Transport Delay Analytics Report: Public Transit in Prague

Nils Wüstefeld, Adam Pasálek

June 15, 2025

Contents

1 Introcution

This project analyzes delay patterns in Prague's public transportation system using schedule and real-time arrival data. The goal was to identify key manifestations of delays and assess their temporal and spatial properties. The resulting data is then visualized in various plots giving insights to the delay structure in Prague. The analysis highlights that peak hours and specific bus and tram routes are responsible for the majority of delays.

2 Application Overview

Our application is a client-side web app built using Streamlit, an open-source Python framework for web interfaces. Streamlit automatically converting Python scripts into shareable web apps without requiring extensive front-end coding. The application consists of five main pages:

- **Landing Page**: Provides an introduction and overview of the application's purpose.
- **Connection Page**: Allows users to upload their credentials to use the App-
- **Dashboard**: Displays key metrics and real-time delay statistics of the trips
- **Stops Page**: Enables users to explore delay data and schedule information for individual transit stops.
- **Heatmap**: Presents a spatial visualization of delays across the transit network, highlighting geographic patterns and hotspots.

2.1 Data Sources and Method

The application relies on two types of GTFS data obtained from the Golemio API endpoint provided by *Pražská integrovaná doprava* (PID). The General Transit Feed Specification (GTFS) is the de facto international standard for digital public transport schedule data. It is generally divided into two categories: static GTFS and GTFS Realtime. The static GTFS feed includes scheduled information such as routes, stops, stop times, trips, and service calendars. This data provides the structural foundation of the transit system and changes relatively infrequently. In contrast, GTFS Realtime contains dynamic updates, including live vehicle positions, service alerts, and trip modifications, allowing for the monitoring of delays and real-time performance. GTFS datasets consist of multiple interlinked tables, connected via primary and foreign keys. For instance, each trip is associated with a specific route, and stop times link trips to individual stops. During initialization, or when refreshing, the application extracts the `routes`, `stops`, and `trips` tables from the static GTFS feed. Real-time GTFS data is utilized when a user interacts with the dashboard or stops page. Since historical GTFS real-time data is not available, we maintain a proxy server with a RESTful API that continuously fetches real-time delay information. This setup allows us to accumulate and store a comprehensive database of historical delays for retrospective analysis and performance monitoring. If the user interacts with the dashboard or stop page, an encrypted request is sent to the proxy server, which in turn returns the provided data as a string for processing.

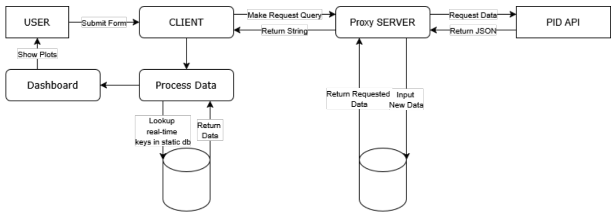


Figure 1: Data Flow of a client request.

If a user submits the form on the stops or dashboard page, a query request is sent to the proxy server which queries the requested delay data from its database. The database is constantly updated with new delays from the PID API. The requested data is returned to the client cia the server's RestfulAPI. Then, the data is aggregated, filtered and enhanced with static information from the client database. Subsequently the data is visualized and can be inspected as plots on the app.

3 Sample Day

To conduct a analysis of public transport delays in Prague, we select Friday, June 13th, 2025 as a representative sample day. The weather on this day was sunny and warm and the city experienced a high volume of tourists, making it an interesting case for examining system performance under typical but intensified usage. For the analysis we split the analysis into two parts: Trip analysis and stop analysis.

3.1 Trips analysis

For the trips, we want to find out which vehicle types are delayed, what the delays are on average and when do they happen, while for the stops we want to examine at which stops the vehicles are most delayed. The analysis begins by setting the start and end date using the form on the dashboard page of the application. By default, the minimum delay threshold is set to 1 minute, but users can modify this parameter based on their individual preferences to refine the results. To maintain a manageable scope for this analysis, we focus on the central tariff zones P, 0, and B, which approximately cover the inner city and its immediate surroundings. We first examine the summary statistics from delay statistics table in Figure 3 It provides a breakdown ordered delays ordered by vehicle type, using three key metrics: the total number of delay observations, the average delay per trip (calculated as the mean of means), and the maximum recorded delay. The minimum is not displayed, since its a user defined input. Buses account for the largest share of delay data, with over 22,000 observations. Their average delay is moderate at around 167 seconds, but they also exhibit one of the highest maximum delays in the dataset—over 5,400 seconds (approximately 1.5 hours). This suggests that while most bus trips run with acceptable delays, occasional outliers can result in significant disruptions. The outliers might be results of traffic, accidents or malfunctioning transponders. Trams, the second-most frequent mode with more than 6,400 entries, show slightly lower average delays (138 seconds) but have the highest maximum delay at nearly 5,900 seconds (about 1 hour and 38 minutes), indicating similar variability in performance. Ferries have the fewest entries (only 45 observations), yet they exhibit the highest average delay per trip—over 400 seconds. This likely reflects the infrequent service and higher sensitivity to disruptions, even though the overall sample size is small.

Metro services (342 entries) and trains (1,141 entries) both show average delays in the

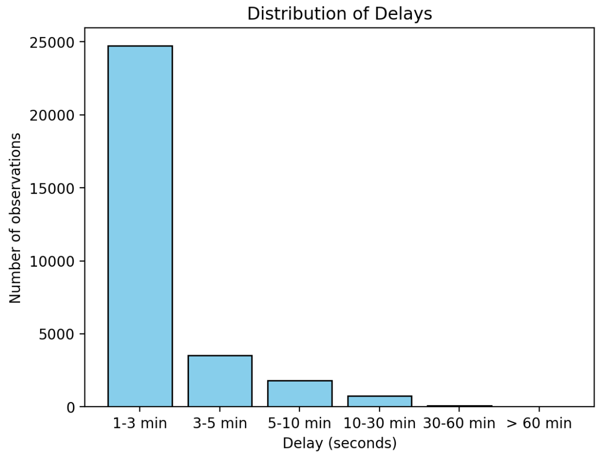


Figure 2: Summary Table of the delay distributions

Delay Statistics

Mean and maximum delay grouped by vehicle type.

route_type	Count	Mean Delay	Max Delay
bus	22272	167.5	5431.2
ferry	45	411	2420.6
metro	342	211.3	2606.1
train	1141	183.6	2303.3
tram	6436	137.6	5881
trolleybus	668	135.9	1644.3

Figure 3: Summary of the delay distribution for Friday the 13th of June 2025

range of 180–210 seconds and have maximum delays just over 2,300–2,600 seconds. This suggests a consistent level of performance for both systems, with occasional, though not extreme, delays. These delays are cause by constructions efforts on line C. Trolleybuses, with 668 entries, show the lowest average and maximum delays—136 seconds and 1,644 seconds respectively—indicating strong reliability within this subset of the public transport network. In summary, buses and trams experience the most frequent and occasionally severe delays, while metro, train, and trolleybus services appear more stable. Ferries are less commonly delayed but tend to have longer delays when they do occur. The shares of delays by type are visualized in Figure 4.
Next we take a closer look at the delay distribution. Overall, we see that by far the majority delays are within 1 to 3 minutes. The other delay bins of 3-5 minutes , 5-10 minutes, 10-30 minutes and more than 60 minutes are way rarer than the one to 1-3 minute delay. The delay distribution is provided in Figure2. Overall we can see that most delays are small an can even be considered within the rage of punctuality, highlighting the service reliability of Prague's public transport system.

Next we examine the temporal distribution of the delays over the whole day. When examining the plot in Figure 5 it becomes apparent that the different frequency of the vehicles results in different smooth plots. For example the plots for bus and tram are rather smooth, while for trolley bus and ferry the plots are spiky. The Ferries that day have among the largest individually delayed trips. The ferry trips at 13:05 have the biggest average delay. This large delay can be a results of a high demand or malfunctioning transponders. When only examining the busses and trams, we get a clearer picture of the overall dynamics. The buses and trams have the highest service frequency of all

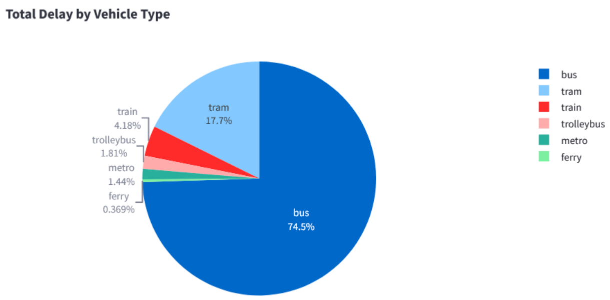


Figure 4: Pie Chart for the share of delays, ordered by vehicle type

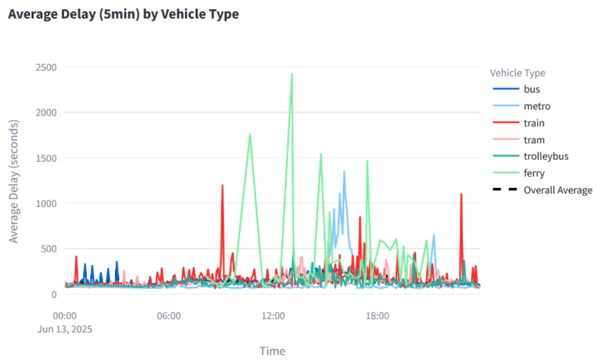


Figure 5: Average Delay for Friday the 13th of June aggregated to 5 minutes rolling windows.

vehicle types and have to share the roads with other road users like cars, pedestrians, bikes and trucks. The resulting plot is visualized in Figure 6. When examining the plot we can clearly see some spikes in the night and systematic increase in the average delay for the buses an trams at roughly 12:00 which gradually goes down till 20:00. In this time frame, the trams encounter two major spikes a 13:40 (403 sec) and 15:10 (562 sec). The degree further decreases in the reminder of the evening. This is in line with regular commuter travel on a Friday and traffic cause by tourists. Next, we examine the ten trips with the highest recorded delays. While extremely high outliers have been excluded from previous visualizations for aesthetic clarity, we still analyze them here, as they may reflect genuine disruptions rather than systematic measurement errors. Figure 7 presents

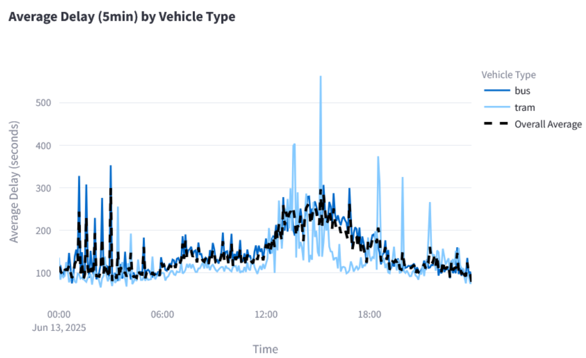


Figure 6: Aggregated delays of the busses and trams over a rolling 5 minute window

Top 10 Delays

Top 10 Delayed Trips by Maximum Delay

This page shows the top 10 delayed trips based on maximum delay.

	gtfs_trip_id	delay	route_short_name	route_long_name
0	7_13533_250602	5881	7	Radlická - Lehovec
1	306_1165_250106	5431.25	306	None
2	306_220_250106	5315.75	306	None
3	240_737_250415	4914	240	None
4	306_1350_250526	4570.7222	306	None
5	7_13520_250602	4528.5	7	Radlická - Lehovec
6	306_217_250106	4514.5	306	None
7	306_216_250526	4433.6667	306	None
8	1732_293_250508	4316.4167	1732	None
9	371_69_241202	4066.8571	371	None

Figure 7: Top Ten delays

the top ten delays observed on the selected date. The longest recorded delay occurred on Tram Line 7, with a delay exceeding 1.5 hours. This was followed by two instances of Regional Bus Line 306, both with delays of similar magnitude. Notably, five out of the ten most delayed trips belong to Line 306, which may indicate recurring issues specific to this service, vehicle or its route. However, due to limited data consistency and gaps in metadata, it is difficult to draw firm conclusions. Even a manual lookup using the internal client-side database did not yield further relevant information.

3.2 Stops Analytics

To complement the trip–level view, we analyse stop-level behavior on **13 June 2025** too. All screenshots and statistics in this subsection were produced with the *Stops Analytics* page. Again, we focus only on the Prague zone and this time also excluding the metro as the stops analytics for this type of transport wouldn't make that much sense.

3.2.1 Overview tab

Figure 8 presents the sortable table of all Prague zone-P stops. For each stop we display its geographic coordinates and the number of individual platforms that belong to the parent station. In total, 1393 parent stops and 3440 platforms are listed.

Prague Stops (zone P)

	Stop	Latitude	Longitude	Number of Platforms
0	Albertov	50.0679	14.4208	3
1	Ametystová	49.9882	14.3622	1
2	Amforová	50.0418	14.3273	2
3	Anděl	50.0711	14.4034	10
4	Antala Staška	50.0416	14.4444	2
5	Apolinářská	50.0715	14.4268	2
6	Arbesovo náměstí	50.0763	14.4048	4
7	Archeologická	50.0441	14.3378	2
8	Arena Libeň jih	50.1029	14.4946	2
9	Argentinská	50.103	14.4434	3

Figure 8: Prague Stops table

The interactive map in Figure 9 visualises those stops. As a default it shows *All Stations.* after clicking on a stops, the number of its is displayed. Clicking on *All Stations* shows every parent stop, selecting (or searching for) a specific parent stop shows its platforms on the map. After clicking on the stop pin, a platform name is shown.

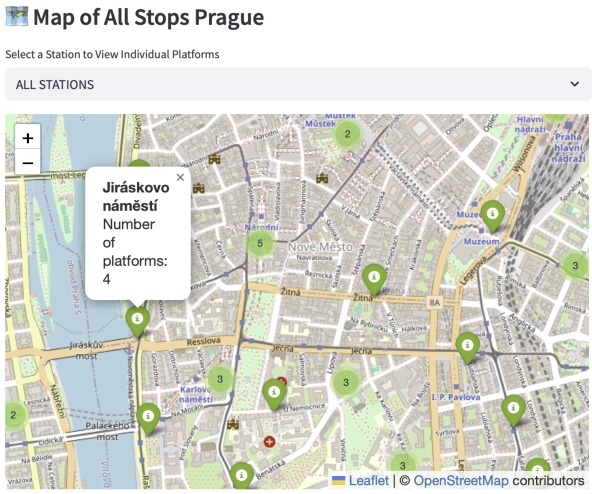
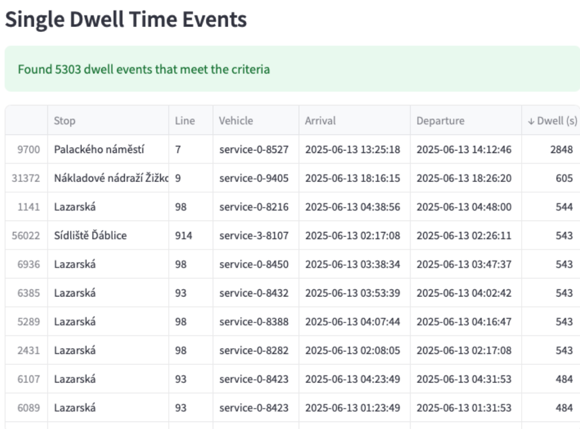


Figure 9: Prague Stops map

3.2.2 Dwell Time Analysis tab

After choosing the sample period and a minimum dwell threshold of 60 seconds (because vehicle-position snapshots are collected only once per minute), the application returns a list of single dwell events which can be sorted for example by the dwell duration as shown on figure 10. Thanks to that we can see the longest dwell of the chosen time period. On June 13th, a total of 5,303 dwells longer than 60 seconds were recorded. One notable outlier was tram 7, which stood at a stop in Prague's centre for approximately 45 minutes. Our investigation suggests two possible explanations: either there was an error in the tram's system connection with the PID servers, or an exception caused traffic to halt. Looking at the bottom of the table's preview, we can see night-time tram lines waiting for other night-lines at the central night transfer point Lazarská.



Single Dwell Time Events

Found 5303 dwell events that meet the criteria

	Stop	Line	Vehicle	Arrival	Departure	↓ Dwell (s)
9700	Palackého náměstí	7	service-0-8527	2025-06-13 13:25:18	2025-06-13 14:12:46	2848
31372	Nákladové nádraží Žižkc	9	service-0-9405	2025-06-13 18:16:15	2025-06-13 18:26:20	605
1141	Lazarská	98	service-0-8216	2025-06-13 04:38:56	2025-06-13 04:48:00	544
56022	Sídliště Ďáblice	914	service-3-8107	2025-06-13 02:17:08	2025-06-13 02:26:11	543
6936	Lazarská	98	service-0-8450	2025-06-13 03:38:34	2025-06-13 03:47:37	543
6385	Lazarská	93	service-0-8432	2025-06-13 03:53:39	2025-06-13 04:02:42	543
5289	Lazarská	98	service-0-8388	2025-06-13 04:07:44	2025-06-13 04:16:47	543
2431	Lazarská	98	service-0-8282	2025-06-13 02:08:05	2025-06-13 02:17:08	543
6107	Lazarská	93	service-0-8423	2025-06-13 04:23:49	2025-06-13 04:31:53	484
6089	Lazarská	93	service-0-8423	2025-06-13 01:23:49	2025-06-13 01:31:53	484

Figure 10: Dwell table

Aggregating by stop yields the cumulative dwell-time bar chart in Figure 11. The top stop is Lazarská probably thanks to the reasons mentioned above.

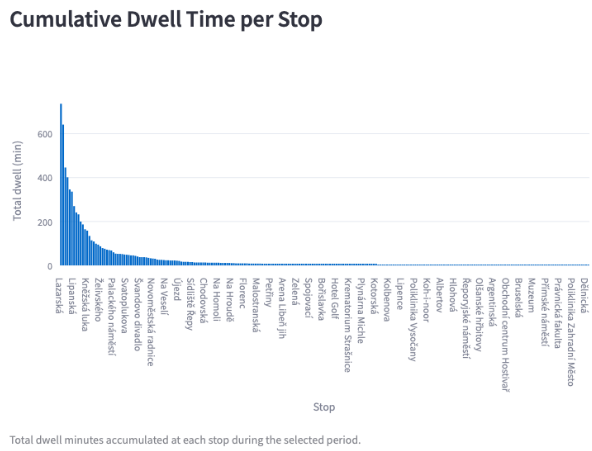


Figure 11: Dwell bar plot

3.2.3 Stop Throughput tab

The throughput tab quantifies how frequently vehicles traverse each stop, independent of dwell duration.

Average vehicles per hour. The first plot (Figure 12) sorts stops by average number of vehicles per hour. Anděl (not surprisingly) leads with more than 25 vehicles per hours (an average, which also includes nights), reflecting its role as a major tram interchange.

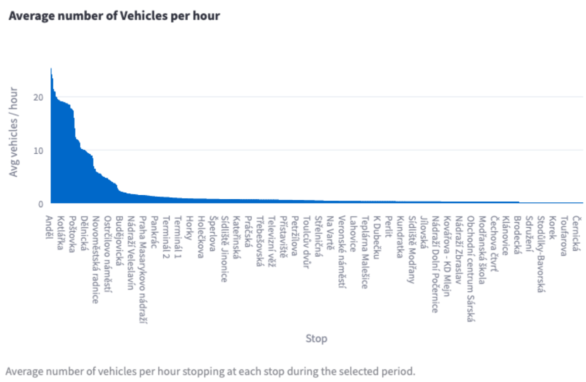


Figure 12: Average number of vehicles per hour per stop bar plot

Throughput heat-map. Figure 13 overlays throughput intensity on the city map. Corridors along the tram axes most notably in the Národní třída area appear in warmer colours.

Overall, the stop-level perspective complements the trip analysis by revealing where delays accumulate (*dwell*) and where operational pressure is highest (*throughput*).

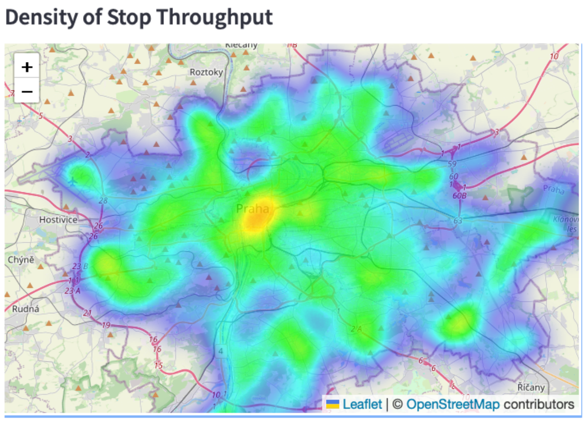


Figure 13: Map of Stop Throughput

4 Limitations

One major limitation of this analysis is the absence of data on punctual trips. As a result, it is not possible to determine what proportion of all trips experienced delays, nor can we confidently assess the overall punctuality of the transport system. This also restricts our ability to develop and validate predictive models for delays, as doing so would require access to a complete dataset of all trips including time—which data, which would exceed available computational and storage resources. Another limitation is the presence of missing or inaccurate GPS data during certain time intervals. These inconsistencies may lead to gaps in tracking, inaccurate delay calculations, or misclassification of vehicle positions, ultimately affecting the reliability of the analysis. Moreover not all data from the real-time feed is properly labeled, complicating analytics further.

5 Conclusion

Overall the analytics and report provided by the app can give a overview on the punctuality of the Prague public transport network. Based from the insights obtained from the sample day, the public transport network appears to be efficient, since most delays are within acceptable range below 5 minutes.