OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Advanced Programming for Business Analytics
MS5114
Group Assignment 2022-2023

Objective	The objective of this assignment is to access your understanding with respect to all learning outcomes of the module.
Lecturer	**Name** **Office** **E-mail** Dr Umair ul Hassan CA369 umair.ulhassan@universityofgalway.ie
Marks Awarded	This assignment carries **60%** of the overall marks for the module. Each group will receive a grade and peer evaluation will be used to determine the individual grade each student receives. Grading descriptors are posted on Blackboard along with this file.
Submission Process	You should submit your completed assignment through Blackboard's Assignment tool. Any member of a group can submit on behalf of the whole group. 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)	**3rd February 2023 at 5:00pm Irish time:**
	    • Sign-up as groups on Blackboard. Each group should consist of 5 members.
	
	20th February 2023 at 5:00pm Irish time:
	    • Interim Deliverable (one per group)
	    • Signed Group Charter (one per group)
	
	3rd April 2023 at 5:00pm Irish time:
	    • Video presentation (one per group)
	    • Python code and data (one per group)
	    • Individual Reflection (one per student)
	
	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	**Group charter:**
	Each group must submit a **group charter**, signed by all members, in a Microsoft Word file by above mentioned deadline**.** The charter should follow the associated template available on Blackboard. The naming of your file should be as follows: "*GROUPNAME_Charter.docx*" e.g., Group1_Charter.docx.
	
	Inform the lecturer through email about issues of non-participation or contribution by a group member **as early as possible for resolution**. Affected students will do the assignment individually if the problem persists. However, it is advised that group members should exercise mutual understanding and contribute to the assignment appropriately.
	
	Interim deliverable:
	A 2-page document outlining your chosen dataset and your objectives for this project. Feedback will provided to you confirming the suitability of your proposal. The deliverable should follow the associated template available on Blackboard. The naming of your file should be as follows: "GROUPNAME_Interim.docx".
	
	Final deliverables:
	Upload 2 deliverables via Blackboard when you submit the group assignment.
	
	The **Python code and data** in an archive file, as "*GROUPNAME_code.zip*" file. If dataset is large, then provide a link to downloadable data file on the Web but still submit the code in an archive file.
	
	A **10-minute video presentation**, screencast while navigating through your Python files and results, that clearly explaining the Python code:
	  a. Present the background of this project (e.g., using Jupyter Notebooks), describe your group decision-making (e.g., in selecting a dataset, functions, choice of visualisations, etc.), various methods, and analytical findings.
	  b. Present your results, implications of analytical findings, and some key lessons learned for this project.
	Further requirements:
	  a. The first slide of the presentation should include: (i) group name, (iii) student names, (iv) student IDs, (v) module title. Include the University logo on the video and indicate that this is a student project.
	  b. All team members must play a part in the final deliverables (i.e., code, analysis, and presentation).
	  c. Introduce all team members in the video (this may be a group/individual profile photo for members voiceover section).
	  d. Ensure you are satisfied your work does not breach the University plagiarism policy or any copyrighted material (e.g. taking material from GitHub, etc).
	  e. This file should name as "*GROUPNAME_presentation.mp4*".
	
	Each student must submit, by the assignment deadline, a separate **individual reflection** of project contribution and peer evaluation using the provided template. This reflection should include a percentage estimation of every member's contribution. This file should name as "*GROUPNAME_YOURNAME_ reflection.docx*".

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.

MS5114 Group Assignment 2022-2023

The objective of this assignment is to apply your Python skills to an open dataset (or dataset which you already have access to) and report on the business insights, opportunities, trends, value and recommendations.

Guidelines & Evaluation

Your submission must meet the following criteria:

- Define and explore a clear problem within the domain of business analytics within a particular domain or sector.
- Demonstrate competence in topics covered within the lab material, for example: data structures; (ii) comparison operators, statements, methods, and functions; (iii) object orientated programming; (iv) NumPy and Pandas; (v) data visualisation.
- Investigate the application of machine learning techniques to predict future trends and the business potential of your insights.

Following factors will be considered in evaluating the final submission:

- Sufficiency of work for the team as a whole and by each individual covering lab and lecture material (with evidence of team collaboration)
- Determining suitable questions to exploit the dataset (e.g., explain group decision to create several questions which your dataset can adequately address and why these questions are worth exploring)
- Explanation of the dataset (e.g., decision to select the dataset, context of dataset, preparing the dataset, etc.), i.e. data literacy.
- Demonstrate use of course material throughout the code, for example:
  - data structures
  - comparison operators, statements, methods, and functions
  - object orientated programming
  - NumPy and Pandas
  - data visualization
  - machine learning.
- Quality and technical difficulty of Python code
- Originality and innovativeness of the project and application of Python
- Evidence of putting skills and theories in practice (combination of labs and lectures)
- Use of additional approaches, beyond course material, to analyse and present findings
- The business case – how does your analysis inform potential stakeholders of business insights, opportunities, trends, value, and recommendations.
- Demonstrated capacity for creative thinking and analytical problem-solving
- Clarity of presentation, communication, and visualisations

Indicative Sources of Datasets

Below you may find some examples of links to find openly available datasets. You are free to select any dataset of your choice, even if it is not included in the following suggestions. You are also free to select a smaller sample of the dataset under examination in case you face any issue with the processing capabilities of your PC/laptop.

Ireland's Open Data Portal (https://data.gov.ie/data)
Promoting innovation and transparency through the publication of Irish Public Sector data in open, free and reusable formats

European Data Portal (https://data.europa.eu/en)
Thousands of datasets from EU member states regarding issues of public interest such as Transport, Agriculture, Economics, Environment, etc.

Kaggle (https://www.kaggle.com/)
Includes over 50,000 public datasets

United Nations Data Sets (http://data.un.org/Explorer.aspx)
Several databases containing over 60 million records covering a wide range of themes including Agriculture, Crime, Education, Employment, Energy, Environment, Health, HIV/AIDS, Human Development, Industry, Information and Communication Technology, National Accounts, Population, Refugees, Tourism, Trade, as well as the Millennium Development Goals indicators.

UK Government Data (https://data.gov.uk/search)
Datasets available from all UK central government departments and a number of other public sector bodies and local authorities.

Mendeley Data (https://data.mendeley.com)
It includes 28.9 million datasets from domain-specific and cross-domain repositories Irish

Organisation for Geographic Information (http://www.irlogi.ie/datasets.php)
Irish GIS data sources and links to other portals

Spatial Data & Other Datasets for Ireland (https://libguides.ucd.ie/gisguide/FindSpatialData)
Links to hosts of datasets for such topics as:
Administrative Boundaries; Agriculture; Air Quality; Archaeology; Bathymetry; Birds; Built Heritage; Conservation Areas; Crime; Economics; Education; Elevation Data; Extractive Industries; Flooding; Flora; Forestry; Geology; Habitats; Health; Housing; Hydrology; Insects; Ireland_Shapefiles; Labour Market; Land Cover; Land Use; Mammals; Marine Data; Mining.

Central Statistics Office Databases (https://www.cso.ie/en/databases/)
Links to statistical databases of the Irish CSO, other public sector bodies, and international organisations.

EPA Geoportal site (https://gis.epa.ie)
This geoportal is designed to make data about the environment easier to find, browse and understand. You can download GIS data and find other portals to download datasets.

Hadoop Illuminated
(http://hadoopilluminated.com/hadoop_illuminated/Public_Bigdata_Sets.html)
List of publicly available big data sets

KDnuggets (https://www.kdnuggets.com/datasets/index.html)
Links to several business data sets from North America, Europe, and worldwide databases