



Assignment 1 Comp7037

Final Submission Due : Week 7 - 8th March

Submission location : Canvas

Submission Details : Single Document – PDF – which includes a screen shot of each query and its output, with appropriate titles

Worth : 35% of final Course Mark

This assignment work as it progresses will be reviewed in the lab with the student, please ensure that as you complete each of part of the assignment 1,2 and 3 you demonstrate your work to your lecturer in the lab.

The student's work will be reviewed in the lab and the student must be able to explain logic and implementation of all work submitted. No mark will be given to student for work which has not been reviewed with the lecturer in the lab.



The Data Set

Titanic Data Set – csv and json – on canvas

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

Variable Notes

pclass: A proxy for socio-economic status (SES)

1st = Upper

2nd = Middle

3rd = Lower

age: Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

sibsp: The dataset defines family relations in this way...

Sibling = brother, sister, stepbrother, stepsister

Spouse = husband, wife (mistresses and fiancés were ignored)

parch: The dataset defines family relations in this way...

Parent = mother, father

Child = daughter, son, stepdaughter, stepson

Some children travelled only with a nanny, therefore parch=0 for them.



Part 1 : Create mysql, neo4j, Mongodb database using the titanic data set – marks will be given for design and set up of the database. (15%)

Part 2 : Do the following queries on each of the three databases created in Part 1 (15%)

- List alphabetically the names of all female passengers that survived
- Output by class, the names of children who survived
- Output by class, the names of children who did not survive
- Count of children in each class that survived
- Count of children in each class who did not survive

Part 3 : In the lab review of your work explain to your lecturer which database was more effective for each query (5%)

Note : Standard CIT penalties apply for plagiarism and late submissions.