Problem Statement:

The purpose of this assignment is to assess your understanding of the fundamental concepts in data science and to test your ability to explore, analyze, and visualize a dataset. By completing this assignment, you will demonstrate your skills in data cleaning, data manipulation, and data visualization, as well as in basic statistical analysis.

Dataset:

You will be using the "Students Performance in Exams" dataset, available on Kaggle: https://www.kaggle.com/spscientist/students-performance-in-exams

Tasks:

1. Download the dataset and create an IPython Notebook or a Google Colab notebook for this assignment.

2. Load the dataset into a pandas dataframe and display the first 10 rows of the dataframe to gain familiarity with the data structure.

3. Perform exploratory data analysis:
- Check for missing values and handle them accordingly (if any). -
Check for outliers and handle them accordingly (if any).
- Provide a brief summary of your data, such as descriptive statistics and data types.

4. Perform data visualization:
- Create a bar plot to show the distribution of male and female students in the dataset.
- Create a histogram to show the distribution of the scores in math, reading, and writing.
- Create a bar plot to show the average scores in math, reading, and writing for each gender.

- Create a scatter plot to show the relationship between math, reading, and writing scores.

5. Perform basic statistical analysis:
- Calculate mean, median, and mode for math, reading, and writing scores.
- Calculate the standard deviation for math, reading, and writing scores.
- Test if there's a significant difference between male and female students' performance in math, reading, and writing using an appropriate statistical test (e.g., t-test).

6. Document your findings:
- Provide explanations and interpretations for each plot and observation. - Clearly state any conclusions you can draw from your analysis.

7. Save your work:
- If using an IPython Notebook, export your work in .ipynb and .pdf formats. - If using Google Colab, save your work in .ipynb format and download it to your local machine, then convert it into .pdf format.

Evaluation Criteria:

Your assignment will be evaluated based on the following criteria:

1. Completeness - You have answered all parts of the assignment and have submitted both .ipynb and .pdf files.

2. Code Quality - Your code is well-written, properly documented, and easy to understand.

3. Analysis and Interpretation - Your plots, analysis, and interpretation are clear, concise, and backed up with supporting evidence.

4. Creativity - You have used various visualization techniques and demonstrated a good sense of data exploration and storytelling.

Submission:

Submit the final .ipynb and .pdf files through the specified submission platform or email them to the specified email address.

Resources:

1. Python Pandas Documentation - https://pandas.pydata.org/pandas-docs/stable/
2. Matplotlib Documentation - https://matplotlib.org/documentation/index.html
3. Seaborn Documentation - https://seaborn.pydata.org/documentation.html
4. NumPy Documentation - https://numpy.org/documentation/
5. SciPy Documentation - https://docs.scipy.org/doc/scipy/reference/index.html

Submission Link: https://forms.gle/FtU2ovnxPSCFWBv36

Deadline: 30th June, 11:59 PM.

Note:

Please make sure to comment your code and explain your thought process in the Jupyter notebook or Python script. Additionally, you may be asked to explain your reasoning behind your data cleaning steps during an interview.

MasterCourse holds the right to disqualify an assignment submission if there are some moral and ethical issues like plagiarism, hate speech, etc. are found.