edX   **Microsoft: DAT203x Data Science and Machine Learning Essentials**

▸ Before You
   Start

▸ Module 1:
   Introduction
   and Data
   Science Theory

▸ Module 2:
   Working with
   Data

▸ Module 3:
   Visualization,
   and Building
   and Evaluating
   Models

▸ Module 4:
   Regression,
   Classification,
   and
   Unsupervised
   Learning

▸ Module 5:
   Recommenders
   and Publishing
   Your Work

▾ **Final Exam**

Exam
Final Exam due Oct 30,
2015 at 00:00 UTC ✎

**Post-Course
Survey**

QUESTION 9 (1/1 point)

You are creating an experiment based on real estate transactions data in which each sale of a property is indicated by a unique transaction ID field. The data is entered into the source database manually, and some records may have been entered more than once. You plan to use the data to train a regression model that predicts property prices.

What should you do to reduce bias in your data that might affect the predictive model?

○ Use only the first 50% of the rows to train, score, and evaluate the model.

○ Exclude the transaction ID column from the training set for the model.

⊙ Remove rows containing duplicate transaction IDs, retaining only the first instance.   ✔

○ Scale the transaction ID column to normalize the data.

You have used 1 of 1 submissions

POWERED BY
OPEN edX