## Data Scientist Case Study Information

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
	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.	Good  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).	Needs Improvement  There is no final summary of the results obtained.
code or presentation	The outcome clearly relates back to the original problem and defined users, making clear how the analysis helps to solve the	The outcomes are stated and some attempt is made to connect back to the original problem statement and how the problem has been	Results restate findings earlier in the report (e.g., stating which is the best model, which	There is no final summary of the
Code or presentation  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.  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	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.  Future actions consider the analysis and data collection but may not connect to the end	Results restate findings earlier in the report (e.g., stating which is the best model, which features are most relevant).  Future improvements to the analysis conducted	There is no final summary of the results obtained.  No consideration has been
Code or presentation  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.  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	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.  Future actions consider the analysis and data collection but may not connect to the end	Results restate findings earlier in the report (e.g., stating which is the best model, which features are most relevant).  Future improvements to the analysis conducted	There is no final summary of the results obtained.  No consideration has been
Communication -	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.  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.	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.  Future actions consider the analysis and data collection but may not connect to the end user or the original problem statement.	Results restate findings earlier in the report (e.g., stating which is the best model, which features are most relevant).  Future improvements to the analysis conducted are provided.	There is no final summary of the results obtained.  No consideration has been given to future actions or work.
Communication - shown in presentation	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.  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.  Excellent  The work is clearly motivated and demonstrates why the work adds value to a specific person/group. The end user is the	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.  Future actions consider the analysis and data collection but may not connect to the end user or the original problem statement.  Good  The work is clearly motivated and shows a connection to practical application. The end	Results restate findings earlier in the report (e.g., stating which is the best model, which features are most relevant).  Future improvements to the analysis conducted are provided.  Satisfactory  A basic motivation has been stated that shows some connection to the analysis being used in	There is no final summary of the results obtained.  No consideration has been given to future actions or work.  Needs Improvement  The presentation provides no motivation for why the analysis has been conducted or why the audience may be interested in the



he presentation includes all of the relevant

order or introduce information too soon/too

information but may occasionally seem out of

Presentation lacks organization or

structure to convey the

appropriate message clearly

The presentation has a logical flow that

typically follows the analysis conducted

Presentation is structured to clearly tell a

the right time to progress that story

story and provides the right information at

Organization