

# **Solution engineer/specialist task**

*Date: 2024-05-21*



## Solution engineer/specialist task

*This task is intended for us to have an enlightened technical discussion and get an idea of how you approach a new task. Please note that we do not expect that you have an answer ready for every question before the interview and it is OK if you cannot answer everything! We do not expect you to use more than 3 hours on this task. If you have any questions regarding this task send an email to [jesper@mindway.ai](mailto:jesper@mindway.ai)*

This task will be the starting point of the second interview. It is up to you how you solve the task and how thoroughly you choose to do so. The main purpose of the task is to get a relevant conversation going and understand how you approach, solve and especially present problem solutions.

It's not necessary that you have a prepared answer for every part of this task, but it is important that you've considered each question and can delve into a theoretical discussion about how you might have approached each of them.

You can choose whichever tool or coding language that you prefer to solve this task. The focus is the content of your solution and especially how you choose to present your solution.

You do not have to submit anything prior to the interview, but please bring your laptop with the solution to the interview so we can review it together.

Consider that your solution is made to be presented. Quality and presentability is in this case far superior to quantity. Please make the assumption that you must present for an audience composing of both technical and non-technical recipients.

### Get started

Along with this problem formulation, you should get a .csv file: Gambling\_data.csv.

This is a slightly cleaned and shrunk version of a [publicly available dataset](#). You can also read about the data in the [dataset description](#). The dataset is shrunk in the sense that we've removed the modified “\_nonzero” and “\_sqrt” columns.

The data has a (binary) target column “rg\_case” (Responsible gambling case), a “user\_id” column and the rest of the columns represent the players and their gambling behaviour.

### Part 1 - Data analysis

The idea with this part is for you to play around with some data and show us how you approach a new dataset. We are both interested in the approach and the considerations that you might have had.

1. Choose a column of interest in the dataset. Produce an appropriate plot that illustrates the difference between rg\_cases (1) and non rg\_cases (0). Prepare an interpretation.
2. Identify the most important variable(s) for distinguishing between rg\_cases and non rg\_cases.
3. The data contains information about usage of 3 different groups of game types: fixed odds, live action and casino.

- Each game type has the same types of data points, but the mechanics of the games are very different. Is it possible to identify differences for `rg_cases` between two of these game types?
- 4. Some players only play one type of game, while others play multiple. How would you take that into consideration when measuring risks associated with problematic gambling behaviour?
  - Could you build datapoints which takes all types of play into consideration? What could be problems to consider when doing that?

## Part 2 - Responsible gambling risk

The idea with this part is for us to get an insight into how you approach the issue of problem gambling. It's not a completely defined concept, and thus often requires us to be creative when making interpretations.

1. Build a function that finds the objectively worst gambler in the dataset.
  - Does this player look like he's at risk in the sense of responsible gambling? Why?
2. When measuring gambling risk, we work through the so-called "Markers of harm", which is different indicators that show different types of problematic behaviour. One Marker that's much debated is the so-called "Loss chasing" marker. DSM-5 for gambling disorder defines it simply: After losing money gambling, (the player) often returns another day to get even ("chasing" one's losses).
  - Build a function that evaluates loss chasing (based on DSM-5 or your own interpretation), and determine the most at risk player based on that specific marker.
  - Does this player look like he's at risk in the sense of responsible gambling? Why?

## Part 3 - System integrations

The following is for considerations and for us to have a discussion around the topic of system integrations. We do not expect you to have any code for this exercise.

- Imagine that we are deploying a piece of containerized software containing a machine learning model on the operators (the gambling providers) cloud. The software requires the operator to setup a data pipeline, a database and a compute environment to run the ML model.
  - How do you see us facilitating this process to make sure that we create a seamless integration?
- The same operator has requested an integration into their existing PAM (Player account management) system with the results of our ML models.
  - Consider how we might integrate into their system with our ML model results and ensure that they interpret the output appropriately.