Simility Data Challenge

We have suggested using a dataset on travel insurance claims (see attachment; under open database license). It would allow us to assess candidates' abilities to work with data as well as to present key findings and results. Below is a short description of the data attributes of the data set.

Analyse the data with the tool of your choice and build a machine learning model with a statistical language of your choice. Please present insights/patterns and key findings you have found in the data as well as your model results in a 15mins presentation. Assume we are your clients, so set your focus on the business outcome.

The questions for the presentation:

- To present interesting patterns/insights found in the data set
- What business value is your solution deliver?
- How would you enhance your solution going forward?
- What additional data and features would help you to improve? What kind of additional features could you derive from the data?
- What practical suggestions would you give for our business based on your analysis?

In addition, we would like to ask you a few questions about your methodology. We will focus on your decisioning process and why you made particular choices. Please send us your code and the presentation within 2 days of receiving this document.

We will not assess you on model performance, but methodology and problem solving. Should you have any questions, don't hesitate to ask them in advance. All the very best :)

Data Description

- Target: Claim Status (Claim.Status)
- Name of agency (Agency)
- Type of travel insurance agencies (Agency.Type)
- Distribution channel of travel insurance agencies (Distribution.Channel)
- Name of the travel insurance products (Product.Name)
- Duration of travel (Duration)
- Destination of travel (Destination)
- Amount of sales of travel insurance policies (Net.Sales)
- Commission received for travel insurance agency (Commission)
- Gender of insured (Gender)
- Age of insured (Age)