Project: Investigate a Dataset

Code Functionality

Criteria	Submission Requirements
Does the code work?	• All code is functional and produces no errors when run. • The code given is sufficient to reproduce the results described.
Does the project use NumPy and Pandas appropriately?	• 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.
Does the project use good coding practices?	• The code makes use of at least 1 function to avoid repetitive code. • The code contains good comments and meaningful variable names, making it easy to read.

Quality of Analysis

Criteria	Submission Requirements
Is a question clearly posed?	The project clearly states one or more questions, then addresses those questions in the rest of the analysis.

Data Wrangling Phase

Criteria	Submission Requirements
Is the data cleaning well documented?	The project documents any changes that were made to clean the data, such as merging multiple files, handling missing values, etc.

Exploration Phase

Criteria	Submission Requirements
Is the data explored in many ways?	• The project investigates the stated question(s) from multiple angles. • The project explores at least three variables in relation to the primary question. This can be an exploratory relationship between three variables of interest, or looking at how two independent variables relate to a single dependent variable of interest. • The project performs both single-variable (1d) and multiple-variable (2d) explorations.
Are there a variety of relevant visualizations and statistical summaries?	• The project's visualizations are varied and show multiple comparisons and trends. • At least two kinds of plots should be created as part of the explorations. • Relevant statistics are computed throughout the analysis when an inference is made about the data.

Conclusions Phase

Criteria	Submission Requirements
Has the student correctly communicated tentativeness of findings?	• The Conclusions have reflected on the steps taken during the data exploration. • The Conclusions have summarized the main findings in relation to the question(s) provided at the beginning of the analysis accurately. • The project has pointed out where additional research can be done or where additional information could be useful. • The conclusion should have at least 1 limitation explained clearly. • The analysis does not state or imply that one change causes another based solely on a correlation.

Communication

Criteria	Submission Requirements
Is the flow of the analysis easy to follow?	• The code should have ideally the following sections: Introduction; Questions; Data Wrangling; Exploratory Data Analysis; Conclusions, Limitation. • Reasoning is provided for each analysis decision, plot, and statistical summary. • Interpretation of plots and application of statistical tests should be correct and without error. • Comments are used within the code cells. • Documented the flow of analysis in the mark-down cells.
Is the data visualized using appropriate plots and parameter choices?	Visualizations made in the project depict the data in an appropriate manner (i.e., has appropriate labels, scale, legends, and plot type) that allows plots to be readily interpreted.

Suggestions to Make Your Project Stand Out

- Use Markdown cells to report your findings.
- Utilize NumPy or Pandas functionality that goes beyond what was covered in the course.
- Use statistical tests to draw rigorous conclusions where appropriate.