Week 5 – Feature Engineering

Peer-Graded Assignment

Preparation Instructions:

1) Download the property_sales.csv file and accompanying instruction files
2) Create a table in PostgreSQL based on the table structure provided in the property_sales.csv file. Kindly use the 'property_sales' as the table name for the raw data.
3) Import the Titanic dataset into the newly created table

Required Tasks:

1) Perform One-Hot Encoding on the variable on one or more Categorical Variables in the dataset
2) Perform Ordinal or Label encoding on one or more Categorical Variables in the dataset
3) Perform Mean encoding on one or more Categorical Variables in the dataset (hint: you may want to use OVER() and PARTITION() commands)
4) Perform Mean Normalization on all the numeric variables to rescale these variables (you may add new columns for this)
5) Perform Standardization on all the numeric variables to rescale these variables (you may add new columns for this)

Submission Instructions:

1) Save your script for each task in text file with the following file names:
    a. feature_engineering_task1.txt
    b. feature_engineering _task2.txt
    c. feature_engineering _task3.txt
    d. feature_engineering _task4.txt
    e. feature_engineering _task5.txt
2) Export your final cleaned table in CSV file with the following filename: property_data_newfeatures.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 maybe, 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 Feature Engineering lesson to perform the required tasks; all necessary steps were performed using SQL scripts and the variables were correctly encoded or rescaled
2	Was able to use the appropriate scripts/command as taught in the Feature Engineering lesson to perform the required tasks and the variables were correctly encoded or rescaled with either of the following deficiencies: ● However, some of the steps were performed and/or computed manually, instead of using SQL scripts. ● The variable chosen was not appropriate for the Feature Engineering Technique
1	Was able to use the appropriate scripts/command as taught in the Feature Engineering lesson to perform the required tasks with either of the following deficiencies: ● However, the variables were not encoded or rescaled properly ● There were some errors in the script that requires 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 variables were not properly encoded or scaled.

Maximums score: 3 x 5 = 15