

Investigate a Dataset

REVIEW

HISTORY

Requires Changes

1 SPECIFICATION REQUIRES CHANGES

Great job so far. I am impressed that you have done almost perfect work in your very first attempt. There is a minor improvement needed to fulfill the requirement of the rubric. I have made comments to help you understand the changes needed. I hope you would be able to make those changes very quickly and resubmit soon a perfect project. Keep up the good work!

Code Functionality

✓	All code is functional and produces no errors when run. The code given is sufficient to reproduce the results described.
Code ran without any errors. Meet the specification.	
✓	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.
You correctly used Pandas series and dataframes throughout the project. Also, great use of inbuilt function like groupby. Here are a few Pandas built-in methods that are very useful for exploring variables in this project: <ul style="list-style-type: none">Boolean-IndexingGroup-byValue-CountsSeries.mapWorking-with-text-data	
✓	The code makes use of functions to avoid repetitive code. The code contains good comments and variable names, making it easy to read.
The notebook has well commented code, the variable names are apt and related to the project.	

Quality of Analysis

✓	The project clearly states one or more questions, then addresses those questions in the rest of the analysis.
Great questions and great answers! I loved the way you framed the questions and proceeded to answer them, It follows a very smooth flow and well complemented by the visualizations. 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.
---	---

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 makes use of both single and multiple variable explorations to investigate different features and the relations between these features in the dataset.	
✓	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.
Conclusion is well documented but limitations of the dataset are not discussed. Even though there is a hint of the limitations in your writing, this rubric item requires to make a subsection inside the conclusion section. In this subsection discuss the limitations of the data which might have hindered your analysis. Examples would be null or missing value, erroneous values like those zero values of budget and revenue, and also whether this dataset can properly represent the population or not.	

Communication

✓	Reasoning is provided for each analysis decision, plot, and statistical summary.
The analysis follows a logical flow, the discussion includes reasonings, explanations about the analysis and relevant statistics to quantify the results and insights.	
✓	Visualizations made in the project depict the data in an appropriate manner that allows plots to be readily interpreted.
Good job! All visualizations are properly labeled and titled.	

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)