LIS 640 FALL 2025-2026

Data Tagging System for UW–Madison Transportation Services Final Project Report

Krish Prakash Bhat | Rohini Ramesh | Srivarshini Ankaiah Krishnappa | Zhongchao Pan | Ziyue Li

1. Introduction

Universities generate large volumes of time-dependent data related to campus operations, including academic schedules, public events, sports activities, facility occupancy, weather conditions, and operational closures. While each of these data sources provides valuable insight independently, they are typically managed and accessed through separate systems. This fragmentation makes it difficult to understand how campus activities interact on a given day or across longer time periods.

As a result, planners, administrators, and analysts often lack a unified view of campus activity. Questions such as how weather affects attendance, how events align with parking demand, or how academic schedules overlap with facility usage require manual data aggregation, if they can be answered at all. The absence of an integrated, date-centered system limits both operational decision-making and historical analysis.

To address this challenge, this project develops a centralized Campus Calendar Analytics System covering the years 2020 to 2024. The system integrates multiple heterogeneous datasets into a unified data architecture and presents them through an interactive calendar-driven interface . By using the calendar date as the primary organizing dimension, the platform enables users to explore daily activity, identify patterns, and apply contextual interpretation through custom tagging.

This report documents the full lifecycle of the project, including the motivation behind the system, the data sources and preparation steps, the technical architecture and implementation, and an evaluation of the system's effectiveness and limitations. Practical applicability and potential future extensions are also discussed.

2. Background and Related Work

Calendar-based information systems and temporal data visualization have been widely studied in the fields of information science, data analytics, and human–computer interaction. Prior research emphasizes that time is one of the most effective organizing dimensions for complex datasets, particularly when multiple domains must be analyzed together (Aigner et al., 2011). Calendar and timeline visualizations allow users to quickly identify patterns, overlaps, and anomalies that are difficult to detect in tabular formats.

In the context of university campuses, existing systems typically focus on individual operational areas such as event listings, academic calendars, parking dashboards, or weather reports. While these systems serve specific purposes, they rarely support cross-domain analysis. For example, campus event calendars may display event schedules but do not provide insight into parking congestion or weather conditions on those dates. Similarly, parking or occupancy dashboards often lack contextual information about why certain days experience higher demand.

Research on smart campus and urban analytics highlights the importance of integrating contextual data when interpreting operational metrics. Studies show that factors such as weather, academic schedules, and special events significantly influence mobility patterns, facility usage, and resource demand (Zhang et al., 2020). However, these factors are often analyzed in isolation rather than within a unified temporal framework.

Several visualization tools and dashboards attempt to aggregate operational data, but they often rely on static reports or predefined filters. Such approaches limit exploratory analysis and personalization. In contrast, interactive temporal interface allow users to shift between overview and detail, enabling both high-level pattern recognition and granular inspection.

This project builds on existing research by applying temporal integration principles to a campus-wide context. Rather than focusing on a single data type, the system aligns diverse datasets using a shared date key and presents them through an interactive calendar interface. This approach enables users to understand not only what happened on a given day, but also how different types of campus activity intersected, providing richer insight than traditional siloed systems.

3. Project Motivation and Objectives

3.1 Project Motivation

University campuses operate as dynamic environments where multiple activities occur simultaneously. Academic schedules, campus events, athletic competitions, facility usage, and operational constraints all influence daily campus life. Despite this complexity, information about these activities is often distributed across independent systems, making holistic analysis difficult.

During the development of this project, it became evident that answering even basic questions—such as identifying high-activity days, understanding periods of peak parking demand, or assessing how weather conditions align with campus events required manual cross-referencing of multiple data sources. This process is time-consuming, error-prone, and not accessible to non-technical users.

The motivation of this project was therefore to design a single, date-centered system that brings together diverse campus datasets into one coherent and explorable interface. By anchoring all information to a shared calendar, the system aims to improve visibility, context, and interpretability of campus data for both analytical and practical use.

3.2 Project Objectives

The primary objectives of the Campus Calendar Analytics System are as follows:

1. Data Integration:
   Combine heterogeneous datasets including events, academic calendars, sports schedules, weather records, parking occupancy, and closures—into a unified relational structure.
2. Temporal Organization:
   Use calendar dates as the primary organizing dimension to enable seamless comparison and aggregation across data sources.
3. Interactive Visualization:
   Provide an intuitive, interactive calendar interface that supports both high-level overviews and detailed daily exploration.
4. Contextual Awareness:
   Enable users to understand how different types of campus activity intersect on specific dates through color-coded indicators and drill-down views.
5. User Personalization:
   Support custom tagging of date ranges to allow users to annotate periods of interest without modifying the underlying data.
6. Extensibility and Practical Use:
   Design the system such that it can be extended with additional datasets, analytical methods, or deployed in operational and planning contexts.

Together, these objectives guide the technical and design decisions of the project and establish clear criteria for evaluating its effectiveness.

4. Data Sources and Preparation

4.1 Data Sources

The Campus Calendar Analytics System integrates multiple datasets that capture different aspects of campus activity between 2020 and 2024. Each dataset represents a distinct operational or contextual domain, and together they provide a comprehensive view of daily campus conditions. The primary data sources used in this project include:

- Campus Events: Records of public and internal university events, including titles, dates, and event types.
- Academic Calendar: Official academic scheduling information such as semester start and end dates, instructional periods, breaks, and holidays.
- Sports Schedules: Dates and metadata for university athletic events.
- Weather Data: Daily weather observations, including temperature and precipitation indicators.
- Facility Closures: Information on partial or full campus facility shutdowns.
- Parking and Occupancy Data: Aggregated measures of parking usage and facility occupancy levels.

These datasets were originally collected in different formats and at varying levels of granularity. Some datasets were event-based, while others were recorded as daily summaries, making direct comparison non-trivial without preprocessing.

4.2 Data Cleaning and Standardization

To enable reliable integration, all datasets were processed through a data cleaning and standardization pipeline. The most critical step in this process was the creation of a shared date_key field using the format YYYYMMDD, which serves as the primary join attribute across all tables.

Key preprocessing steps included:

- Converting date fields into a consistent format
- Validating date ranges and removing records outside the 2020–2024 period
- Standardizing categorical labels (e.g., event types and closure categories)
- Handling missing or incomplete records through filtering or normalization
- Removing duplicate entries where applicable

This preprocessing ensured consistency across datasets and reduced ambiguity during aggregation and visualization.

4.3 Date-Centered Data Modeling

A core design decision of the project was to adopt a date-centered data model, where a central `dates` table functions as the backbone of the system. All domain-specific tables reference this table using the `date_key`.

This approach offers several advantages:

- Simplifies cross-domain joins
- Supports efficient aggregation of daily counts
- Enables calendar-based visualization without complex query logic
- Allows new datasets to be integrated with minimal restructuring

By anchoring all records to a single calendar dimension, the system maintains temporal consistency while supporting flexible analytical queries.

4.4 Data Readiness for Visualization

After cleaning and integration, the prepared datasets were validated through exploratory queries and sample visual checks to confirm correct alignment across domains. Each date in the calendar could then be reliably associated with zero or more records from each dataset category.

This ensured that the frontend calendar interface could accurately reflect daily activity through aggregated counts and detailed drill-down views.

5. System Architecture and Implementation

5.1 Overall System Architecture

The Campus Calendar Analytics System follows a modular, client–server architecture that separates data preparation, data logic, and user interaction within a browser-based application. The system is designed to clearly separate data storage, business logic, and user interaction, allowing for scalability and ease of maintenance.

At a high level, the system includes:

- A unified, preprocessed data layer used for client-side aggregation and analysis
- A frontend client that retrieves and visualizes data
- A logical connection layer that maps relational data to calendar-based views

This architecture ensures that complex data integration and querying occur on the backend, while the frontend focuses on user interaction and visualization.

5.2 Data Storage and Processing

The Campus Calendar Analytics System is implemented as a client-side, data-driven web application. All datasets are preprocessed prior to deployment and loaded into the application at runtime using JavaScript. This design choice simplifies deployment, eliminates backend dependencies, and ensures that the system can be hosted using static web hosting services.

The data is structured around a unified date-based format, allowing records from different domains to be aggregated and joined within the application logic. By performing aggregation and filtering on the client side, the system avoids the complexity of managing a live database while still supporting rich exploratory analysis.

This approach is well suited for historical analysis scenarios, where datasets are stable and do not require real-time updates or transactional operations.

5.3 Frontend Implementation

The frontend interface is built using HTML, JavaScript, and TailwindCSS, with Day.js used for date manipulation and formatting. The calendar view dynamically renders monthly layouts and populates each day with aggregated activity indicators retrieved from the backend.

Each calendar cell displays color-coded badges representing the number of records in each category for that date. This design allows users to quickly assess daily activity levels while maintaining a clean and readable interface.

When a user selects a specific date, the system loads a detailed panel showing all associated records for that day. This drill-down capability enables users to transition seamlessly from high-level patterns to granular inspection.

5.4 Custom Tagging Feature

To support contextual interpretation and personalization, the system includes a custom tagging mechanism that allows users to apply labels to specific dates or date ranges. Examples include periods such as “Finals Week,” “Spring Break,” or other user-defined categories.

Tags are stored locally in the user’s browser using web storage, ensuring that personalization does not modify the underlying dataset or require backend authentication. While this approach limits cross-device persistence, it simplifies the system design and preserves data integrity.

5.5 Data Flow and Interaction

When the application loads, the frontend processes the preloaded datasets and computes aggregated daily counts for each activity category. User interactions such as month navigation, date selection, filtering, and tagging are handled entirely on the client side. This client-side data flow keeps the application responsive while avoiding the complexity of maintaining a separate backend service.

This division of responsibility ensures responsiveness, reduces unnecessary data transfer, and maintains a clear separation between data management and user interaction logic.

6. Features and Functionality

6.1 Calendar-Based Visualization

The primary interface of the Campus Calendar Analytics System is a monthly calendar view that spans the years 2020 to 2024. Each day in the calendar represents a single date and serves as a container for aggregated campus activity across multiple domains.

Rather than displaying raw records, the system summarizes daily activity using small, color-coded indicators. Each indicator corresponds to a specific category, such as campus events, academic calendar items, sports activities, facility closures, parking and occupancy data, and weather records. This design allows users to quickly identify days with high activity or notable overlaps without needing to navigate away from the calendar.

6.2 Multi-Category Daily Aggregation

For each calendar date, the system computes and displays the count of records in each category. This aggregation is performed at the date level using the shared `date_key`, enabling consistent and efficient comparison across datasets.

By presenting multiple categories within a single calendar cell, the system allows users to visually assess how different types of activities align temporally. For example, a user can easily observe whether high parking occupancy coincides with major events or whether academic breaks correspond with reduced campus activity.

6.3 Drill-Down and Detailed Views

When a user selects a specific date, the interface provides a detailed view listing all associated records for that day. This drill-down panel includes event titles, activity types, and relevant metadata depending on the dataset.

This feature enables a smooth transition from overview to detail. Users can begin by identifying patterns at the monthly level and then examine individual dates to understand what factors contributed to those patterns.

6.4 Filtering and Exploration

The system supports interactive filtering that allows users to focus on specific categories of interest. Users may choose to view only certain types of activity—such as sports events or closures—while temporarily hiding others.

This functionality supports exploratory analysis by allowing users to isolate variables and examine their temporal distribution without modifying the underlying data.

6.5 Custom Tagging and Annotation

A key feature of the system is the ability for users to apply custom tags to individual dates or date ranges. These tags allow users to annotate meaningful periods, such as exam weeks, long weekends, or major campus events, adding qualitative context to quantitative data.

Custom tags are stored locally in the browser, enabling personalization without requiring authentication or database modification. This approach maintains system simplicity while allowing flexible interpretation of calendar data.

6.6 User-Centered Design Considerations

The interface is designed to balance information density with readability. Color-coding, minimal text within calendar cells, and consistent layout patterns help reduce visual clutter while maintaining access to detail through interaction.

The system emphasizes ease of use for non-technical users, making it suitable for planners, students, administrators, and analysts alike.

7. Evaluation and Practical Applicability

7.1 Evaluation Criteria

The Campus Calendar Analytics System was evaluated based on three primary criteria: usability, data integration effectiveness, and analytical usefulness. Rather than relying on formal user studies, the evaluation focused on whether the system successfully met its stated project objectives and whether it enables meaningful exploration of campus activity over time.

These criteria were chosen to reflect the practical goals of the project, emphasizing clarity, reliability, and applicability rather than predictive accuracy or performance benchmarking.

7.2 Usability Evaluation

From a usability perspective, the calendar-based interface provides an intuitive entry point for users. The familiar calendar layout reduces the learning curve and allows users to immediately interpret activity patterns at the daily and monthly levels. Color-coded indicators and summary counts enable quick scanning without overwhelming the user with detail.

Interactive elements such as date selection and filtering support deeper exploration while maintaining interface simplicity. The drill-down view allows users to access detailed information only when needed, reducing unnecessary cognitive load during initial exploration.

7.3 Effectiveness of Data Integration

The use of a shared `date_key` across all datasets proved effective in enabling seamless integration and aggregation. Daily alignment of heterogeneous data sources allows users to compare and contextualize different types of campus activities without manual reconciliation.

The system consistently reflects correct associations between dates and activity categories, confirming that the data preparation and modeling approaches support accurate temporal alignment. The date-centered structure also simplifies integration of additional datasets, demonstrating the system's extensibility.

7.4 Analytical and Operational Applicability

The platform supports several practical use cases. Campus planners and administrators can use the system to identify historically high-activity periods, assess the impact of major events on parking demand, or review how weather conditions influence campus operations. Academic administrators can examine how academic schedules align with broader campus activity.

For example, the system makes it possible to quickly identify dates where high parking occupancy coincides with large campus events and favorable weather conditions. Similarly, academic break periods can be visually distinguished by reduced activity across multiple categories. These types of observations demonstrate how the integrated calendar view supports rapid situational understanding without requiring complex analytical queries.

From an analytical perspective, the system serves as a foundational exploratory tool. It allows users to generate hypotheses—such as relationships between events and occupancy—that could later be tested using statistical or predictive methods. By providing context-rich historical views, the system improves both descriptive analysis and decision support.

7.5 Overall Assessment

Overall, the Campus Calendar Analytics System successfully meets its intended objectives. It demonstrates that temporal integration of diverse campus datasets can significantly improve interpretability and accessibility of complex information. While the system is not intended to replace domain-specific dashboards or predictive tools, it fills an important gap by offering a unified, contextual view of campus activity.

8. Limitations

Despite successfully meeting its primary objectives, the Campus Calendar Analytics System has several limitations that should be considered when interpreting its results and potential use cases.

First, custom user-defined tags are stored locally within the user's browser. While this design decision simplifies implementation and preserves personalization without affecting shared data, it limits persistence across devices and prevents collaborative tagging. Users accessing the system from different browsers or computers will not see the same custom annotations.

Second, the system relies on historical datasets that vary in completeness and accuracy. Any inconsistencies, omissions, or errors present in the original data sources are inherently reflected in the system. Although preprocessing steps were applied to standardize and validate records, the system does not perform automatic data quality correction beyond basic cleaning.

Third, the platform is designed primarily for exploratory and descriptive analysis. It does not currently support predictive modeling, automated trend detection, or advanced statistical inference. As a result, users can visually identify patterns and generate hypotheses but must rely on external tools for formal analytical validation.

Finally, the system does not incorporate real-time data feeds. All information is historical, covering the period from 2020 to 2024. While this is appropriate for retrospective analysis and planning, it limits immediate operational decision-making in live scenarios.

These limitations reflect deliberate trade-offs made to prioritize clarity, usability, and scope control within the project timeline. Addressing these constraints presents meaningful opportunities for future development.

9. Future Work

The current implementation of the Campus Calendar Analytics System provides a strong foundation for integrated, calendar-based campus analysis. However, several avenues exist for extending the system's functionality and analytical value.

One important area for future improvement is persistent storage support for custom tags. Storing user-defined tags externally rather than only within local browser storage would enable cross-device access, collaborative annotation, and long-term retention of contextual information.

Another extension involves the integration of real-time data feeds. Incorporating live updates for events, parking occupancy, or weather conditions would allow the system to support operational monitoring in addition to retrospective analysis. Real-time integration could transform the platform from a historical explorer into a decision-support tool.

From an analytical perspective, future versions could incorporate predictive modeling and trend analysis. Machine learning or statistical methods could be used to forecast high-activity days, predict parking demand, or identify seasonal patterns across multiple years. These analytical layers would build on the system's existing date-centered structure.

Additionally, expanding support for external analytics and visualization tools—such as Tableau or Power BI—would enhance interoperability and allow users to perform deeper quantitative analysis. Providing curated data views or APIs would facilitate integration with institutional reporting workflows.

Finally, usability improvements such as advanced filtering, accessibility enhancements, and role-based interfaces could further broaden the system's audience and applicability.

10. Conclusion

This project demonstrates the value of a date-centered approach to integrating and visualizing complex campus datasets. By unifying events, academic schedules, sports activities, weather conditions, facility closures, and parking occupancy into a single interactive calendar, the Campus Calendar Analytics System provides users with a comprehensive and contextual view of campus activity from 2020 to 2024.

The system addresses a common challenge in institutional data use: fragmentation across independent systems. Through careful data preparation, relational modeling, and user-focused design, the platform enables both high-level pattern recognition and detailed exploration of daily records. The use of calendar-based visualization makes temporal relationships immediately visible and accessible to non-technical users, improving interpretability and practical usability.

While the system is primarily intended for exploratory and descriptive analysis, it establishes a scalable foundation for more advanced analytics and operational integration. The identified limitations and proposed future enhancements highlight opportunities for extending the platform's analytical depth and real-time applicability.

Overall, the Campus Calendar Analytics System illustrates how temporal integration and interactive visualization can enhance understanding of complex environments such as university campuses. The project contributes a practical framework that can support planning, retrospective analysis, and data-informed decision-making, and it offers a solid basis for future research and development in campus analytics.

References

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, *Visualization of Time-Oriented Data*. London, U.K.: Springer, 2011.

[2] Y. Zhang, M. Chen, and X. Li, "Smart campus analytics: Integrating temporal and contextual data for decision support," *Journal of Urban Computing*, vol. 12, no. 3, pp. 45–60, 2020.

[3] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proc. IEEE Symp. Visual Languages*, 1996, pp. 336–343.

[4] C. Ware, *Information Visualization: Perception for Design*, 4th ed. San Francisco, CA, USA: Morgan Kaufmann, 2020.

[5] S. Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, 2nd ed. Burlingame, CA, USA: Analytics Press, 2012.

[6] J. Heer, N. Kong, and M. Agrawala, "Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations," *IEEE Trans. Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1674–1683, Dec. 2014.

[7] Mozilla Developer Network (MDN), "Web Storage API," 2024. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/API/Web_Storage_API

[8] Day.js, "Day.js documentation," 2024. [Online]. Available: https://day.js.org

[9] M. Bostock, V. Ogievetsky, and J. Heer, "D³ data-driven documents," *IEEE Trans. Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, Dec. 2011.

[10] ISO, "ISO 8601: Date and time format," International Organization for Standardization, 2019.

Appendix

A. System Architecture & Deployment

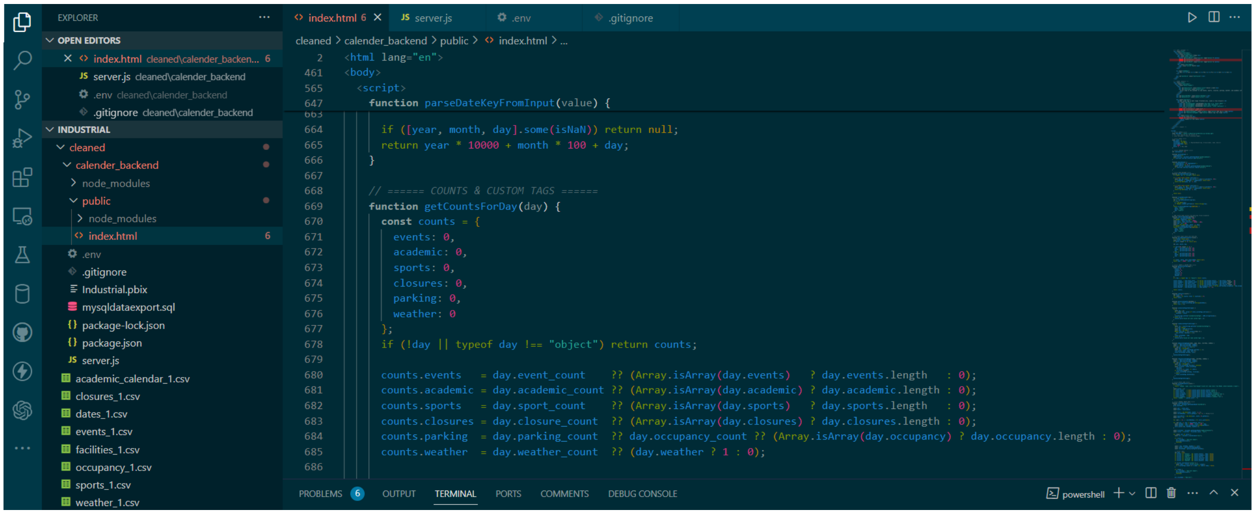

Figure A1: Frontend Code Structure:
Shows the project’s HTML, JavaScript, and stylesheet hierarchy, including calendar rendering logic, tag-management scripts, and dataset-loading components.

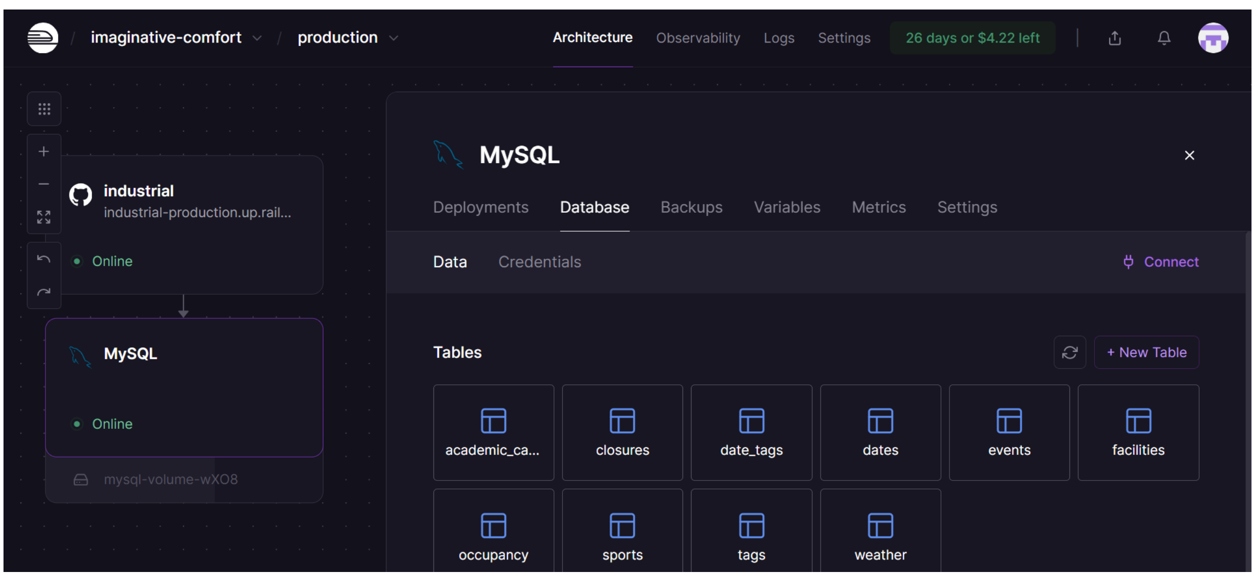

Figure A2. Railway Cloud MySQL Instance:
Screenshot of the cloud-hosted MySQL database containing all cleaned and standardized datasets, including events, weather, closures, occupancy, sports, and academic calendar tables.

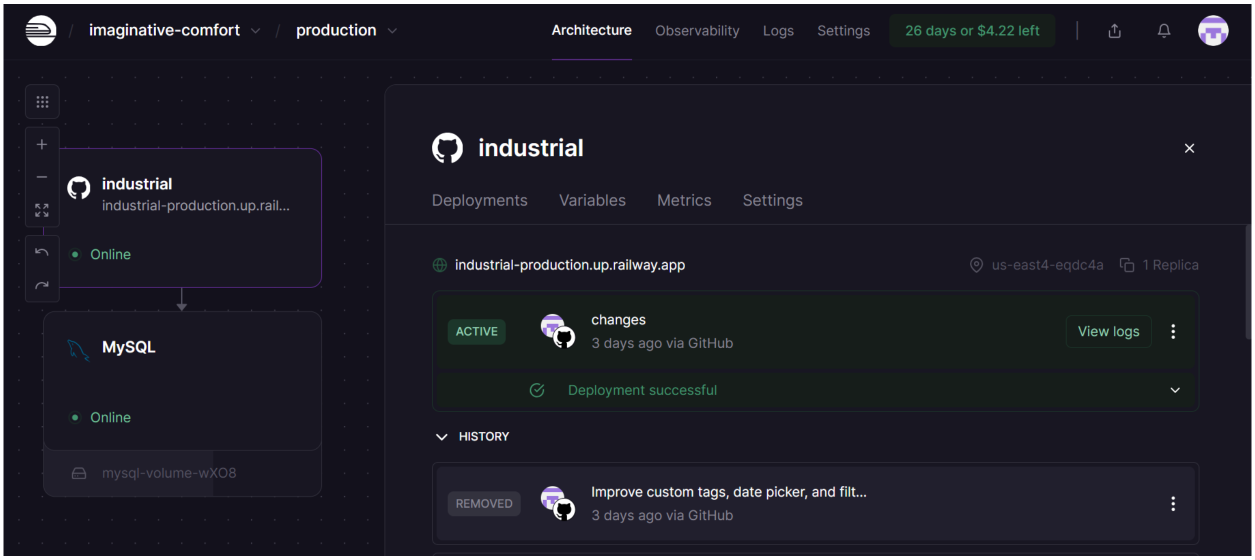

Figure A3. Railway Deployment Instance:
Depicts the secondary Railway instance that links the MySQL database to the deployed GitHub-hosted frontend. This instance serves as the connector enabling the static frontend to read from the cloud database.

A. Application Interface

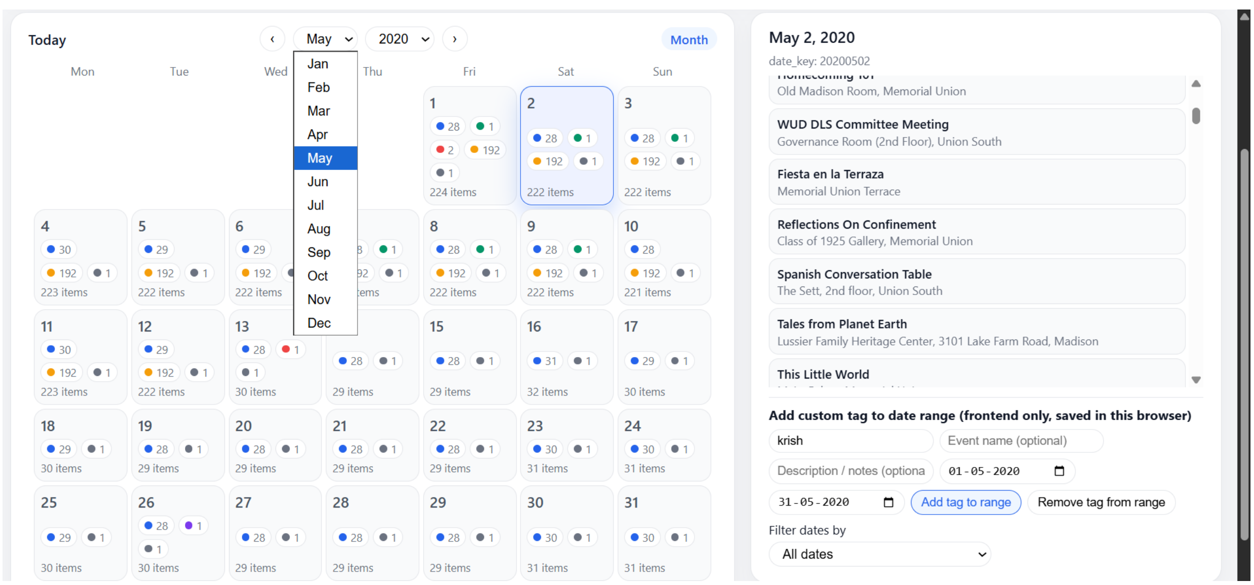

Figure B1. Calendar Interface with Default Tag Indicators:
Displays the color-coded monthly calendar view with system-defined contextual tags (e.g., events, closures, weather indicators), along with the metadata panel describing selected-day details.

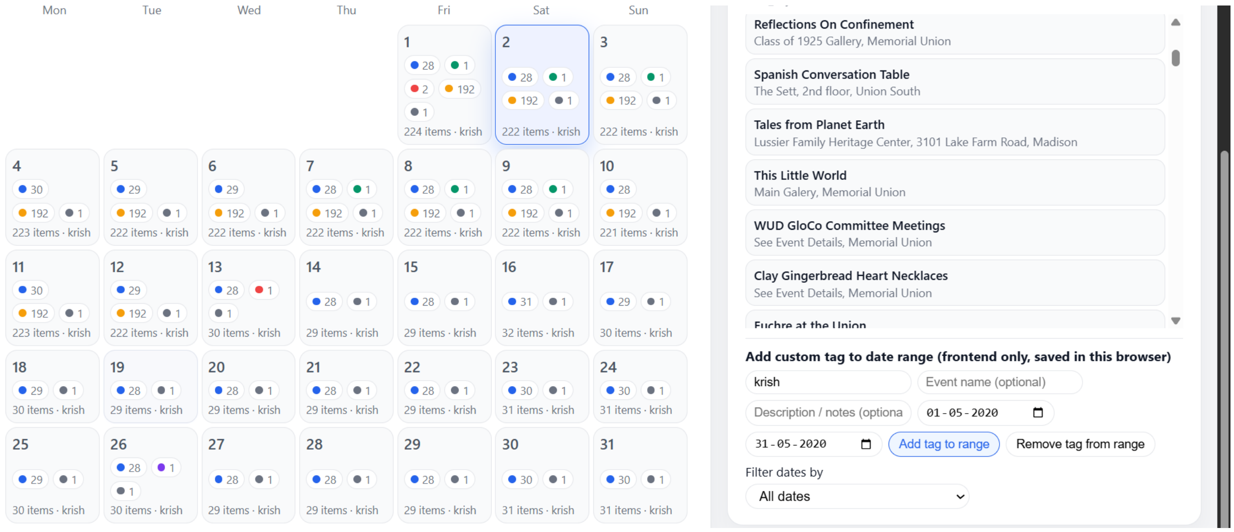

Figure B2. Custom Tag Creation for Data Ranges:
Shows the user interface where local, browser-stored custom tags are created and applied to single or multiple consecutive dates.

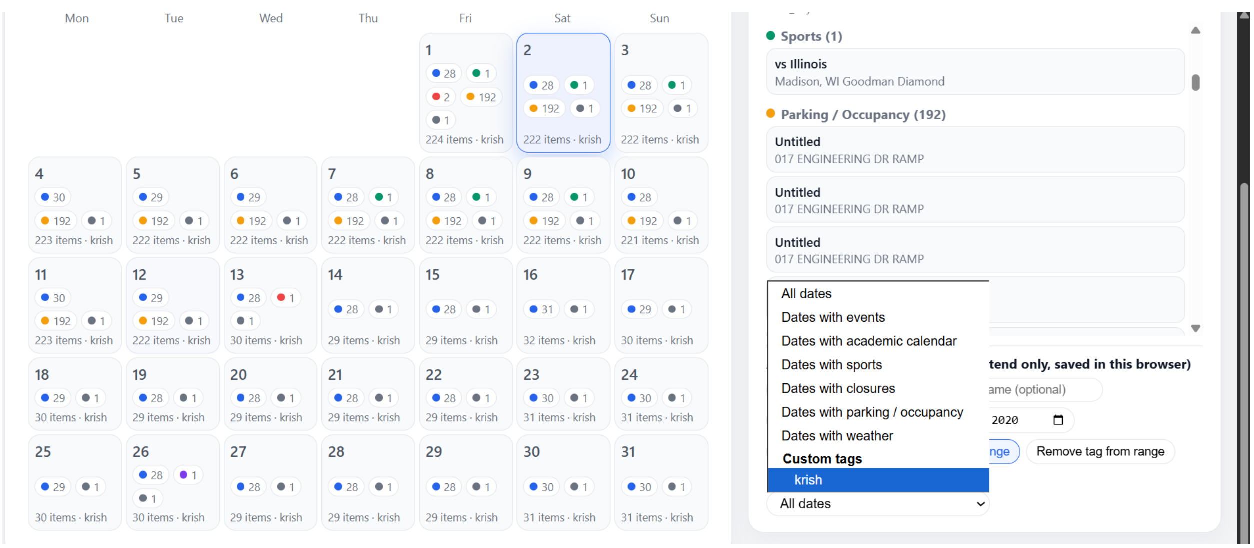

Figure B3. Tag Filtering Panel:
Highlights the filtering interface that allows users to toggle visibility of specific tag categories to support focused exploratory analysis.