Investigate a Dataset

REVIEW | HISTORY

Meets Specifications

Greetings Student,

This was a good implementation and I congratulate you for passing all rubric items with this submission. It was delightful reviewing your work as it was well thought-out. I encourage you to keep up the good work as it will make you a great Data Analyst. Way to go! 👏🏽

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.

Excellent work using Pandas for this submission. Here are some important/regularly used pandas operation -

- Value-counts
- Boolean-Indexing
- Group-by
- Pandas dummies

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

Good work using descriptive names and comments in your code which makes it easier for other programmers to follow-up on the work.

Learning Notes

- Why to use functions?
- Why to use comments?

Quality of Analysis

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

Good job on stating questions beforehand. It helps set the tone of the project.

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 questions were thoroughly investigated from various angles, and both 1d and 2d explorations were used for several variables investigated.

Learning Notes

This link summarises the difference between bivariate and univariate 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.

Rate this review

Visualizing data requires a lot of patience and determination because it's not easy selecting the best visualization to match with a given data type. Well enough, the project rightly builds descriptive visualizations using a variety of plots.

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.

Good work presenting the results of the analysis while showing its limitations clearly.

Learning Notes

- A description of limitations typically identifies either a shortcoming of the dataset that has caused difficulty (e.g. missing data) or a shortcoming of the methods of analysis (e.g. a statistical approach which may not be ideal given the characteristics of the data set).

Communication

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

Great job on providing reasoning for every plots and analysis. Awesome.

Visualizations made in the project depict the data in an appropriate manner that allows plots to be readily interpreted.

Awesome! The plots are well labeled and easy to interpret.

DOWNLOAD PROJECT

RETURN TO PATH