

## ***Punctuality on the Roslagsbanan:*** **How Line Stop Amount Influence Punctuality**

### Introduction

Public transportation networks are constantly evolving in response to changing demographic needs, and there is a critical need to understand their complex operations. Analyzing operations of the systems can generate important findings for planning future enhancements. This in-depth analysis is directly relevant to understanding the transportation network of Stockholm, specifically the Roslagsbanan network. The Roslagsbanan is a unique, isolated light-rail in Sweden, being the only narrow-gauge railway in Sweden including service concepts that blend standard and express services. The study will explore the impact of different service patterns on punctuality using the Roslagsbanan network as a case study.

Through an operational data analysis, the study aims to understand the factors that drive system delays and how the type of service affects those factors. Understanding these can lead to improvements in the scheduling plan, the services offered, and the overall reliability of the Roslagsbanan network. In short, SL (Storstockholms Lokaltrafik) can ensure that the services provided better align with Stockholm's commuters' needs while enhancing the efficiency of the overall transportation network.

The Roslagsbanan network is a unique public transport segment in Stockholm. To study the influence of the number of stops on punctuality, it is crucial to understand the Roslagsbanan network and define the research scope. The Roslagsbanan network is located in northeast Stockholm and has three central lines, 27, 28, and 29 (SL, 2024). Overall, the Roslagsbanan network has 39 stations with three branches starting at Stockholm's Östra and going to Kårsta, Österskär, and Näsby park.

The network dates back to the late 19th century (1885) when it was created as a narrow-gauge tram line connecting Stockholm with the Roslagen region (Victory Stockholm AB, 2024) (SL, 2024). During the 20th century, multiple lines merged into what we see today, a system that connects what used to be rural villages with the capital of Sweden. Not only did the Roslagsbanan network contribute to the economic viability of the Roslagen region, but it is also an early example of organic transit-oriented development. Eventually, these private lines became part of the government's SL network in the mid-20th century (SL, 2024).

When acquired, officials recognized that the Roslagsbanan needed significant modifications, and this responsibility was given to Region Stockholm. In the 2010s, many further improvements were done to the system, including expanding the network, adding double tracks and new trains. In the future, the system will be connected to T-Centralen, the central public transportation hub for Stockholm city (Region Stockholm, 2024). Today, the lines' transit hubs include Stockholms Östra, Mörby, Täby Centrum, and Åkersberga, and they account for an estimated 13 million journeys every year (Transdev, 2021). As noted above, in addition to its standard service, the Roslagsbanan light rail now offers express service (27S, 28S, and 28X). The express services make fewer stops than the standard trains and only run during busy commuting hours. Understanding how line characteristics influence punctuality when comparing standard

and express services will provide additional valuable insights into the operations of SL's Roslagsbanan lines.

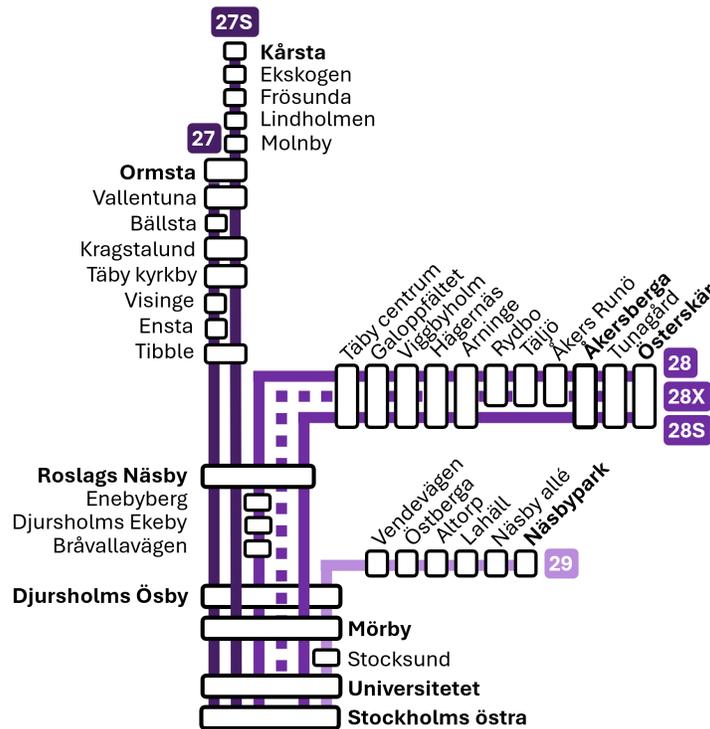


Figure One: The Roslagsbanan line network.

The problem statement to understand how the number of stops affects punctuality for the Roslagsbanan network was chosen to illustrate a two-factor public transit analysis. Punctuality impacts both passenger experience and network performance. In addition, the authors' personal daily experience with the Roslagsbanan line provides an additional real-world connection to the analysis. As a result, insights regarding delays on the network could directly affect our behavior and use of the network. Because the Roslagsbanan line connects to KTH, its performance and operations directly affect students' lives. More broadly, the network analysis, including how passenger interactions across multiple lines (standard and express), could potentially inform the choices of the broader public. The core hypothesis is that more stops and distances can expose a line to additional delays, whether caused by technical factors or passenger behaviors. Looking in detail at isolated features of station punctuality across the network can provide additional insights into exactly how stop counts affect punctuality. This will also give more information on the benefits of express service on the Roslagsbanan network and user experience.

## Methodology

The analytical study of the Roslagsbanan line uses a quantitative approach to assess the impact of the number of station stops on the punctuality of the trains. In evaluating the punctuality differences between the express services (27S, 28S, and 28X) and the standard services (27 and 28), the data collection effort focused on the primary commuting times when both services are offered. We focused on Monday to Friday evening commuting hours in 2024, specifically between 16:00 and 19:00 on April 30 and May 6, 8, 16, and 17. The variety of analysis will reveal critical information on the network lines using both

automated and secondary data. Relying primarily on the SL Transport real-time automated data source (captured via the API) to obtain automated real-time data, our research is also based on stated operational and timetable data from SL. Using this combined dataset, we can analyze the different lines to explore the relationship between the number of stops and punctuality outcomes. Note that when conducting the analysis, we did our best to control for outside variables that could affect service times, including (but not limited to) days of the week, data collection times, and special scheduling changes (e.g., holiday service). Using the mixed-methods approach in data analysis, we can statistically analyze the influence of the number of stops on punctuality for Stockholm's Roslagsbanan network, focusing on comparing standard and express services.

In any data-related study, many independent, dependent, and controlled variables exist. Diving into the variables in our study, let us start with independent variables. Independent variables are variables that can affect the dependent variable being studied. In our case, the independent variables include the number of stops on the line, stations where data is collected, bunching and the type of line (standard or express). Dependent variables are variables that change because of changes in the independent variables. The dependent variable for the study is punctuality or, as measured here, the average delay times. This also, in turn, includes speed as a factor. The last category of variables is controlled variables. Controlled variables can inadvertently influence the study's outcome and, thus, must be controlled to the best of one's abilities. Controlled variables in the study include which days of the week and times for which data is collected, seasonal variations, and special events like holidays (e.g., Valborg, Labour Day, and Ascension Day) or maintenance periods. Train frequency, which depends on the controlled variables, will be considered. The origin of much of the data affecting service punctuality is the real-time station arrival/departure data, which the APIs can be programmed to collect at corresponding stations. Overall, identifying the variables enables one to identify what factors are most relevant to the study, and controlling outside factors to produce clear results.

Several limitations had to be overcome when working through the data. First, there were significant limitations regarding the possible project scope, given the time allocated for the effort. Another such limitation was the lack of data for the 28X line. The line connects Stockholm's Östra to Österskär and has 20 stops. However, as can be seen in the timetable, there were minimal trips of the line during the days and times for which we collected data. Additionally, the results relied on SL data, which posed a challenge given the absence of arrival times in the API data. Because of this, it was not possible for us to analyze the impact of dwell times on punctuality.

On the other hand, we analyzed bunching as we expected it to be of crucial importance on Roslagsbanan reliability. Under optimal conditions that include being free of the influence of weather and other traffic as well as having a route equipped with moving block signaling, the minimum time interval possible for trains is around 70 seconds, as that is considered physically possible (Menzel, 2021). However, the Roslagsbanan line does not meet these optimal conditions due to factors such as weather and its block-based signaling system. Instead, the minimum realistic interval for the Roslagsbanan network is closer to 90 to 120 seconds, or 1.5 to 2 minutes. The time between trains on the Roslagsbanan network is meaningfully longer, as seen in the table below. For the sake of this project, the minimum threshold between two trains was set to 150 seconds or 2.5 minutes to account for bunching. Assuming from the official timetable of the Roslagsbanan network and our threshold, we can see that the network is operating

almost at maximum capacity during peak hours, as the time between many trains is often only slightly above our threshold (3 minutes with a threshold of 2,5 minutes).

Line	Time to next train
28S	3'
27	3'
28	3'
29	5'
27S	6'
28S	3'
27	3'
28	3'
29	4'
27S	7'
28S	3'
27	3'
28	3'
29	5'
27S	

Figure Two: The running order of Roslagsbanan at 16:00 from Universitetet and time to the following train.

For our data collection, we used the new SL API to gather real-time information about the Roslagsbanan. This data was collected in the afternoon-peak (from 16:00 - 19:00) as mentioned before in the Methodology section.

Data was collected at four stations, which can be seen in Figure 4: Universitetet, Roslags Näsby, Vallentuna, and Åkersberga. All four stations are stops on both standard and express service lines. Moreover, the station Universitetet is closer to the start of the lines and has the 27 and 28 and their respective express services going through. Roslags Näsby is a central station where the lines to Österskär and Kårsta branch off. Furthermore, it marks the end of the corridor from Stockholms Östra (see Figure Three), in which some stops are only served by stopping line 28, making this point ideal for analysis.

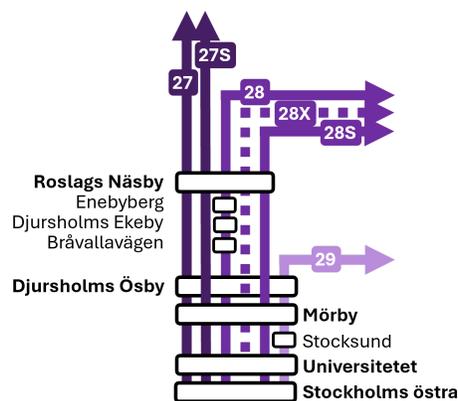


Figure Three: The corridor from Stockholms östra to Roslags Näsby.

This compares to Vallentuna and Åkersberga, which are towards the end of their respective branches. Vallentuna is a stop on the 27 line to Kårsta, whereas Åkersberga is a stop on the line 28 to Österskär. At these stops, the outgoing trains have already had a significant travel time including differences in stopping

patterns amongst the stopping and express services, making them ideal for analysis. The collection points can be seen in Figure 4, which also shows the difference in stopping patterns among the different services.

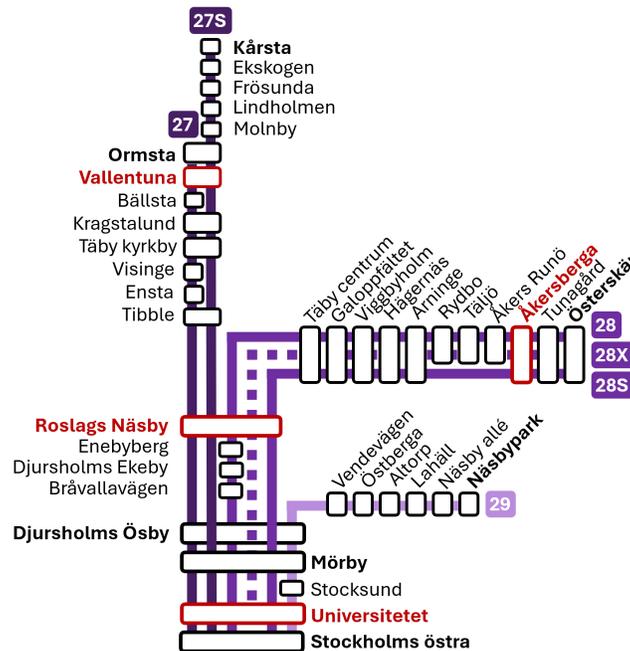


Figure Four: The Roslagsbanan line network with collection points highlighted.

The data was collected using the SL real-time API in a 0.5-minute time interval, with a focus on departure times from stations. Ensuring that the limitations mentioned beforehand were accounted for systematically and methods were in place to overcome the challenges ensured the most accurate results regarding the relationship of train punctuality and frequency variations across times and stations.

Our Python analysis tested whether line service type, distance, bunching and the number of stops influenced delays. The difference between actual and scheduled departure times was calculated using SL real-time data API and Python coding. During the analysis, we took note of the direction of travel, i.e. towards or away from Stockholm city center, which was essential since it determined how many stops had occurred on the line before the station being measured. As noted above, the minimum threshold of bunching was set at 2.5 minutes and the occurrence of Bunching was analyzed through the time between trains.

Using Python code, the data was cleaned and filtered for each of the respective lines. We used box-plot graphs to visualize the difference between expected and actual departure times and scatter plots to visualize the buffer times and bunching, providing insights into the relationship between delays and time differences and bunching.

## Results

Going into the study, the group's expected that the longer lines with more station stops would lag in punctuality (i.e., have longer delays). This expectation was informed by various assumptions and a review of similar studies. These previous studies inspired the analytical approach and variables chosen to study.

One such study is *Some Influencing Factors for Passenger Train Punctuality in Sweden*, conducted at Lund University. The study found that trains that travel across larger geographic spaces experience a "correlation coefficient of 0.47" related to variation in punctuality based on distance traveled (Palmqvist, 2017, p. 7). Palmqvist also found that for a journey, the interactions at each station lead to a decrease in punctuality by "1 %-point per interaction," thus suggesting that a greater number of stops would lead to more delays (Palmqvist, 2017, p. 9). Palmqvist used a similar analytical approach when it came to train movement tracking through the use of timetables as one of the data sets. Moreover, the analysis dove deeper into other variables affecting punctuality, covering weather, timetable, operations, and infrastructure. Thus, the critical difference between our two studies is the complexity of what was considered as it relates to punctuality. However, as seen below, the studies used similar outputs, such as scatter diagrams and correlation matrices, to analyze data. Furthermore, as Hansson recounts in *Effects of Rural Bus Stops on Travel Time and Reliability*, studies have shown that "the number of stops will make the service slower and travel times less predictable" (Hansson, 2021). Likewise, Hansson discovered that longer travel times increase service variability and delays. For the study, the use of dwell time, run time, time variability, and on-time performance was observed. In Hansson's study, dwell time could be known, but the SL API does not include such data. Furthermore, Hansson's study used similar time-point segment-level measurements and route-level data. Because of the similarities to our research study, both studies informed our expected results and data analysis approach. However, it is critical to point out that both studies involved substantially a larger effort than our study, which means our research could serve as a launch point for a full-fledged research project.

After gaining insights about past research, now is time to dive into the results from the SL API. First, the total results averaged over the four selected stops are investigated, which can be seen in figure five. Here a surprising trend occurs; the express services seem to perform worse than the 'regular' lines 27 and 28. The biggest outliers have been removed, but both with and without them the trend stays clearly defined; express services perform significantly poorer than the stopping services. The boxplots demonstrate this clearly by moving up on the y-axis respectively, demonstrating a significant difference in delay time between the 25th and the 75th percentiles, but also the median lines show a difference for both line groups.

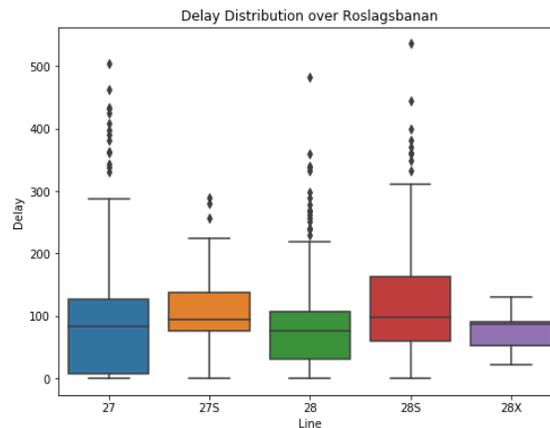


Figure Five: Total delay distribution over all stops

Line 28X performs relatively well, better than the other express services (27S and 28S). This result might be less reliable however, as that line is by far the least frequent one. The data sample is simply smaller for

this line, so more unforeseen errors could creep in. This is why most comparisons for the rest of the report will be between the regular and the ‘S’ services.

This overall result demanded a more detailed look at specific stations and lines; therefore, the next results are specifically aimed at finding differences for line 28 and its express variants. This is done by looking at trains going from Stockholms Östra in the direction of Österskär, and following them through three stops: Universitetet, Roslags Näsby, and Akersberga (as seen in figures six, seven, and eight). Note that 28X did not reach Akersberga during the times data was collected. From these figures a preliminary idea is born; it may be the case that the extra stops act as a buffer, and make line 28 more reliable than the express versions. This is because the differences really start to emerge at Roslags Näsby, and that is the first stop which differences in stop schedule have preceded.

By not stopping at Djursholms Ösby, Bråvallavägen, Djursholms Ekeby, and Enebyberg, lines 28S and 28X relinquish the opportunity to recover some time which may have been lost at the start. These smaller stops are generally not as busy as the main hubs where the express service stops, so the embarking and disembarking times might be shorter than the scheduled stop times. The delay times improve so vastly over more stops that, by the end of the line at Akersberga, there is little to no delay left for line 28. For 28S the delays have remained almost constant over the whole line, with a median value of about 110 seconds.

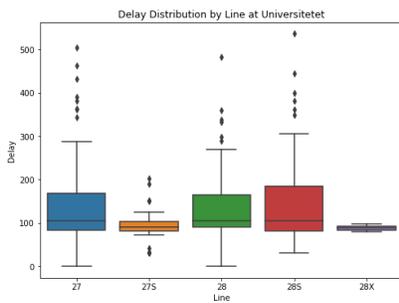


Figure Six: Delay at Universitetet

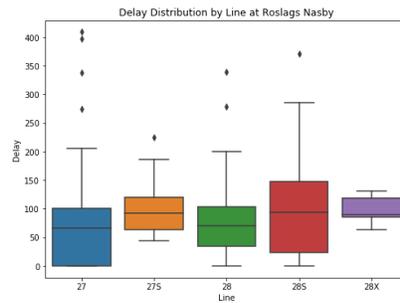


Figure Seven: Delay at Roslags Näsby

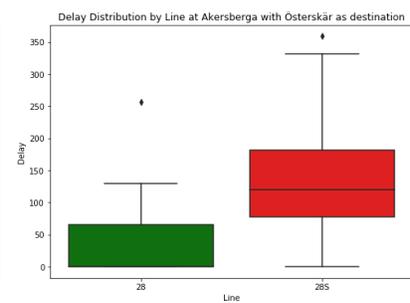
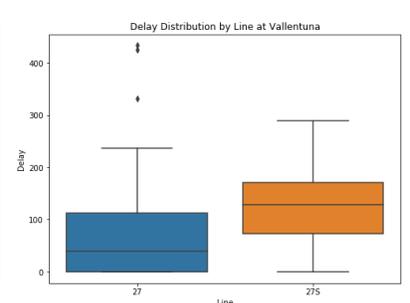
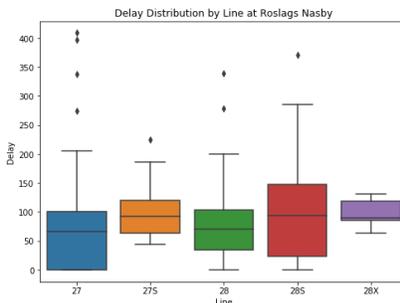
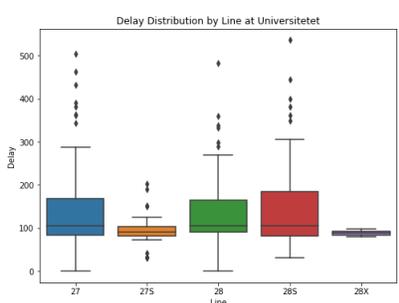


Figure Eight: Delay at Akersberga

For lines 27 and 27S a similar trend seems to occur, as seen here in figures nine, ten, and eleven. The median delays are close at the beginning of the line, at Universitetet, but the lines diverge the more distance is traveled, and the more stop differences occur. For line 27 one of the last stations is Vallentuna, here the difference is as clear as it was for line 28 at Akersberga. By skipping the stations Ensta, Visinge, and Bällasta, line 27S does not get as close to the scheduled times as the regular line 27 does.



Besides looking into line performance, it is also relevant to mention a general performance indicator of the Roslagsbanan; at Roslags Näsby, a station central in the network, almost no trains arrive on time during peak hours. Figure twelve demonstrates this, showing that all lines have a median delay of around 100 seconds after traveling for around 15 minutes. It can be noted for this graph that line 27, 27S, 28S, and 28X all act as express services, with only line 28 making more stops in between Universitetet and Roslags Näsby. And whilst the differences are not big, it is line 28 that seems to perform best out of the 5 alternatives seen here. This strengthens again the idea that perhaps the amount of stops plays a part in delays.

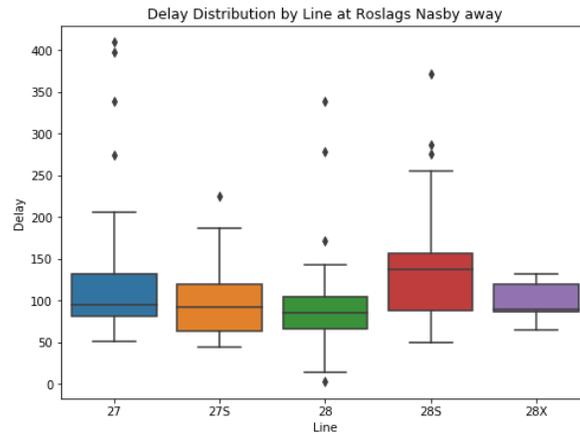


Figure Twelve: Delay distribution at Roslags Näsby

After these findings some more in depth research was carried out, with the goal of finding out what was causing the delays. Here the initial hypothesis was that the express services might bunch up close to other trains more quickly, causing them to suffer more delays that way. It could be the case that the amount of stops and the level of bunching are correlated.

To figure this out the bunching per line type at each stop has been calculated, and is shown in figure thirteen. Using the assumption that trains that follow closer than 150 seconds encounter negative effects, this became the threshold to classify bunching. It should be noted that bunching at Universitetet could happen due to the timetable settings, as some trains follow 180 seconds after the previous. This small margin gives a few recordings of bunching, but these are not expected to interfere much with the planned running. The stops Vallentuna and Akersberga find very little bunching because there are simply fewer trains serving the ends of branches over the main artery.

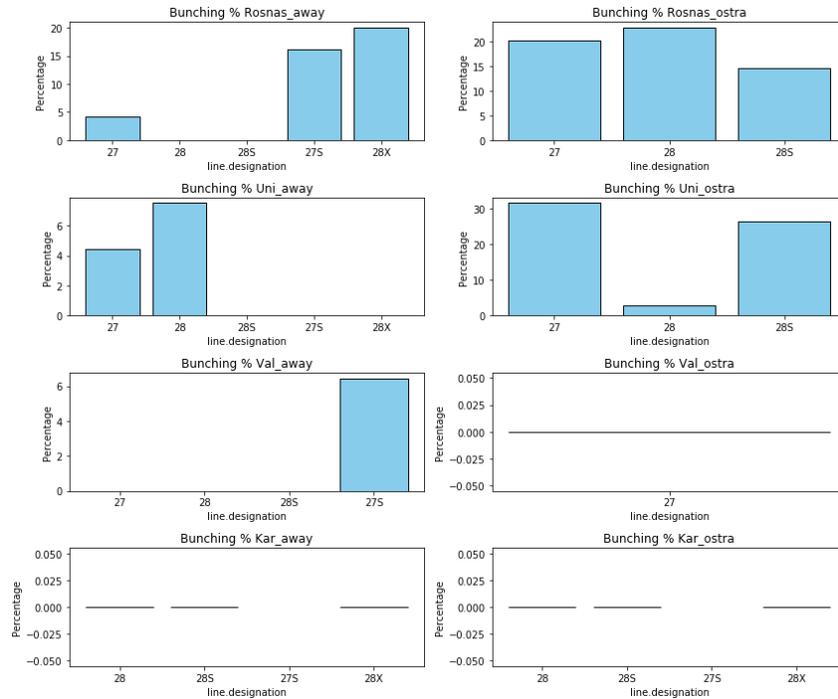


Figure Thirteen: Bunching levels per stop.

The most relevant changes from these graphs are found between Universitetet and Roslags Näsby. Trains leaving from Stockholm Östra are sporadically bunched. This happens around 5% of the time for line 27 and 28. It can be seen that this level of bunching stays on line 27, which operates on the same stations as the express line, but it disappears for line 28, which makes extra stops. Lines 27S and 28X also gained a high level of bunching, with significantly larger values of 15 and 20% emerging. These degrees of bunching, further along the path, suggested that perhaps the express trains may be experiencing delays because of bunching.

This pattern of line 28's bunching levels decreasing with extra stops also occurs for trains in the direction of Stockholms Östra. Bunching levels are for all three recorded lines around 20% before the stop differences, but line 28 loses almost all of the bunching, whilst it even increases for 27 and 28S. The fewer stops seem to correlate with more bunching.

To find the effects of this perceived bunching on delay times, two scatterplots have been created, shown in figures fourteen and fifteen. This is done separately for both directions because a time difference is always recorded to the previous train in the same direction. What is surprising here is that the relation seems to be positive; larger time difference between trains (especially in the direction away from Stockholms Östra) seems to cause more delays. This is counterintuitive and in opposition with the hypothesis, as it was expected that bunching effects (trains close together) would cause delays.

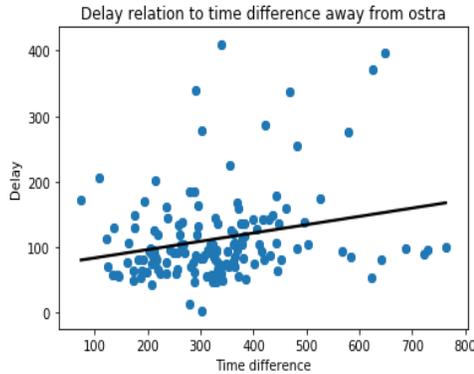


Figure Fourteen: Scatterplot relation away

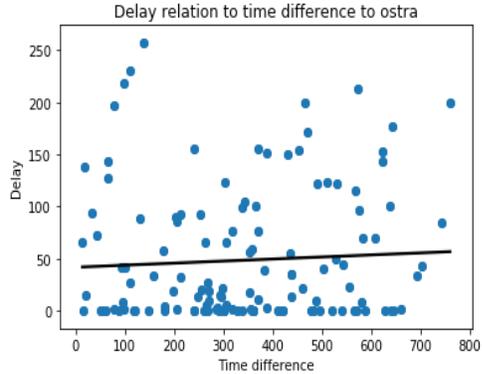


Figure Fifteen: Scatterplot relation to Östra

This effect has been checked by also creating a correlation matrix for time differences and delays, which can be seen in figure sixteen. This correlation matrix also produces small, but positive values, confirming the ideas from the scatterplots.

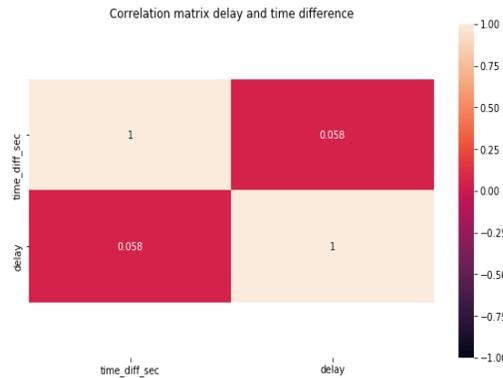


Figure Sixteen: Correlation matrix

## Conclusion

The main finding from the results is clear: the express services on the Roslagsbanan generally perform worse (during peak hours) than the ‘regular’ lines 27 and 28. This contrasts the hypothesis, which stated that more stops were likely to cause more delays, implying that the express services with fewer stops should have performed better than their regular counterparts. This compares to the expected results, which were informed by Hansson’s study. The significant difference that could explain the drastic results is that the Roslagsbanan runs on an isolated network, not being influenced by occurrences happening far away, as it can be the case in the main rail network of Sweden. Furthermore, it should also be addressed that the dataset used is of a small count size, further possibly impacting the result and conclusion of the study.

While the result of the conducted study was surprising initially, there can be many explanations for why this might be the case. The main reason that was explored was the effect of bunching on delay times. Because the express services make fewer stops, they are effectively faster than the regular trains, and bunching could start to occur, having an impact on reliability. Trains in close proximity to each other can experience negative effects and delays; therefore, the hypothesis was that bunching could be the cause of worse performance from the express services. According to the results, this appeared not to be the case.

While bunching did occur more often for the express lines, the correlations that were found were again not in line with expectations. Scatterplots and a correlation matrix have shown that the relationship between delay and time difference to the previous train is positive. This suggests that the trains riding close to each other did not interfere, or not significantly enough, to create a visible effect on delays.

For future research it would be crucial to collect more data, to find out if the preliminary findings of this relatively small scale study are relevant. This larger study should then also include more hypotheses, and dive into other variables that could be the causes for both delays and bunching. This is because both are not completely understood for the Roslagsbanan system after this research. It could be relevant to look into both the dwell times at stations and crowdedness, and differences occurring here between the service types. This SL API did not offer that data, so these areas could not be investigated.

Besides that, future researchers could explore the physiological aspects of passengers' behavior regarding the Roslagsbanan Express and standard service. Understanding how service type on the Roslagsbanan's 27/27S or 28/28S/28X lines influences passengers' choices may provide SL valuable insights into the advantages and disadvantages of the new express service introduced in 2023. An observational study to understand passengers' choices could explore whether riders were actively choosing the express service, simply taking the next train available, or deciding based on some other factor. The behavioral study could also provide information on passenger satisfaction. Furthermore, it could provide insight into line efficiency when used conjointly with this study above. Understanding factors that lead to greater efficacy and productivity helps inform future public transportation operational decisions.

## References

- Hansson, J., Pettersson-Löfstedt, F., Svensson, H. et al. Effects of rural bus stops on travel time and reliability. *Public Transp* 14, 683–704 (2022). <https://doi.org/10.1007/s12469-021-00281-1>
- Här går Roslagsbanan. (n.d.). SL. Retrieved April 6, 2024, from <https://sl.se/reseplanering/var-trafik/roslagsbanan>
- Menzel, C. Schieneninfrastruktur (2021). RheinMain University of Applied Sciences. 7060 ÖPNV Infrastrukturplanung Mobilitätsdienstleistungen
- Palmqvist, C.-W., Olsson, N., & Hiselius, L. (2017). Some Influencing Factors For Passenger Train Punctuality In Sweden Palmqvist, Carl-William; Olsson, Nils; Hiselius, Lena. Lund University Research Portal. Retrieved May 22, 2024, from [https://lucris.lub.lu.se/ws/portalfiles/portal/34403908/Palmqvist\\_et\\_al\\_2017\\_IJPHM.pdf](https://lucris.lub.lu.se/ws/portalfiles/portal/34403908/Palmqvist_et_al_2017_IJPHM.pdf)
- Region Stockholm. (2024). Det här är Roslagsbanan. Region Stockholm. Retrieved May 8, 2024, from <https://www.regionstockholm.se/kollektivtrafik/projekt-inom-kollektivtrafiken/sparvagnlokalbana/roslagsbanan/>
- Transdev will start operating the Roslagsbanan in Stockholm next year. (n.d.). Transdev. Retrieved April 6, 2024, from <https://www.transdev.com/en/success/transdev-start-roslagsbanan-stockholm-next-year/>
- Victory Stockholm. (2014, April 17). Roslagsbanan. Victory Stockholm. Retrieved May 9, 2024, from <https://victorystockholm.se/en/roslagsbanan/>