

Laboratory Activities for Week 1: Basic Pandas

SC310005 Artificial Intelligence
Khon Kaen Business School

(10 Points) Exploring Titanic Dataset with Pandas

Introduction:

The Titanic dataset from Kaggle contains information about passengers aboard the Titanic. In this exercise, we'll utilize Pandas, a powerful data manipulation library in Python, to explore and derive insights from this dataset.

Objective:

The primary objective of this exercise is to familiarize ourselves with fundamental data analysis using Pandas by addressing various questions related to the Titanic dataset.

Dataset Description:

The dataset includes passenger information (e.g., age, sex, class), survival status, fare, and more.

Data:

https://raw.githubusercontent.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2023s1/main/dataset/titanic_dataset.csv



Q1	What is the total number of passengers in the dataset?
Q2	How many passengers survived, and how many did not survive?
Q3	What was the average age of passengers on board?
Q4	How many male and female passengers were aboard the Titanic?
Q5	What was the survival percentage of passengers by gender?
Q6	What is the distribution of passenger classes (Pclass) on the Titanic?
Q7	What was the average fare paid by passengers in each class?
Q8	What was the survival rate among passengers in each class?
Q9	What was the percentage of passengers with siblings or spouses aboard?
Q10	What was the average age of passengers who survived and those who did not?

Appendix:

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