UDACITY

< Return to Classroom

Investigate a Dataset

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

Very impressive work ! your project reflects your hard work and I have to congratulate you for that 😄 Your code is very solid as well, you only need some modifications order to continue. Good luck in your next submission !

Don't hesitate to reach your study hall mentor or use the slack channel in order to get help, we are here to help you succeed 🏆

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.

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.

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 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.

Hi Karine, you usually speak about your conclusions first and after that you point at the limitations. Please look at the following structure:

Conclusions

> **Results**: Our data suggest that
>
> 1. There is not big difference between the distribution of Age between patients who showed up for the appointment versus the patients that did not show up for the appointment.
> 2. There is a higher percentage of people that received an SMS and did not show up when compared to people who received an SMS and did show up.
> 3. People that have a disease are 3% more likely to show up for the appointment than people who do not have a disease.
> 4. Handicap patients specifically,however, are more likely to show up to the appointment compared to people who are not Handicap.
> 5. Being enrolled in the Scholarship program does not seem to make people more likely to show up to the appointment.

> **Limitations**: There are a couple of limitations with our data:
>
> 1. Most of our variables are categorical, which does not allow for a high level of statistical method that can be used to provide correlations etc
> 2. The statistics used here are descriptive statistics, not inferential, meaning that we did not create any hypotheses or controlled experiments or inferences with our data.
> 3. We do not have a lot of details for certain factors to draw conclusions. For the SMS_ received example, the data shows that no-showers are more likely to receive an SMS. This may seem counter intuitive, but we do not have information on the conditions of when the SMS is sent. For example they may target No-showers with SMS, or they may send the SMS once the Patient has not checked in 30 minutes prior to their appointment etc.
> 4. Cannot show strong correlations between factors since most of our data is categorical.

Just change that and you will meet expectations 😉

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 project