UDACITY

☰

‹ Return to "Data Analyst Nanodegree" in the classroom

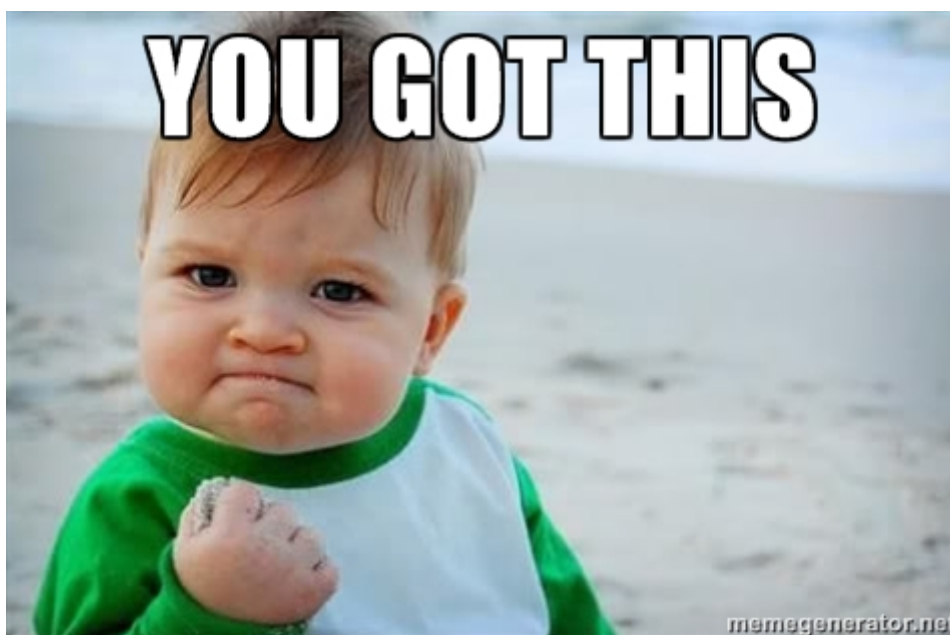⊡ DISCUSS ON STUDENT HUB ›

Investigate a Dataset

REVIEW
HISTORY

Meets Specifications

Congratulations, Suraj! You passed in this project!

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.

Very well done using built-in functions. Here is the documentation and tutorial for several other functionalities that we can find within the Pandas module that are very useful.

✓

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

Excellent work creating three functions to use them in your exploration! The choice of variable names are also good ones and correctly represent the object that those variables are storing.

Quality of Analysis

✓

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

Five questions were proposed and answered along the exploration.

Data Wrangling Phase

✓

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

All necessary changes during the data wrangling phase were documented using markdown and inline comments.

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.

Both kinds of exploration were done using several different kinds of plots.

✓

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.

I liked how you use several kinds of plots in your exploration. It should be noted that there are many more kinds of plots and the more kinds you know, the better you become in identifying the best one to your specific use case. Check here for 44 types of plots that you may find interesting! You can also look here to matplotlib plots with their implementations in Python!

Tips:

- For `countbudget` and `countprofit` it is better to use a barplot for visualization. Whenever you can, try and use plots rather than raw tables or statistical reports. Plots have much more visual appeal and are easier and faster to understand.
- In general, horizontal bar plots have their highest bar on top followed by the others in decreasing size. That is, using `ascending = True` in `countgen=colgen.value_counts(ascending=False)`.

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.

The conclusions summarize well all your findings in the exploration phase.

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.

Now that you are familiar with matplotlib, you should feel free to learn about other plotting libraries. One of them is called plotly in which you can make interactive plots rather than the static ones obtained using matplotlib and seaborn. Check it out!

⬇ DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

★ ★ ★ ★ ★