UDACITY

PROJECT

Investigate a Dataset

A part of the Data Analyst Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

One of the best reports I've graded. 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 and SQL databases. We will go over all this next. Good luck on the next course. It was a pleasure to look over your project. Keep up the great work!

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.

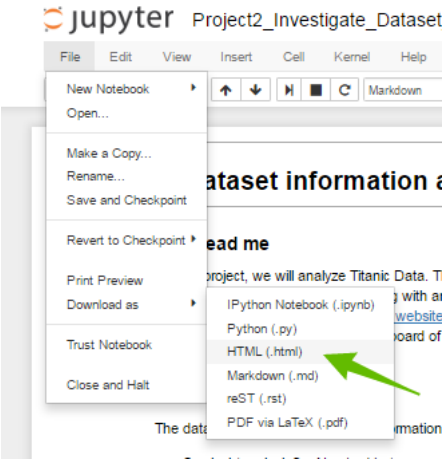
Submission

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).

IPython notebook instructions

If you used IPython notebook to create your analysis, you can include your code directly in the notebook and do not need to submit it separately. To download your notebook as an HTML file, click on File -> Download.As -> HTML (.html) within the notebook. If you get an error about "No module name", then open a terminal and try installing the missing module using `pip install <module_name>` (don't include the "<" or ">" or any words following a period in the module name).



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

Since this was the only area that did not meet submission requirements and you have proven your understanding of course material, I have decided to pass the project. Please remember to thoroughly look over the submission requirements for future projects.

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.

Great work on documentation. I would highly recommend getting accustomed to wrangling procedures related to creating more descriptive labels. Here is what I mean.

For each of these variables, I've created another column with more descriptive labels.

```
# Create Survival Label Column
titanic['Survival'] = titanic.Survived.map({0 : 'Died', 1 : 'Survived'})
titanic.Survival.head()
0        Died
1    Survived
2    Survived
3    Survived
4        Died
# Create Pclass Label Column
titanic['Class'] = titanic.Pclass.map({1 : 'First Class', 2 : 'Second Class', 3 : 'Third Class'})
titanic.Class.head()
0    Third Class
1    First Class
2    Third Class
3    First Class
4    Third Class
```

```
# Create Embarked Labels Column
titanic['Ports'] = titanic.Embarked.map({'C' : 'Cherbourg', 'Q' : 'Queenstown', 'S' : 'Southampton'})
```

- One of the main reasons for this is that it helps when creating interpretable visualizations. It will save a lot of steps later on in trying to modify the plots.

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/stats 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. The statistical testing is above and beyond. I love it. This will look amazing in your portfolio.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.

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

Communication

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

I really appreciate the level of depth with your communication. You do a really great job guiding the reader.

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. The only thing I would recommend is to create more descriptive labels for items like Pclass and Ports.

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

Student FAQ