UDACITY

PROJECT

Investigate a Dataset

A part of the Data Analyst Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

Dear Student,

Overall your project had improved from the previous review. You are on the right path and your project is very well structured. Even though you have successfully passed this review. I still left some suggestions to help you tweak it further for future development.

CONGRATULATIONS

I wish you all the best for the future
Cheers!

Code Functionality

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

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

SUGGESTION

Here are a few Pandas built-in methods that are very useful for exploring variables in this project:

- Boolean-Indexing
- Group-by
- Value-Counts

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.

SUGGESTION

The question asked is a broad one, but it would be nice to actually state the other questions in the introduction. This can be a very helpful agenda for readers and add an appeal to the introduction of the report. Based on the report and the analysis done, here are a few questions that could be asked:

- How many persons survived the Titanic?
- Which age group survived ?
- Which P class had the most survivors?
- Which gender survived with a higher count, Male or Female?

Data Wrangling Phase

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

Good work in implementing a Data Wrangling Phase

SUGGESTION

The most important aspect of Data Wrangling is to clean or transform the data preparing it for analysis.

One main issue is having missing data while conducting analysis, which can provide skew/bias results. Luckily there are a few methods that Pandas provide to deal with these issues:

- The first thing to do is to always Identify the missing values within the dataset. The few steps after this explain how to deal with the missing data
- If there are columns with a few rows of missing data the Dropna methodcould be used to drop the missing rows.
- If there are rows with missing data the Fillna-method can be used instead of dropping them completely (This method can vary with the data and the project)
- The final option is if there are way too many missing values within a column it is best to drop the column completely using the Drop-column-method

Data Wrangling does not only involve Identifying and dealing with missing values but also involves in transforming the data to a more effective state to target the analysis. Here are other wrangling methods:

- Binning or Cutting Groups continuous or numerical values into smaller groups or 'bins'
- Pandas-DummiesTransforms categorical data into dummy/indicator variables

All these methods are applicable to this project!

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.

Good work here

SUGGESTION

Within the conclusion it is good to include statical figures that was found within the findings for final assurance. For example all the analysis done for each of the graphs could be summarized and placed within the conclusion.

Here are some other limitations to explore as well:

- Missing data
- Was the data sufficient to prove your findings

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

Rate this review

Student FAQ