UDACITY

Investigate a Dataset

REVIEW

HISTORY

Requires Changes

3 SPECIFICATIONS REQUIRE CHANGES

Amazing job so far!

Thank you for all the hard work and dedication you have put in this project. I have tried to help you with the missing specifications as best as i could. I have added detailed instructions on how to fix it and additional links if you feel like you need more information.

I'm so Proud of you and your Journey so far and I Hope you will like my review and will find it helpful in getting through the project!
Wish you Good Luck on your Journey Ahead, Stay Udacious and Happy Learning!

Code Functionality

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

The code cells produces no errors when run, both the html and the ipynb files are included.

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.

The project uses NumPy and Pandas appropriately

TIPS

For future reference here's some useful links on using Numpy:

- https://www.kaggle.com/paraspatidar/numpy-tricks-from-zero-to-hero
- https://www.w3schools.com/python/numpy_creating_arrays.asp
- https://towardsdatascience.com/hitchhiker-tips-on-effectively-using-python-numpy-arrays-66649b7bd5fb

And some other links on Pandas:

- https://towardsdatascience.com/10-python-pandas-tricks-to-make-data-analysis-more-enjoyable-cb8f55af8c30
- https://www.dataschool.io/python-pandas-tips-and-tricks/
- https://realpython.com/python-pandas-tricks/

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

REQUIRED:

The udacity specification requires that you use functions to avoid repetitive code.

TIPS

Here's some links that will help you out:

- https://www.programiz.com/python-programming/function
- https://www.w3schools.com/python/python_functions.asp
- https://www.tutorialspoint.com/python/python_functions.htm

Quality of Analysis

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

The project states the questions first then addresses it appropriately

We will be examining two primary research questions in this project:

1. Does receiving the SMS have a positive impact on people keeping their appointments? This is to see whether the SMS system is effective at getting patients to keep their appointments.
2. Is the appointment keeping rate same for different ailments of the patients? This is to understand whether there is a difference amongst the various ailments which leads to differences in appointment keeping rates. It is possible that some ailments may lead to patients not keeping their appointments more often and may need to be addressed.

Data Wrangling Phase

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

Well done, The changes in the data cleaning section is well documented!

TIPS:

I highly recommend that you use Markdown cells for documenting the changes. This will ensure that your audience are more focused and show how comfortable you are with your notebook.

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.

REQUIRED:

the udacity specification requires that you investigates both single-variable (1d) and multiple-variable (2d) explorations.

TIPS:

single-variable (1d) explorations are plots with only one variable used in it like histograms, bar plots, char plots, etc
multiple-variable (2d) explorations are are plots with more than one variable used in it like scatter plot,

heatmaps, boxplots.

kindly consider adding at least one plot with single variables.
here's a great guide on the difference between the Univariate, Bivariate:

Examine the differences between univariate and bivariate data.

| Univariate Data	Bivariate Data
• involving a **single variable**	• involving **two variables**
• does not deal with causes or relationships	• deals with causes or relationships
• the major purpose of univariate analysis is to describe	• the major purpose of bivariate analysis is to explain
• central tendency - mean, mode, median • dispersion - range, variance, max, min, quartiles, standard deviation. • frequency distributions • bar graph, histogram, pie chart, line graph, box-and-whisker plot	• analysis of two variables simultaneously • correlations • comparisons, relationships, causes, explanations • tables where one variable is contingent on the values of the other variable. • independent and dependent variables
Sample question: How many of the students in the freshman class are female?	**Sample question:** Is there a relationship between the number of females in Computer Programming and their scores in Mathematics?

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.

There are more than two kinds of visualizations present in the project

TIPS:

For future reference here's a really great guide to help you with graphs:
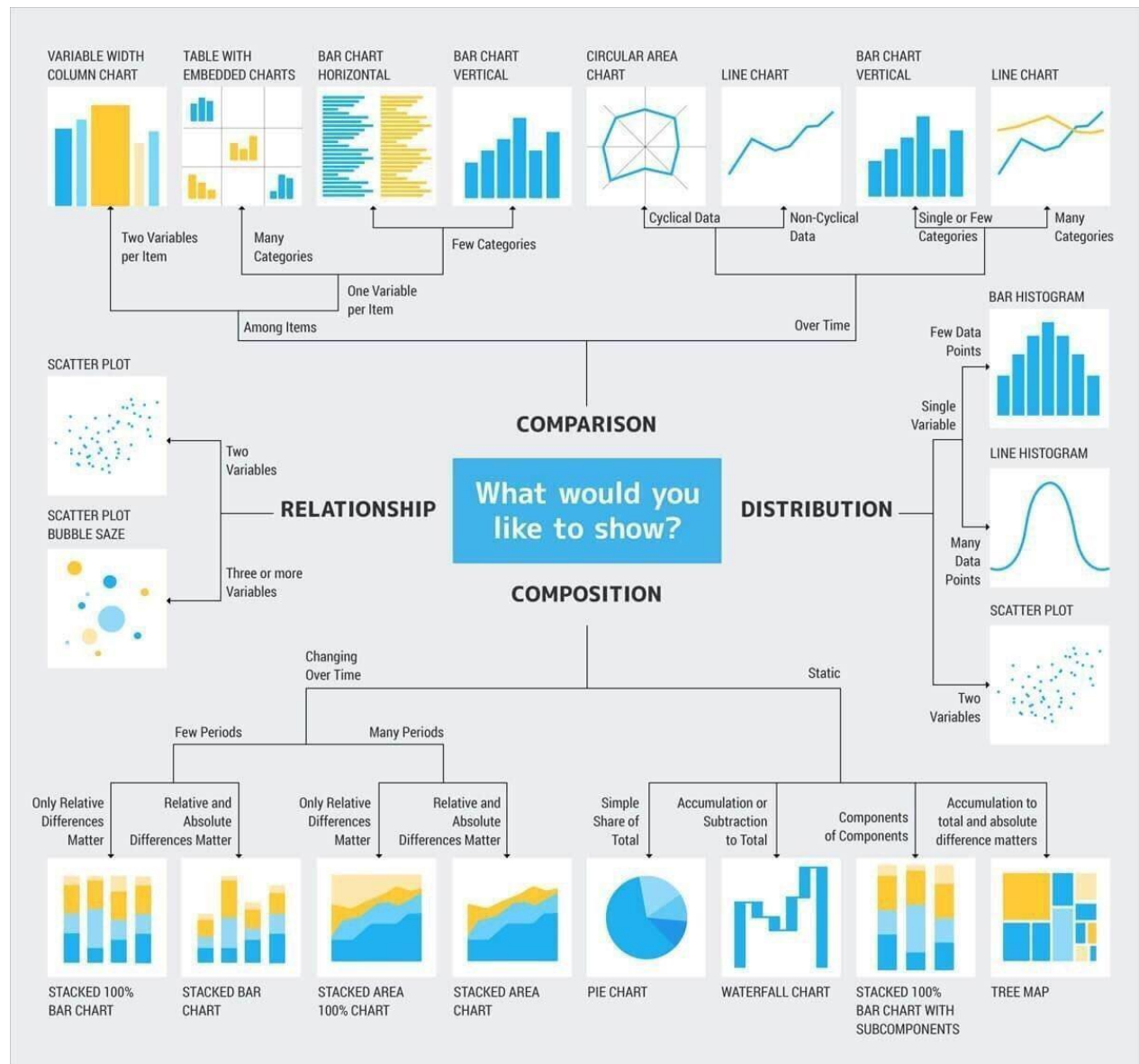
Conclusions Phase

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.

REQUIRED:

The udacity specification requires that you add limitation subsection to the conclusions section.

TIPS:

In the limitation subsection you can write about the restrictions and the limitations that you have faced during the implementation of your project. like the null and unwanted values.

here's the conclusions and the limitations section structure:

Conclusions

Finally, summarize your findings and the results that have been performed. Make sure that you are clear with regards to the limitations of your exploration. If you haven't done any statistical tests, do not imply any statistical conclusions. And make sure you avoid implying causation from correlation! This was a very interesting data analysis. We came out with resulting some facts about movies. After this analysis process we can take conclusion that:

Profitable movies have criteria

1. Average duration of the movie must be 114 minutes
2. Genre must be : Comedy, Action, Drama, Adventure, Thriller

Limitations: The analysis process was done which had a significant amount of profit of around 70 million dollar on various runtime, genres. As per our observation before the budget and revenue column do not have any currency unit or documentation representing what exactly currency mechanism used, it might be some possibility of different movies have budget in different currency unit. Moreover we are not sure if the data provided to us is complete corect and up-to-date. Dropping or discarding the rows with missing values in our case is zero value also affected the overall analysis. The result of analysis process might not be completely error free but by following these criteria can increase the probability of a movie to becoming profitable.

Communication

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

There is reasoning properly 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.

All plots are well made and easy to read!

☑ RESUBMIT

⬇ 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