

DATA WRANGLING AND DATA CLEANING MODULE EVALUATION

INSTRUCTIONS

- This test is divided into two parts: theory and practical
- It has 20 marks in total
- Marks you score will be converted into a percentage
- Answers for the theory assessment should be typed in word
- To be fair to those who haven't been able to install MySQL yet, I only need a written code in MS Word (For the MySQL evaluation).
- Those with MySQL installed should practice the queries in MySQL workbench but submit the codes in word.
- For the Excel assessment, you will attach the final cleaned data
- In summary, submit your theory and MySQL work as a word document and attach the final cleaned Excel data set

DATA CLEANING AND DATA WRANGLING ASSIGNMENT

1. THEORY

1. Briefly explain the six steps involved in data wrangling (6 marks)
2. Why is data wrangling important? (3 marks)

PRACTICAL

1. MySQL

1. Write a query which retrieves all the details of customers whose points are below 500 (1 mark)
2. Write a query which retrieves details of customers from the state of VA only (1 mark)
3. Write a query which updates salary for employee with employee id 37270 to 100000 (1 mark)
4. Write a query which removes duplicate values from the customer's table (1 mark)
5. Write a query which deletes from employees table where employee id is 63196 (1 mark)

2. Excel

Download the excel dataset attached and perform the following operations on the Excel data. Once you are done, save and attach the new dataset here.

1. Remove all the duplicates from this data (1 mark)
2. Add a validation criterion to the points column so that customers can only be awarded a maximum of 400. Add this warning to your criteria: "A customer can only have a maximum of 400 points". (3 marks)
3. Freeze the first row of the dataset (1 mark)
4. Sort the points column from smallest to largest (1 mark)

