Germany’s Railway Paradox

“Efficiency Measured in 6 Minutes”

By- Enosh Paul Niju GH1026595

Introduction

Germany, Europe’s own land of opportunities. A land where efficiency and punctuality are taken to the most extreme gravity, yet one of the biggest challenges to their national value lies in its’s own heartbeat, the railway system.

The colossal that operates these tracks is Deutsche Bahn or simply “DB”. For many residents of the land of Germania, DB is both a necessity and a frustration more like a love hate relationship built around its role in the daily life of the people of Germany.

This single company is responsible for keeping Germany moving bringing people to work, reuniting them with loves ones, and connecting the nation. Ironically, in a Country that prides itself in efficiency and punctuality, one of the biggest public concerns is DB itself, more specifically, the daily caused by its trains. These delays ripple through the workforce and daily routines, disrobing the very rhythm of the country. To outsides, the issue may seem small, but for the millions who depend on DB daily, it’s indeed and unavoidable source of frustration.

“It’s the same daily misery”- (Patridge and Cologne, 2023)

Yes, that is indeed a remark that was talked about in the article that was published by “The Guardian”. In this very article most of the very daily commuters shares their frustrations with the German railway service and talk about how unreliable the system is altogether.

“Just 56,1 percent of DB trains were punctual in July” - (Iamexpact in Germany, 2025)

This was the exact headline of another article that was published by Olivis Logan talking about the “Chronic delays at Deutsche Bahn”. The major question “Will the situation ever improve?”

"Deutchse Bahn is on track for more delays and cancellations"- (Kinkartz, 2025)

Another headline article which was published in Germany's biggest state-funded international public broadcaster Deutsche Welle (DW). It seems that entire country as well is quite frustrated by its own dire problem.

One of the factors that was discussed in this very article was the fact that in Germany a train is "considered punctual if it is delayed by less than six minutes".

So yes that phrase is the reason for this very project to come to life and is by asking the question; how many trains are exactly punctual in other words how many trains come after the very "six minutes", which would be done by measuring the delays caused in different train stations.

The dataset used for the project is from Kaggle, which the author sourced from "DB Open Data Portal". :-

https://www.kaggle.com/datasets/nokkyu/deutsche-bahn-db-delays

GitHub link to the project details:-

https://github.com/ENORIC/DB_DELAY_PROJECT/blob/main/README.md

Data set link from google drive as the dataset is quite big to upload directly into git hub :-

https://drive.google.com/drive/folders/15mGyWyDkivdMMjkLCiz_2qd_OQlHtyAq?usp=sharing

The DB dataset records the operation that happens in the span of a complete week that are the even days of the week furthermore the particular dataset has five train station categories or "Bahnofskategorie" which are classified based on the level of service, passenger volume, infrastructure and various other reasons.

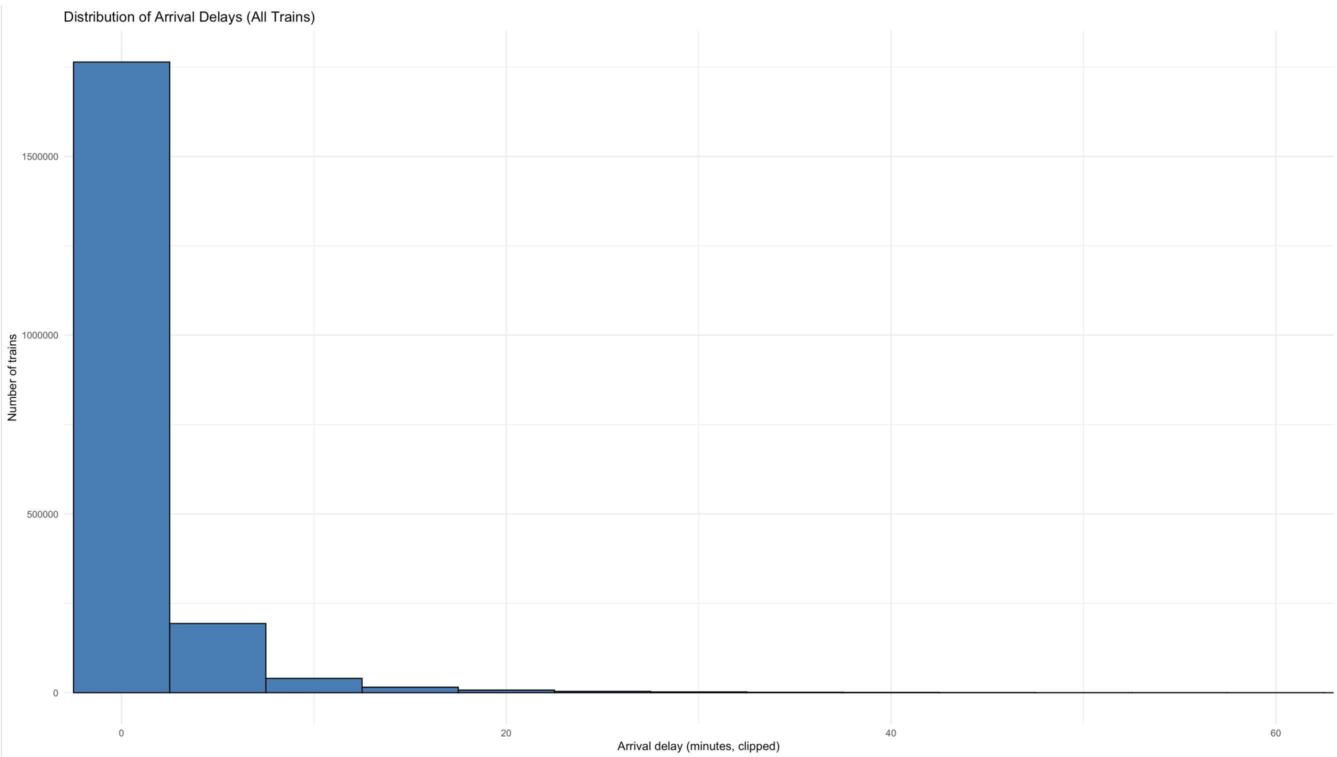
Category	Role Description of the station
1	Considered to be the major traffic hubs. "Hauptbahnhof" usually connection hub of the big cities in Germany.
2	These are also fairly larger stations often used for long-distance traffic.
3	Regional hubs are the intermediate stations with few shops or so around it.
4	Local or quite regional stations, most of the trains that stops over here are the regional trains
5	Smaller, suburban or rural stations. Lower passenger counts and usually quite deserted compared to the other stations.

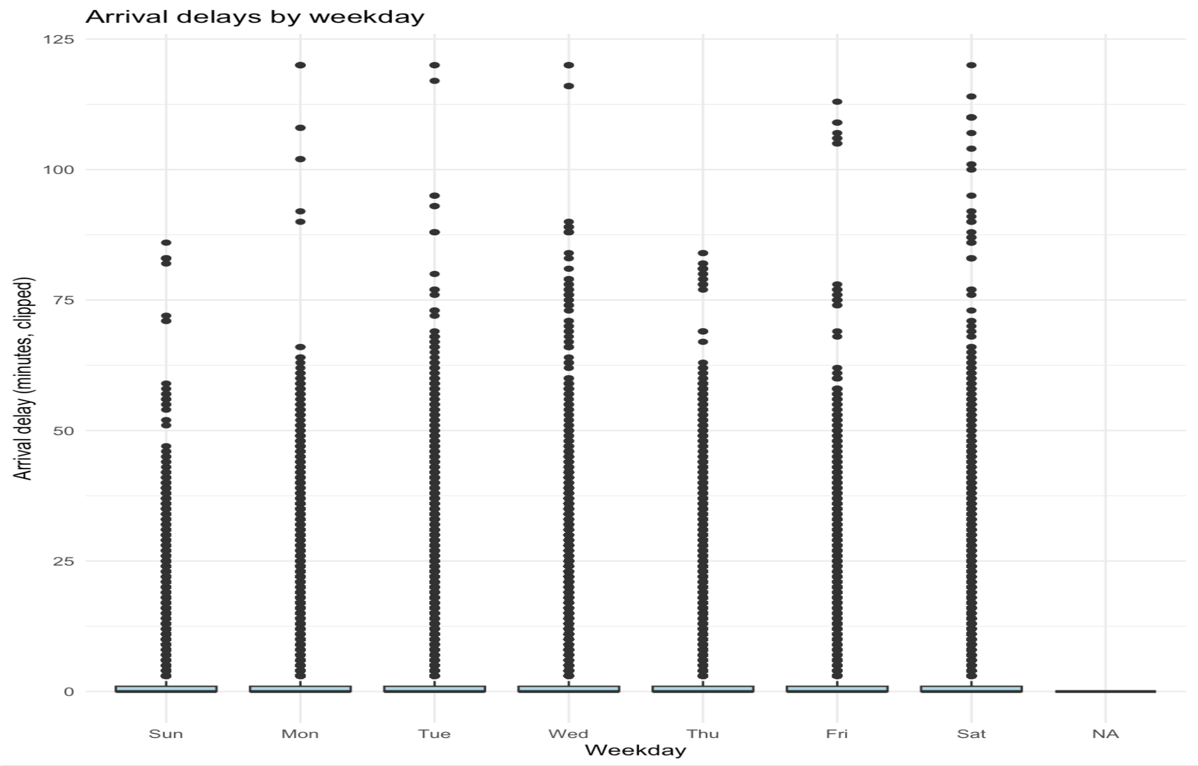
Business problem

Passenger satisfaction and operational efficiency are compromised by Deutsche Bahn's frequent delays. The study investigates the location and timing of delays as well as whether station category, weekdays and peak hours have significant impact on punctuality. A major question being what kind of factors do strongly influence train delays at Deutsche Bahn? Which considering the features like station size, time of day, weekdays as well as arrival and departures.

Before addressing about the main business problem, I was tasked to clean and prepare the data along with the data exploration. During the Exploratory Data Analysis (EDA) I was able to position and interpret the data for "better answering" the main hypothesis questions we would be tackling later. Starting of the initial EDA questions which are;

- What do train delays look like overall?
- Do some weekdays have systematically worse delays?
- How are delays distributed across weekdays not for just the medians?

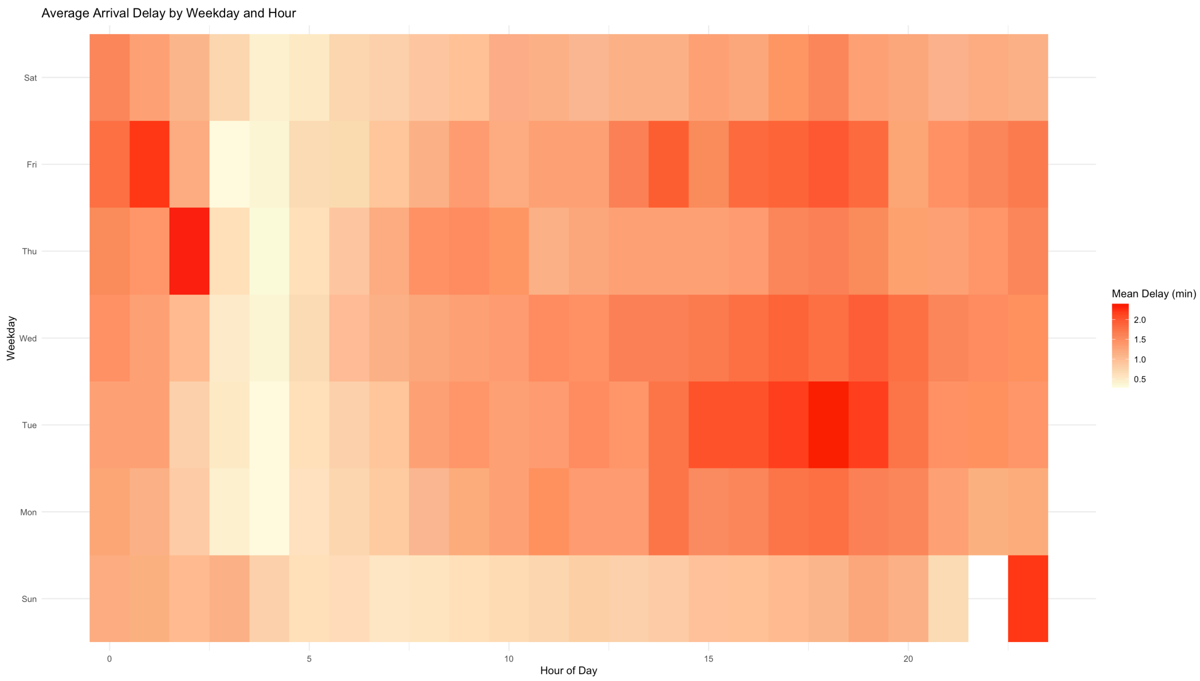




The Histogram as you see shows the overall distribution of the arrival delays, as depicted in the graph most train tend to arrive on time or at least with very minimum delays that is between 0-5 minutes. Also, upon reading the graph there is a sharp spike at 0 minutes which indicates the larger punctuality of the trains and as the delays increases that is between 10-20 minutes there is a significant drop of number of trains, only few trains occasionality experience much larger delays that is up to 60+ minutes. DB's main problem is frequent small delays and not the very rare extreme ones, as DB officially counts a train as "delayed" if it's more than 6 minutes late.

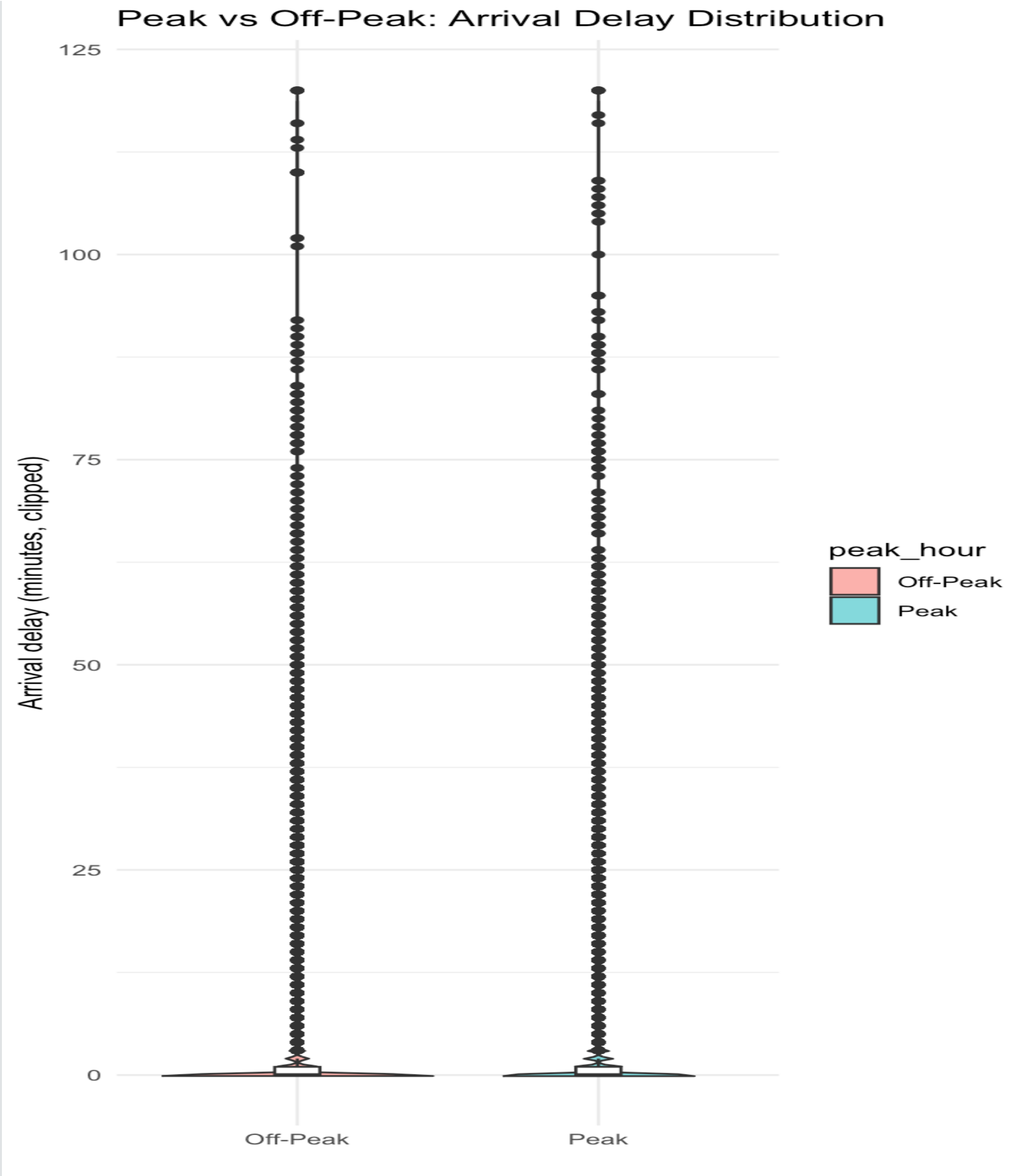
Boxplot of arrival delays are kept by weekday the boxes are all low and narrow. Also indicating that most delays are small as well as showing the same across the days, and the outliers which are the dots above the graph are the train that are delayed 60-120 minutes across all the weekdays. A point to note is that the median line inside each box is slightly higher on weekdays that is between Monday-Friday that on the weekends which are Saturdays and Sundays seems better when it comes to delays, directly indicating that weekday trains are more delayed supposably because its considered to be the "Working weekdays".

- "When during the week do delays tend to be worst?



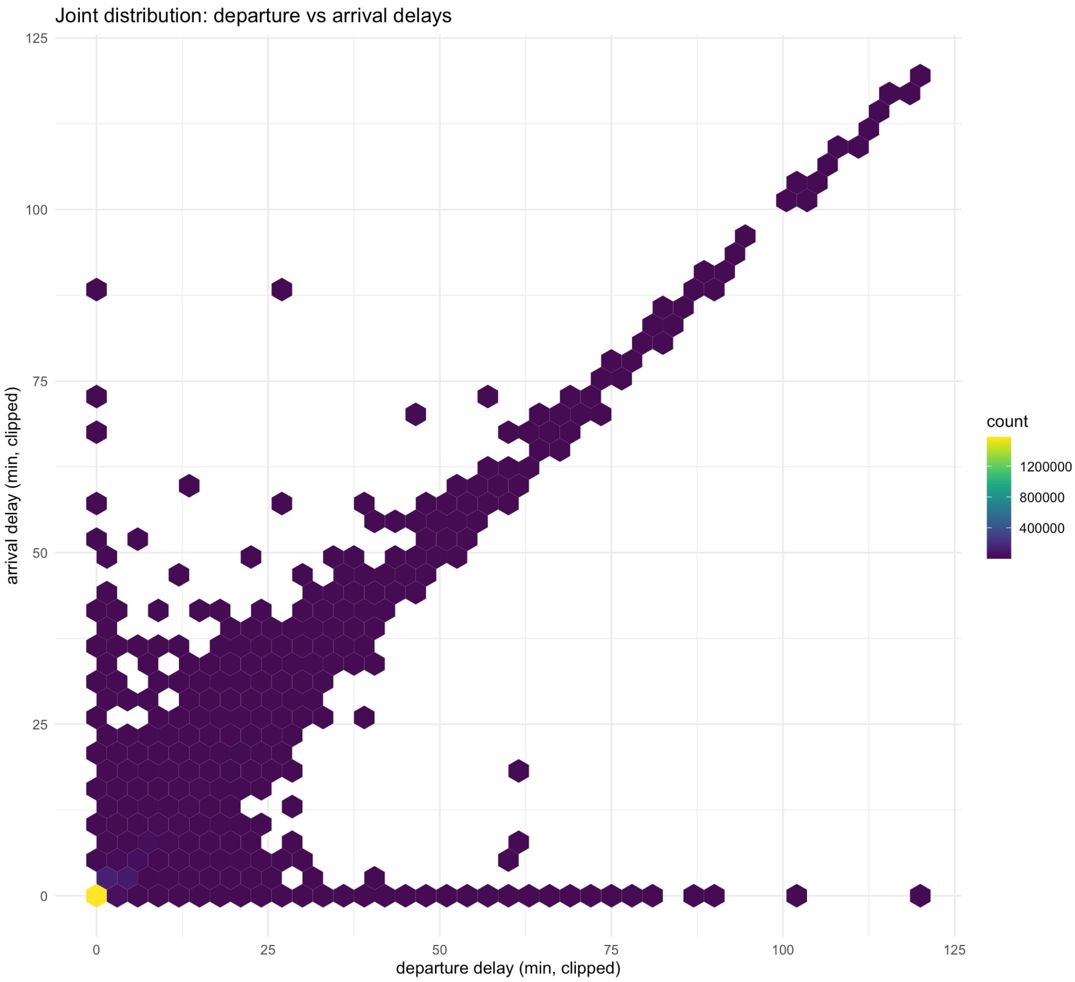
From the indicate we learn that darker red spots around 7~9 and 16~19 show a significate mean delay, also considering the fact that Mid or late week is the worst that is Tuesday-Friday showing the darkest peak hour blocks specifically around the evening. Meanwhile early mornings are light which is between 03-06 are pale across all the days considerably the shorter delays. Also, weekend is calmer especially Sunday and much of Saturday are lighter overall, although there is a late-night spike of delays around 23;00 which are likely due to the sluggish operation.

- Do trains experience more delays during peak commute hours let’s say between 7–9am, 4–7pm compared to off-peak times?



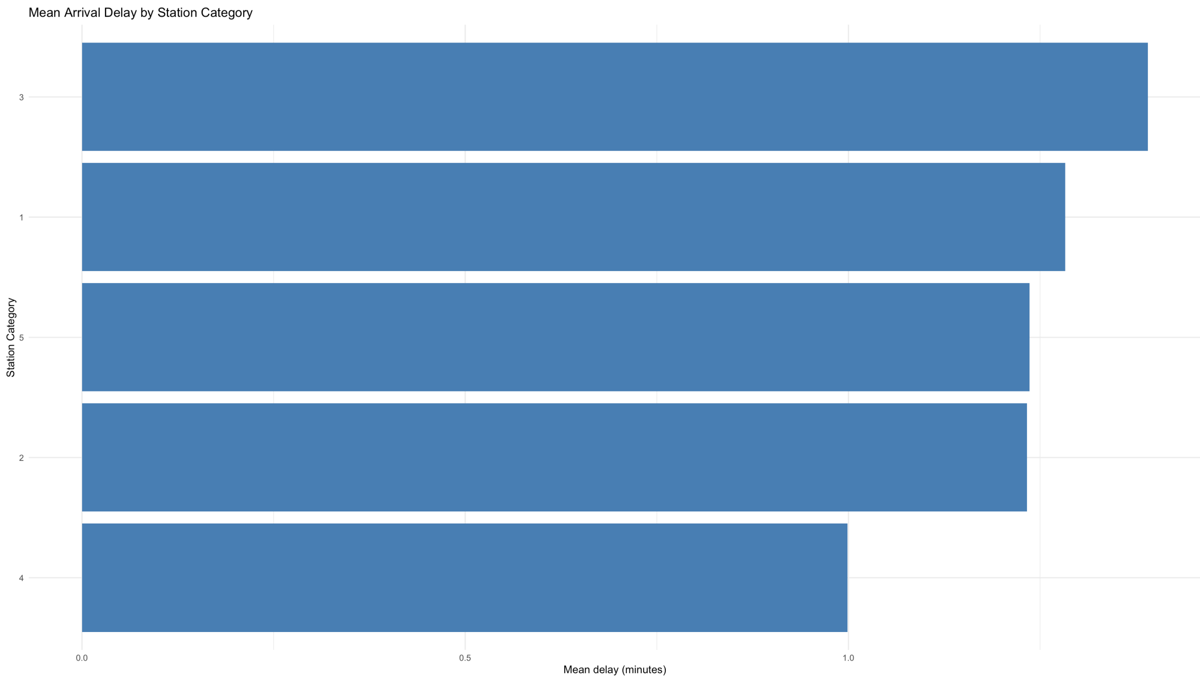
So according to the graph both the peak and off-peak hours have most trains clustered near to 0-5 minutes delay. As the violin for peak hours is slightly taller which showing heavier tails which indicates more extreme delays. So the final explanation would be trains that run during peak commute hours 7-9 am and 4-7pm show slightly higher and more variable delays when compared to off-peak which suggest congestion play a role for the delays.

- Do departure delays lead to arrival delays or vice versa ?



As one can easily interpret from the frap that there is a clear diagonal pattern when departure delays increase along with the arrival delays increase as well. Most of the trains are to be seen clustering near the origin that is between the 0-5 minutes mark but as we move respectively to the right and left you can see the dots of both departure and arrival delays moving growing too. Hence the assumption can be confirmed that departure and arrival delays are strongly linked, which means late departures almost always translate into late arrivals.

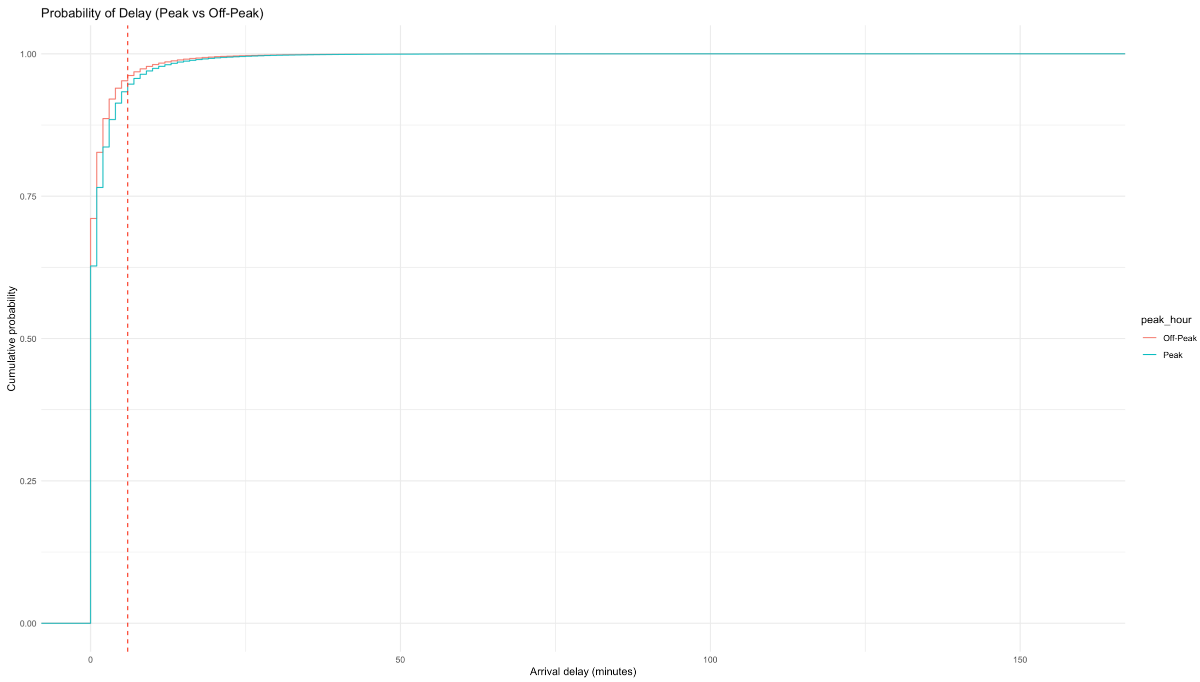
- Do bigger station categories have higher average arrival delays?



	category	arrival_delay_m
4	4	0.9986725
2	2	1.2329527
5	5	1.2363898
1	1	1.2826783
3	3	1.3907002

Category 4 stations which are the local stations have the lowest mean delay approx. 1.0 min. Category 3 stations medium or regional stations have the highest mean delay with an approx. of 1.39 min. The bigger stations such are the categories 1 and 2 still have higher delays compared to the small local ones. Hence the patter is “not always” linear that is bigger stations don’t always mean bigger delays, but higher category hubs do tend to experience most of the delays than the smaller or local hubs.

- What is the probability of crossing DB’s 6-minute official delay threshold, and does this differ between peak and off-peak times?



From the ECDF we can understand that most trains that is well about 90% of the trains are on time withing DB’s 6-minute rule. Peak hour trains are slightly words a higher share of them cross the 6-minute line compared to off peak.

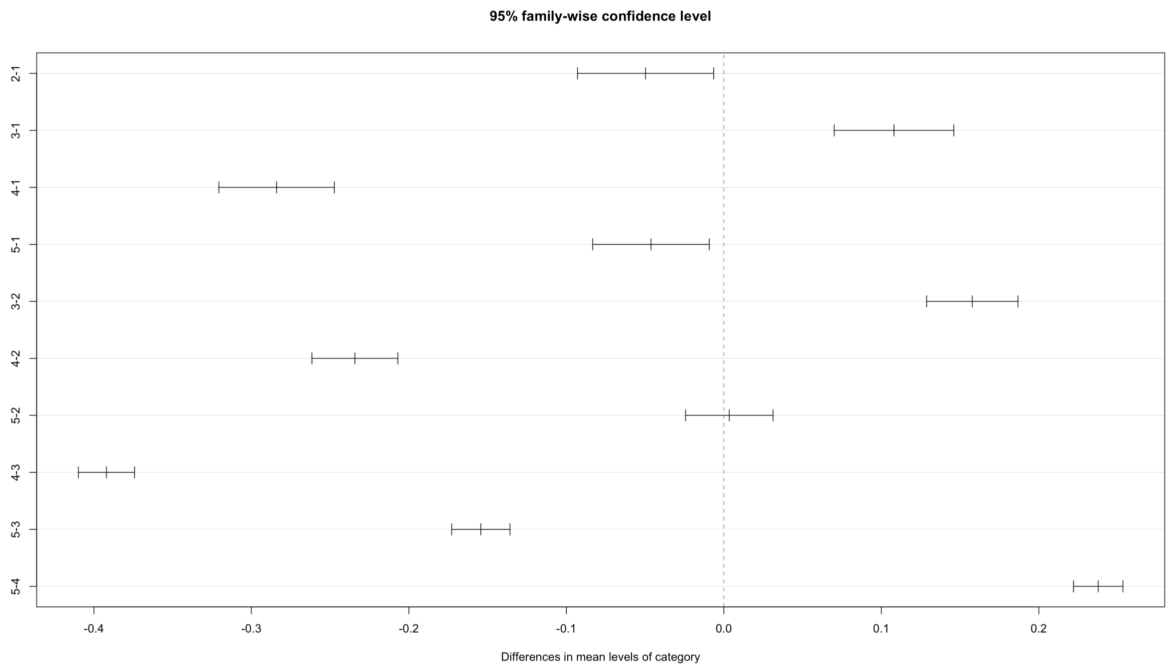
Hypothesis Testing

The business problem will be further divided into five sub-questions which are hypothesis based. Which are each categorized by null hypothesis and alternate hypothesis, which are:-

1. **Hypothesis 1: Station Size (ANOVA) Arrival delay by station category.**

   H_0 (null): Mean arrival delays are equal across all station categories.

   H_1 (alt): At least one station category has a different mean arrival delay.




```
                 Df   Sum Sq Mean Sq F value Pr(>F)    
category          4    47065   11766    1019 <2e-16 ***
Residuals   2029889 23429830      12                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = arrival_delay_m ~ category, data = db_data)

$category
             diff         lwr          upr     p adj
2-1 -0.049725652 -0.09294698 -0.006504327 0.0146672
3-1  0.108021933  0.07007290  0.145970965 0.0000000
4-1 -0.284005829 -0.32066556 -0.247346100 0.0000000
5-1 -0.046288458 -0.08328463 -0.009292288 0.0057992
3-2  0.157747585  0.12873508  0.186760091 0.0000000
4-2 -0.234280177 -0.26158461 -0.206975742 0.0000000
5-2  0.003437194 -0.02431732  0.031191709 0.9972019
4-3 -0.392027762 -0.40984849 -0.374207036 0.0000000
5-3 -0.154310390 -0.17281334 -0.135807440 0.0000000
5-4  0.237717371  0.22202770  0.253407044 0.0000000
```

ANOVA result show p < 2e-16 which **rejects the *H_0*.** Station categories differ significantly in mean delays. According the to the Turkey HSD shows that Category 3 > 1,2,4,5 has the largest delays. Category 4 < all others lowest delays. Category 5 tends to be very close to Category 2 not significantly different. Hence ***H_1* is accepted** bigger hub stations have higher mean delays than smaller regional ones which have lower delays.

2. **Hypothesis 2: T-Test.**

   H_0 (null): Mean delays are the same in peak vs off-peak hours.

   H_1 (alt): Mean delays are higher during peak hours.

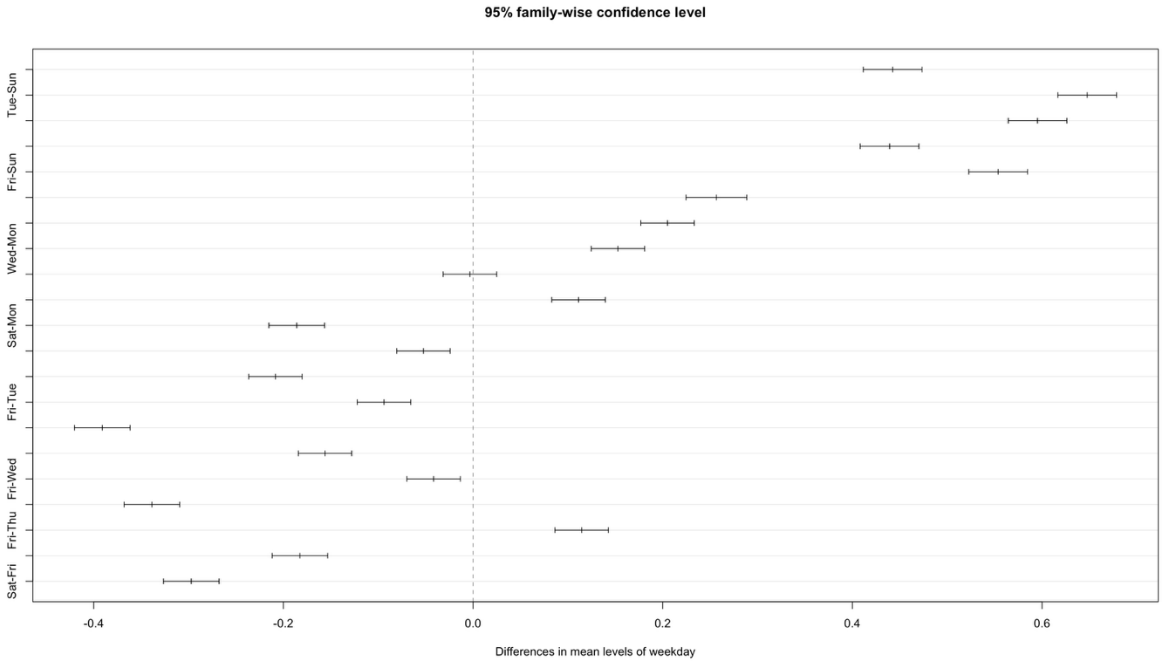
```
	Welch Two Sample t-test

data:  arrival_delay_m by peak_hour
t = -73.122, df = 1240625, p-value < 2.2e-16
alternative hypothesis: true difference in means between group Off-Peak and group Peak is not equal to 0
95 percent confidence interval:
 -0.3955871 -0.3749340
sample estimates:
mean in group Off-Peak     mean in group Peak 
               1.04571                1.43097
```

T-Test result p < 2.2e-16 **rejects the *H_0*.** Mean off-peak = -1.05 min, mean peak = 1.43 min. Trains in peak hours are tend to be delayed about 0.4 minutes which more than off-peak hence the difference is statistically significant hence ***H_1* is accepted.**

3. Hypothesis 3: ANOVA

H_0 (null): Average delays are equal across weekdays (Mon–Sun).

H_1 (alt): At least one weekday has a different average delay.




```
                 Df   Sum Sq Mean Sq F value Pr(>F)
weekday           6    68276   11379   897.8 <2e-16 ***
Residuals   1821493 23086105      13
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
208394 observations deleted due to missingness
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = arrival_delay_m ~ weekday, data = db_data)

$weekday
                diff         lwr         upr     p adj
Mon-Sun  0.442498887  0.41150235  0.47349542 0.0000000
Tue-Sun  0.647607125  0.61665208  0.67856217 0.0000000
Wed-Sun  0.595268798  0.56434696  0.62619064 0.0000000
Thu-Sun  0.439199694  0.40824624  0.47015315 0.0000000
Fri-Sun  0.553719820  0.52272557  0.58471407 0.0000000
Sat-Sun  0.256614362  0.22462499  0.28860373 0.0000000
Tue-Mon  0.205108238  0.17689605  0.23332042 0.0000000
Wed-Mon  0.152769911  0.12459416  0.18094566 0.0000000
Thu-Mon -0.003299193 -0.03150963  0.02491124 0.9998671
Fri-Mon  0.111220933  0.08296574  0.13947612 0.0000000
Sat-Mon -0.185884524 -0.21522788 -0.15654117 0.0000000
Wed-Tue -0.052338327 -0.08046843 -0.02420823 0.0000009
Thu-Tue -0.208407431 -0.23657228 -0.18024259 0.0000000
Fri-Tue -0.093887305 -0.12209698 -0.06567763 0.0000000
Sat-Tue -0.390992762 -0.42029228 -0.36169324 0.0000000
Thu-Wed -0.156069104 -0.18419745 -0.12794076 0.0000000
Fri-Wed -0.041548978 -0.06972221 -0.01337575 0.0002766
Sat-Wed -0.338654435 -0.36791887 -0.30939000 0.0000000
Fri-Thu  0.114520126  0.08631220  0.14272805 0.0000000
Sat-Thu -0.182585332 -0.21188317 -0.15328749 0.0000000
Sat-Fri -0.297105457 -0.32644639 -0.26776452 0.0000000
```

ANOVA results F = 897.8, $p < 2e\text{-}16$ **rejects the H_0.** Which mean the average delay are not really equal across the weekdays. The results shown by the Tukey's test tells us that Monday-Friday in genral have higher delays than the weekend which is Sunday as well as Saturday which is much closer because of the lower delays. The test also shows Tuesday has the highest delays compared to most days which might be due to some unprecedented events. hence **H_1 is accepted.**

4. **Hypothesis 4: Correlation**

   H_0 (null): Arrival delays and departure delays are unrelated (correlation = 0; mean difference = 0).

   H_1 (alt): Arrival delays strongly predict departure delays (positive correlation > 0).

```
	Pearson's product-moment correlation

data:  db_data$arrival_delay_m and db_data$departure_delay_m
t = 4617.6, df = 2029892, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9554305 0.9556696
sample estimates:
      cor
0.9555502
```

Correlation results $r = 0.956$, $p < 2e\text{-}16$ **rejects the *H_0*.** very strong positive correlation. Test statistic $t = 4617.6$, $df = 2029892$ which are $p < 2e\text{-}16$ highly significant. 95% CL confirming correlation is consistently strong. Hence this directly corelated that if a train departs late, it is in-fact guaranteed to also arrive late which seem quite logical that is also considering the size of the departure delay closely mirrored in the arrival delay ***H_1*** **is accepted.**

5. **Hypothesis 5**

H_0 (null): Probability of delay > 6 min is independent of station category, weekday, and time of day.

H_1 (alt): Probability of delay > 6 min depends on these factors.

```
 OnTime Delayed
1941962   87932

	Pearson's Chi-squared test

data:  table(db_data$arrival_delay_check, db_data$weekday)
X-squared = 4422.6, df = 6, p-value < 2.2e-16


	Pearson's Chi-squared test

data:  table(db_data$arrival_delay_check, db_data$category)
X-squared = 1767.4, df = 4, p-value < 2.2e-16

	Pearson's Chi-squared test with Yates' continuity correction

data:  table(db_data$arrival_delay_check, db_data$peak_hour)
X-squared = 2513.5, df = 1, p-value < 2.2e-16
```

Results from number jotted above we can tell that out of 2029892 trains 87932 were delayed more than 6 minutes compared to the larger amount of On time trains it might not be a lot, but the frustrations of the commuters in real time is evident. When it comes to Delay in weekdays is it significant as delay likelihood varies by respected weekday. Delay in station category is also substantial as larger or smaller stations have different probabilities of delay. Delay during Peak and Off-Peak times are also momentous as peak hours have a higher risk of delays that is more than the 6 minutes golden rule. Hence all three **rejects the *H_0*.** *Which frankly means delays that happen more than 6 minutes are not random they systematically depend on when, where and during what time the train runs*. ***H_1* is accepted.**

Conclusion

The analysis of Deutsche Bahn delay or testing the 6 minutes golden rule reveals that punctuality is not random but systematically influenced by several structural and operational factors. As discussed in the very start of the report Germany's heartbeat runs on its railways, yet this very pulse is constantly interrupted by delays. The evidence from the analysis done above makes it exceptionally clear, these delays are not simply random mishaps but follow clear and predictable patterns.

Station size: The larger stations like the "Hauptbahnhöfe", carry the burden of heavier delays compared to their smaller reginal cousins.

Time of day: Rush hours is a big factor as the entire nation is on the move to work and back which logically amplifies the chaos with higher and more frequent delays.

Weekday patterns: Upon the analysis we were able to interpret that especially during the weekdays more specifically the middle of the week, stands out as the true testing ground of DB's punctuality while worked really do provide some relief.

Arrivals VS Departures: Perhaps the most striking but honestly not that much of a surprise is that once a train departs late, it almost certainly arrives late which is the very domino effect echoing across the network.

6 – Minute Threshold: Last but not the least is the 6 – minute threshold mark the very reason for this report to come to life. The shows that peak hours, larger stations and weekdays all significantly raise the risk of a train crossing that threshold.

But while the considerable story is strong, we must also admit the very cracks in this very report. This dataset though large, may not capture the full reality of every German railway line, as this dataset is simply the records that was taken of a random week in a year. There would also be other factors that can change the view of the report as well, external factors like weather, strikes or any technical failures were not included meaning in entirety the picture might be incomplete.

In short, the love-hate relationship the people of Germany have with DB is not unfounded. To fix punctuality, DB must simply look squarely at tits busiest stations, formulate the rush-hour scheduling and the chain reaction of delays that ripple across the system. Yet, as with all analyses, these results should be seen as a strong indication, not the final truth.

"A first step towards solving a problem that touches millions every single day"

Bibliography

- *IamExpat in Germany. (2025). Just 56,1 percent of DB trains were punctual in July. [online] Available at: https://www.iamexpat.de/expat-info/germany-news/just-561-percent-db-trains-were-punctual-july*
- *Kinkartz, S. (2025). Deutsche Bahn is on track for more delays and cancellations. [online] dw.com. Available at: https://www.dw.com/en/deutsche-bahn-is-on-track-for-more-delays-and-cancellations/a-73304129*
- *Partridge, J. and Cologne, J.P. (2023). 'It's the same daily misery': Germany's terrible trains are no joke for a nation built on efficiency. The Observer. [online] 14 Oct. Available at: https://www.theguardian.com/business/2023/oct/14/its-the-same-daily-misery-germanys-terrible-trains-are-no-joke-for-a-nation-built-on-efficiency*