Pandas DataFrame Assignment – 3

Note: Use titanic dataset to answer the below questions

Dataset Link:

- https://www.kaggle.com/datasets/brendan45774/test-file

1. Display the names of all passengers who paid a fare greater than $100 and were under the age of 30.

2. Find the top 5 oldest passengers. What are their names, ages, and whether they survived?

3. Count how many passengers have a missing value for their cabin number (Cabin).

4. Extract and create a new column called 'Title' from the 'Name' column (e.g., Mr., Mrs., Miss, etc.). After creating the column, display the first 5 unique titles.

5. Identify all passengers whose age is missing (NaN). Replace the missing age values with the average age of all passengers in the dataset.

6. Sort the dataset based on fare in descending order and display the top 10 passengers with the highest fares.

7. Find the name of the youngest passenger who did not survive.

8. Display the first 5 rows where the passenger's name starts with the letter 'A'.

9. Create a new column called ‘FamilySize’, which is the sum of 'SibSp' and 'Parch'. Then, display the first 10 rows where the family size is greater than 3.

10. Create a new column ‘FarePerPerson’ by dividing the 'Fare' by the sum of 'SibSp' + 'Parch' + 1 (to account for the passenger themselves). Then, find the highest ‘FarePerPerson’ and display the name and 0fare of that passenger.