School of IT & Business Technologies

Graduate Diploma in Data Analytics

Course Code & Title:	GDDA612 – Data Transformation and Management		
Assignment Title	Assessment 2	**Assessment Type**	Project 2
Level	6	**Credits**	15
Term & Cohort:			
Due date:			
Overall Weighting:	40%		
Total marks available	100		
Tutor:			

Course aim

The course provides students with the skills to work with various data sources to collect and prepare data for processing by cleaning and transforming it to appropriate format, import the pre-processed data to data stores, query and manipulate the data sets and data structures, and export data to an appropriate format for subsequent use.

Purpose

The purpose of this assessment is to manipulate the datasets and data structures and export the resultant data into various formats for further use to support and improve business decision-making.

Learning Outcomes (LOs)

This assessment is mapped to the following learning outcomes for this course:

LO3	Apply programming skills to query and manipulate the datasets and database structures.
LO4	Export data in appropriate form for further use, using programming language.

Graduate profile outcome

The above stated learning outcomes are mapped to the graduate profile outcomes shown below for this programme:

GPO 05	Critically analyse data using data collection and transformation methods, analysis methods, data and decision modelling concepts and visualisation tools to address complex business situations ensuring ethical professional considerations are met.
GPO 06	Apply problem solving and strategic thinking skills to establish and create data warehouses, apply data mining techniques, and find data patterns to enhance business decision making.

GPO 07	Apply machine learning, AI and big data analytical technologies to data analytic systems that support and improve business decision-making through critical thinking, programming and research skills.

Assessment Information

- This is an individual open-book assessment. The assessment is worth 60% of your total assessment weighting of the entire course.
- You must complete the Student Declaration and attestation on Canvas LMS.
- You must submit your completed assessment to Canvas.
- You must ensure you are familiar with NZSE's academic policies regarding assessments, and the relevant resubmission regulations that apply to this course. These policies and regulations can be found on your Course Outline on Canvas LMS.
- Your assessment will be marked on Canvas, and your assessment result will be provided on Canvas too.
- Resources and/or equipment which may be used for this activity:
    - NZSE issued material; and
    - your own course notes.
- A maximum of 15% of the content may be quoted or paraphrased from other sources provided you acknowledge and cite the original source of material you use.
- Use APA 7th edition referencing for all quoted or paraphrased material.
- All cases of plagiarism and/or cheating will be investigated and dealt with according to A08: Misconduct in Assessment Policy

Submission Instructions

You are required to ensure you have carried out the following before submitting your assessment:

- Naming convention of submitted file/s must be adhered to:

CourseCode_AssessmentNumber_ AssessmentName_StudentNumber_DocumentNumber

For example, "GD604_A1_Project_7647XXXX_1"

- All written answers are in your own words.
- Proofread and spell check the assessment work carefully.
- Submission should include word processed document or pdf that includes screenshots of all programming scripts and the results of the scripts.
- Compressed files of the program scripts and data store must also be uploaded on Canvas.
- DO NOT email your document/s to your tutor; it must be uploaded to the Canvas LMS.
- Check that all evidence required has been uploaded to the link provided on the NZSE LMS (Canvas).
- The GitHub link for your source code needs to be added into your report (must be accessible by your lecturer).
- The template for your academic report is available in Canvas.

Resits and Resubmissions of Assessments

Students enrolled in this programme have the opportunity for resubmission with the following conditions:

a. One resubmission in Semester One only. No resubmissions will be permitted during Semester Two.

A resubmission can only be offered where the original grade was at least 45%. No further teaching or specific feedback should occur between the final submission date and the resubmission.

The maximum grade allowable for any resubmitted assessment event is 50%.

The student shall make a request for a resubmission within three working days of the return of the graded assessment and negotiate a resubmission date. The agreed resubmission date must be adhered to, and no further extensions will be given.

In all cases, the grade achieved on the resubmission will be the grade used to calculate the final course grade.

Assessment Description

The objective of this assessment is to perform data manipulation on datasets and data structures and then export the resulting data into multiple formats. This process aims to facilitate and enhance business decision-making by providing valuable insights and analysis-ready data.

Assessment Guidelines

You may select the scenario described in the following section. Or, if you have any real projects from industry clients, you can choose any of them. This, however, MUST be approved by your tutor. Therefore, you are required to discuss the project details with your tutor and obtain approval prior to beginning the assignment.

Learning Outcome

Learning Outcome 3:
Apply programming skills to query and manipulate the datasets and database structures.
Learning Outcome 4:
Export data in appropriate form for further use, using programming language.

Scenario:

You are employed as a data analyst at a fashion retail e-commerce company, where your primary responsibility is to analyse sales data to uncover insights into customer behaviour and refine marketing strategies accordingly. The dataset provided encompasses detailed information regarding customer transactions, including purchase specifics, product attributes, and customer demographics.

In this role, you aim to leverage the extensive dataset to drive business decisions effectively. To realize these objectives, you are tasked with manipulating and exporting datasets into diverse formats conducive to analysis and reporting. This process empowers decision-makers within the company with access to invaluable insights, fostering informed and strategic decision-making essential for supporting business expansion and development.

Task A – Data Preparation and Database Integration [50 Marks][LO3]

a) Use Python or similar tool, to load and analyse the structure of the selected dataset. (4 marks)
b) Utilize data analytic systems knowledge, including machine learning packages and techniques, to transform the messy dataset into a tidy data format. (10 marks)
c) Display the initial rows of the resulting tidy dataset. (2 marks)
d) Employ Python and similar tools to filter the tidy dataset based on specified criteria and retrieve the filtered dataset. (6 marks)

e) Demonstrate proficiency in data analytic systems by establishing a connection with either an SQL or NoSQL database. (4 marks)
f) Import the dataset into a table or collection within the database. (4 marks)
g) Retrieve and display records or documents from the table or collection. (3 marks)
h) Sort the records or documents based on a given condition. (5 marks)
i) Count the number of records or documents present in the table or collection.(3 marks)
j) Perform grouping operations on records or documents within the table or collection. (5 marks)
k) Execute update operations on records or documents within the table or collection. (4 mark)

Task B – Data Export, Migration and Backup [50 Marks][LO4]

a) Export data from a specified table or collection in a database to a specified format file. (10 Marks)

b) Establish a connection to the cloud storage service and upload the locally stored file to a specified cloud storage bucket using a cloud service (GCP/Azure/AWS). (15 Marks)

c) Schedule automated backups of a specified directory to a cloud storage service. (15Marks)
d) Include error handling to handle backup failures gracefully, such as connectivity issues or file upload errors. (10 Marks)

Note: *Check the assessment marking rubric on pages 6-7 for the marks allocation and marking criteria/guidelines to attain optimum marks.*

Assessment Marking Rubric – GDDA – 612- Data Collection and Analysis

Assessment 2 – Project-2 Assignment (Total = 100 marks)

CRITERIA		EVIDENCE					
Task-A							
Task A – 50 Marks (LO-1)	**a) Load and analyse the structure of the selected dataset. (4 marks)**	Comprehensive analysis of dataset structure, including data types, dimensions, and summary statistics. (4 marks)	Adequate analysis of dataset structure, covering most aspects. (3 marks)		Basic analysis of dataset structure, missing some key details (2 marks)	Minimal attempt with incomplete analysis (1 mark)	No attempt made to load or analyze the dataset structure. (0 marks)
	b) Transform the messy dataset into a tidy data format. (10 marks)	Flawless transformation with clear documentation and meticulous handling of data. (8-10 marks)	Effective transformation of messy dataset into tidy format with clear documentation of steps taken. Proper handling of missing values, duplicates, and inconsistent data. (5-7 marks)		Partial transformation with some errors or missing steps. (3-4 marks)	Minimal attempt with incomplete transformation. (1-2 marks)	No attempt made to transform the dataset into a tidy format. (0 marks)
	c) Display the initial rows of the resulting tidy dataset. (2 marks)	Correct display of initial rows of the tidy dataset. (2 marks)	Attempt to display initial rows with errors or incomplete output. (1 mark)		No attempt made to display the initial rows of the tidy dataset. (0 marks)		
	d) Filter the tidy dataset based on specified criteria and retrieve the filtered dataset. (6 marks)	Accurate filtering of dataset based on specified criteria and retrieval of filtered dataset. (5-6 marks)	Partial filtering with some errors or missing criteria. (3-4 marks)		Attempt to filter dataset with errors or incomplete filtering. (1-2 marks)	No attempt made to filter or retrieve the dataset based on specified criteria. (0 marks)	
	e) Establish a connection with either an SQL or NoSQL database. (4 marks)	Successful establishment of connection with the chosen database system. (4 marks)	Successful establishment of connection with minor issues (3 marks)	Basic connection established with some issues. (2 marks)	Minimal attempt with errors or incomplete connection. (1 mark)	No attempt made to establish a connection with a database. (0 marks)	
	f) Import the dataset into a table or collection within the database. (4 marks)	Proper import of dataset into a table or collection within the database, ensuring data integrity. (4 marks)	Successful import into table or collection with minor issues. (3 marks)	Basic import with some issues. (2 marks)	Minimal attempt with errors or incomplete import. (1 mark)	No attempt made to import the dataset into a database table or collection. (0 marks)	
	g) Retrieve and display records or documents from the table or collection. (3 marks)	Correct retrieval and display of records/documents from the database table/collection. (3 marks)	Basic retrieval with some issues (2 marks)	Attempt to retrieve records/documents with errors or incomplete output. (1 mark)	No attempt made to retrieve or display records/documents from the database table/collection. (0 marks)		

	h) Sort the records or documents based on a given condition. (5 marks)	Perfect sorting of records/documents based on specified condition. (5 marks)	Accurate sorting of records/documents based on specified condition with minor issues. (4 marks)	Proper sorting of records/documents based on specified condition. (3 marks)	Basic sorting with some issues. (2 marks)	Attempt to sort records/documents with errors or incomplete sorting. (1 mark)	No attempt made to sort records/documents based on a given condition. (0 marks)
	i) Count the number of records or documents present in the table or collection. (3 marks)	Correct counting of the number of records/documents present in the table/collection. (3 marks)	Basic counting with some issues (2 marks)	Attempt to count records/documents with errors or incomplete count. (1 mark)	No attempt made to count the number of records/documents present in the table/collection. (0 marks)		
	j) Perform grouping operations on records or documents within the table or collection. (5 marks)	Perfect grouping operations with no errors. (5 marks)	Accurate grouping operations with minor issues (4 marks)	Proper implementation of grouping operations on records/documents within the table/collection (3 marks)	Basic grouping with some issues (2 marks)	Attempt to perform grouping operations with errors or incomplete grouping (1 mark)	No attempt made to perform grouping operations on records/documents within the table/collection. (0 marks)
	k) Execute update operations on records or documents within the table or collection. (4 marks)	Accurate execution of update operations on records/documents within the table/collection, ensuring data consistency. (4 marks)	Proper execution of update operations on records/documents within the table/collection (3 marks)	Basic update with some issues (2 marks)	Attempt to execute update operations with errors or incomplete update. (1 mark)	No attempt made to execute update operations on records/documents within the table/collection. (0 marks)	
Task B – 50 Marks (LO-2)	**a) Export data from a specified table or collection in a database to a specified format file. (10 Marks)**	Flawless export of data to the specified format file with clear documentation and no errors. (9-10 Marks)	Proper export of data from the specified table/collection with minimal errors. (6-8 Marks)		Successful export of data, but with minor errors or omissions. (3-5 Marks)	Partial attempt with errors in exporting data. (1-2 Marks)	No attempt made to export data from the specified table/collection. (0 Marks)
	b) Establish a connection to the cloud storage service and upload the locally stored file to a specified cloud storage bucket using a cloud service (GCP/Azure/AWS). (15 Marks)	Seamless establishment of connection and flawless upload of file to specified cloud storage bucket with clear documentation and no errors. (13-15 Marks)	Proper establishment of connection and successful upload of file to specified cloud storage bucket with minor issues. (9-12 Marks)		Successful establishment of connection and file upload, but with some issues or errors. (5-8 Marks)	Partial attempt with errors in establishing connection or uploading file. (1-4 Marks)	No attempt made to establish a connection or upload the file to cloud storage. (0 Marks)
	c) Schedule automated backups of a specified directory to a cloud storage service. (15 Marks)	Seamless scheduling of automated backups with clear documentation and no errors. (13-15 Marks)	Proper scheduling of automated backups with minor issues. (9-12 Marks)		Successful scheduling of automated backups, but with some issues or errors. (5-8 Marks)	Partial attempt with errors in scheduling backups. (1-4 Marks)	No attempt made to schedule automated backups. (0 Marks)

	d) Include error handling to handle backup failures gracefully, such as connectivity issues or file upload errors. (10 Marks)	Comprehensive error handling to gracefully handle all potential backup failures, with clear documentation and no major shortcomings. (9-10 Marks)	Effective error handling to address various backup failures, but with minor shortcomings. (7-8 Marks)	Adequate error handling to address most common backup failures, but with some gaps. (4-6 Marks)	Partial attempt with basic error handling, but insufficient to handle all potential errors. (1-3 Marks)	No attempt made to include error handling. (0 Marks)

School of IT & Business Technologies
Graduate Diploma in Data Analytics
Cover Sheet and Student Declaration

This sheet must be signed by the student and attached to the submitted assessment.

Course Title:	**Data Transformation and Management**	**Course code:**	**GDDA-612**
Student Name:		**Student ID:**	
Assessment No & Type:	Assessment 2- Project-2	**Cohort:**	
Due Date:		**Date Submitted:**	
Tutor's Name:			
Assessment Weighting	40%		
Total Marks	100		

Student Declaration:

I declare that:

- I have read the New Zealand School of Education Ltd policies and regulations on assessments and understand what plagiarism is.
- I am aware of the penalties for cheating and plagiarism as laid down by the New Zealand School of Education Ltd.
- This is an original assessment and is entirely my own work.
- Where I have quoted or made use of the ideas of other writers, I have acknowledged the source.
- This assessment has been prepared exclusively for this course and has not been or will not be submitted as assessed work in any other course.
- It has been explained to me that this assessment may be used by NZSE Ltd, for internal and/or external moderation.
- If I am late in handing in this assessment without prior approval (see student regulations in handbook), marks will be deducted, to a maximum of 50%.

Student signature:

Date:

Tutor only to complete		
Assessment result:	**Mark /100**	**Grade**