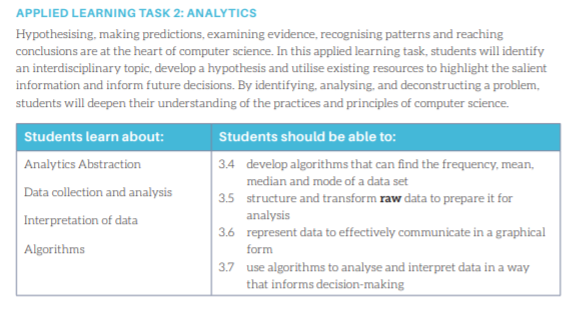
**LEAVING CERTIFICATE COMPUTER SCIENCE**

**APPLIED LEARNING TASK 2: ANALYTICS**

****

**Dates:**

Project Assigned: 17th December 2024

Project Due: 20th January 2025

**Theme:**

You are required to perform a data analysis project. The project should put forward a hypothesis that should be then tested using data analysis tools and a conclusion reached.

The topic can be anything of your choosing.

**The Project:**

1. **Getting Started**

Brainstorm **lots of ideas** based on the topics given. No idea is a bad idea - create a mind map showing all your ideas

Narrow down these ideas into **three** that you would consider doing. **Research** these three ideas further - maybe take one each.

You will need to find a suitable **dataset**. This can be difficult and will probably decide your topic for you.

Decide on your **topic** and **hypothesis**. A hypothesis is an assumption.

Examples of a hypothesis:

There are more people unemployed during an economic downturn

More crimes are committed in cities

A team is more likely to score in home games

Coronavirus cases are higher in urban cities than rural areas

1. **Source, Collect, Clean and Transform Data**

Source/Collect:

Once you have decided on your topic, you need to **gather/collect data**. Make sure to save it somewhere that you have access to it. The data can be obtained from a third party source or collected directly through observation, surveys, experiments

Clean/Transform:

* Check what format it is currently stored in (.xlsx, .csv, .txt) and does it need to be converted?
* Are there any symbols, characters etc that you need to remove?

1. **Algorithms for Processing Data**

When your data is ready, use **algorithms** to find the mean, median and mode (if possible).

1. **Data Visualisation**

Once your data is processed, it must be **visualised** to make it easier to interpret. This is done by creating one or more graphs.

1. **Data Interpretation**

**Interpret** your findings. How does it compare with your hypothesis and what future predictions or decisions will it affect?

**Final Report:**

The final report should be presented as a web page (using Google Sites) and be structured according to the headings below. There should be one report per group.

1. **Research**

Brainstorm - mindmap

Choose three of your ideas and research them further.

* Where did you read about or find out about this topic?
* What websites/videos did you view?
* Is there data available on this topic and is it from a trusted source?
* Is the data suitable for this assignment? Why or why not?

Decide on a topic, hypothesis and dataset.

Why are you choosing this topic?

Who do you think will benefit from this information?

1. **Plan**

Detail what you plan on doing BEFORE CODING including the following:

The plan can be broken into separate parts

**Data**

* Where you got your data
* What format was it in?
* What did you have to do in order to get it into a csv file that you can use?
* Are there any symbols or characters that need to be removed from your data before you use it? If so, how are you planning on doing that
* Where have you stored the file?

**Python**

* Explain in detail what you will do.
  + I will import the csv and matplotlib libraries to use
  + Example: I will set up 2 empty lists called …. that I will use to store my data that I will read from the csv file.
  + I will use the csv reader to open and read my csv file.
  + I will loop through each row in the csv file and store the data in the 2nd column in my list etc
* Do any of your lists need to be converted to int or float - why?
* How are you going to find the mean, median or mode.

**Graph**

* Graphical output - describe the graphical output you chose and why.
  + Example: I am choosing a line graph because I want to graph two sets of data and it will be easier to show the difference between them etc

1. **Development**

Implement your plan.

Keep track of your progress - this can be done like a diary.

Take note of any issues you have and how you went about solving them - include code snippets or pictures.

1. **Evaluation**

**Reflection on meeting the brief**

What have you learned from the data?

How does it compare to your hypothesis?

Did you manage to do all you set out to?

Do you think you met the requirements and if so how or why not?

**Future development**

How might you grow/add to/improve the project and why?

If you had more time, what would be the next thing to work on with this project?

**Submission:**

All of the following must be submitted

* Report as detailed above
* Full code (Thonny file)
* Dataset

**DataSets**

You can source data from lots of places. Here are some websites to try - bear in mind that just because there is a dataset, it doesn’t mean that it will be nice and easy to use so look at it first.

If you are using Google to search make sure you put the word “dataset” in your search.

<https://www.kaggle.com/> (you do not have to sign in to view data)

<https://datasetsearch.research.google.com/>

<https://data.gov.ie/dataset>

<https://docs.google.com/spreadsheets/d/1vg6eg17yNNs70PSy3Ehd3MIqYL5UNhe2aR4OwxQGQtQ/edit#gid=0>

<https://www.cso.ie/en/index.html>

<https://corgis-edu.github.io/corgis/python/>

<https://corgis-edu.github.io/corgis/python/>

<https://www.met.ie/climate/available-data>

<https://knoema.com/WBCC2019/world-bank-climate-change>

<https://data.world/datasets/global-warming>

<https://developers.google.com/earth-engine/datasets/catalog/>

**Groups**

| **Group 1** | Michael | Mahbub | Jamal | Josh |
| --- | --- | --- | --- | --- |
| **Group 2** | Anvith | Azrielle | Victoria | Natas |
| **Group 3** | Roaa | Afsheen | Evan | Abdullah |
| **Group 4** | Jonas | Hermann | Aaron | Qasim |
| **Group 5** | Hassan | Nivedha | Aballah |  |