

Assignment 2-B

An Introduction to Pandas

What is Pandas?

Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Pandas Getting Started:

<http://pandas.pydata.org/pandas-docs/version/0.15/tutorials.html>

Objectives

There are nine tasks to be completed for this assignment. You will be working on the training file of the infamous [Kaggle Titanic Dataset](#) provided in the Data folder. Your outputs must resemble the outputs given in the sample_solution folder.

1. Find the survival rate of passengers from different passenger classes
2. Find the details of the passengers who paid the most money
3. Find the number of missing values in each column
4. Replace the missing values of age with the median age of the passengers
5. Select all the passengers above the age of 60 who survived and order them in ascending order of PassengerId
6. Find the average cost and standard deviation of the ticket price based on Passenger Class and Embark Location
7. Extract the titles of all the passengers (Mr, Mrs, etc)
8. Find the occurrence of the different titles and mean survival rate based on the titles and merge it with the original dataset
9. List all the unique titles of the passengers

Outcome of Assignment

You would have learnt to

- Read and write csv files using Pandas
- Manipulate Pandas DataFrames
- Perform simple calculations on DataFrames
- Use Groupby and Merge
- Perform some basic string manipulation
- Use some handy Pandas Functions