Dr. Justin B. Keeler
BSAN 710

**Assignment #2 – Python Programming**

Assignment #2: Expanding the Analysis of the Titanic Dataset Using Python (No ChatGPT)

In this assignment, you will use Python to analyze the Titanic dataset, which contains information about the passengers who were aboard the Titanic when it sank. You will use dictionaries, Pandas, and visualization using Matplotlib to explore and analyze this dataset.

Dataset Description: The Titanic dataset contains information about 891 passengers, including their age, gender, ticket class, and whether they survived or not. The dataset is available in CSV format and can be downloaded from the Kaggle website.

Tasks:

1. Load the dataset into a Pandas DataFrame.

2. Explore the dataset using Pandas to answer the following questions:
    - How many passengers survived?
    - What was the average age of the passengers?
    - What was the survival rate of male and female passengers?
    - Which passenger class had the highest survival rate?
    - How many passengers had siblings/spouses aboard?
    - How many passengers had parents/children aboard?

3. Create a dictionary to store the data for each passenger class (1st, 2nd, and 3rd) and their respective survival rates.

4. Use Matplotlib to create a bar chart showing the survival rates of each passenger class.

5. Create a scatter plot showing the relationship between passenger age and fare paid. Color code the plot based on survival status (survived or did not survive).

6. Create a histogram of the passenger age distribution.

7. Create a pie chart showing the proportion of passengers who survived and did not survive.

8. Create a box plot showing the distribution of passenger fares by passenger class.

Deliverables:

1. A Jupyter notebook containing your Python code.

2. A brief report (no more than 4 pages double-spaced, inclusive of a cover page) summarizing your findings, including your answers to the questions above and the insights you gained from the visualizations.

**Assignment #2 – Python Programming**

Hints:

1. The Pandas `read_csv` function can be used to load the CSV file into a DataFrame.
2. The Pandas `describe` function can be used to get summary statistics for the numerical variables.
3. The Pandas `groupby` function can be used to group the data by passenger class, gender, or any other variable.
4. The Matplotlib `bar` function can be used to create bar charts, the `scatter` function can be used to create scatter plots, `hist` function can be used to create histograms, `pie` function can be used to create pie charts and `boxplot` function can be used to create box plots.

**Disclaimer: Usage of ChatGPT is Strictly Forbidden**

Attention students: Please note that the use of ChatGPT or any similar AI language model is strictly forbidden and disallowed for Version 1 of the "Designing and Solving Real-World Optimization Problems using Linear and Integer Programming" assignment. Any team found to be in violation of this rule will receive a score of zero for the entire assignment, and all team members will be referred to the Office of Student Conduct and Community Standards (https://www.wichita.edu/about/student_conduct/) for further investigation and potential disciplinary action in accordance with the Academic Integrity Policy.

By choosing to work on this assignment, you and your team members are committing to complete the assignment without the use of ChatGPT or any similar AI assistance. It is the responsibility of each team member to ensure that no such tools are used during the completion of the assignment.

Failure to comply with this disclaimer may have serious consequences for your academic standing and future prospects within the institution. Please make sure to abide by the guidelines and maintain the highest standards of academic integrity.

**Assignment #2 – Python Programming**

Assessment Criteria (Rubric):

Work Quality:

- Excellent (10 points): The code is well-organized, clearly documented, and easy to read. All deliverables are met, and the insights gained from the data analysis are insightful and well-presented in the report.
- Good (7 points): The code is mostly well-organized, documented, and easy to read. Most of the deliverables are met, and the insights gained from the data analysis are presented adequately in the report.
- Fair (4 points): The code is poorly organized, documented, and difficult to read. Some of the deliverables are not met, and the insights gained from the data analysis are presented inadequately in the report.

Task Completion:

- Excellent (30 points): All tasks are completed with a high level of accuracy, and the code is written efficiently and effectively. All deliverables are met, and the insights gained from the data analysis are comprehensive, detailed, and accurate.
- Good (20 points): All tasks are mostly completed accurately, but some code may be redundant or inefficient. Most of the deliverables are met, and the insights gained from the data analysis are presented adequately.
- Fair (10 points): Some tasks are not completed accurately or not completed at all, and the code may be very inefficient. Some of the deliverables are not met, and the insights gained from the data analysis are presented inadequately.

Visualization:

- Excellent (20 points): The visualizations are well-designed, accurately represent the data, and are effectively integrated into the report. The visualizations provide valuable insights into the data.
- Good (14 points): The visualizations are mostly well-designed and accurately represent the data, but they may be lacking in some aspects, such as labeling or color coding. The visualizations are adequately integrated into the report and provide some insights into the data.
- Fair (8 points): The visualizations are poorly designed, do not accurately represent the data, and are not effectively integrated into the report. The visualizations do not provide valuable insights into the data.

Overall Assessment:

- Excellent (50-60 points): The code is well-written, well-organized, and efficient. All tasks are completed accurately, and all deliverables are met. The report is well-written, comprehensive, and insightful.

**Assignment #2 – Python Programming**

- Good (35-49 points): The code is mostly well-written, and most tasks are completed accurately. The report is adequately written, but some aspects may be lacking in detail or insight.
- Fair (20-34 points): The code is poorly written, and many tasks are not completed accurately. The report is poorly written, and the insights gained from the data analysis are presented inadequately.