



[◀ Return to Classroom](#)

# Data Scientist Capstone

## REVIEW

## CODE REVIEW

## HISTORY

### Meets Specifications

**Awesome** 🙌🙌

**Congratulations on passing this project.** 🏆🏆

This project was not easy but you have done it gracefully. This is all because of your hard work and continuous evaluation. But still don't relax, keep exploring and learning from the references provided. Try to solve and practice this kind of data analysis on [Kaggle](#). This will bring a habit and will develop a mental model.

### Extra Materials

The links below, provide more insights to these processes:

- [The Python Requirements File and How to Create it](#) [Article]
- [Python Virtual Environment | Introduction](#) [Article]

You can also share your project on LinkedIn and ask the audience for necessary feedback or open the project for anyone to collaborate. This way you will find many interesting connections and engagement with others.

I wish you good luck. Looking forward to your success.  
For any queries, you can ask on [Knowledge Portal](#) as well.

Stay  ! Stay Safe

DON'T FORGET TO RATE MY WORK AS PROJECT REVIEWER! YOUR FEEDBACK IS VERY HELPFUL AND APPRECIATED.

## Project Definition

Student provides a high-level overview of the project. Background information such as the problem domain, the project origin, and related data sets or input data is provided.

✓ High-level overview of the project

✓ Problem Domain

**Nice! you clearly describe the overview of your project. It helps the to know about the overall project scope.**

Most people dive into technicals, but that does not help them understand the values an application can provide. So it becomes necessary to think from the perspective of a reader while writing an Overview and Project Domain.

**\*\*[How to Create the Perfect Project Overview](#) [Article]**

✓ Related data sets or Input data.

The problem which needs to be solved is clearly defined. A strategy for solving the problem, including discussion of the expected solution, has been made.

✓ The problem which needs to be solved is clearly defined

✓ A strategy for solving the problem, including discussion of the expected solution, has been made.

**You have excellently explained how you incorporate the Supervised classification Machine Learning Algorithms to solve the problem.**

Metrics used to measure the performance of a model or result are clearly defined. Metrics are justified based on the characteristics of the problem.

For example, explain why you want to use the accuracy score and/or F-score to measure your model performance in a classification problem,

✓ Metrics used to measure the performance of a model or result are clearly defined.

**Good Job! You have come up with one of the excellent metrics to measure your classification model.**

- ✓ Metrics are justified based on the characteristics of the problem.

**You have clearly explained the reason for selecting Accuracy for this problem statement.**

## Analysis

Features and calculated statistics relevant to the problem have been reported and discussed related to the dataset, and a thorough description of the input space or input data has been made. Abnormalities or characteristics about the data or input that need to be addressed have been identified.

- ✓ Features and calculated statistics relevant to the problem have been reported and discussed related to the dataset, and a thorough description of the input space or input data has been made.
- ✓ Abnormalities or characteristics about the data or input that need to be addressed have been identified.

Build data visualizations to further convey the information associated with your data exploration journey. Ensure that visualizations are appropriate for the data values you are plotting.

- ✓ Build data visualizations to further convey the information associated with your data exploration journey.

**You have effectively used the Distribution and Bar plot to extract the valuable information.**

- ✓ Ensure that visualizations are appropriate for the data values you are plotting.

**Awesome! correct input data and parameters have been passed to the methods.**

## Methodology

All preprocessing steps have been clearly documented. Abnormalities or characteristics about the data or input that needed to be addressed have been corrected. If no data preprocessing is necessary, it has been clearly justified.

- ✓ All preprocessing steps have been clearly documented.

✅ Abnormalities or characteristics about the data or input that needed to be addressed have been corrected.

**Data has been properly handled before ingesting it into the model.**

The process for which metrics, algorithms, and techniques were implemented with the given datasets or input data has been thoroughly documented. Complications that occurred during the coding process are discussed.

✅ The process for which metrics, algorithms, and techniques were implemented with the given datasets or input data has been thoroughly documented

**Very solid step by step process here, as it is quite clear in how you approached this problem. Your results would definitely be replicable.**

✅ Complications that occurred during the coding process are discussed.

The process of improving upon the algorithms and techniques used is clearly documented. Both the initial and final solutions are reported, along with intermediate solutions, if necessary.

✅ The process of improving upon the algorithms and techniques used is clearly documented.

✅ Both the initial and final solutions are reported, along with intermediate solutions

**Excellent job! You have discussed the ways to improve the existing techniques used in the analysis.**

## Results

If a model is used, the following should hold: The final model's qualities — such as parameters — are evaluated in detail.

Some type of analysis is used to validate the robustness of the model's solution. For example, you can use cross-validation to find the best parameters.

Show and compare the results using different models, parameters, or techniques in tabular forms or charts.

Alternatively, a student may choose to answer questions with data visualizations or other means that don't involve machine learning if a different approach best helps them address their question(s) of

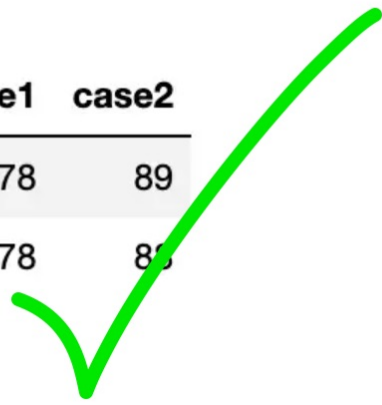
interest.

- ✓ If a model is used, the following should hold: The final model's qualities — such as parameters — are evaluated in detail.
- ✓ Show and compare the results using different models, parameters, or techniques in tabular forms or charts.

**Well Done! You have successfully compared the models.**

### Comparison Table

	RandomForest	case1	case2
train_score	80	78	89
test_score	65	78	85



The final results are discussed in detail. Explain the exploration as to why some techniques worked better than others, or how improvements were made are documented.

- ✓ The final results are discussed in detail
- ✓ Explain the exploration as to why some techniques worked better than others, or how improvements were made are documented.

## Conclusion

Student adequately summarizes the end-to-end problem solution and discusses one or two particular aspects of the project they found interesting or difficult.

- ✓ Student adequately summarizes the end-to-end problem solution
- ✓ Discusses one or two particular aspects of the project they found interesting or difficult.

**Nice work discussing your final end-to-end problem solution as this reads quite well. I can definitely tell that you have spent a long time on this project as it really shows.**

Discussion is made as to how at least one aspect of the implementation could be improved. Potential solutions resulting from these improvements are considered and compared/contrasted to the current solution.

- ✓ Discussion is made as to how at least one aspect of the implementation could be improved
- ✓ Potential solutions resulting from these improvements are considered and compared/contrasted to the current solution.

## Deliverables

If the student chooses to provide a blog post the following must hold: Project report follows a well-organized structure and would be readily understood by a **technical** audience. Each section is written in a clear, concise and specific manner. Few grammatical and spelling mistakes are present. All resources used to complete the project are cited and referenced.

If the student chooses to submit a web-application, the following holds: There is a web application that utilizes data to inform how the web application works. The application does not need to be hosted, but directions for how to run the application on a local machine should be documented.

- ✓ Project report follows a well-organized structure and would be readily understood by a **technical** audience.

### Your post clearly communicates findings in a layman's manner. 👍

This way all the non-technical audience or stakeholders would be able to understand your insights and would be able to interpret the visualization and statistical metrics.

- ✓ Each section is written in a clear, concise and specific manner

### Awesome work! I really liked how you structured your blog post.

- ✓ All resources used to complete the project are cited and referenced.

Student must have a Github repository of their project. The repository must have a README.md file that communicates the libraries used, the motivation for the project, the files in the repository with a small description of each, a summary of the results of the analysis, and necessary acknowledgements. If the student submits a web app rather than a blog post, then the Project Definition, Analysis, and Conclusion should be included in the README file, or in their Jupyter Notebook. Students should not use another student's code to complete the project, but they may use other references on the web including StackOverflow and Kaggle to complete the project.

- ✓ Student must have a Github repository of their project

## Suggestion:

**Always try to keep commits as small and focussed as possible.** When you are trying to fix one particular bug and you spot another one, then first resolve the first bug and commit as soon as you resolved it. Do not dive into another one without committing.

Apart from that, try to write meaningful commit messages. For example:

```
Bugfix: bugfix message
Update: update message
Correction: correction message
Added: files added or whatever
```

### **\*\*Best practices for using Git [Article]**

- ✓ README file communicates the libraries used

You can also create a `requirements.txt` file and mention how visitors can install libraries from it. Here is an example:

#### Installation

This code runs with Python version 3.\* and requires some libraries, to install theses libraries you will need to execute:

```
pip install -r requirements.txt
```

**Command to install the libraries**

- ✓ README file has motivation for the project section
- ✓ README file has files in the repository section
- ✓ README file has summary of the results of the analysis
- ✓ README file has acknowledgement section

## Useful References:

[Manage your data science project structure in early stage \[Article\]](#)

[How to write a good readme for your github project? \[Article\]](#)

Code is formatted neatly with comments and uses DRY principles. A README file is provided that provides. PEP8 is used as a guideline for best coding practices.

Best practices from software engineering and communication lessons are used to create a phenomenal end product that students can be proud to showcase!

- ✓ Code is formatted neatly with comments and uses DRY principles
- ✓ A README file is provided.
- ✓ PEP8 is used as a guideline for best coding practices.

**Good Work! your comments are constructive in understanding the flow of the analysis. 100**

Markdown text is essential in ensuring that our notebooks are easy to use and understand. The Markdown text is not a replacement for line or block comments in the code cells but rather a place to provide a broader context for the code. You have nicely followed the pep8 style guidelines for naming your variable and functions.

## Useful References:

[Python Code Quality: Tools & Best Practices](#) [Article]

 [DOWNLOAD PROJECT](#)

[RETURN TO PATH](#)