

Module: COMP1800 Data Visualisation	Coursework
Contribution: 100% of course	ZIP file required, containing PDF report & Python notebook
Module Leader: Prof. Chris Walshaw	Due date: Thursday 6th April 2023
This coursework should take an average student who is up-to-date with tutorial work approximately 50 hours	
Learning Outcomes: <ol style="list-style-type: none"> 1. Identify and discuss fundamental concepts related to visualisation. 2. Demonstrate an understanding of different types of data visualisation and identify appropriate types of visualisation for various types of data. 3. Design, implement and evaluate interactive visualisation systems. 4. Apply visualisation tools and techniques to obtain insight from datasets. 	

Plagiarism is presenting somebody else's work as your own. It includes: copying information directly from the Web or books without referencing the material; submitting joint coursework as an individual effort; copying another student's coursework; stealing or buying coursework from someone else and submitting it as your own work. Suspected plagiarism will be investigated and if found to have occurred will be dealt with according to the procedures set down by the University.

All material copied or amended from any source (e.g. internet, books) must be referenced correctly according to the reference style you are using.

Your work will be submitted for electronic plagiarism checking. Any attempt to bypass our plagiarism detection systems will be treated as a severe Assessment Offence.

Coursework Submission Requirements

- An electronic copy of your work for this coursework should be fully uploaded by midnight (local time) on the Deadline Date.
- The last version you upload will be the one that is marked.
- For this coursework you must submit a **single zip file** containing your report, in an **Acrobat PDF document**, and supporting code. In general, any text in the document must not be an image (i.e. must not be scanned) and would normally be generated from other documents (e.g. MS Office using "Save As .. PDF").
- There are limits on the file size (currently 2Gb).

- Make sure that any files you upload are virus-free and not protected by a password or corrupted otherwise they will be treated as null submissions.
- Comments on your work will be available from Moodle. The grade will be made available in the portal.
- You must NOT submit a paper copy of this coursework.
- All coursework must be submitted as above.

The University website has details of the current Coursework Regulations, including details of penalties for late submission, procedures for Extenuating Circumstances, and penalties for Assessment Offences. See <https://www.gre.ac.uk/student-services/regulations-and-policies> for details.

Detailed specification

You are to carry out a visual data exploration for ChrisCo, the fictional company whose sales and website data we have been analysing throughout the module, using a Python Notebook (in Colab or Jupyter).

ChrisCo is a fictional, but nonetheless very successful company managing a range of retail outlets across the UK. ChrisCo collects a huge amount of data about individual customers visiting its outlets using its loyalty card scheme but this customer data has been aggregated/averaged to give information about the company's 45 outlets, each identified by a unique 3 letter code (e.g. ABC, XYZ, etc).

Data

Each student on the module has their own, randomised version of the dataset to explore. You will find your data in the following csv files, where **ID** is your 9-digit student ID number (e.g. 001234567):

- <https://tinyurl.com/ChrisCoDV/ID/OutletDailyCustomers.csv>
listing the daily number of customer visits to the company's outlets
- <https://tinyurl.com/ChrisCoDV/ID/OutletMarketing.csv>
the total annual spend on local marketing for each outlet
- <https://tinyurl.com/ChrisCoDV/ID/OutletOverheads.csv>
the total annual cost of overheads for each outlet
- <https://tinyurl.com/ChrisCoDV/ID/OutletSize.csv>
the size (floor space) in metres squared for each outlet
- <https://tinyurl.com/ChrisCoDV/ID/OutletStaff.csv>
the total number of full-time staff employed at each outlet

Please contact the module leader if you cannot find your data files.

You should compile the data into two dataframes: one containing daily customer data (one row for each date); the other containing summary data (one row for each outlet), compiled from **all** of the .csv files, including the daily visitors.

Report

Your task is to investigate the data visually and present some conclusions about any characteristics you discover, including correlations, seasonal behaviour, outliers, etc., together with a suggestion about how the data might be best segmented, based on the total volume of visits at each outlet.

The company is most interested in the high and medium volume outlets but would like a summary of the low volume outlets plus any anomalies you identify in the data. You should also identify new outlets that have been opened during the year or outlets that the company has closed during the year.

You should present your findings in the form of a pdf report for the company, i.e. based on the assumption that the reader knows nothing about data visualisation. The report should include:

- A brief **introduction to data visualisation** (no more than ½ a page).
- A **discussion** of your findings, including a total of 8 visualisations (your mark will be **capped** if you include more – see below under Grading Criteria). Each visualisation should be accompanied by two paragraphs of text in which you should present:
 - a **justification** for including that particular visualisation:
 - a **description** of what the visualisation reveals about the data – do not assume that the reader will recognise and understand correlations, seasonality and anomalies.
- A **critical review** of your work, with a discussion of what you have learned from the module, how you have applied it to the coursework, and how best practices were demonstrated (about ½ a page).
- A summary of the **conclusions** you have drawn (no more than ½ a page). You are **not** required to make business recommendations and the summary may contain conclusions as bullet points.

For the 8 visualisations you include, you should choose your most illuminating charts / plots and paste in a screenshot. It is strongly recommended to use Insert > Screenshot in Word or the Windows snipping tool (or similar) and to carefully crop each screenshot so that it shows **only** the visualisation. Also do not distort the images when you resize them – if you do change the size make sure you maintain the aspect ratio.

The **interactive** visualisations (see below) should be included as 2 of the 8 visualisations in the report and for these the screenshots should **illustrate** some aspect of the interactivity (e.g. zoom, hover tools, etc).

All 8 visualisations should be carefully **numbered** and labelled, with a self-explanatory title and legend (if appropriate) and should be referred to in the text (e.g. "Figure 1 shows that ..."). Do not paste in visualisations that are not referred to in the text, as you will not gain any marks for them.

Finally, the visualisations must demonstrate that you have understood and made use at least some of the techniques taught in the module. Also, the order of the visualisations should be carefully considered, leading the reader through the data exploration, ideally with each visualisation leading on to the next one.

Notebook

Your Colab / Jupyter notebook should contain the details of your data exploration and support the report. The **markdown** should indicate the purpose of each preceding / following code section but you do not have to present your findings here.

The code should be written **efficiently**, so that you do not repeat unnecessary code in each section.

At least 2 of the visualisations in the notebook should be **interactive** and provide functionality to explore the data in more detail. The markdown for these must include a **description** of available user interactions and **instructions** on how to use the interactions to explore the data more fully.

You **may include more than 8 visualisations in your notebook**. You should then select the 8 most useful to paste into your report.

Deliverables

You must upload a **single zip file** (with a .zip extension – do not use other formats) containing:

- The pdf report containing your 8 chosen visualisations.
- A supporting Python notebook (.ipynb) containing your data exploration. You may upload 2 notebooks if you have used both Jupyter & Colab (e.g. for the interactive visualisations).

Marking scheme

The report will be marked on the discussion and analysis, together with both the quality and impact of the visualisations. The notebook will be marked on its organisation, presentation and efficiency of coding. There are also marks for the interactive visualisations.

The marks will be broken down according to the following tasks in the rubric assessment domains of Knowledge, Evaluation and Communication – see below:

Task	Marks
Report text (50%)	
Knowledge: Introduction to data visualisation [10 marks]	
Knowledge: Justification of visualisations chosen [10 marks]	
Evaluation: Description of findings [10 marks]	
Evaluation: Critical review [10 marks]	
Evaluation: Data conclusions [10 marks]	
Report visualisations (20%)	
Communication: Presentation quality (labelling, legends, etc) [10 marks]	
Communication: Value and impact (as part of the exploration) [10 marks]	
Notebook (30%)	
Communication: Organisation and presentation [10 marks]	
Knowledge: Code efficiency (non-duplication) [5 marks]	
Knowledge: Interactive visualisations – functionality [5 marks]	
Communication: Interactive visualisations – instructions [10 marks]	

Capping

The report **must not contain more than 8 visualisations**, although you may use the technique of facetting (i.e. a number of subplots in a single figure). Regardless of the total marks, reports with 9-10 visualisations will be **capped at 60%** and those with 11 or more visualisations will be **capped at 30%**. However, your notebook may contain as many visualisations as you need to carry out the investigation.

Assessment Rubric

For more information see the University Feedback and Assessment Policy

Assessment Domains	0-29% Fail	30-49% Fail	50-59% Acceptable	60-69% Good	70-79% Excellent	80-100% Exemplary
D1 Knowledge: of the concepts and principles of data visualisation	No evidence of the concepts and principles of data visualisation	Little evidence of the concepts and principles of data visualisation	Acceptable evidence of the concepts and principles of data visualisation	Good evidence of the concepts and principles of data visualisation	Excellent evidence of the concepts and principles of data visualisation	Exemplary evidence of the concepts and principles of data visualisation
D3 Evaluation: results, findings and conclusions from the exploratory data visualisation process	Unacceptable quality results, findings and conclusions from the exploratory data visualisation process	Poor quality results, findings and conclusions from the exploratory data visualisation process	Acceptable quality results, findings and conclusions from the exploratory data visualisation process	Good quality results, findings and conclusions from the exploratory data visualisation process	Excellent quality results, findings and conclusions from the exploratory data visualisation process	Exemplary quality results, findings and conclusions from the exploratory data visualisation process
D4 Communication: well-structured presentation of data visualisations with accompanying text in coherent, well-written English	Unacceptable presentation of data visualisations or accompanying text	Poor presentation of data visualisations or accompanying text	Acceptable presentation of data visualisations and accompanying text	Good presentation of data visualisations and accompanying text	Excellent presentation of data visualisations and accompanying text	Exemplary presentation of data visualisations and accompanying text

