**L3DT – Pressure test**

Aims & Objectives

This pressure test has been developed to help you prepare for your upcoming Scenario Demonstration as part of your L3 Data Technician apprenticeship assessment.

Whilst every effort has been made to replicate what might be expected at the assessment it should be considered that what you experience through this pressure test will be different from the actual assessment set by the EPA. The aim of this pressure test is to simulate the conditions of the Scenario Demonstration with regards to following instructions, time restrictions and ability to work unaided (either through instructors or work colleagues).

The test should be considered in two parts. Firstly, appending of the four data sets into one master data set with the associated validation/verification processes. The second part will aim at producing visualisations and summary statistics to understand the characteristics of the variables within the data set.

All the above will be carried out using any tool you feel most comfortable with, although it is recommended that these tasks can be completed simply in MS Excel. The data sets will be provided to you in the form of four csv. files. The pressure test will last for 60 minutes and there will be no conferring during the test – the test will be invigilated so it will be “cameras on” throughout.

Once the test is complete your MS Excel workbook(s) are to be sent via email to the invigilator. They will then be assessed according to criteria which has been mapped out in preparation. This will then provide a guide on how well you have done and any areas for practice and support.

The Pressure Test

You will have received four csv files entitled Data2019, Data2020, Data2021 & Data2022. Please check that you can open these files on your computer prior to the test – if there are any problems then please raise this as soon as possible and an alternative will be found.

When the test is started, you are required to:

Assume you have been tasked by a data scientist in your organisation (a market leading wine producing company) to provide a clean and complete data set for further analysis. The data scientist has provided instruction on what data variables are required and that they would like to have four years’ worth of data. Assume that you have identified that they are available but across four sets of yearly data and extracted them. The data scientist would like you to:

Produce one master workbook which has a master worksheet organised, formatted and complete. Remember to validate and verify the data set.

The data scientist would also like to see some summary statistics of the data to describe the characteristics of the variables and visualisations should also be used (boxplots to identify outliers, or histograms for distribution for example). They are particularly interested in the Residual Sugar, Alcohol and Sulphate content.

The data scientist is also interested in the number of units produced per month over the four years and the ratio of quality across the data set. Please consider appropriate visualisations to demonstrate insight.

When the 60 minutes has elapsed, you will be asked to save the master workbook in the correct format (Excel workbook or equivalent) and send with all relevant files (i.e., any other files you have used for working out) via email for assessment.

Each submission will be assessed according to criteria already prepared over the next couple of days and then a follow up session will be arranged individually (if required).