

PROJECT

Investigate a Dataset

A part of the [Data Analyst Nanodegree Program](#)

PROJECT REVIEW

CODE REVIEW

NOTES

Meets Specifications

SHARE YOUR ACCOMPLISHMENT



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.

Please note The project instructions require the analysis be presented in either a PDF or HTML document. I have reviewed your project through the submitted `.ipynb` file in order to save unnecessary submission attempts, however, `.ipynb` is not a universal file format for documentation and cannot be opened on all machines. It is fortunate that I am able to open, however, this may not be the case for other reviewers. If this was a GitHub repo link, please note that even though GitHub has native support for `.ipynb` files, reviewers do not get access to the repo. Instead, Udacity downloads the contents of the repo and sends us an archive. This is for the privacy of your other repos. For future submissions, please submit scripts as well as the analysis in PDF or HTML formats. Since this was the only issue in the project, I've overlooked it.

Submission

What to include in your submission

1. A PDF or HTML file containing your analysis. This file should include:

- A note specifying which dataset you analyzed
  - A statement of the question(s) you posed
  - A description of what you did to investigate those questions
  - Documentation of any data wrangling you did
  - Summary statistics and plots communicating your final results

2. If the code you used to perform your analysis is not included in the above, you should submit the code separately in `.py` file(s).

3. A list of Web sites, books, forums, blog posts, github repositories, etc. that you referred to or used in creating your submission (add N/A if you did not use any such resources).

<https://www.udacity.com/course/viewer#!c-nd002/l-3176718735/m-5464560482>



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.

You have postulated several meaningful questions 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.

Communication Phase



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

Reasoning and logic flow is sound during the investigation. Thank you for adding the notes and summarizations for each section of the analysis. It really helps us understand your thought process.



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](mailto:review-support@udacity.com) and include the link to this review.

RETURN TO PATH