**Meets Specifications**

**Congratulations Student....!!! 🎉**

* This was a great 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 really like the documentation part very well explained**
* I encourage you to keep up the good work as it will make you a great Data Analyst. Way to go! 

All the best for your upcoming projects..!!! 

**Code Functionality**

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

**Good Work..!!!** 👍

The code works well as it doesn't produce errors during the run. Also, it's sufficient to reproduce the results described.

**TIPS:  **

* It is always recommended that you handle your errors by segregating the erroneous block of codes into the singular ones run them line by line to pinpoint the main issue. This is frequently suggested and practised by top-notch coders.
* Jupyter notebook is a very powerful tool to document your codes and comments alongside.
* It helps you to segregate different blocks of code for better error handling along with suitable headings, comments and conclusions in different types of cells. This provides a focused approach and helps to establish a better connection with your audience.

You have truly developed this skill and the submission portrays it clearly..!! Well done 

* **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.**

**Nice work in using Pandas library to facilitate the work for this submission!!!👍🎉**

**TIPS and SUGGESTIONS:**

* Pandas is a very handy and powerful python library for handling data frames and various tedious tasks as hand.
  + [Link1](https://towardsdatascience.com/10-python-pandas-tricks-that-make-your-work-more-efficient-2e8e483808ba)
  + [Link2](https://realpython.com/python-pandas-tricks/)
* Here's are two links on a number of tips and tricks which we can use when using pandas....!! I encourage you to check it out in your free time! 

**Learning Notes **

Some important Pandas built-in functions:

* [Value-Counts](https://towardsdatascience.com/getting-more-value-from-the-pandas-value-counts-aa17230907a6)
* [Indexing and Selecting data](https://pandas.pydata.org/pandas-docs/version/0.15/indexing.html)
* [Apply, Map](https://medium.com/@evelynli_30748/map-apply-applymap-with-the-lambda-function-5e83028be759)
* [Group-by](https://realpython.com/pandas-groupby/)

**It is really interesting to see that the submission includes above-mentioned practices...Good Job!!**👍

* **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.**

**The project nicely avoids repetitive blocks of code by using pre-defined and user-define functions in the submission!!**

* Comments and appropriate variable names are essential for a good coder.
* These not only guide the viewer through the code but also helps in understanding it easily. You have portrayed these skills well... Keep up this good work in future too...

**Quality of Analysis**

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

**The submission does a nice job in stating and addressing multiple insightful questions in your analysis..!!!**👍🎉

All the questions are relevant and have been addressed to in the analysis and relevant visualisations have been framed.... 

**Data Wrangling Phase**

* **The project documents the steps that were taken to clean the data, such as merging multiple files, handling missing values, etc.**

**The submission contains a separate section of data wrangling and proper steps have been taken to identify missing values, duplicates, datatype issue or non - important columns and resolving them relevantly. Good Job...!!** 

**Suggestions and comments:**

Data Wrangling is aimed at cleaning the data and also transforming it into a state which can be easily analyzed. Note that, having uncleaned data could invalidate the analysis or provide inaccurate results. These are a few steps to take before analysis.

* Identify missing values in the dataset
* Decide what to do with missing values
* Identify fields which are relevant to the analysis and eliminate any fields that will not be useful in the analyses.
* Identify data fields which do not have proper data types and decide better data types for these columns.
* Make sure to check the data before and after the data wrangling is applied to make sure any changes have been done.

**Good work in looking into all the above points!!**

**Some Helpful documentation and blogs:**

* [Pandas.isnull](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.isnull.html)
* [Pandas.dataframe.info](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.info.html)
* [Dropna function](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.dropna.html) to drop any rows with missing values
* [Fillna function](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.fillna.html) to fill missing values
* [pandas.DataFrame.drop](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.drop.html) to drop whole column
* [Handling missing values in dataset](https://towardsdatascience.com/the-tale-of-missing-values-in-python-c96beb0e8a9d)

**Exploration Phase**

* **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.**

**The questions were thoroughly investigated from various angles, and both 1d and 2d explorations were used for several variables investigated..!! Well done..!! **

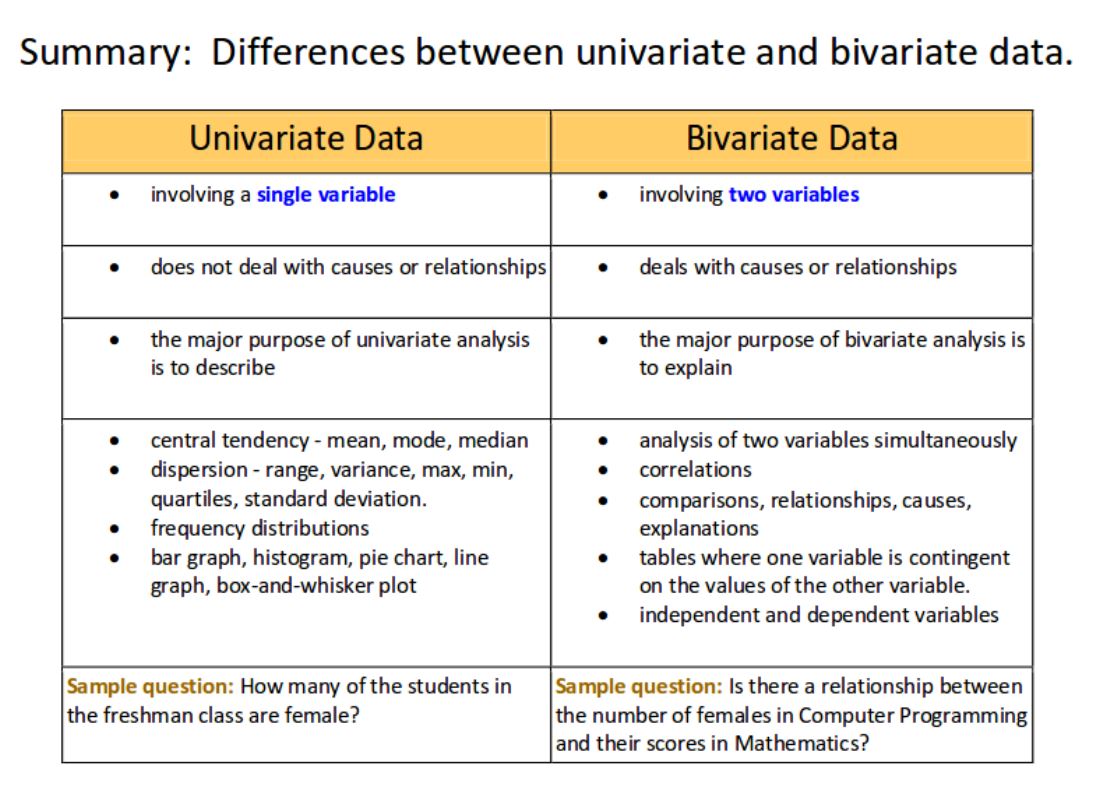
**COMMENTS: **

**Exploratory Data Analysis (EDA) is an approach for data analysis that employs a variety of techniques (mostly graphical) to maximize insight into a data set. The graphical techniques employed in EDA are often quite simple, consisting of various techniques of:**

* 1) Plotting the raw data such as histograms, histograms, probability plots, lag plots, block plots, scatter plots.
* 2) Plotting simple statistics such as mean plots, standard deviation plots, box plots, and main effects plots of the raw data.

**Remember:**

* What is very important when you analyze data is to stay focused on your questions. Build plots or statistical summaries which answer your questions, and not just because they are nice.

**Below are the key differences between univariate and bivariate analysis:**  
[](https://udacity-reviews-uploads.s3.us-west-2.amazonaws.com/_attachments/251099/1578135833/uni_vs_bivariate.JPG)

* **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.**

GOOD JOB!!! 

**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. The project rightly builds descriptive visualizations using multiple types of plots...!!**

**COMMENTS**

* Data visualization is the presentation of data in a pictorial or graphical format. It enables decision-makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns.
* Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

**In general data visualization can also**:  
• Identify areas that need attention or improvement.  
• Clarify which factors influence certain trends.  
• Help you understand which products to place where.  
• Make predictions.

<https://seaborn.pydata.org/tutorial/categorical.html>

**Conclusions Phase**

* **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.**

**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**

* **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.**

**Well done..  **

* The Documentation is very well done in **Note and conclusion.py** FIle.
* You have done a great job describing every analysis decision, and plot stating the results obtained and the limitations of that analysis...!!!
* **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.**

**The analysis and visualizations throughout the report are well drafted. 🎉**  
Awesome! The plots are well labelled with appropriate comments and easy to interpret. 