Python for Data Analysis Assignment

Problem Statement:
This assignment focuses on utilizing DataFrame operations such as loc, iloc, indexing, slicing, aggregations, groupby, orderby, etc., to analyze the Titanic dataset. By exploring the dataset and employing these operations, students are expected to extract meaningful insights and perform various data manipulations to derive valuable information.

Guidelines:
1. Foundational Knowledge:
   - Understand the fundamentals of DataFrame operations in Python using libraries like Pandas.
   - Familiarize yourself with loc and iloc methods for data selection, as well as other DataFrame manipulation techniques.
2. Data Exploration:
   - Analyze the Titanic dataset's structure and characteristics using descriptive statistics, visualizations, and exploratory techniques.
   - Gain insights into different attributes and their distributions to guide subsequent operations.
3. Data Manipulation:
   - Utilize loc and iloc methods for indexing and slicing operations to extract subsets of data based on specific criteria.
   - Perform aggregations to compute summary statistics and derive insights from the dataset.
   - Apply groupby operations to group data based on one or more categorical variables.
   - Sort the DataFrame using orderby operations to arrange data in ascending or descending order based on specified columns.
4. Task Execution:
   - Execute the provided tasks using DataFrame operations and document your approach.
   - Validate your results and ensure the correctness of operations performed.

Step-by-Step Approach to Perform DataFrame Operations:
1. Setup and Data Loading:
   - Import necessary libraries: pandas, numpy.
   - Load the Titanic dataset into a DataFrame.
2. Data Exploration:
   - Explore the structure and contents of the DataFrame using functions like info(), describe(), head(), and tail().
   - Visualize distributions of relevant variables using histograms, box plots, or scatter plots.
3. Indexing and Slicing:
   - Use loc and iloc methods to perform indexing and slicing operations to extract specific subsets of data based on conditions or positions.
4. Aggregation Operations:
   - Compute summary statistics like mean, median, count, etc., for numerical variables.
5. Groupby Operations:
   - Group the data based on categorical variables and perform aggregations within each group.

6. Orderby Operations:
   - Sort the DataFrame based on specific columns in ascending or descending order.
7. Documentation and Interpretation:
   - Document the steps followed for each operation and provide interpretations of the results obtained.

Assignment Questions:
1. Indexing: Retrieve the details of the passenger in the 10th row.
2. Slicing: Extract the information of passengers from the 5th to the 15th row.
3. loc: Retrieve the age and sex of passengers with PassengerId 100 to 150.
4. iloc: Extract the details of passengers from the 3rd to the 6th row and the 2nd to the 4th columns.
5. Aggregation: Calculate the average fare paid by passengers.
6. Groupby: Group the data by 'Pclass' and calculate the median age for each passenger class.
7. Orderby: Sort the DataFrame by 'Fare' in descending order.
8. Indexing: Retrieve the name and ticket number of passengers who survived.
9. Slicing: Extract the subset of passengers who embarked from 'C' and paid a fare greater than $50.
10. Groupby: Group the data by 'Embarked' and 'Sex', and calculate the average fare for each combination of embarkation point and gender.

Link to Dataset for the Assignment:
- Titanic Dataset: [https://www.kaggle.com/c/titanic/data]