Investigate a Dataset

REVIEW | HISTORY

Requires Changes

4 specifications require changes

Great submission!

You have met most of the requirements, only the explorations need to add.

Keep working on it!

Code Functionality

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

All code works well without any errors, great job!

Further Consideration

Working through some examples of Automate the Boring Stuff with Python will level up your skills in Python.

- **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.**

Numpy and Pandas have been used, well done!

Below is an article about the Pandas performance, you may have a look for your reference:Pandas Performance

- **The code makes use of at least 1 function to avoid repetitive code.**
- **The code contains good comments and meaningful variable names, making it easy to read.**

A function has been defined, good job!

All codes have good comments and are easy to read.

For you to further improve, you may refer to the document below from google which introduces how to make docstring, comments, and variable names.

https://google.github.io/styleguide/pyguide.html

Quality of Analysis

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

Questions have been listed at the beginning and addressed in the later analysis, well done!

Data Wrangling Phase

- **The project documents the steps that were taken to clean the data, such as merging multiple files, handling missing values, etc.**

You did a perfect data wrangling, you have looked into the numeric statistic, data type, missing value, duplicates and outliers. Each step has been well documented.

Exploration Phase

- **The project investigates the stated question(s) from multiple angles.**
- **The project explores at least three variables in relation to the primary question. This can be an exploratory relationship between three variables of interest, or looking at how two independent variables relate to a single dependent variable of interest.**
- **The project performs both single-variable (1d) and multiple-variable (2d) explorations.**

Single-variable explorations have been done, fantastic.

Required Improvement:

In order to meet the rubric, you still need to do multiple-variable explorations. Line plot or scatter plot will be a good choice.

Below is a very good article to introduce these two explorations: Data Exploration

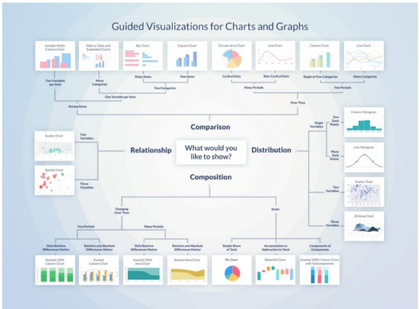
- **The project's visualizations are varied and show multiple comparisons and trends.**
- **At least two kinds of plots should be created as part of the explorations.**
- **Relevant statistics are computed throughout the analysis when an inference is made about the data.**

Required Improvement:

To meet the rubric you need to use at least two kinds of plots, in the submission only pie plot has been used.

Please add another kind of plot, line plot, scatter plot and etc.

Below is a guideline for graphing, just for your future reference.



Conclusions Phase

- **The Conclusions have reflected on the steps taken during the data exploration.**
- **The Conclusions have summarized the main findings in relation to the question(s) provided at the beginning of the analysis accurately.**
- **The project has pointed out where additional research can be done or where additional information could be useful.**
- **The conclusion should have at least 1 limitation explained clearly.**
- **The analysis does not state or imply that one change causes another based solely on a correlation.**

You did conclude the findings with limitations, fantastic!

You may refer to the below article for more information about limitations.
What are the limitations of a study and how to write them?

Communication

- **The code should have ideally the following sections: Introduction; Questions; Data Wrangling; Exploratory Data Analysis; Conclusions, Limitation.**
- **Reasoning is provided for each analysis decision, plot, and statistical summary.**
- **Interpretation of plots and application of statistical tests should be correct and without error.**
- **Comments are used within the code cells.**
- **Documented the flow of analysis in the mark-down cells.**

Validate this point once multiple variable explorations added

- **Visualizations made in the project depict the data in an appropriate manner (i.e., has appropriate labels, scale, legends, and plot type) that allows plots to be readily interpreted.**

Validate this point once multiple variable explorations added

RESUBMIT PROJECT

DOWNLOAD PROJECT

Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

Watch Video (3:01)

RETURN TO PATH