

# Data Analyst Intern – Sportradar

## Challenge



Hi there!

We are happy to invite you to the next stage of our application process.

### TASK 1: Programmatic Advertising

In programmatic advertising, ads (commonly referred to as *impressions*) are shown on websites with the help of DSPs (Demand Side Platforms). These platforms want to make sure that they are showing the right *impressions* to the right users to generate as many *conversions* as possible. If a user clicks on the ad and lands on the advertiser's website, we store this event in a table called *clicks* (schema shown below). If the user successfully places a deposit after landing on the advertiser's website, we store this event in a table called *conversions* (schema shown below). For each conversion there exists at least one impression, but not all impressions have a conversion or a click.

You are given 3 tables with the following schemas, defined by column names and types:

- **impressions**

impression\_id: string

url\_address: string

user\_id: string

request\_country: string

tracking\_type: string [this is the tracking type (fingerprinted or cookie-based)]

dynamic\_display: boolean [this is whether the impression was served through Dynamic Display]

dynamic\_display\_variables: string [content served in the impression, i.e soccer vs baseball]

request\_browser\_name: string

timestamp: date

- **clicks**

impression\_id: int

user\_id: int

timestamp: string

- **conversions**

conversion\_id: string  
user\_id: string  
dval: integer [this is the deposit value of the conversion]  
curr: string [this is the currency of the deposit value]  
timestamp: date

Using the tables defined above, use a code of your choice to answer the following questions to the best of your knowledge

- **Q1:** What is the CTR (%) for impressions served in "Austria"?
- **Q2:** For each converted user, find out how many impressions they were served. Specifically capturing the timestamp for the first & last impression.
- **Q3:** What other insights could you present which the audience might find interesting from the above schemas? **Do not answer this question with code, just write about what else could be interesting and why (think how to present this in Part 3!)**

## **TASK 2: World Cup 2022**

With Euros underway, it is a repetitive task for the team to produce insights around the importance of sporting events in Ads. Using the attached data, feel free to generate any insights you think is relevant around what we have seen at the World Cup in 2022 & how this could be interesting, below are some hints to help you get started

**Q1:** Which countries have the largest reduction in Cost per Acquisition (CPA) for different conversions during the World Cup compared to other date ranges available?

**Q2:** What insights could you present which clients might find interesting & encourage them to spend more on the next major sporting event?

**Q3:** Euro 2024 is currently underway; how could you adapt your analysis for this competition instead? Are there any insights or questions that would be more or less relevant?

## **TASK 3: Presentation**

**Please prepare a presentation of your findings for the above tasks (aim for around 10 minutes). The presentation should include both Task 1 & Task 2 results. Upon submission, feel free to attach any other relevant work, good luck & have fun working on the challenge.**