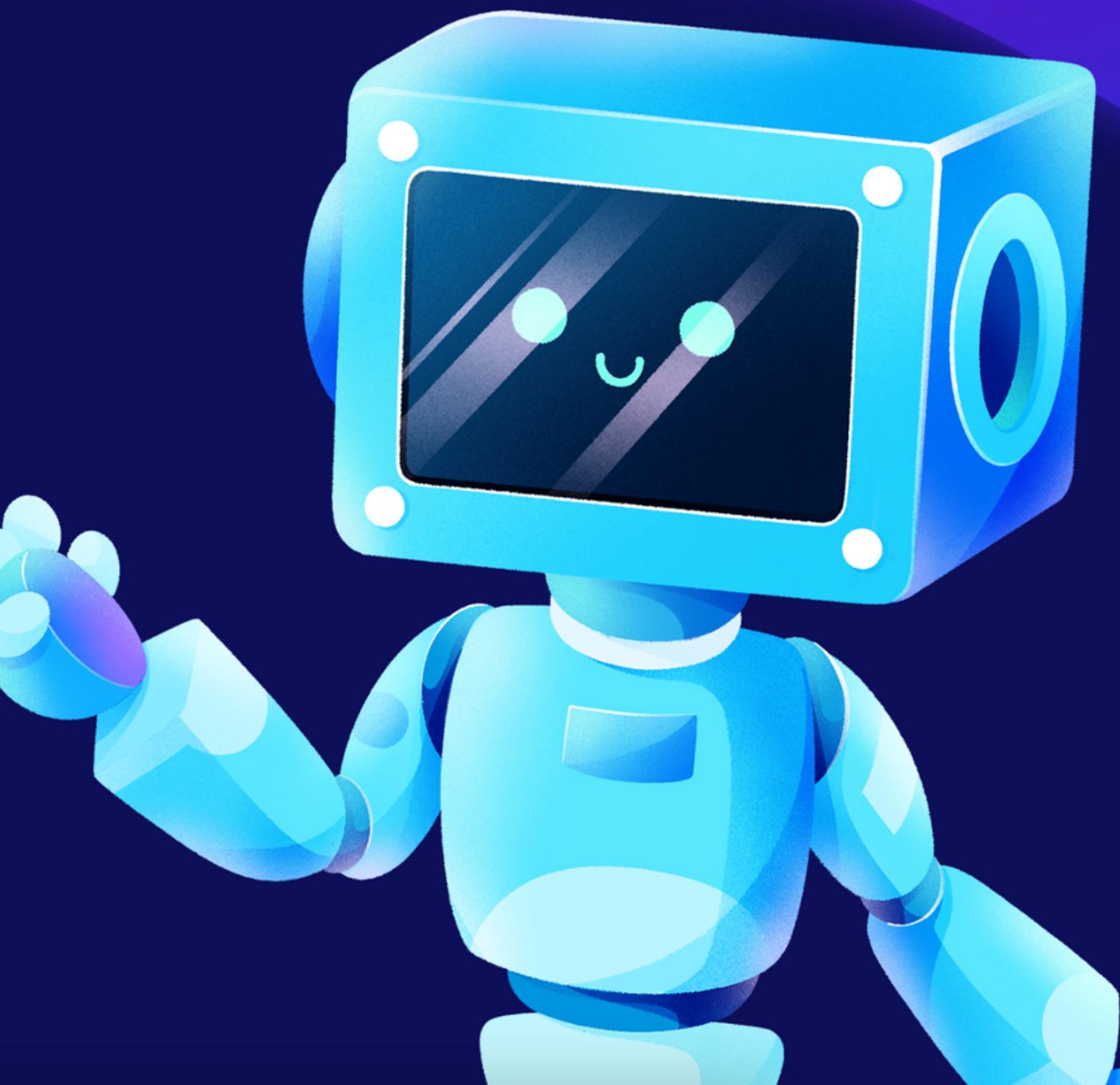




Datathon 2023

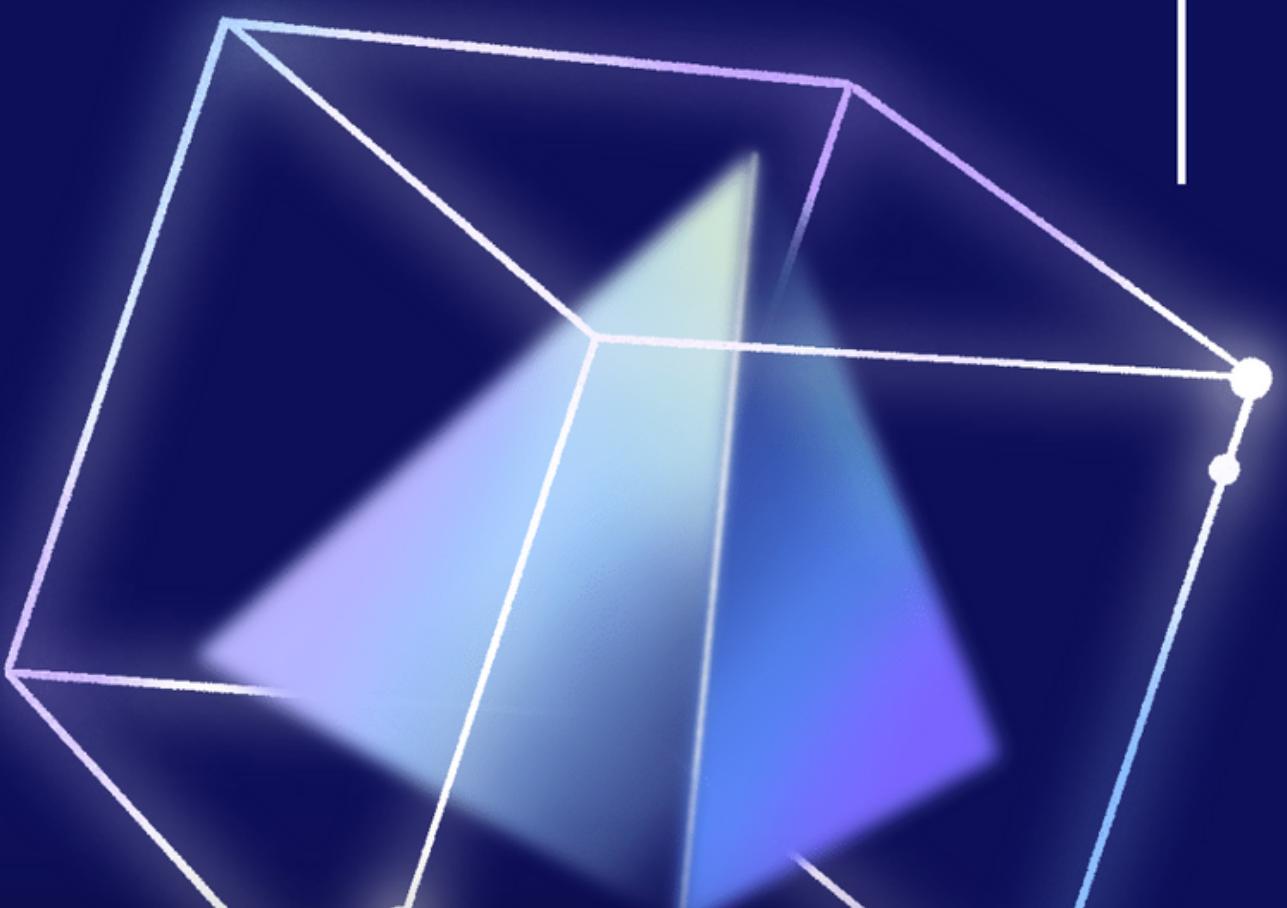
# Case Brief





# Table of Contents

• Instructions	03
• Case	04
• Dataset	06
• Data Dictionary	07
• Marking Rubric	11





# Instructions

## Rules

- 4 hours to work, starts at **10am**
- 2 rounds of presentations
  - Heats: 5 min each, 1 Q&A, 6 finalists
  - Finals: 5 min each, 3-4 Q&A, 3 winners

## Submission

- Deliverables: only submit findings and insights through presentation
- Submission deadline: **2pm, no submissions allowed after**  
***Submission link provided in discord***

# The Case

You are working for SparkCognition as a Data Scientist. SparkCognition has been commissioned by an insurance company to develop a tool to **optimize their marketing efforts**. They have given us a dataset as a result of an email marketing campaign. The dataset includes customer information, described below, as well as whether the customer responded to the marketing campaign or not.

---

**Aim:** *Develop a model that can be used to better target customers in future email campaigns. In other words, predict the responded target variable described above based on all the input variables provided. The **targeted stakeholder** that you present your analysis to will be a semi-technical marketing manager.*

## In your solution, you should consider:

- *How will your model be used?*
- *Who do you want to target?*
- *How can we make sure customers don't receive mail if they don't want it?*

## You may use the following questions to guide you through your work:

- *Describe your model and why did you choose this model over other types of models?*
- *Describe any other models you have tried and why do you think this model performs better?*
- *How did you handle missing data?*
- *How did you handle categorical (string) data?*
- *How did you handle unbalanced data?*
- *How did you test your model?*

# Dataset

## Files you'll be working with:

**Marketing\_training.csv**: This file contains the training set that you will use to build the model. The target variable is responded.

- [https://drive.google.com/file/d/1OGxsy47DxJqlm\\_QDyYM8P5GvgfHVqyF2/view?usp=sharing](https://drive.google.com/file/d/1OGxsy47DxJqlm_QDyYM8P5GvgfHVqyF2/view?usp=sharing)

**Marketing\_test.csv**: This file contains testing data where the input variables are provided but not the responded target column.

- [https://drive.google.com/file/d/1inIVmjaj9\\_GU8rtGPu37xNA\\_gkaN3xkg/view?usp=sharing](https://drive.google.com/file/d/1inIVmjaj9_GU8rtGPu37xNA_gkaN3xkg/view?usp=sharing)



# Data Dictionary



Type	Name	Description
Input Variables	custAge	The age of the customer (in years)
Input Variables	profession	Type of job
Input Variables	marital	Marital status
Input Variables	schooling	Education level
Input Variables	default	Has a previous defaulted account?
Input Variables	housing	Has a housing loan?
Input Variables	contact	Preferred contact type
Input Variables	month	Last contact month

Type	Name	Description
Input Variables	day_of_week	Last contact day of week
Input Variables	campaign	Number of times the customer was contacted
Input Variables	pdays	Number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previous contacted)
Input Variables	previous	Number of contacts performed before this campaign for this client
Input Variables	poutcome	Outcome of the previous marketing campaign
Input Variables	emp.var.rate	Employment variation rate - quarterly indicator

Type	Name	Description
Input Variables	cons.price.idx	Consumer price index - monthly indicator
Input Variables	cons.conf.idx	Consumer confidence index - monthly indicator
Input Variables	euribor3m	Euribor 3 months rate - daily indicator
Input Variables	nr.employed	Number of employees - quarterly indicator
Input Variables	pmonths	Number of months that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previous contacted)
Input Variables	pastEmail	Number of previous emails sent to this user
Target Variables	responded	Did the customer respond to the marketing campaign and purchase a policy?

# Marking Rubric

---

## Note

***This is just a rough understanding of some criteria that judging could be based on, final decisions are ultimately up to the Atlassian Representatives.***



# Technical



## Depth of Analysis & Understanding of Problem

*Evaluates the extent to which the presenter demonstrates a thorough understanding of the problem.*



## Insights & Analysis

*Assess the depth, quality, and innovativeness of the analysis, with a focus on providing insights that go beyond surface-level examination of the data.*



## Results

*Evaluates if conclusions are explicitly tied to analysis and context, information is conveyed correctly with adequate and appropriate facts.*



## Relevance to Business Context

*Evaluates ability to make recommendations for analytic approaches based on business goals.*

# Presentation



## **Adaptation to the Target Audience & Background**

*Evaluates the presenter's ability to adapt the content and language to suit the target audience's background.*



## **Clarity & Organisation of Presentation**

*Evaluates the clarity and organization of the presentation structure.*



## **Engaging & Persuasive Delivery**

*Assesses the presenter's delivery style, confidence, and ability to engage and persuade the audience.*



## **Effective Use of Visuals & Supportive Materials**

*Examines the effective use of visuals and supporting materials to enhance the presentation.*



# Best of Luck!!!

~ DataSoc Careers 2023