Project Specification for Investigate a Dataset

|  |  |
| --- | --- |
| **CRITERIA** | **MEETS SPECIFICATIONS** |
| **Code functionality** |  |
| 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 functions to avoid repetitive code. The code contains good comments and variable names, making it easy to read. |
| **Quality of analysis** |  |
| 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** |  |
| 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** |  |
| Is the data explored in many ways? | 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. |
| Are there a variety of relevant visualizations and statistical summaries? | 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** |  |
| Has the student correctly communicated tentativeness of findings? | 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. |
| **Communication** |  |
| Is the flow of the analysis easy to follow? | Reasoning is provided for each analysis decision, plot, and statistical summary. |
| Is the data visualized using appropriate plots and parameter choices? | Visualizations made in the project depict the data in an appropriate manner 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.