



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.

Code reflects the work done in the analysis and produces no errors.

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.

Great work applying the `numpy` and `pandas` libraries appropriately throughout the report!

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.

You have postulated several points of exploration and have provided a thoughtful investigation in turn. Asking the right questions is, arguably, the most important part in data analysis. Well done!

Data Wrangling Phase

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

You have documented all the changes made to the dataset as well as addressing then handling of missing values. Outstanding work.

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 use of univariate and multivariate plots to investigate the answers to your questions from several perspectives is thorough and compelling.

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.

Truly outstanding work on the visualizations. The plots are varied and intuitively make sense with the idea being portrayed. Also, relevant statistics are computed to supplement the plots whenever necessary. Well

done!

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.

The statistical testing was out of this world! Well done! It adds a great deal of validity to your conclusions and is well out of the requirements of the class.

Communication Phase

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

Outstanding work on presenting the investigation in a way that makes the limitations clear.

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

Visualizations are easily interpretable and does not need supplemental material for explanation.

 [DOWNLOAD PROJECT](#)

Have a question about your review? Email us at review-support@udacity.com and include the link to this review.

[RETURN TO PATH](#)

Rate this review

[Student FAQ](#)