Case Study

General Information

Congratulations on making it to the final stage of the Data Scientist Professional certification!

So far we have assessed a range of the skills required of a data scientist. In this case study we will be assessing your skills in model development and communication. We want to see that you can take a business problem, select and implement appropriate techniques and then give a relevant summary of what you have found to the defined audience.

24 hours before your presentation...

We will send you:

- A DataCamp Workspace template including the project brief and data
- A link to join the presentation session

You will then be able to prepare for the presentation session. You must perform your analysis in the DataCamp Workspace - this will be graded. Your presentation can be in any format that you feel appropriate. We expect you to spend around 1.5 hrs to 2 hrs preparing.

During the presentation session

We will let you know the running order. You will have ten minutes to give your presentation. It should be targeted at the audience described in the project brief. You will be stopped if you reach 12 minutes.

> ⓘ **Please note:**
> This session will be recorded so that your case study can be graded by a senior member of the team and for quality assurance purposes.

After the session:

We will grade your presentation, and if you are successful subsequently grade your analysis as well. We will grade against the rubric included below

We will let you know the outcome of your case study within 3 to 5 working days.

Grading

The case study will test your skills in modelling, and communicating your findings.
You are not expected to fit a perfect model but you will need to show that you have chosen appropriate methods for the problem you are solving. You will have to show that you can communicate at the right level for the target audience.

In order to pass you have to reach “Satisfactory” in all of the criteria in the table below.

Analysis - shown in script/notebook	**Excellent**	**Good**	**Satisfactory**	**Needs Improvement**
Exploratory Data Analysis (EDA)	The EDA included has been refined to reflect the most relevant elements to the problem and elements that may impact further analysis or decisions.	Some refinement of EDA has been considered, with the most relevant elements being reported. Does not always demonstrate a clear connection to the problem.	Large number of graphics and tables created in the EDA, it is not always clear as to the connection to the problem under consideration	There is no evidence that EDA has been conducted
Visualizations/Tables	Graphics and tables reflect good practices (e.g., use of color, when to use a table instead of a plot), have been clearly labelled, and add value.	Consideration has been given as to when to use each graphic or table. Some good practices have been followed (e.g., visualizations are clearly labelled and titled).	Graphics do not follow good practices but are appropriate to the data type. Tables are extensive and not well formated to make them easy to read.	Graphics/tables have been included that are not appropriate to the data or analysis.
Analysis notes	Notes connect any results/graphics/ tables to the problem being considered and the approach being taken. It is clear from the notes how each element of the analysis relates to the overall problem.	Notes are clear and go beyond restating what is seen in the results, with a clear connection to the analysis and problem.	Notes on results/graphics/tables are basic and simply restate what is shown.	Results, tables, and graphics are created with no additional notes.
Model fitting	The choice of model has a clear connection to the problem and the author has shown a clear understanding of modelling good practices, selecting a model that is also suitable for the size of the data as well as the type of problem.	Modelling good practices have been demonstrated, the choice of model is suitable for the problem.	A suitable model has been selected and fitted.	Models selected are not appropriate for the data or problem statement.
Evaluation	The model evaluation is clear, appropriate, and connected to the problem.	The model(s) have been evaluated using appropriate techniques and a clear interpretation of the evaluation has been given.	Model evaluation techniques appropriate to the model have been demonstrated.	The model fit has not been evaluated in any way.

Results - shown in either code or presentation	**Excellent**	**Good**	**Satisfactory**	**Needs Improvement**
Outcome	The outcome clearly relates back to the original problem and defined users, making clear how the analysis helps to solve the defined problem and the impact it has.	The outcomes are stated and some attempt is made to connect back to the original problem statement and how the problem has been solved with these results.	Results restate findings earlier in the report (e.g., stating which is the best model, which features are most relevant).	There is no final summary of the results obtained.
Future work	Future actions are not limited to the analysis conducted but also consider the end user, how they may be able to make use of the work conducted and any actions they may be able to take to improve the work.	Future actions consider the analysis and data collection but may not connect to the end user or the original problem statement.	Future improvements to the analysis conducted are provided.	No consideration has been given to future actions or work.

Communication - shown in presentation	**Excellent**	**Good**	**Satisfactory**	**Needs Improvement**
Motivation	The work is clearly motivated and demonstrates why the work adds value to a specific person/group. The end user is the focus rather than the analysis.	The work is clearly motivated and shows a connection to practical application. The end user may not have been fully considered.	A basic motivation has been stated that shows some connection to the analysis being used in a practical application.	The presentation provides no motivation for why the analysis has been conducted or why the audience may be interested in the outcomes.
Timing	Within the expected timing, appropriate weighting provided to each aspect of the presentation	May have exceeded timing by up to 1 minutes. In general a well balanced presentation with good weighting to each aspect	May have been between 1 and 2 minutes over time. Some aspects are given more focus than necessary	Exceeds timing by more than 2 minutes
Audience	Generally chooses not to use technical terminology, but where absolutely necessary provides simple to understand explanations, often using analogies that can easily	Uses minimal technical terminology and as much as possible provides clear explanations to terms used	May include technical details (such as model information, metrics etc) but makes an attempt to explain relevant information	Frequently uses technical terms with little to no non-technical explanation
Organization	Presentation is structured to clearly tell a story and provides the right information at the right time to progress that story	The presentation has a logical flow that typically follows the analysis conducted	he presentation includes all of the relevant information but may occasionally seem out of order or introduce information too soon/ too late	Presentation lacks organization or structure to convey the appropriate message clearly

Booking a case study session:

Case study sessions can be booked directly through the calendly link on your certification dashboard.
In case you can’t find an available slot, please note that we consistently add more sessions each week.