



CFGDEGREE

DATA ASSESSMENT MATERIAL RELEASE

THEORY QUESTIONS

SECTION	MARK
1. Theory Questions	25
2. Pandas Questions	25
3. Matplotlib Challenge	25
4. Numpy Questions	25
TOTAL	100

Important notes:

- This document shares the first section of the Data Assessment which is composed of 5 Data Theory Questions
- The answers do not have to be long, but they have to answer each of the mention points for each question
- It is worth a quarter of your assessment mark
- You have 24 hours before the assessment to prepare.
- If any plagiarism is found in how you choose to answer a question you will receive a 0 and the instance will be recorded.
- Consequences will occur if this is a repeated offence. You can remind yourself of the plagiarism policy [here](#).
- You are allowed to use any online images to support your answers.

Section 1: Theory Questions [25 points]

1.1 In your own words, what does the role of a data scientist involve?	2 points
1.2 What is an outlier? Here we expect to see the following: <ul style="list-style-type: none">a. Definitionb. Examplesc. Should outliers always be removed? Why?d. What are other possible issues that you can find in a dataset?	4 points
1.3 Describe the concepts of data cleaning and data quality. Here we expect to see the following: <ul style="list-style-type: none">a. What is data cleaning?b. Why is data cleaning important?c. What type of mistakes do we expect to commonly see in datasets?	4 points
1.4 Discuss what is Unsupervised Learning - Clustering in Machine Learning using an example. Here we expect to see the following: <ul style="list-style-type: none">a. Definition.b. When is it used?c. What is a possible real-world application of unsupervised learning?d. What are its main limitations?	7.5 points
1.5 Discuss what is Supervised Learning - Classification in Machine Learning using an example. Here we expect to see the following: <ul style="list-style-type: none">a. Definition.b. When is it used?c. What is a possible real-world application of supervised learning?d. What data do we need for it? Is there any processing that needs to be done?	7.5 points