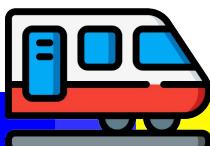


Bahnverspätungen

Welche Faktoren haben einen Einfluss?
(Bahnhof, Temperatur, Bahntyp)

Daten

- Bahnverspätung
- Wetter
- Positionen von Bahnhöfen



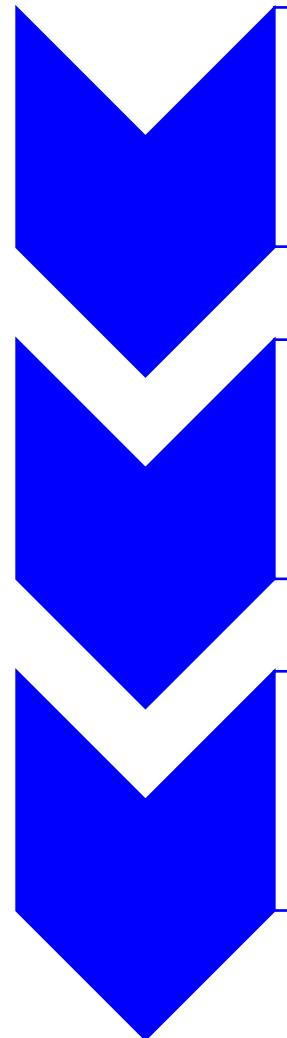
Bahnverspätung

- <https://github.com/piebro/deutsche-bahn-data>
- Seit Juli 2024: größte 100 Bahnhöfe
- Seit November 2025 alle Bahnhöfe



Bahnverspätung

4x am Tag



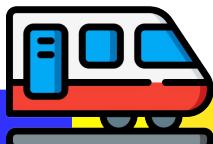
Stationendaten abfragen
mit StaDa



DB Timetables
der Stationen



Upload
nach Huggingface



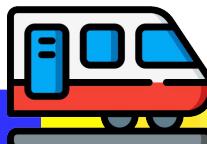
Bahnverspätung

10 rows | 10 rows x 6 cols

timestamp in ns



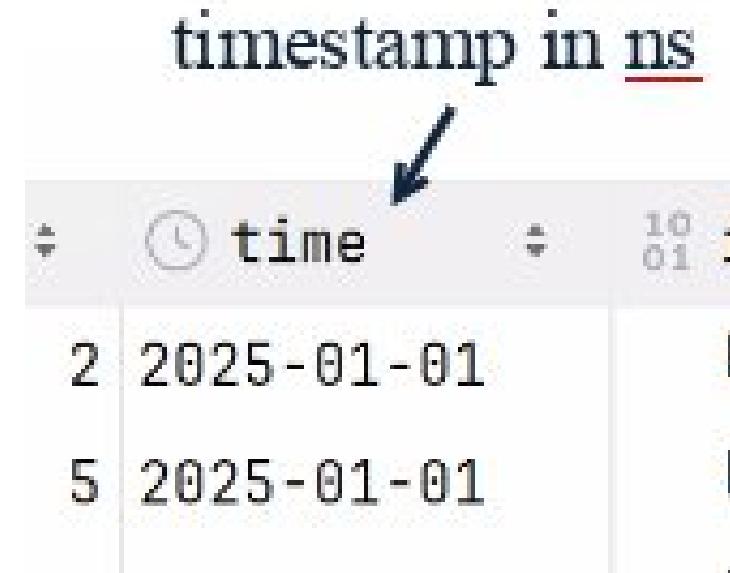
	station_name	train_name	delay_in_min	time	isCanceled	train_type
0	Berlin Ostkreuz	S 5	2	2025-01-01	False	S
1	München Hbf	S 1	5	2025-01-01	False	S
2	München Ost	RB 40	10	2025-01-01	False	RB
3	Hamburg Dammtor	S 5	1	2025-01-01	False	S
4	Berlin-Lichtenberg	RB 12	1	2025-01-01	False	RB
5	München Hbf	S 1	5	2025-01-01	False	S
6	München-Pasing	S 6	4	2025-01-01	False	S
7	Mainz Hbf	S 8	1	2025-01-01	False	S
8	Plochingen	MEX 16	0	2025-01-01	False	MEX
9	Bietigheim-Bissingen	MEX 17a	1	2025-01-01	False	MEX



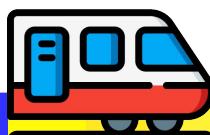
Bahnverspätung

- PySpark akzeptiert keine Nanosekunden
- Casten zu Mikrosekunden

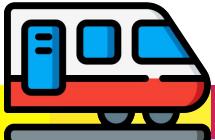
timestamp in ns



	time	
2	2025-01-01	
5	2025-01-01	



Haltestellen



Haltestellen

- Haltestellendaten DELFI
- Spalten u.A.:
 - Name
 - Breitengrad
 - Höhengrad



<https://zhv.wvigmbh.de>

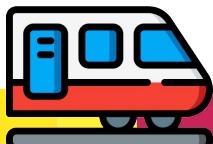


Haltestellen

- JOIN an Stationsname
- Normalisierung der Namen:
 - Erstes Wort
 - Lowercase
 - Nur a-z

10 rows × 6 cols

	station_name
0	Berlin Ostkreuz
1	München Hbf
2	München Ost
3	Hamburg Dammtor
4	Berlin-Lichtenberg
5	München Hbf
6	München-Pasing
7	Mainz Hbf
8	Plochingen
9	Bietigheim-Bissingen



Haltestellen

10 rows | 10 rows × 2 cols

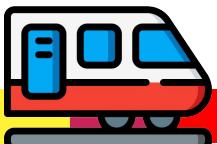
	station_name	norm_name
0	Berlin Ostkreuz	berlin
1	München Hbf	mnchen
2	München Ost	mnchen
3	Hamburg Dammtor	hamburg
4	Berlin-Lichtenberg	berlin
6	München-Pasing	mnchen
7	Mainz Hbf	mainz
8	Plochingen	plochingen
9	Bietigheim-Bissingen	bietigheim
10	Dresden Hbf	dresden
96	Flughafen BER	flughafen



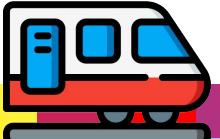
Haltestellen

	norm_na...	station_na...	delay...	is_cance...	1	time	Latitude	Longitude
0	berlin	Berlin-Lichtenb...	19	False	S	2025-03-0...	52.5064125061...	13.3953971862...
1	bamberg	Bamberg	40	False	ICE	2025-03-0...	49.8945693969...	10.8944530487...
2	frankfurt	Frankfurt (Main...	12	False	RB	2025-03-0...	50.1249504089...	8.66991806030...
3	berlin	Berlin Ostkreuz	1	False	S	2025-03-0...	52.5064125061...	13.3953971862...
4	berlin	Berlin Potsdame...	1	False	S	2025-03-0...	52.5064125061...	13.3953971862...
5	berlin	Berlin Ostbahnhof	4	False	S	2025-03-0...	52.5064125061...	13.3953971862...
6	mainz	Mainz Hbf	1	False	S	2025-03-0...	49.9994049072...	8.26036834716...
7	berlin	Berlin Zoologischer G...	0	False	S	2025-03-0...	52.5064125061...	13.3953971862...

Median



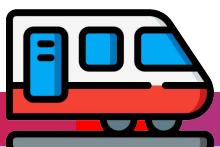
Wetter



[https://opendata.dwd.de/climate_environment/CDC/grid
s_germany/daily/hyras_de/air_temperature_mean/](https://opendata.dwd.de/climate_environment/CDC/grid_s_germany/daily/hyras_de/air_temperature_mean/)

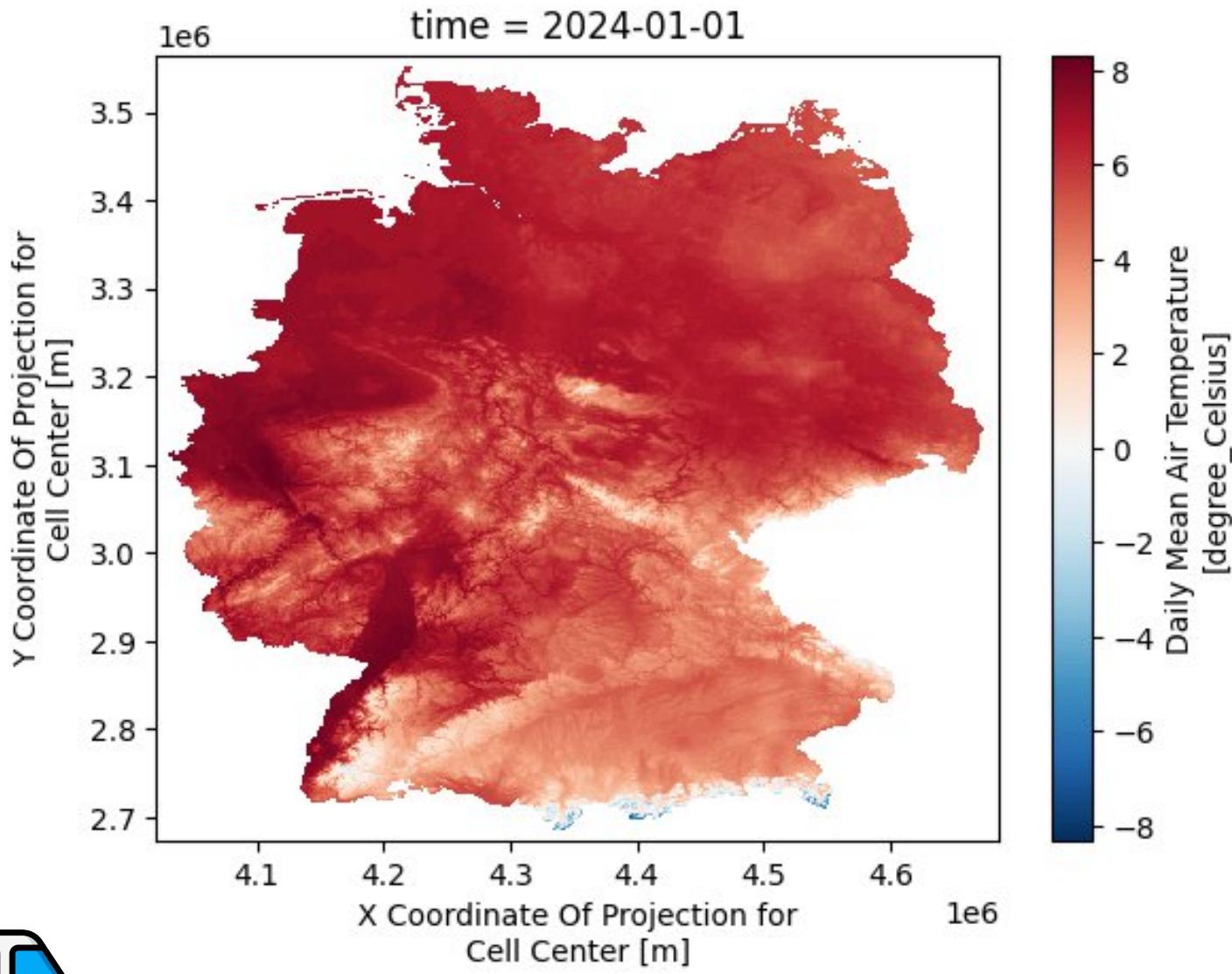
Wetter

- HYRAS-DE-TAS vom Deutschen Wetterdienst
- 1x1 km Raster Temperaturen interpoliert
- Tagesmittel 1951-heute
- NetCDF Datenformat



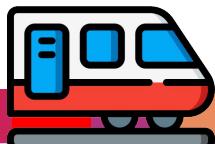
[https://opendata.dwd.de/climate_environment/CDC/grid
s_germany/daily/hyras_de/air_temperature_mean/](https://opendata.dwd.de/climate_environment/CDC/grid_s_germany/daily/hyras_de/air_temperature_mean/)

Wetter



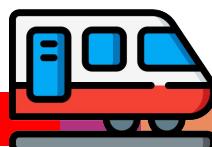
Wetter

- Erste Idee: JOIN über Zeit, Breiten-, Höhengrade
 - Aber: Wetterdaten riesig



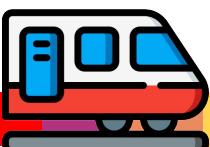
Wetter

- Erste Idee: JOIN über Zeit, Breiten-, Höhengrade
 - Aber: Wetterdaten riesig
- Bessere Performance
 - NetCDF Array über Zeit, Ort
 - Ortkoordinaten in ETRS89 LAEA (3035)
 - Lookup mit Spark mapInPandas



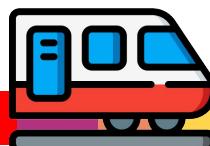
- 60.701.100 Zeilen

	norm_name	station_name	delay	is_cancelled	1	time	Latitude	Longitude	temperature
0	berlin	Berlin-Lichtenb...	19	False	S	2025-03-0...	52.5064125061...	13.3953971862...	3.7000000476...
1	bamberg	Bamberg	40	False	ICE	2025-03-0...	49.8945693969...	10.8944530487...	1.5
2	frankfurt	Frankfurt (Main...	12	False	RB	2025-03-0...	50.1249504089...	8.66991806030...	4.7000002861...
3	berlin	Berlin Ostkreuz	1	False	S	2025-03-0...	52.5064125061...	13.3953971862...	3.7000000476...
4	berlin	Berlin Potsdame...	1	False	S	2025-03-0...	52.5064125061...	13.3953971862...	3.7000000476...
5	berlin	Berlin Ostbahnhof	4	False	S	2025-03-0...	52.5064125061...	13.3953971862...	3.7000000476...
6	mainz	Mainz Hbf	1	False	S	2025-03-0...	49.9994049072...	8.26036834716...	4.7000002861...
7	oberhausen	Oberhausen Hbf	0	False	Bus	2025-03-0...	51.4976882934...	6.86074781417...	3.2000000476...
8	hamburg	Hamburg Hbf	0	False	S	2025-03-0...	53.5549545288...	10.0237770080...	3.7000000476...
9	berlin	Berlin-Spandau	3	False	S	2025-03-0...	52.5064125061...	13.3953971862...	3.7000000476...
10	frankfurt	Frankfurt (Main...	16	False	RE	2025-03-0...	50.1249504089...	8.66991806030...	4.7000002861...
11	freiburg	Freiburg (Breis...	0	False	S	2025-03-0...	47.9968528747...	7.84072017669...	3.1000001430...
12	hanau	Hanau Hbf	1	False	S	2025-03-0...	50.1272354125...	8.92194223403...	3.6000001430...
13	bietigheim	Bietigheim-Biss...	1	False	MEX	2025-03-0...	48.9542102813...	9.13451433181...	3.0

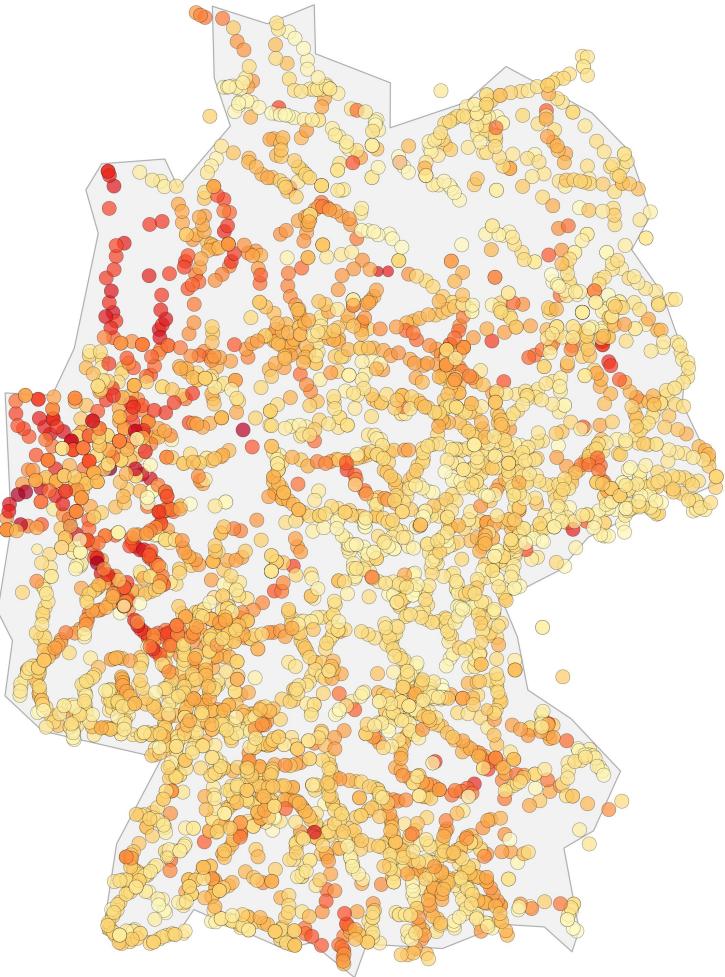


Data Analysis

- Aggregieren in Spark
- Grafiken lokal mit Pandas, Geopandas, Matplotlib

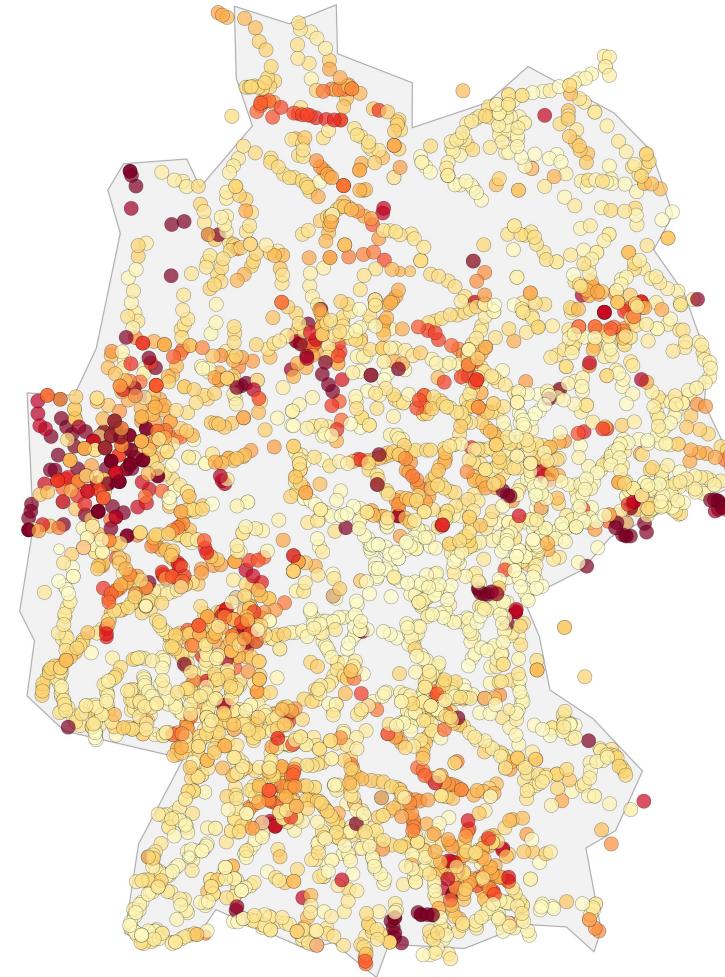


Verspätungen Bahnhöfe



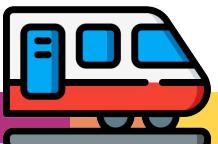
Avg Delay (min)

0	2	4	6	8	10
---	---	---	---	---	----

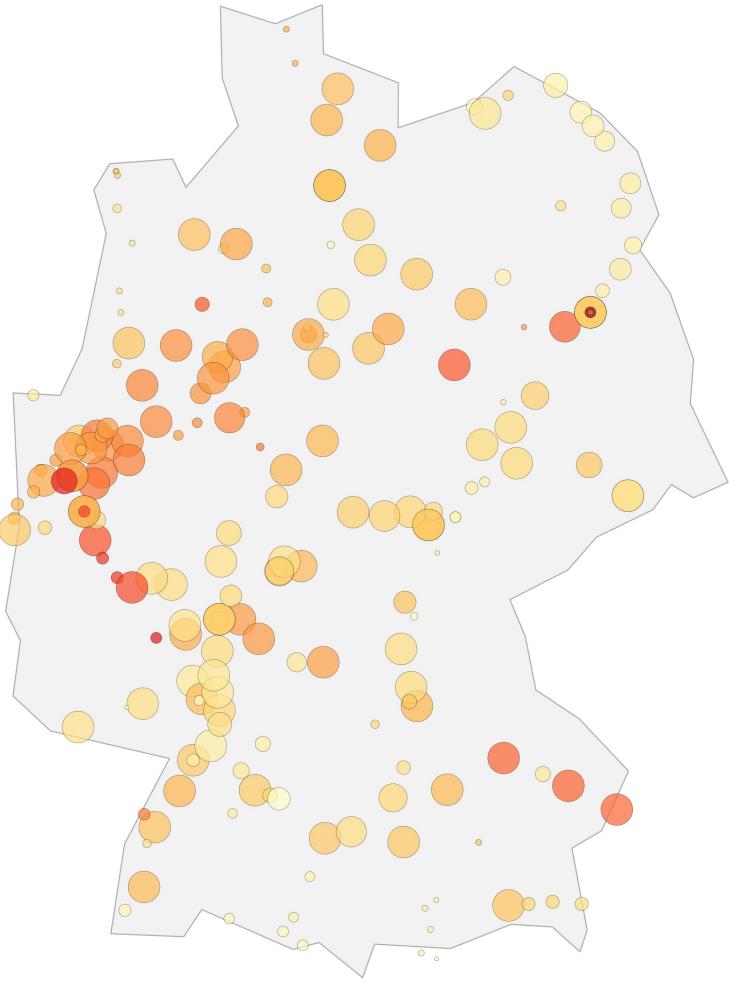


cancelation_rate

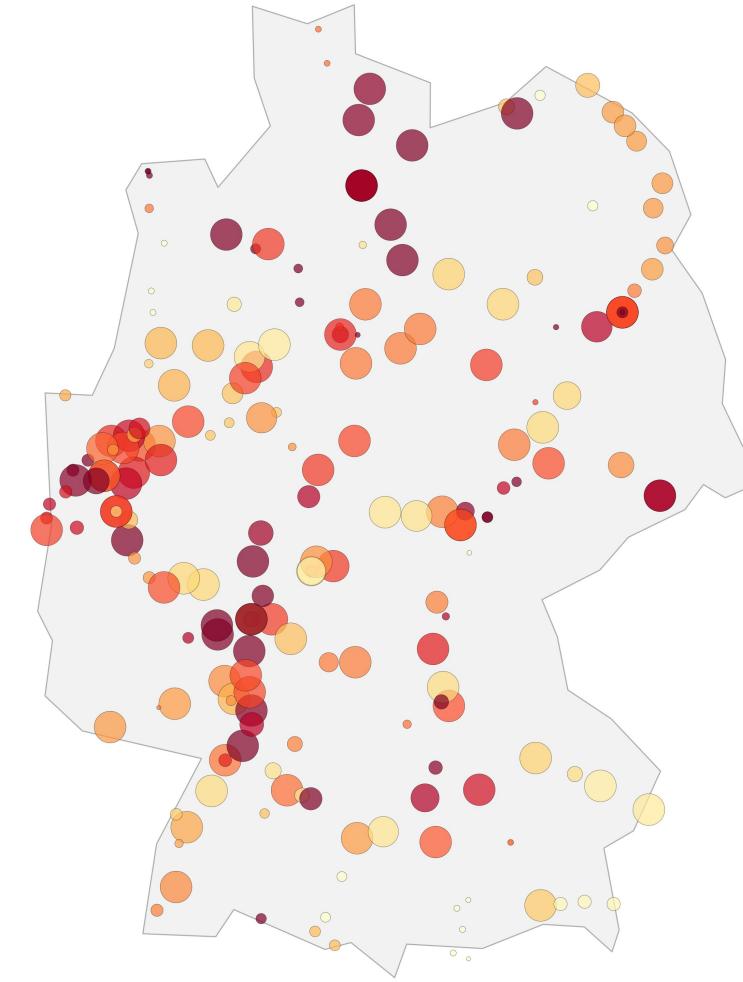
0.00	0.02	0.04	0.06	0.08	0.10
------	------	------	------	------	------



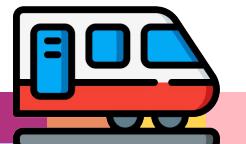
Verspätungen Bahnhöfe ICE



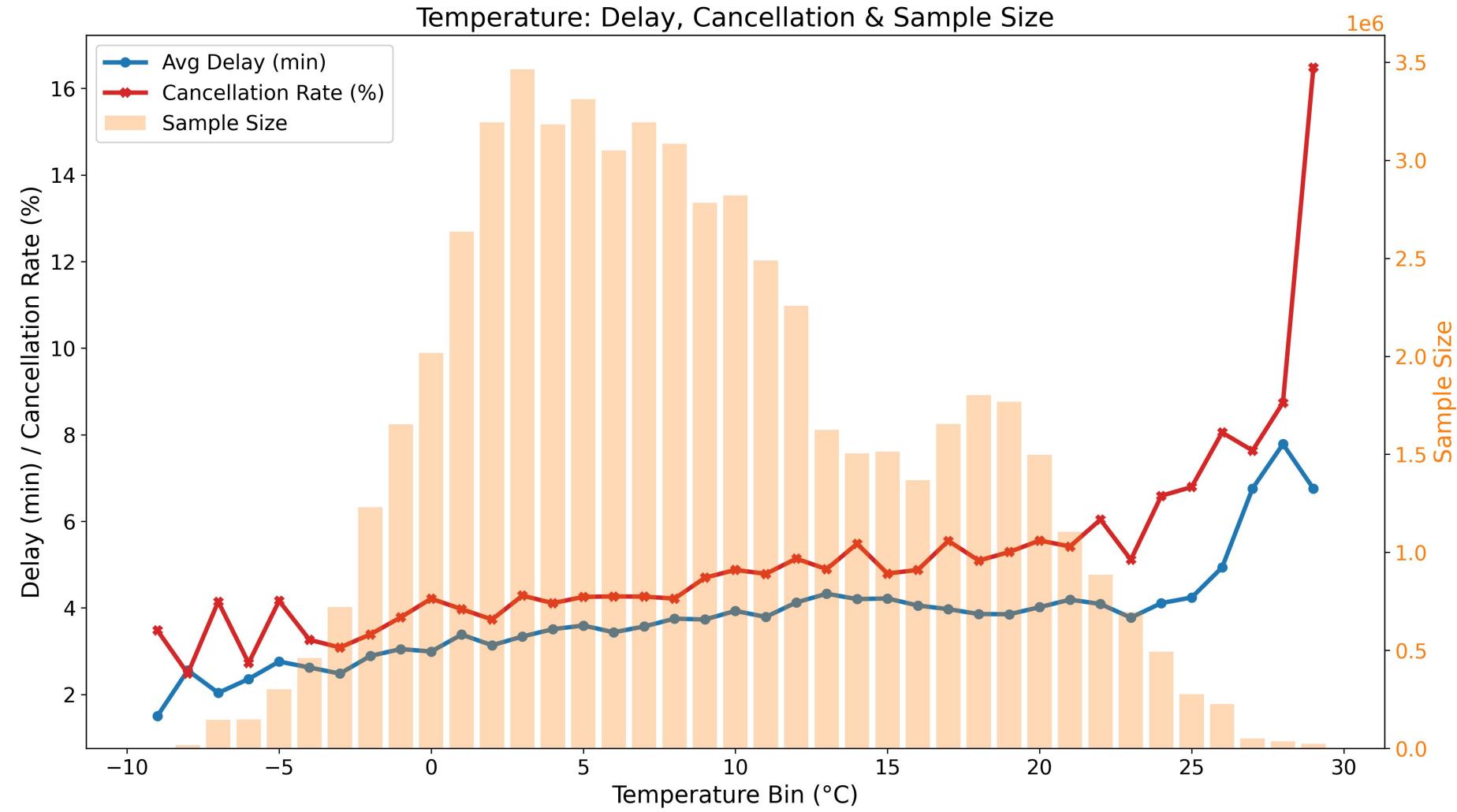
Avg Delay (min)



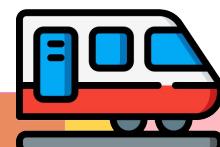
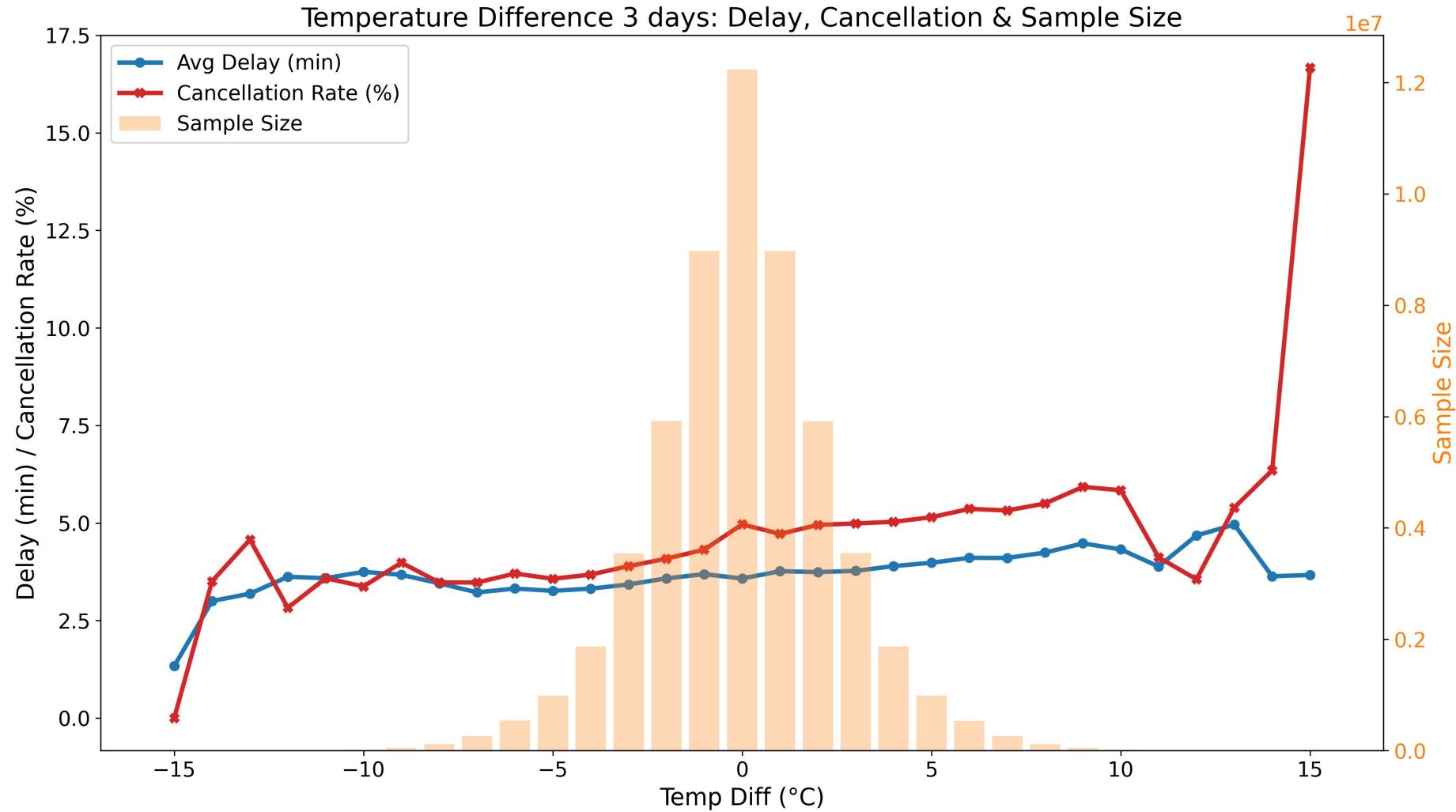
Cancellation Rate



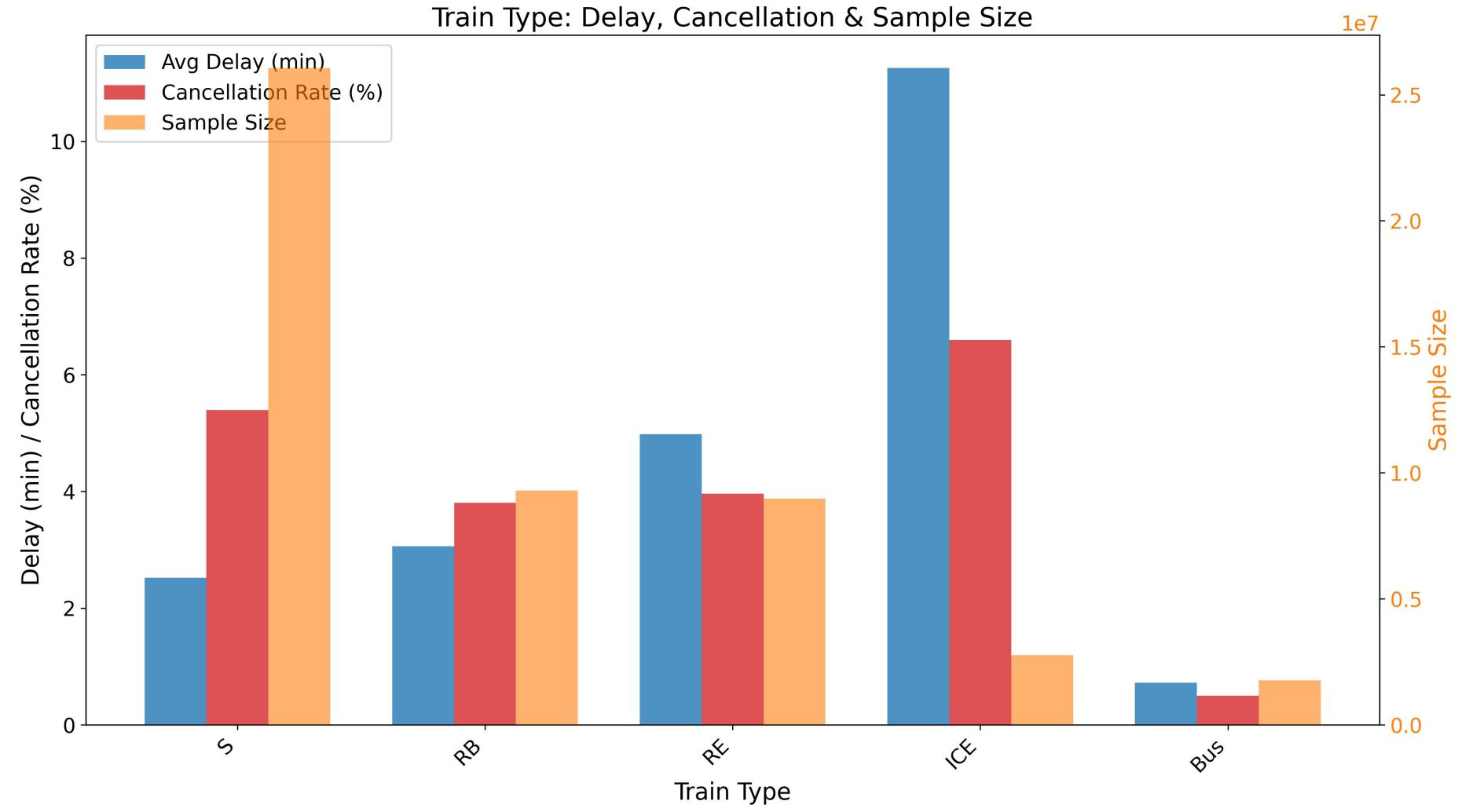
Verspätungen Temperatur



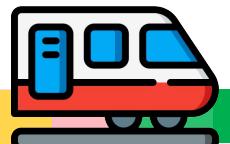
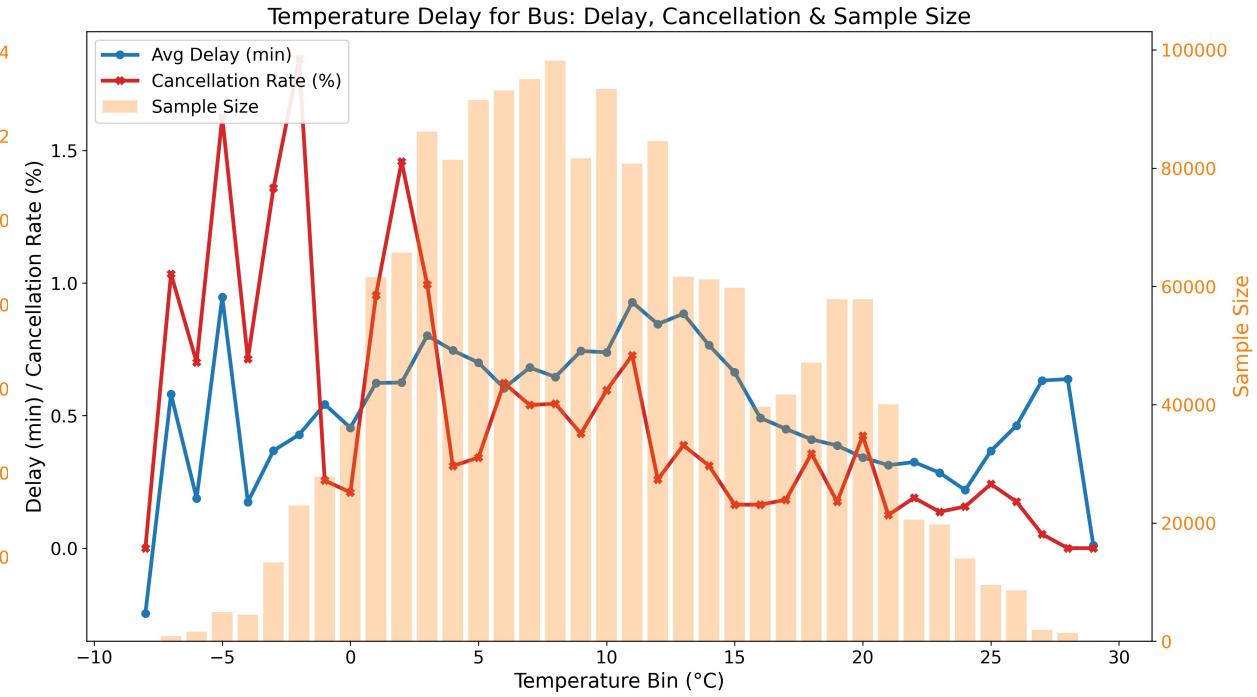
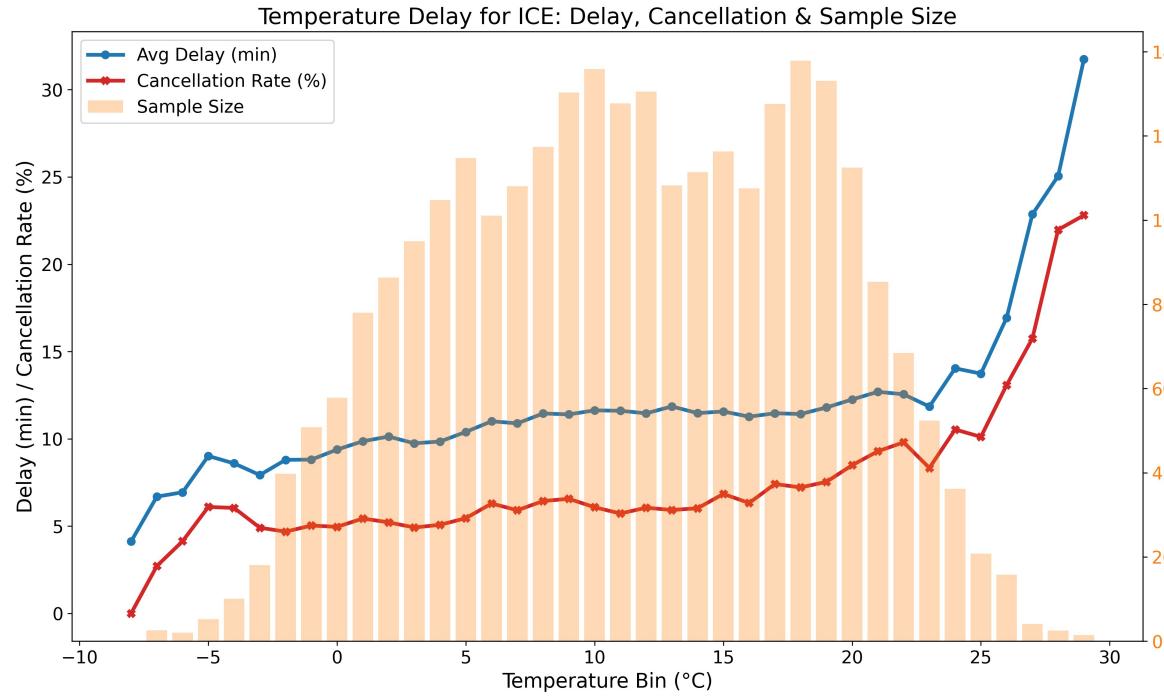
Verspätungen Δ Temperatur



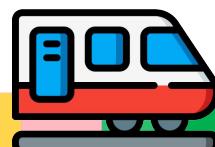
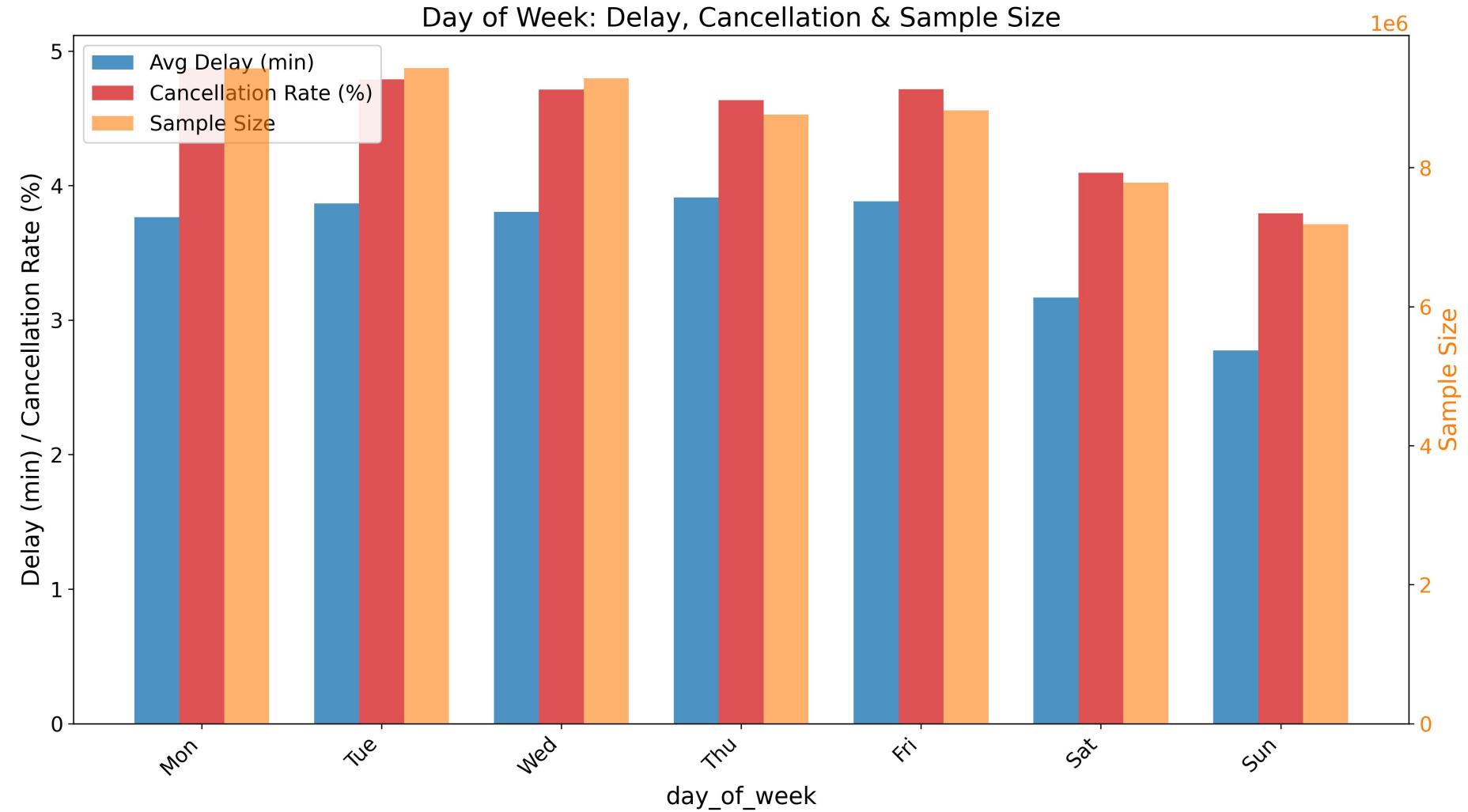
Verspätungen Typen



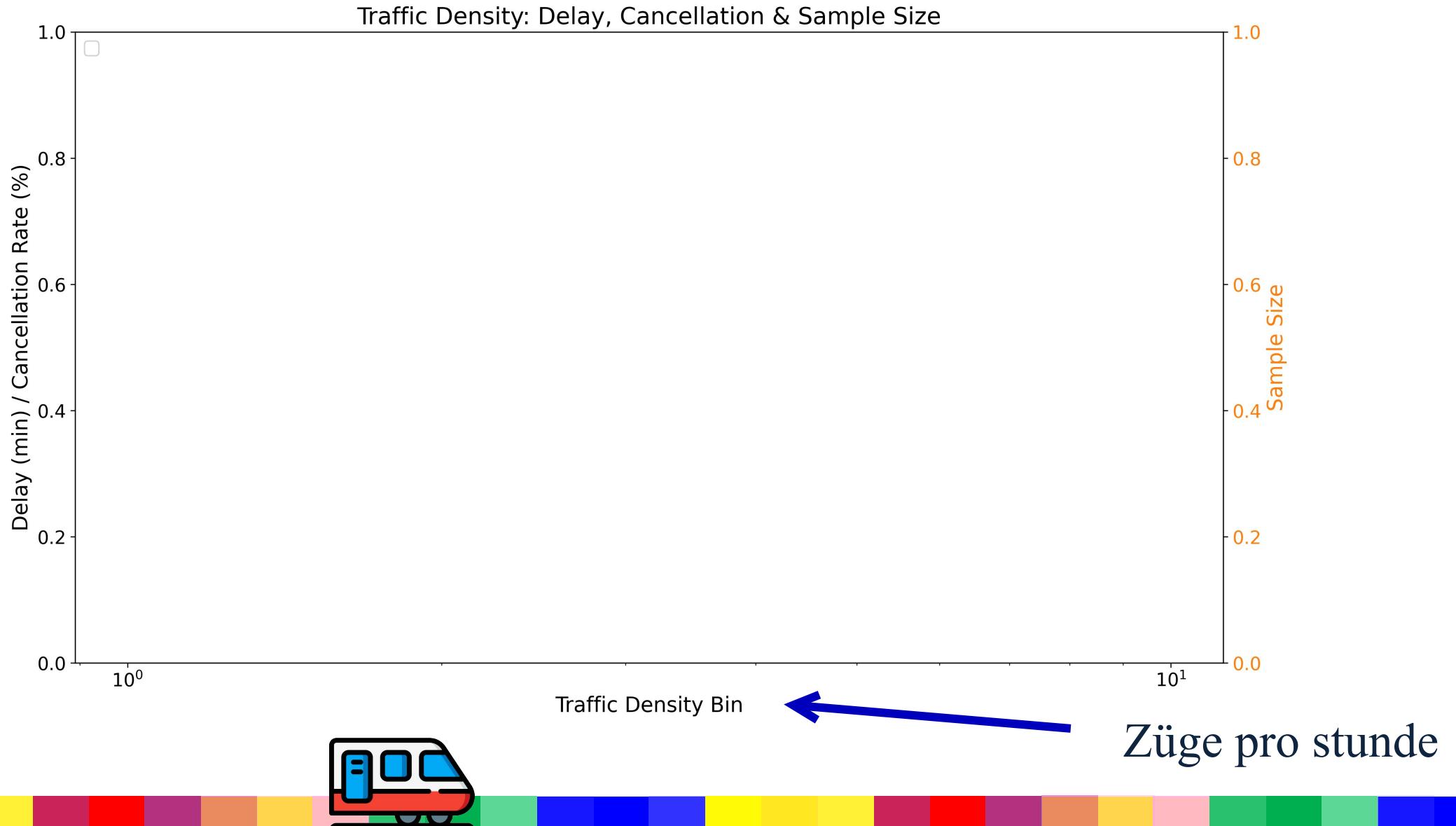
Verspätungen Typen Temperatur



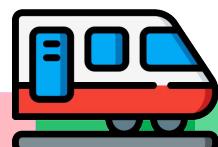
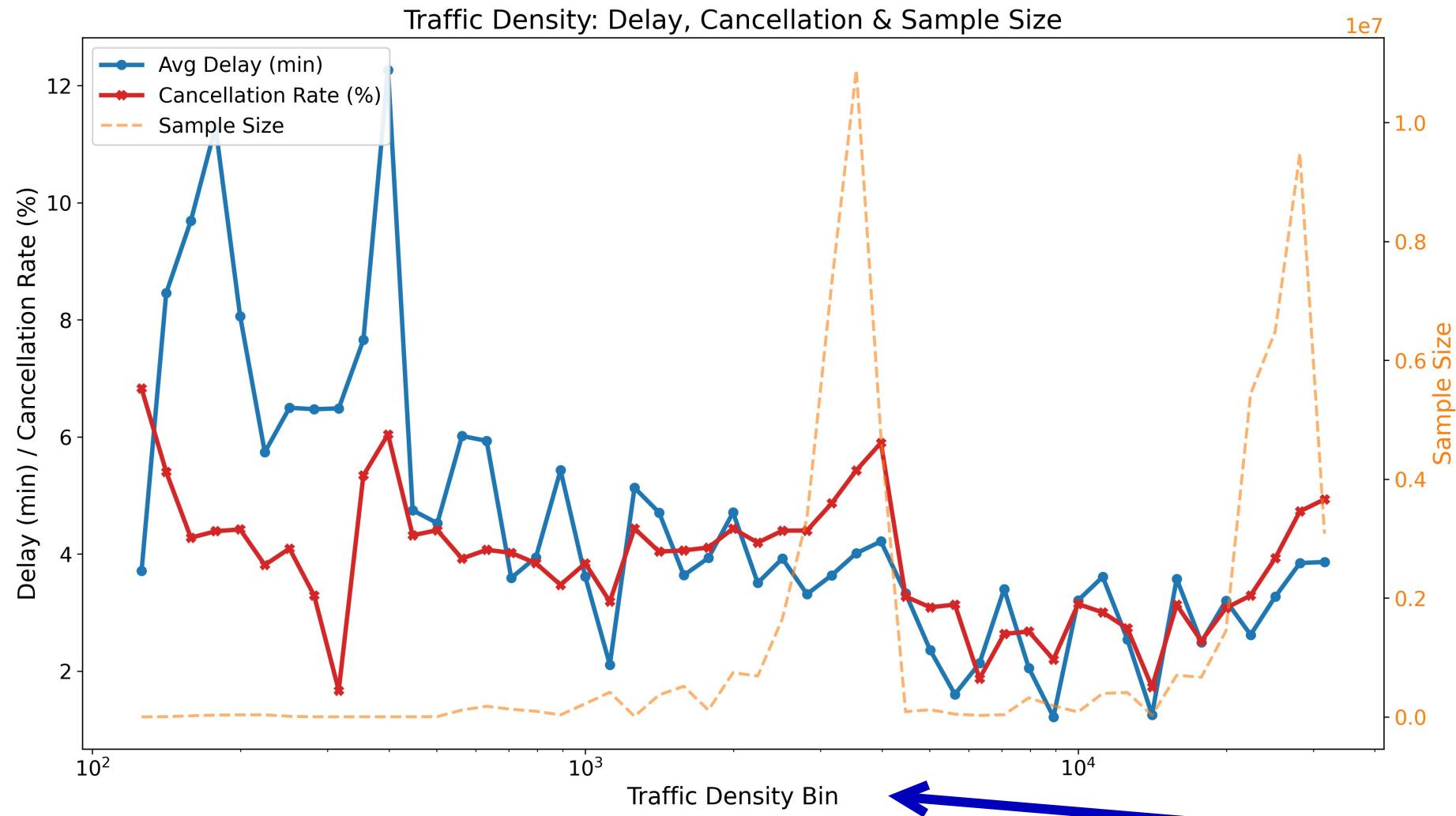
Verspätungen Wochentage



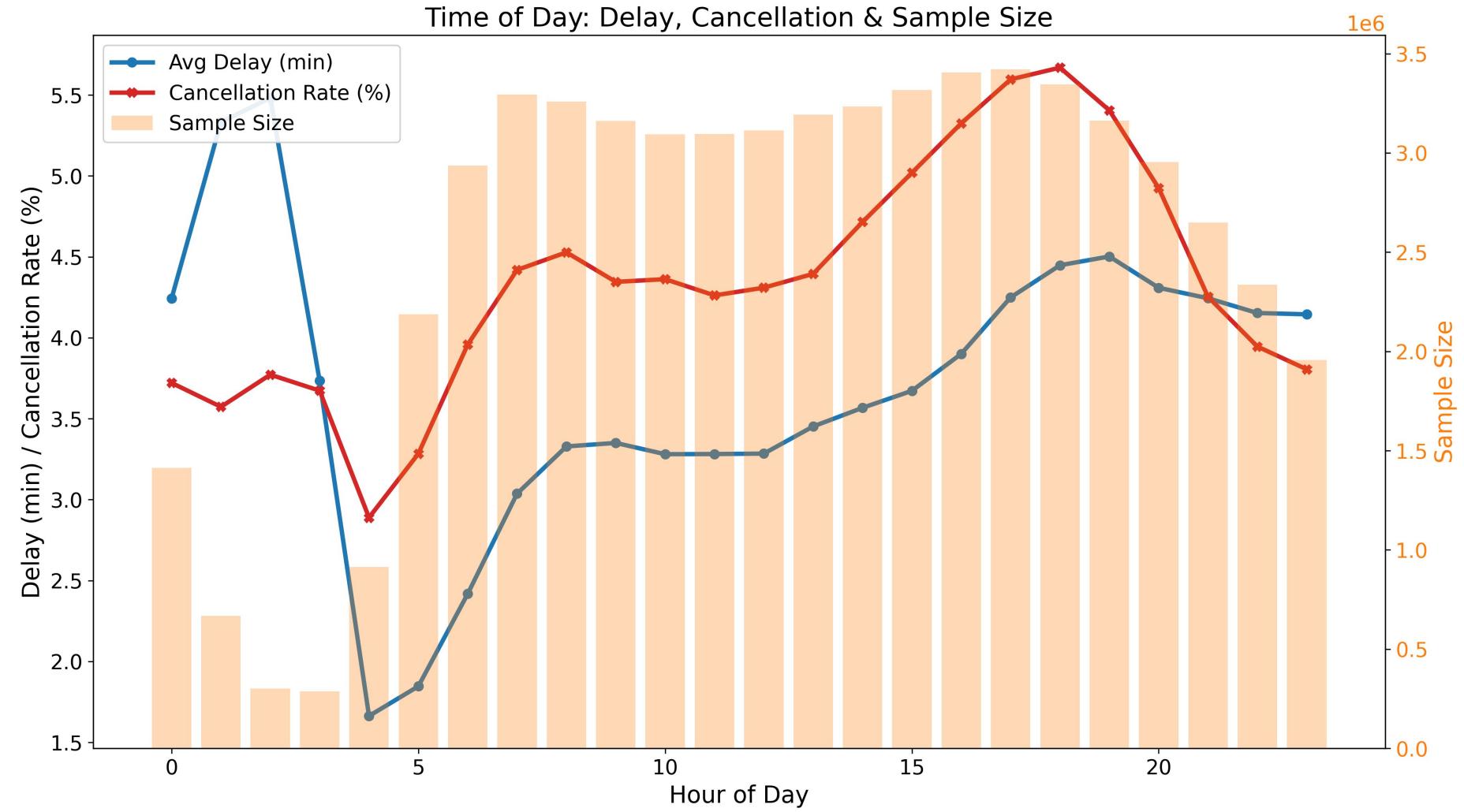
Verspätungen Verkehrsaufkommen



Verspätungen Verkehrsaufkommen

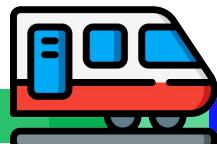


Verspätungen Uhrzeit



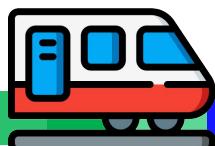
Takeaway

- Wo sollte ich reisen?
- Wie sollte ich reisen?
- Wann sollte ich reisen?



Takeaway

- Wo sollte ich reisen?
 - Nicht ins Ruhrgebiet
- Wie sollte ich reisen?
 - Bus
- Wann sollte ich reisen?
 - Bei niedrigen Temperaturen!
 - Am Wochenende
 - Früh Morgens



Anhang

Wetterdaten Herkunft

- “Die Temperatur (TAS, TASMIN, TASMAX) und die relative Luftfeuchte (HURS) beruhen auf einer Kombination nichtlinearer Temperaturprofile mit nichteuklidischer Residueninterpolation (Krähenmann et al., 2018)”

Stationen Daten

- 285.400 Haltestellen
 - 128.800 Bereiche
 - 525.400 Masten
-
- Exportdatum:
 - 05.01.2026

Flughäfen

seqNo	Name	Latitude	Longitude
501712	Flughafen Rostock-Laage	53,913546	12,292043
491757	Flughafen Trossenhangen	53,604658	13,295417
744820	Flughafendirektion	52,46171	9,698823
289804	Flughafen BER	52,364606	13,509874
842201	Flughafen Leipzig/Halle, Zen...	51,423315	12,223639
577464	Flughafen Köln/Bonn Bf	50,878903	7,119301
785747	Flughafen Hahn Terminal,Hahn...	49,94527	7,268314
825585	Flughafen, Saarbrücken Enshe...	49,220416	7,113278
8032	Flughafen/Messe	48,690749	9,193413
131975	Flughafen, Terminalbereich A...	48,356488	11,784666
131990	Flughafen, Terminal 2	48,356011	11,789921
131698	Flughafen, P41/P43/Novotel	48,355157	11,758327
132461	Flughafen, Briefzentrum	48,354906	11,745607
131669	Flughafen, P44/Lab-Campus	48,354709	11,752515
133282	Flughafen, Terminalstr. Mitte	48,353826	11,787091
133269	Flughafen München	48,353778	11,786238
133261	Flughafen, MAC/Terminal 1	48,353247	11,786148
132439	Flughafen, Besucherpark	48,352733	11,764076
132009	Flughafen, Terminalbereich D...	48,350823	11,785654