

# ASSESSMENT GUIDE

**MIT5512**

**Methods of Data Science**

**Semester 2 2020**



## Assessment overview

Assessments	Overview	Weight	Due date	Unit Learning Outcomes (ULO)
Assessment 1: Case Study	In this assignment you will be given a small case study and you will need to apply your knowledge to Identify, make appropriate analytics for a Data Science project.	10%	Session 5	1, 2, 3
Assessment 2: Research Report	In this assessment you will write a critique or report on an academic paper approved by your lecturer in the field of Data Science Methods.	10%	Session 9	1, 2, 5
Assessment 3: Major Assignment	In this assessment you will work in groups on a major practical based case to devise and implement Data Science based solution.	30%	Session 12	1, 2, 3, 4,
Assessment 4: Final Assessment		50%	During the end of semester exam period	1, 2, 3, 4

## Referencing guides

You must reference all the sources of information you have used in your assessments. Please use the **IEEE referencing style** when referencing in your assessments in this unit. Refer to the Library's referencing guides for more information.

- Guide 1.
- Guide 2.

## Academic misconduct

VIT enforces that the integrity of its students academic studies follow an acceptable level of excellence. VIT will adhere to its [VIT Policies, Procedures and Forms](#) where it explains the importance of staff and student honesty in relation to academic work. It outlines the kinds of behaviours that are "academic misconduct", including plagiarism.

## Late submissions

In cases where there are no accepted mitigating circumstances as determined through [VIT Policies, Procedures and Forms](#), late submission of assessments will lead automatically to the imposition of a penalty. Penalties will be applied as soon as the deadline is reached.

## Short extensions and special consideration

Special Consideration is a request for:

- Extensions of the due date for an assessment, other than an examination (e.g. assignment extension).
- Special Consideration (Special Consideration in relation to a Completed assessment, including an end-of-unit Examination).

Students wishing to request Special Consideration in relation to an assessment the due date of which has not yet passed must engage in written emails to the teaching team to Request for Special Consideration as early as possible and prior to start time of the assessment due date, along with any accompanying documents, such as medical certificates.

For more information, visit [VIT Policies, Procedures and Forms](#).

## Inclusive and equitable assessment

Reasonable adjustment in assessment methods will be made to accommodate students with a documented disability or impairment. Contact the unit teaching team for more information.

## Contract Cheating

Contract cheating usually involves the purchase of an assignment or piece of research from another party. This may be facilitated by a fellow student, friend or purchased on a website. Other forms of contract cheating include paying another person to sit an exam in the student's place.

Contract cheating warning:

- By paying someone else to complete your academic work, you don't learn as much as you could have if you did the work yourself.
- You are not prepared for the demands of your future employment.
- You could be found guilty of academic misconduct.
- Many of for pay contract cheating companies recycle assignments despite guarantees of "original, plagiarism-free work" so similarity is easily detected by Turnitin.
- Penalties for academic misconduct include suspension and exclusion.
- Students in some disciplines are required to disclose any findings of guilt for academic misconduct before being accepted into certain professions (e.g. law).
- You might disclose your personal and financial information in an unsafe way, leaving yourself open to many risks including possible identity theft.
- You also leave yourself open to blackmail - if you pay someone else to do an assignment for you, they know you have engaged in fraudulent behaviour and can always blackmail you.

## Grades

We determine your grades to the following Grading Scheme:

Grade	Percentage
A	80% – 100%
B	70% – 79%
C	60% – 69%
D	50% – 59%
F	0% – 49%

# Assessment details

## Assessment 1: Case Study

### Overview

Weight	Length	Due date	ULO
10%	1200-1500	Session 5	1, 2, 3

### Introduction

### Task

Suppose you are working for a company/organisation. Your manager gives you a data and ask you what you can do with the data in terms of adding some values to the company goals and future operational research plans.

Three tasks you need to do: 1) choose a proper data related to a business case study, 2) choose a proper software to open the data and do data visualisation and exploratory analytics and 3) write a clear and accurate report and put all findings in the report.

Your report should have 1200-1500 words addressing the following: information on the data, type of features, literature review on the data and methodology you are going to apply, what you are going to solve and how, plots and recommendations. The report should have some plots (4-6 screenshots) from your findings with explanations.

#### 1. Choose a data

Choose a data from Kaggle website, <https://www.kaggle.com/datasets>, or a government open source data. The data should be related to a business case study, such as house marketing, climate change, patients records and banking data. You need to add information on data in your report, including reference where you downloaded the data, information of data type and features.

## 2. Visualisation and Exploratory Analysis

Select any data science tools to open the data. Look at the data and find out how you can improve quality of the data. You must provide some data visualisation using selected software. Do an exploratory data analysis on the data that you have gathered. Exploratory data analysis is an approach for analysing data sets to summarize their main characteristics, often with visual methods. These analytics should be in your report.

### Submission Instructions

All submissions are to be submitted through turn-it-in. Drop-boxes linked to turn-it-in will be set up in the Unit of Study Moodle account. Assignments not submitted through these drop-boxes will not be considered.

Submissions must be made by the due date and time (which will be in the session detailed above) and determined by your Unit coordinator. Submissions made after the due date and time will be penalized at the rate of 10% per day (including weekend days).

The turn-it-in similarity score will be used in determining the level if any of plagiarism. Turn-it-in will check conference web-sites, Journal articles, the Web and your own class member submissions for plagiarism. You can see your turn-it-in similarity score when you submit your assignment to the appropriate drop-box. If this is a concern you will have a chance to change your assignment and resubmit. However, re-submission is only allowed prior to the submission due date and time. After the due date and time have elapsed you cannot make re-submissions and you will have to live with the similarity score as there will be no chance for changing. Thus, plan early and submit early to take advantage of this feature. You can make multiple submissions, but please remember we only see the last submission, and the date and time you submitted will be taken from that submission.

Your report should be a single word or pdf document containing your report.

## Marking criteria

You will be assessed on the following marking guide

Total Marks: 50

Task	Description	Marks
<b>Introduction</b>	An <b>introduction</b> and <b>objective</b> for the analysis.	10
<b>Data Description</b>	<b>Brief overview and summary of the dataset.</b> Clearly identify the <b>data type</b> , <b>features</b> and their <b>relationships</b> relevant to the chosen dataset.	10
<b>Data Analysis</b>	Successfully extract and analyse the dataset and demonstrate the insightful information by applying different data science methods & techniques.	10
<b>Data Visualization</b>	<b>Develop (4-6) visualization graphs</b> , charts or any other relevant visual representations.	10
<b>Conclusion &amp; Recommendation</b>	A <b>summary</b> , <b>findings</b> and <b>recommendation</b> of the analysis discussing the main characteristics of the data and improvement techniques.	10

## Assessment 2: Research Report

### Overview

Weight	Length	Due date	ULO
10%	1500	Session 9	1, 2, 5

### Introduction

This assessment item relates to the unit learning outcomes as in the unit descriptor. This assessment is designed to improve student research skills and to give students experience in researching a topic and writing a report relevant to the Unit of Study subject matter.

### Task

For this component you will write a report or critique on a recent academic paper related to Data Science or Data Science Methodologies. Some possible topic areas include but are not limited to:

- Supervised Learning
- Unsupervised Learning
- Semi-Supervised Learning
- Anomaly Detection
- Association Analysis
- Regression Analysis
- Classification Analysis
- Pattern Recognition
- Feature Selection - (aka Dimensionality Reduction)
- Ensemble Methods
- Neural Nets and Deep Learning
- Transfer Learning
- Reinforcement Learning
- Natural Language Processing



- Applications of Data Science

The paper you select must be directly relevant to one of the above topics or another topic and be related to Data Science. The paper must be approved by your lecturer and be related to what we are studying this semester in Data Science Course. The paper can be from any academic conference or other relevant Journal or online sources such as Google Scholar, or Academic department repositories. All students must select a different paper. Thus, the paper must be **approved by your lecturer** before proceeding. In case two students are wanting to present on the same paper, the first who emails the lecturer with their choice will be allocated that paper. Please note that popular magazine or web-site articles are not academic papers. The paper you chose should be published in the **last 5 years**.

The report should be limited to approx. 1500 words (not including references). Use 1.5 spacing with a 12-point Times New Roman font. Though your paper will largely be based on the chosen article, you should use other sources to support your discussion or the chosen papers premises. Citation of sources is mandatory and must be in the IEEE style.

## Report Content

Title Page: The title of the assessment, the name of the paper you are reporting on and its authors, and your name and student ID.

Introduction: Identification of the paper you are critiquing/ reviewing, a statement of the purpose for your report and a brief outline of how you will discuss the selected article (one or two paragraphs).

Body of Report: Describe the intention and content of the article Document a critical analysis on the selected business case, brief overview of the dataset, data type and variables. You may assume the details of the dataset if not considered in your chosen paper. Moreover, critically describe the adopted data science models and decision-making tools which has been used and applied in your chosen paper. In addition to that, report the outcomes of the recommended business strategic directions. If such recommendation is not outlined in your chosen paper, discuss and justify your own view.

Conclusion: A summary of the points you have made in the body of the paper. The conclusion should not introduce any 'new' material that was not discussed in the body of the paper. (One or two paragraphs)

References: A list of sources used in your text. They should be listed alphabetically by (first) author's family name. Follow the IEEE style. The footer must include your name, student ID, and page number.

Note: reports submitted on papers which are not approved or not the approved paper registered for the student will not be graded and attract a zero (0) grade.

## Submission Instructions

Submit your report to the Moodle drop-box for Assignment 2. Note that this will be a Turnitin drop box and as such you will be provided with a similarity score. This will be considered when grading the assignment. Note that incidents of plagiarism will be penalized. If your similarity score is high you can re-submit your report, but resubmissions are only allowed up to the due date. If you submit your assignment after the due date and time re-submissions will not be allowed.

Please Note: All work is due by the due date and time. Late submissions will be penalized at the rate of 10% per day including weekends.

## Marking criteria

You will be assessed on the following marking guide

Total Marks: 50

Task	Description	Marks
<b>Introduction</b>	A statement of the purpose for your report and a brief outline of how you will discuss the selected article.	10
<b>Case Analysis</b>	Document a critical analysis on the selected business case, brief overview of the dataset, data type and variables. You may assume the details of the dataset if not considered in your chosen paper.	10
<b>Analytical Tools/Techniques</b>	Critically describe the adopted data science models and decision-making tools which has been used and applied in your chosen paper.	10
<b>Insights/ recommendation</b>	Report the outcomes of the recommended business strategic directions. If such recommendation is not outlined in your chosen paper, discuss and justify your own view.	10
<b>Conclusion and Reference (If any)</b>	A summary of the points you have made in the body of the paper. If references have been used, a list of sources used in your text. They should be listed alphabetically by (first) author's family name. Follow the IEEE style.	10

# Assessment 3: Major Assignment

## Overview

Weight	Length	Due date	ULO
30%	2000	Session 12	1, 2, 3, 4

## Introduction

This assessment item relates to the unit learning outcomes as in the unit descriptor. This assessment is designed to improve student analytic skills and to give students experience in problem solving, decision-making and presentation skills in data science methods and tools.

## Task

For this assignment, you are required to work in a group of **maximum 4 students** and **two files** are required to be uploaded in the Moodle (provided links) by one of the group members. The first file is a report containing maximum 2000 words with 6-10 screenshots from your findings. The second file is a presentation file of your analytics and findings. Note that both files need to be uploaded by only one of the group members.

### 1. Data

The required dataset is available in the Moodle. You need to add a section in your report and talk about the data and challenges there. What kind of issues available there? What are the features? Add some information on the data in this section of your report.

### 2. Data Analytics & Visualization

Identify what kind of data it is and what you can do with this data if this data passes to you by your company. To apply such method, you need to explore the data and apply data processing, such as data cleaning and feature engineering, if it is required. Then choose a proper data science method/s to analysis the data. Suppose this is a company data that you are working for them. What are the issues available there and what you can recommend for your manager in company to enhance their objectives and to the benefits of the company?

### 3. Report (Weightage 20%)

Your report should have 1500-2000 words, excluding references, addressing the business questions, challenges, analytics, recommendation and visualisation related to the data. It should cover what are the issues in the data, you are going to solve and how, plots and recommendations. The report should have important plots (6-10 screenshots) from your findings. Note that plots need to be labelled and explained inside the report.

All coding, including data uploading, cleaning, analytics and visualisation should be coded in Python. The python code should be included at the end of your report in a section called Appendix.

Note: Structure and font of your report should follow the word file template provided in the Moodle. Your report should be a single word or pdf document containing your report and need to be submitted through Moodle. One submission per group and make sure all group members participate and add their names in the report. Your report should have a contribution table at the end of the report.

### 4. Presentation (Weightage 10%)

The presentation should be a maximum of 10 minutes for the whole team. Each member must participate in video presentation file and talk at least 2 minutes in the video related to the methodology used, findings, contribution or recommendation.

Note: Presentation file should be submitted by the same person who submitted the report. Your presentation file should have a standard video format and it should not exceed 200 MB. Slides and your faces should be clear in the video file. The same person, who submitted the report, needs to submit the presentation file (not link to your video) in the provided video submission link. **Only video file** will be marked; the link for your video is not accepted and will not be considered for marking.

## Submission Instructions

All submissions are to be submitted through turn-it-in. Drop-boxes linked to turn-it-in will be set up in the Unit of Study Moodle account. Assignments not submitted through these drop-boxes will not be considered.

Submissions must be made by the due date and time (which will be in the session detailed above) and determined by your Unit coordinator. Submissions made after the due date and time will be penalized at the rate of **10% per day** (including weekend days).

The turn-it-in similarity score will be used in determining the level if any of plagiarism. Turn-it-in will check conference websites, Journal articles, the Web and your own class member submissions for plagiarism. You can see your turn-it-in similarity score when you submit your assignment to the appropriate drop-box. If this is a concern you will have a chance to change your assignment and resubmit. However, re-submission is only allowed prior to the submission due date and time. After the due date and time have elapsed you cannot make re-submissions and you will have to live with the similarity score

as there will be no chance for changing. Thus, plan early and submit early to take advantage of this feature. You can make multiple submissions, but please remember we only see the last submission, and the date and time you submitted will be taken from that submission.

Note: You must find group by yourself and **provide group member names by week 6**. It is your responsibility to find group members. If you could not find any students to create a group, you have to work by yourself and provide a quality report and presentation as others.

## Marking criteria

You will be assessed on the following marking criteria (will be publish soon).