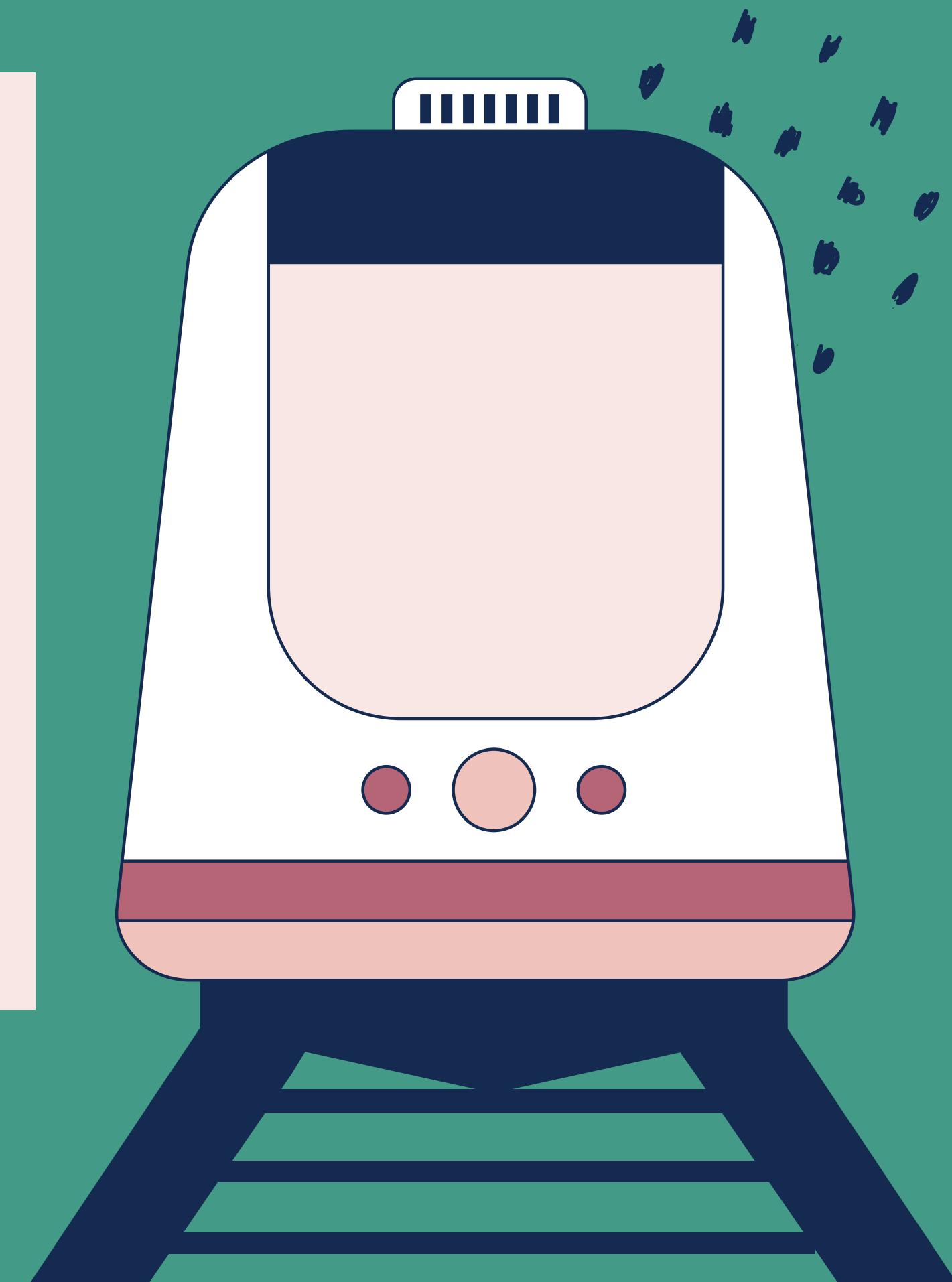


Data Harvesting

Train Ticket Pricing Analysis

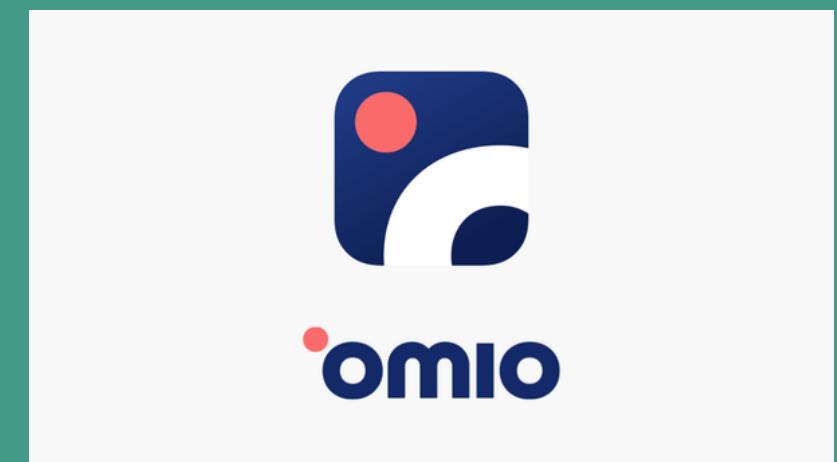


Yijia Lin, Jia Lin and Linghan Zheng





Research motivation

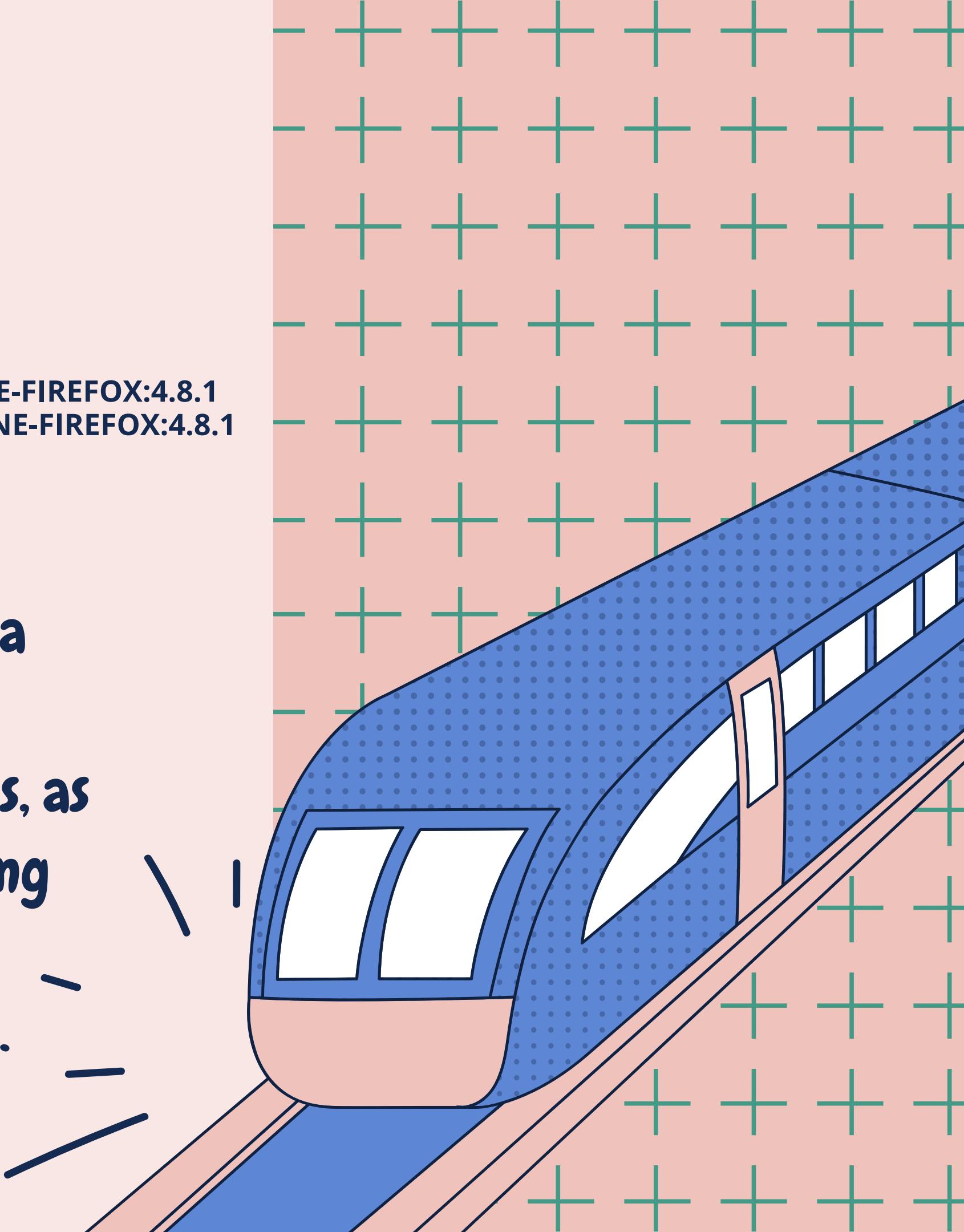




Workflow management

1. The same docker image version: compatible with Windows, MacOs and Linux system
2. Consistent table structures, data formats, and data cleaning standards for each company's data.
3. Loops to automatically create folders and CSV files, as well as to automatically save new data into existing files.
4. Here() + R Project: ensure stable file-saving paths.

SELENIUM/STANDALONE-FIREFOX:4.8.1
SELENIARM/STANDALONE-FIREFOX:4.8.1





**What are our
highlights?**

Renfe

Trenes al mejor precio | Renfe Viajes

ORIGEN: MADRID (TODAS)

DESTINO: BARCELONA (TODAS)

FECHA IDA: mié., 26/03/25

PASAJEROS: 1 adulto

Buscar billete

OUIGO

AGENCIES | OUIGO DESTINATIONS | OUR PRODUCT | FAQ |

From: Madrid - All stations

To: Alicante - Terminal

Depart: 26/03/25

Travellers: 1 Adult (26-59)

SEARCH

trainline

From: Madrid (Any)

To: Pouxeux

+ Via

One-way

Out: Wed 26 Mar • 19:45

1 Adult (26-59)

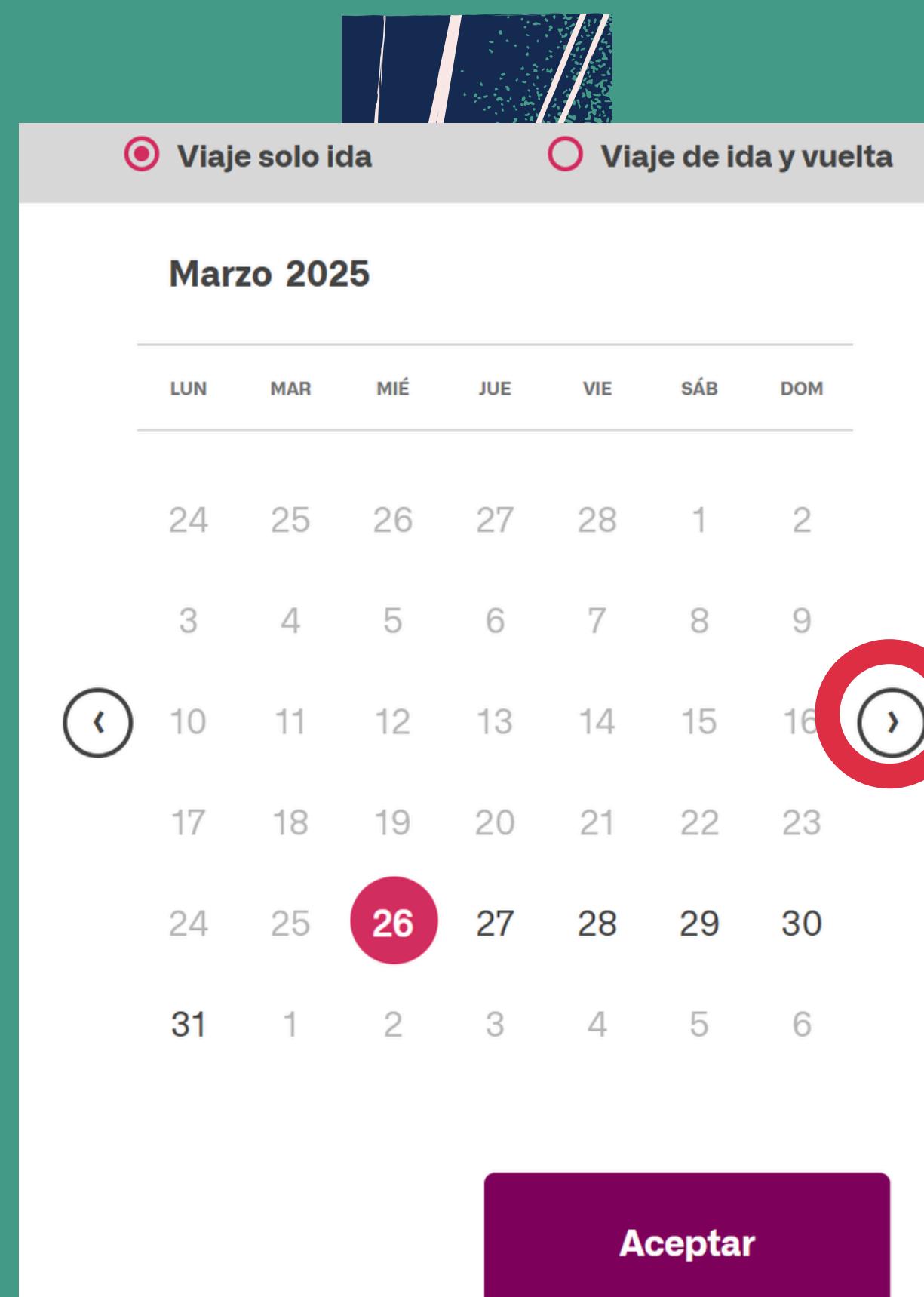
Add railcards

+ Add a voucher code

Open places to stay

Booking.com

Get cheapest tickets



1. Get the current month and year shown in the calendar.
2. Check if the target date is in the future.
3. If yes, click the "next month" button until the target month appears.
4. Find all available dates on the calendar.
5. Match and click the target date.

```
for (i in 1:10) {  
  dates <- remDr$findElements(using =  
    "xpath", "//div[contains(@class, 'lightpick__day is-start-date')]")  
  if (length(dates) > 0) {  
    selected_date <- dates[[1]]  
    selected_text <- selected_date$getElementText()[[1]]  
    if (selected_text == format(search_date, "%d")) {  
      break }  
  Sys.sleep(1)}  
}
```



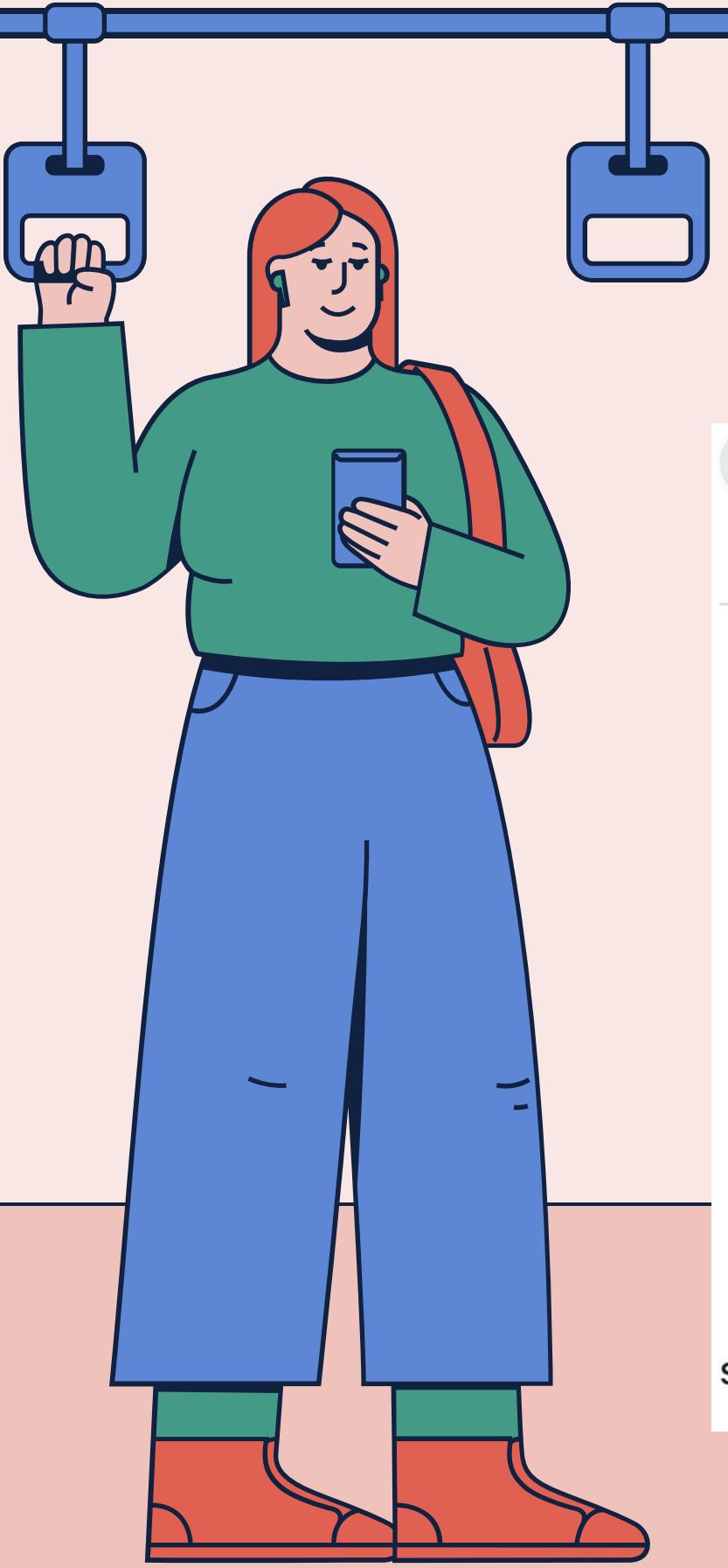
- This loop checks if the correct date was successfully selected on the calendar.
- It tries up to 10 times to find the selected date element.
- If the selected date matches the target date, the loop stops.

```
remDr$executeScript("arguments[0].click();", list(date))
```



- Add wait and retry steps to avoid loading errors.
- If normal click fails, use JavaScript to click the date.





marzo 2025

lu	ma	mi	ju	vi	sá	do
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

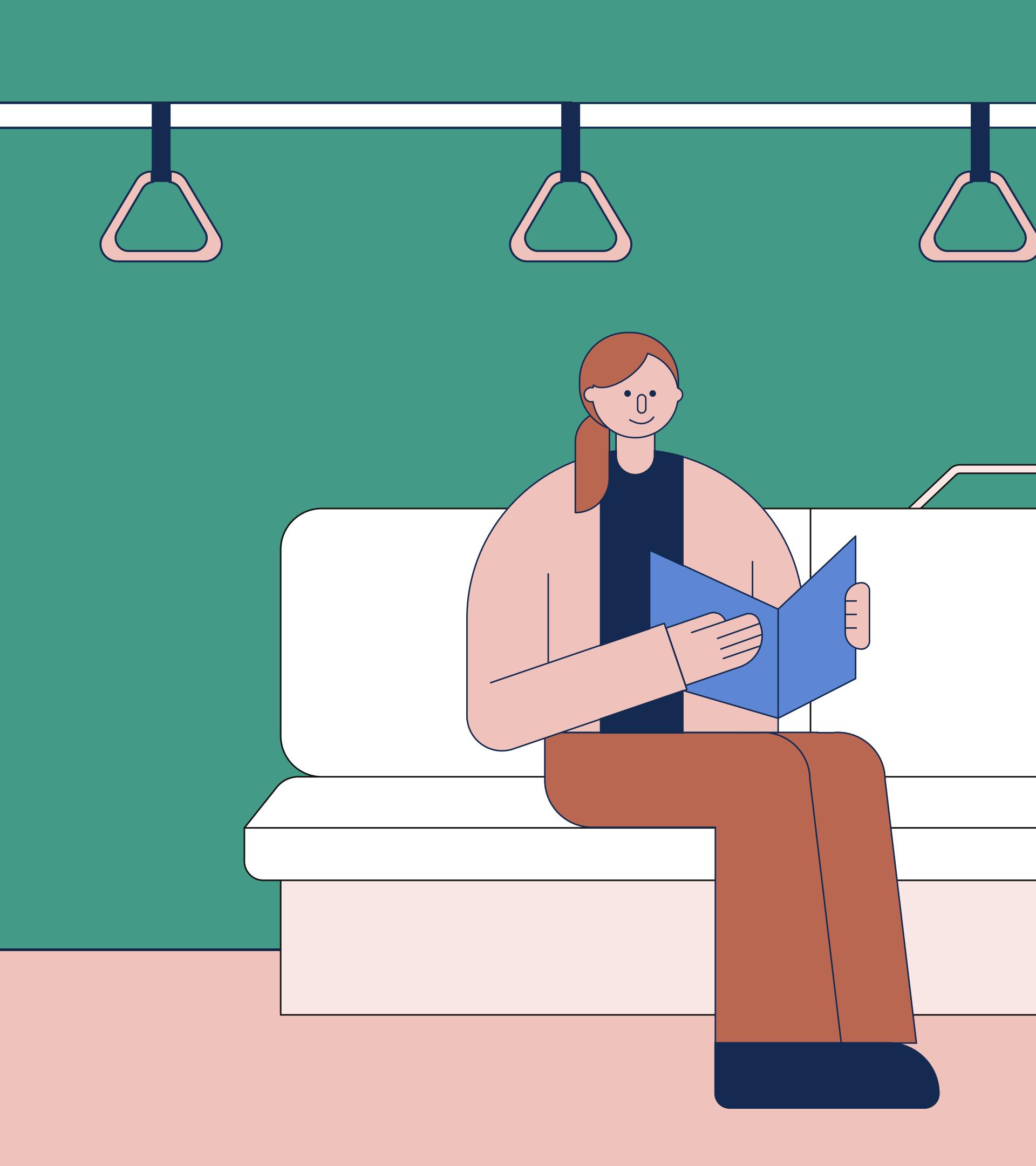
Salida a las **20** ✓ **15** ✓ (i) Solo ida, tarifa para 1 adulto

Extra work on Trainline :)

Recomendamos este viaje basándonos en el precio y duración.

12:57 → 15:34 2h 37m, 0 cambios **54,25 €** **126,25 €**

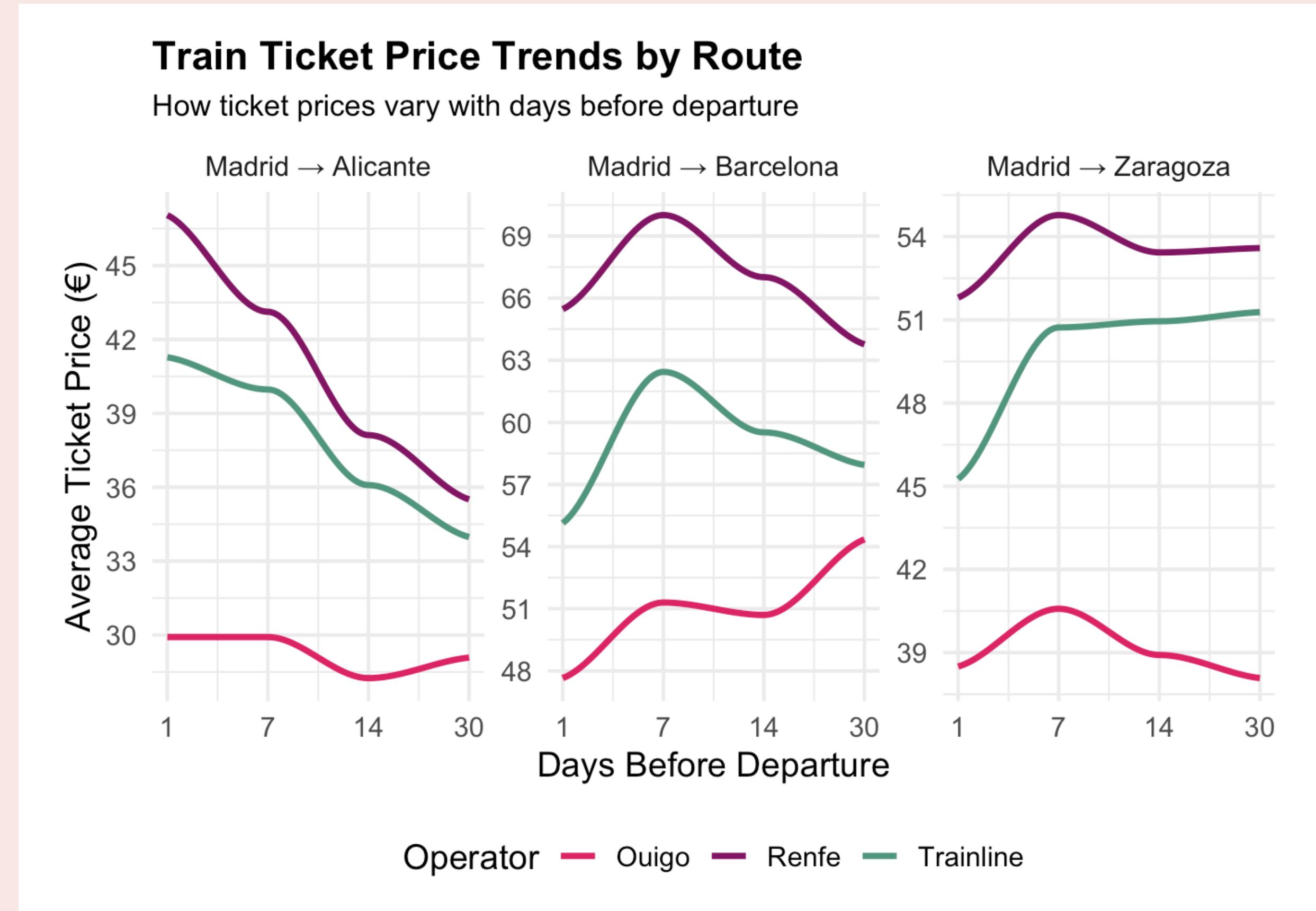
Más tarde

An illustration of a woman with brown hair tied back, wearing a dark blue top and brown pants, sitting on a white bus seat. She is holding and reading a blue book. The bus interior includes overhead handrails and a vertical pole. The background is a solid teal color.

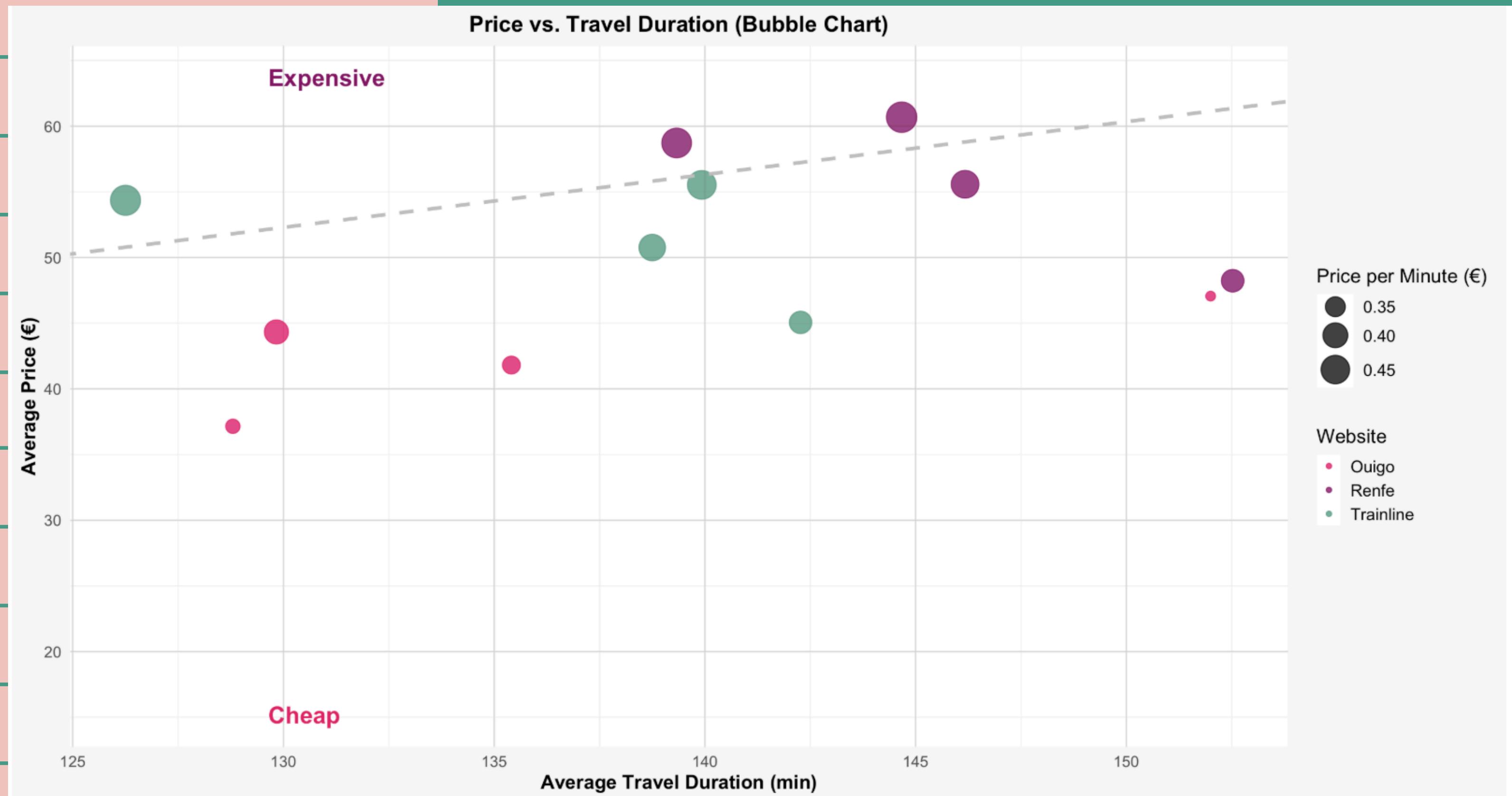
Analysis & Insights



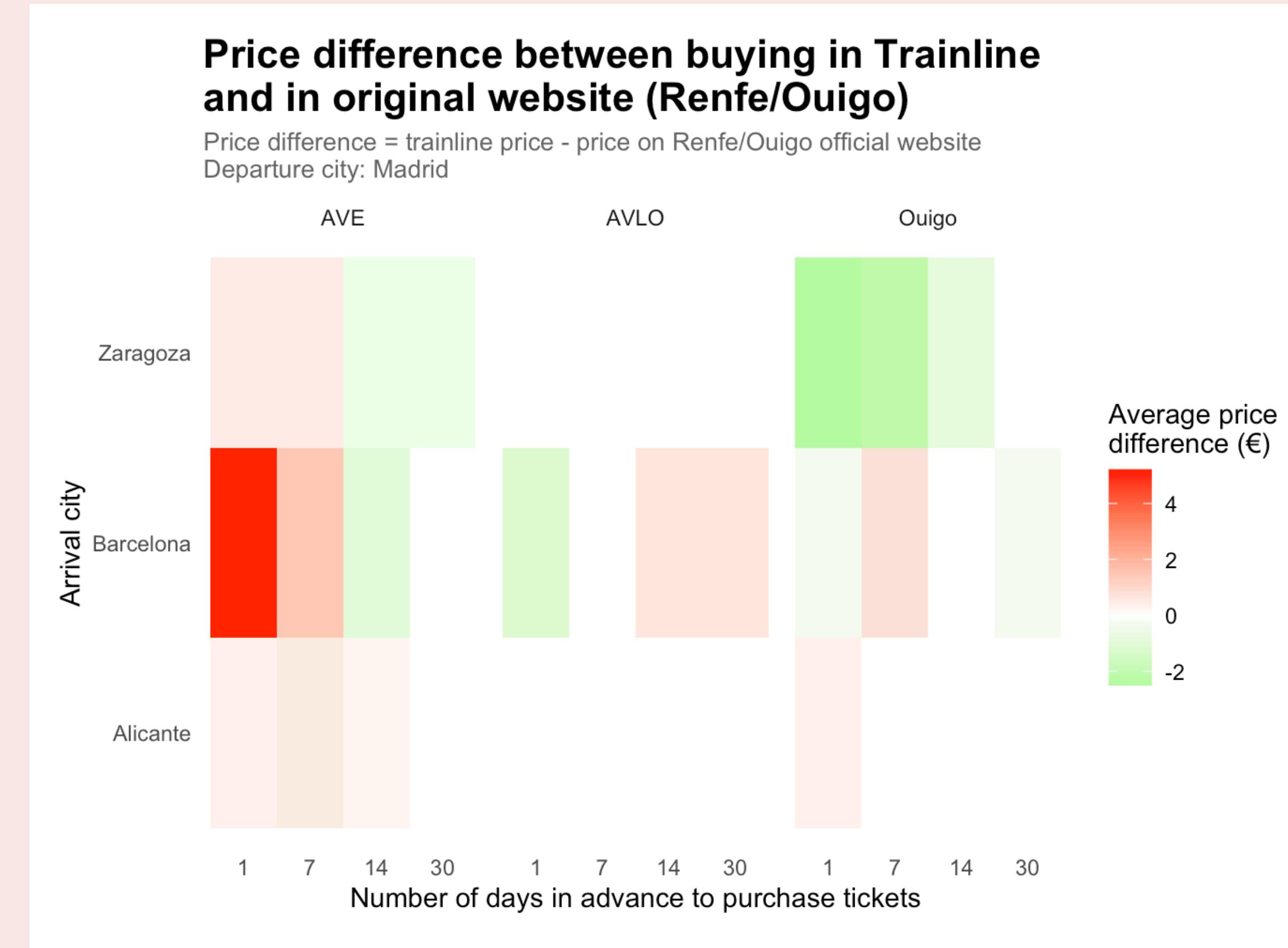
1. Analyze the price trends of train tickets on the different routes and study how the prices vary with the number of days before departure.



2. Train Ticket Pricing, Travel Duration, and Cost-Effectiveness Analysis by Time Period & Booking Website



3. Analyze the price difference between prices on Trainline and prices on official websites (Renfe/Ouigo)



Thank You!

