

## Week 4 – Cleaning Data in SQL

### Peer-Graded Assignment

#### Preparation Instructions:

- 1) Download the Titanic.csv file and accompanying instruction files
- 2) Create a table in PostgreSQL based on the table structure provided in the Titanic.csv file. Name the table as 'titanic'.
- 3) Import the Titanic dataset into the newly created table

#### Required Tasks:

- 1) Handle the missing data under columns "Age" and "Cabin" by replacing the appropriate values
- 2) Standardize the values found under column "Sex"
- 3) Create a column "Last\_Name" and populate with values derived from "Name" column
- 4) Create a column "Title" and populate with values derived from "Name" column
- 5) Create a column "First\_Name" and populate with values derived from "Name" column (do not remove values in parenthesis)

#### Submission Instructions:

- 1) Save your script for each task in text file with the following file names:
  - a. data\_cleaning\_task1.txt
  - b. data\_cleaning\_task2.txt
  - c. data\_cleaning\_task3.txt
  - d. data\_cleaning\_task4.txt
  - e. data\_cleaning\_task5.txt
- 2) Export your final cleaned table in CSV file with the following filename:  
Titanic\_cleaned.csv
- 3) Submit all your output files

#### IMPORTANT NOTE:

- You may or may not choose to use temporary tables to arrive at your final output or just simply use nested queries for the same result
- Whatever the case may be, please include all necessary steps and queries in the text file arranged in the correct order of execution.
- Keep in mind that this is a peer-graded assignment and that it is important for the peer-reviewer to be able to follow the steps you have performed to validate your test results
- If you are the reviewer, perform the queries in your own database to validate your peer's submission

**Rubric for Scoring Peer-Graded Assignment:**

For each task of the 5 assigned task, please follow the scoring rubrics below:

Points	Measure
3	Was able to use the appropriate scripts/command as taught in the Data Cleaning lesson to perform the required tasks. The resulting data was perfectly cleaned.
2	Was able to use the appropriate scripts/command as taught in the Data Cleaning lesson to perform the required tasks. However, there were some values left uncanceled.
1	Was able to use any of the data cleaning scripts/commands demonstrated in the lesson but there were some errors in the script themselves that needed correction for them to work. (Note that errors do not include those that are caused by differences in table or field name. The errors referred to here are syntax or logical errors.)
0	Scripts used were not appropriate for the task and the data was not cleaned

Maximum score:  $3 \times 5 = 15$