OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Applied Customer Analytics
MS5108 2022-2023
Individual Assignment 2

Objective	The objective of this assignment is to access your understanding of the basics of R programming language.
Lecturer	**Name**　　　　　　　　**Office**　　　　　**E-mail** Dr Umair ul Hassan　　　CA369　　　umair.ulhassan@universityofgalway.ie
Marks Awarded	This assignment carries **20%** of the overall marks for the module.
Submission Process	You should submit your completed assignment through Blackboard's Assignment tool. If you are unable to submit your assignment via Blackboard, please email it to your lecturer whilst also copying business@universityofgalway.ie in the same email and clearly state the issue that you have faced while uploading to Blackboard.
Assignment Deadline(s)	**9th March 2023 at 5:00pm Irish time:** 　• R code file 　• Report (max 5 pages) To avoid technical issues, it is strongly advised that you **upload your submission well in advance of the deadline**. You may submit at any time on any day prior to the deadline. You will have multiple attempts to submit, and last submitted version will be graded.
Late or No Submission	Blackboard will record late submissions. Except in extenuating circumstances, late submissions will carry a penalty up to 24 hrs after the deadline, after which time any submissions will not be marked. Non submission of assignment will carry a mark of zero in determination of overall marks for this sitting. There is no opportunity to resubmit continuous assessment before the next offering of the module, should student(s) fail to submit by specified deadlines.
Deliverables	The **R code** file, as "*MS5108_Assignment2_YOURNAME_code.R*" file. Include your student name, ID, and module code in the start of your code. A **report** of results that discusses you visualizations and outputs, as "*MS5108_Assignment2_YOURNAME_report.docx*" file. Include your student name, ID, and module code at the start your report. Limit the report size to 5 pages with font size not more than 11 points.
Academic Integrity	Each module instructor reserves the right to follow up with a student by interview if there is any concern in relation to the integrity of the assignment.

	For any assignments not submitted via Turnitin, we reserve the right to check it using Turnitin where required.
Plagiarism	Plagiarism is the use of another person's ideas or work without appropriate acknowledgement or credit. Plagiarism may be intentional or unintentional. Intentional plagiarism is the clear intent to pass off another person's work or ideas as your own for your own gain. Unintentional plagiarism may occur if you do not understand the appropriate way to acknowledge the source of your ideas and information. If you are unsure of the acceptable methods of acknowledgment you should refer to the University of Galway Code of Practice for Dealing with Plagiarism, consult with your lecturer or the library staff. Proven plagiarism is a very serious matter which may result in severe disciplinary action and/or exclusion from the University. Ensure all assignment submissions include a signed plagiarism statement.
Referencing & Citation	Correct referencing and citation avoids plagiarism. There are varying referencing styles available but the most popular is the Harvard Referencing Style. Details on how to reference journal articles, books, electronic information and various other supports is available from the NUI, Galway Library at the following link: http://libguides.library.nuigalway.ie/c.php?g=543943&p=4591416
Blackboard Ally	Blackboard Ally supports you to access more user-friendly file formats. Please contact the lecturer if you experience any accessibility issues for this module material. Should you have a visual impairment and require the document in another format, please contact the lecturer to explore alternative format options.
Special Requirements	If you are registered with the Disability Support Service (DSS), you will find recommended accommodations listed on your Learning and Educational Needs Summary (LENS) report. If the alternative assessment offered for this module does not fully meet the recommendations in your LENS report, please email your lecturer as well as whilst also copying business@universityofgalway.ie, stating clearly how you feel the recommendations are not being met. Please ensure you attach a copy of your LENS report to this email.

MS5108 Individual Assignment 1
2022-2023

Please ensure all your visualisations are labelled correctly.

Question 1

From the https://data.gov.ie/ website, pick a dataset of your choice. In R Studio, represent any aspect of your dataset as a histogram using;

   a)   The R base graphics, then
   b)   GGPLOT2

Using the same dataset, represent another aspect of the data as a scatter plot using;

   c)   The R base graphics, then
   d)   GGPLOT2 with colours

Question 2

Using the "Client Analysis" dataset on Blackboard that contains data about customer spending on different platforms, perform a time series plot of spend changes from month to month for all customers using;

   a)   The R base graphics
   b)   Any other R functions which will enable you to produce a more professional and higher quality graphic.

Question 3

You are hired by an Analytics team to analyse the Tweets of Mr. XYZ. Tweets data is provided in CSV format and has four attributes (id, date, text, source). The source column is about the platform from where the tweet was created (like iPhone, Android, Twitter web client, etc). The dataset contains 4575 tweets. Using any R package, load these tweets in a corpus. After removing English stop-words and other necessary cleanings, provide and discuss the following:

   a)   A frequency plot of the top 25 words
   b)   A word cloud of 40 most common words
   c)   Can you check if he was using Social media during breakfast time (6-10 AM) or not? (Hint: Plot by the hour)
   d)   What was the usage per month?
   e)   What were the top-15 words for source='iPhone' and source='Media Studio'. Any interesting about this? Do think both sources were handled by one person only?
   f)   What are the six words he did not use in the last six months of the data but were frequently used in the first six months