UDACITY

Investigate a Dataset

A part of the Data Analyst Nanodegree Program

PROJECT REVIEW
CODE REVIEW
NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

Overall, great work!. Excellent analysis and well done writing your report in explanatory terms.

Congratulations on passing your exam!

Code Functionality

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

Your code perfectly works. As a side comment, I think this is an excellent reference. Note notebooks are a great tool and markdown notation is becoming more and more popular, so it is definitely worth to invest some time to learn more about it.

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.

Good work with `Pandas` . The code built-in functions and operations to avoid for-loops or manual operations.

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

The code demonstrates the use of a function that reduces repetitions. The non trivial code lines are commented and explained.

As a suggestion, you can calculate survival rates using a function like this:

```python
def change_of_survival(variables):
    '''Returns the survival rates for the list of variables'''
    return df.groupby(variables).mean().loc[:,'Survived']

print change_of_survival(['Sex','Pclass'])
```

Quality of Analysis

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

The report states interesting questions being addressed throughout 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.

Excellent auditing for the data including documentations and clear explanations. Consider to include also a discussion about other options and the shortcoming for the method you suggested. As a suggestion, Pandas .fillna(method='pad') fill gaps while preserving variable's distribution.

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 analysis implements 1d and 2d visualizations allowing the audience understand how variables are distributed and how are related.

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.

The report includes many figures that depict comparisons and trends, with relevant statistics computed throughout the analysis. Awesome work!!

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.

Results are clearly presented and dataset shortcomings described and handled. Also, great work deploying statistical tests to estimate the significance of your results.

Communication

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

As a suggestion, it is important that the report will include a discussion about the limitations or shortcomings of the methods that you are using. For example, Age column includes several missing values, how this affected your analysis? does it include any limitation? how did you handle these missing values?. Since there is an uneven distribution between the genders might tilt the results. Do you miss any other feature that would be interesting to include in the analysis? for example, crew members details.

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

Nice work!. Visualizations included are properly formatted and help to better understand the different data exploration conclusions.

⬇ DOWNLOAD PROJECT

RETURN TO PATH

Student FAQ