Analysis of Train Delays in Switzerland  
Executive Summary:  
This report presents a comprehensive analysis of train delays in the Swiss railway system, focusing on patterns, causes, and impacts across various routes, times, and transportation types. The findings reveal insights into the punctuality of Swiss Federal Railways (SBB) and highlight areas for potential improvement.  
Key Findings:  
  
Delay Patterns:  
  
Long-distance trains experience more significant delays.  
Delays are most frequent during peak hours (morning commute, lunch break, and evening rush).  
Trams have the highest frequency of delays, though these are typically short in duration.  
Departure and arrival delays show a strong correlation, as expected.  
  
  
Delay Statistics:  
  
Average departure delay: 1.58 minutes  
Average arrival delay: 1.16 minutes  
Percentage of arrivals with delays: 8.51%  
Percentage of departures with delays: 11.88%  
  
  
Transportation Type Analysis:  
  
S-Bahn (suburban rail) is the most common transportation type (40,132 instances), followed by regional trains (R) with 10,449 instances.  
Long-distance trains (IC, EC, ICE, TGV, RJX, NJ) have fewer instances but longer delays.  
  
  
High-Risk Stations:  
The following stations have the highest probability of delays:  
  
Düdingen (100%)  
Wynigen (100%)  
Flamatt (100%)  
Ambrì-Piotta (75%)  
Colmegna (61.54%)  
  
  
Geographical Trends:  
  
Central Swiss stations tend to be more punctual.  
Border areas and major junction points experience disproportionately high delays.  
  
  
Overall Performance:  
  
SBB demonstrates relatively good punctuality, with less than 10% of 63,000 individual departures experiencing delays.  
  
  
Station-Specific Data:  
  
Zürich HB: Highest number of train movements (1,909)  
Zürich Oerlikon: Second highest number of train movements (919)  
Olten: Third highest number of train movements (722)  
  
  
Additional Observations:  
  
328 out of 63,047 trains were additional services (Zusatzfahrt).  
652 out of 63,047 trains were cancelled.  
Only 27 out of 63,047 trains were pass-through services without stops.  
  
  
Cluster Analysis:  
Four main clusters were identified:  
  
Cluster 0: 44,940 instances (likely representing on-time or minor delays)  
Cluster 1: 4,409 instances  
Cluster 2: 28 instances (likely representing extreme delays)  
Cluster 3: 2,285 instances  
  
  
Temporal Distribution:  
  
Trains operate from 00:20 to 02:11 the next day.  
Peak arrival times are at 07:54 (75 instances) and 17:54 (73 instances).  
Peak departure times are at 07:49 and 07:34 (74 instances each).  
  
  
  
Conclusions:  
  
The Swiss railway system generally maintains good punctuality, with only a small percentage   
of trains experiencing significant delays. However, there are clear patterns in delay occurrence,   
with long-distance trains and peak hours being more susceptible to delays. Certain stations and   
border areas also show higher risks of delays. The S-Bahn system, while experiencing the most   
delays in absolute numbers, tends to have shorter delay durations compared to long-distance services.  
  
  
Recommendations:  
  
Focus on improving punctuality for long-distance trains, particularly IC, EC, ICE, TGV, RJX, and NJ services.  
Investigate and address issues at high-risk stations such as Düdingen, Wynigen, and Flamatt.  
Enhance capacity and efficiency during peak hours to reduce delays.  
Conduct further analysis on specific stations and routes to identify localized issues and solutions.  
Implement targeted strategies for reducing both departure and arrival delays, especially for S-Bahn services   
which have the highest volume of traffic.  
  
Further Research:  
  
Additional statistical analysis focusing on specific stations, routes, and time periods could provide more   
detailed insights. Examining the causes of delays at problematic stations and for long-distance services   
could inform targeted improvement strategies. A more in-depth analysis of the identified clusters could   
also reveal patterns that may not be immediately apparent from the summary statistics.