

Grading Rubric Data Analyst

July 2024 additions / changes are highlighted in yellow

Category	Competency	Sufficient	Insufficient
Data Validation	Assess data quality and perform validation tasks	Has validated all variables and where necessary has performed cleaning tasks to result in analysis-ready data.	Has not conducted all the required checks and/or has not cleaned the data. May have removed data rather than performed cleaning tasks.
<p>There should be either a written description of the data validation performed or a mix of code and written description.</p> <p>For writing only: there must be a description for every column (8 or 9 columns) in the data.</p> <p>For code and text: the text should be the primary grading method. Code can be taken into account where text has not been included for a column. The code must show that they have checked the distinct values (for categorical) or range (for continuous) for the columns. They must have highlighted in text the problems that were encountered. Only checking the data type is not sufficient.</p> <p>All columns must be checked/described.</p> <p>They must have resolved the issues that were included in the data as described in the project for associate or with their own justification for professional.</p> <p>Code is not required for any role or level.</p> <p>Information article provided to candidates</p>			

Data Visualization	Create data visualizations in R or Python to demonstrate the characteristics of data and represent the relationships between features	<p>Has created at least two different visualizations of single variables (e.g. histogram, bar chart, single boxplot)</p> <p>Has created at least one visualization including two or more variables (e.g. scatterplot, filled barchart, multiple boxplots)</p> <p>Has used visualizations that support the findings being presented</p> <p>Data visualizations displayed in the presentation are clear and readable onscreen (e.g. components such as axis labels are large enough to be read onscreen, points and lines can be distinguished, key is included if needed, any chartjunk / clutter does not inhibit accurately reading the visualization)</p>	<p>Has used the same visualization throughout.</p> <p>Has not included graphics to represent single variables and relationships.</p> <p>Has not used visualizations that support the findings being presented.</p> <p>Visualizations are unreadable: too small, missing information (e.g. keys, axis labels), or unclear in other ways</p>
<p>There must be two visualizations of single variables (any graphic that includes only one variable i.e. only a variable defined on one axis).</p> <p>There must be one visualization that includes multiple variables (i.e. both an x and y axis variable). This plot can include any number of variables.</p> <p>There must be at least two different types of visualization i.e. they cannot include 3 bar charts, but two bar charts and a</p>			

scatter plot is ok. The questions given will usually direct to a categorical and continuous plot as well as a relationship plot but this is not a requirement.

The graphics included **should support their analysis description**. For instance, if they state that the data shows that sales increase over time, there should be a plot that shows sales increasing over time.

The graphics can appear anywhere in the report.

Code is not required.

The readability / clarity criterion is for graphics presented; it is very reasonable to create graphics during analysis that are not meant to be seen and can therefore be too small or incomplete

We **don't** expect candidates to demonstrate

- understanding of accessibility features e.g. appropriate choice of color
- refinements of layout e.g. order of bins
- logical color choice e.g. warmer temperatures as red, colder as blue

[Information article provided to candidates.](#)

Data Communication	Employ multiple tactics (written and verbal) to communicate to business leaders	<p>For each analysis step, has provided a written explanation of their findings and/or reasoning for selecting approaches</p> <p>Delivers a presentation addressing the business goals, outcomes and recommendations</p> <p>Delivers a presentation with a recognisable narrative (i.e. a logical</p>	<p>Has not provided a written summary for each step</p> <p>Has not delivered a presentation</p> <p>Delivers a presentation with no apparent narrative or no connection with the findings of the data analysis</p>
--------------------	---	---	---

		progression between points) that can be followed without significant effort, and is supported by the findings of the data analysis	
<p>They need to have completed a report with written text summaries throughout.</p> <p>In the report, steps in data cleaning are clearly understandable and have been presented through a short piece of text at minimum (e.g. 3 sentences)</p> <p>In the report, findings are presented in short piece of text at minimum (e.g. a few sentences)</p> <p>Data Scientist: The report must include all code for model development and evaluation</p> <p>They must also have delivered a presentation. There is no requirement for slides or other visual aids. The presentation must include the business goal, their approach to the problem, what they found and their recommendations.</p>			
Business Focus	Make recommendations for analytic approaches based on business goal	<p>Has described at least one of the business goals of the project</p> <p>Has explained how their work has addressed the business problem</p> <p>Has provided at least one business recommendation for future action to be taken by the company based on the outcome of the work done</p>	<p>Has not identified any business goals</p> <p>Has not explained how their work has addressed the business problem</p> <p>Has not provided any recommendations for future actions for the company</p>
<p>We expect to see/hear a description of the business goal, how they addressed the problem and at least one recommendation for next steps.</p>			

This criteria can be **graded from the report or presentation** (note all these things need to be included in the presentation to pass the communication criteria).

Repeating the information exactly as we provide is permitted. Repeating the customer questions we provide is permitted.

The recommendation can relate to either something the customer does or the individuals doing the work.

The recommendation should be an explicit statement on the best course of action; the reader should not need to infer this from the results of the analysis

Business Metrics	Benchmarks, monitors, and evaluates business processes	Enables continual monitoring by defining a metric that can be used by the business in the future to measure success in solving the problem Has evaluated the metric using the existing data to provide a baseline measure for the problem	Has not identified a metric for continual monitoring
------------------	--	--	--

They should **state a way to monitor the business goal**. This could be a metric or graphic.

They should have **used the existing data to demonstrate** the metric. E.g. if they state they will monitor the proportion of product sales, they should show what the proportions are based on the data we have given.

The report must specify why the metric was chosen