Homework 3

Jupyter Notebooks and Data Analytics in Python

In this assignment you do some exploratory data analysis in Jupyter Notebook on the titanic dataset. The titanic dataset holds information on the passengers on the Titanic and whether they survived or perished. It includes the following:

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

Set up

1. Launch Jupyter Notebook and create a new Python 3 file.
2. Download titanic.csv and store in the root folder of Jupyter Notebook.

Data Analysis

1. Read titanic.csv into a Pandas dataframe. Display the first five rows. Display the data types of each column.
2. Calculate and display the mean and the median of the Age column and the fare column.

3. Calculate and display the mean and median age of passengers by survival. In other words, what is the mean age of passengers who survived and what is the mean age of passengers that perished.
4. Calculate and display the mean and median fare paid by passengers by survival.

Data Visualization

1. Create a scatter plot. Put Age on the X-axis and fare on the Y-axis.
2. Create another chart of your choice that you think would be interesting. You can look at the matplotlib documentation to learn about other charts or google matplotlib. There are many tutorials and articles on chart building using matplotlib.

Submission

You should submit the hw3.ipynb file to Blackboard.

Grading

100 points – Titanic notebook

The Titanic notebook is correctly implemented.

Your name: Add your name to the top of the .ipynb file.

Compilation: If your code does not compile, you will receive a 0 on the assignment. Once notified, you will have 3 days to fix the code to receive partial credit, up to 75%.

Late Policy: There is a 3 point deduction for each day the assignment is late. You do not have to ask for an extension.