our Lady of Aparecida* holiday week. Using OpenAI's GPT-4o mini.

Potential Explanation	Increased crime during the **Corpus Christi** holiday period due to heightened social activities and gatherings.
Spatial Relationships	Node (id: 547) **Rua Conselheiro Crispiniano** is in proximity to Node (id: 513) **Praça Ramos de Azevedo**, a central area with cultural venues and public spaces, likely attracting larger crowds during holidays, potentially leading to increased opportunities for crime. References: *[REF_001]*, *[REF_002]*
Functional Relationships	The area around Node (id: 547) **Rua Conselheiro Crispiniano** is known for its vibrant cultural scene, including theaters and public spaces, which may draw larger crowds during holidays, increasing opportunities for crime. References: *[REF_001]*, *[REF_002]*
Timing	The anomaly coincides with the **Corpus Christi** holiday period, a time when social activities and gatherings typically increase, potentially leading to higher crime rates. References: *[REF_001]*, *[REF_002]*
Conclusion	The increased crime incidents during the **Corpus Christi** holiday period in the area around Node (id: 547) **Rua Conselheiro Crispiniano** are likely due to heightened social activities and gatherings, leading to more opportunities for crime.
References	**[REF_001]** **Gazeta de São Paulo - 6 programas para curtir São Paulo durante o feriado de Corpus Christi** Source "Article discussing various events and activities in São Paulo during the Corpus Christi holiday period, highlighting increased social gatherings and cultural events." Date: 2024-05-21 **[REF_002]** **Guide of the Week - What to do in São Paulo during the Corpus Christi holiday 2023** Source "Guide listing various events and activities in São Paulo during the Corpus Christi holiday period, indicating increased public gatherings and events." Date: 2023-05-30

Fig. 13. Parsed LLM Response. Anomaly during *Corpus Christi* holiday week. Using OpenAI's GPT-4o mini.

G. Links Referenced in LLM Responses

Regarding Carnival:

REF_001 *Sao Paulo, Sao Paulo, Brasil. 21st Feb, 2023. (INT) Street Carnival in Sao Paulo - Alamy*.

REF_002 *RISKS DURING THE CARNIVAL | MOVI NEWS*.

Regarding Pre-Carnival:

REF_001 *Ilu Obá de Min abre carnaval paulistano com ópera de rua - Agência Brasil - EBC*.

REF_002 *Sao Paulo, Sao Paulo, Brasil. 21st Feb, 2023*.

Regarding Our Lady Aparecida's holiday:

REF_001 *Feast of Our Lady Aparecida in São Paulo*

REF_002 *São Paulo 2022 Criminal Statistics*

Regarding Corpus Christi's holiday:

REF_001 *Gazeta de São Paulo - 6 programas para curtir São Paulo durante o feriado de Corpus Christi*

REF_002 *Guide of the Week - What to do in São Paulo during the Corpus Christi holiday 2023*