



PROJECT

Investigate a Dataset

A part of the Data Analyst Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!  

Meets Specifications

Next, Data Wrangling and MongoDB

Congratulations on completing this difficult course. Next we will be doing some data wrangling with python - cleaning up datasets so they can be analyzed meaningfully. This is an incredible skill considering the bulk of Big Data is not standardized. You may have even noticed this about the Subway Dataset. Another incredible skill which will become even more important as we acquire more and more data is the ability to work with NoSQL databases, like with MongoDB. We will go over all this next. Hope you enjoy!

Code Functionality

All code is functional and produces no errors when run. The code given is sufficient to reproduce the results described.

Hello again! I was your previous reviewer, Miadad. I am really glad I get to review your project again and see it grow. I hope my feedback has helped. Let's take a look!

The project uses NumPy arrays and Pandas Series and DataFrames where appropriate rather than Python lists and dictionaries. Where possible, vectorized operations and built-in functions are used instead of loops.

The code makes use of functions to avoid repetitive code. The code contains good comments and variable names, making it easy to read.

Quality of Analysis

The project clearly states one or more questions, then addresses those questions in the rest of the analysis.

Data Wrangling Phase

The project documents any changes that were made to clean the data, such as merging multiple files, handling missing values, etc.

Please note that there are 177 missing age values that should be discussed as well as the cabin values.

Exploration Phase

The project investigates the stated question(s) from multiple angles. At least three variables are investigated using both single-variable (1d) and multiple-variable (2d) explorations.

The project's visualizations are varied and show multiple comparisons and trends. Relevant statistics are computed throughout the analysis when an inference is made about the data.

At least two kinds of plots should be created as part of the explorations.

Conclusions Phase

The results of the analysis are presented such that any limitations are clear. The analysis does not state or imply that one change causes another based solely on a correlation.

Apologies for the confusing quotation in the previous submission. I am glad you understood what I was trying to communicate. Simply outstanding work on the limitation discussion. Your project is truly outstanding and ready to go into your job-ready portfolio.

Communication

Reasoning is provided for each analysis decision, plot, and statistical summary.

Visualizations made in the project depict the data in an appropriate manner that allows plots to be readily interpreted.

 [DOWNLOAD PROJECT](#)

[RETURN TO PATH](#)

[Student FAQ](#)