

Module: COMP1800 Data Visualisation	Coursework
Contribution: 70% of course	ZIP file required, containing PDF report & Python notebook
Module Leader: Dr Chris Walshaw	Due date: Friday 26th March 2021
This coursework should take an average student who is up-to-date with tutorial work approximately 35 hours	
Learning Outcomes: <ol style="list-style-type: none"> 1. Identify and discuss fundamental concepts related visualization in general and visual analytics in particular. 3. Design, implement and evaluate Interactive Visualization Systems and visual analytics solutions. 4. Apply visualization tools and techniques to obtain insight from data sets. 	

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 data exploration for ChrisCo, the fictional company whose sales and website data we have been analysing throughout the course, using a Python Notebook (in Colab or Jupyter) and producing visualisations of store / customer data.

The dataset concerns the company's 40 stores, each identified by a unique 3 letter code (e.g. ABC, XYZ, etc). However, each student on the course has their own, randomised dataset to explore, and the codes are randomised so that a store code in one student's dataset is very unlikely to represent the same store in another student's.

Data

You will find your data in the following csv files, where **BannerID** is your student ID number (e.g. 001234567):

- <https://tinyurl.com/ChrisCoDV/BannerID/DailyCustomers.csv>
listing the daily number of customer visits to the company's 40 stores
- <https://tinyurl.com/ChrisCoDV/BannerID/StoreMarketing.csv>
the total annual spend on local marketing for each store
- <https://tinyurl.com/ChrisCoDV/BannerID/StoreOverheads.csv>
the total annual cost of overheads for each store
- <https://tinyurl.com/ChrisCoDV/BannerID/StoreSize.csv>
the store size (floor space) in metres squared for each store
- <https://tinyurl.com/ChrisCoDV/BannerID/StoreStaff.csv>
the total number of full-time staff employed at each store

Please contact your tutor if you cannot find your data files.

You should compile your data into two dataframes: one containing daily customer data (one row for each date); the other compiled from all of the .csv files into a dataframe of summary data (with a row for each store).

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.

The company is most interested in the large and medium sized stores but would like a summary of the small stores plus any anomalies you identify in the data. You should also identify new stores that have been opened during the year or stores 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 (no more, no less). 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 how best practices were demonstrated and applied (about ½ a page).
- A summary of the **conclusions** you have made about the data points (no more than ½ a page). You are **not** required to make any business recommendations and the summary may contain conclusions as bullet.

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.

Each visualisation 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.

The order of the visualisations should be carefully considered, leading the reader through the data exploration step by step and ideally with each visualisation leading on to the next one.

Notebook

Your Python 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 **clear description of available user interactions**.

Deliverables

You must upload a **single zip file** containing:

- The pdf report containing your 8 chosen visualisations
- A supporting Python notebook (.ipynb) containing your data exploration

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.

Task	achieved well	partially achieved	poorly/not achieved	marks
Report text (50%)				
Introduction to data visualisation				/10
Discussion – justification of visualisations chosen				/10
Discussion – description of findings				/10
Critical review				/10
Data conclusions				/10
Report visualisations (20%)				
Presentation quality (labelling, legends, etc)				/10
Impact (as part of the exploration)				/10
Notebook (30%)				
Organisation and presentation				/10
Code efficiency (non-duplication)				/ 5
Interactive visualisations – functionality				/10
Interactive visualisations – description				/ 5

Grading criteria

70-100%	All requirements completed to an excellent standard
60-69%	All requirements completed. However, there are a number of minor deficiencies in significant areas.
50-59%	All requirements completed. However, significant improvements could be made in many areas.
40-49%	All requirements completed. However, significant improvements could be made in all areas.
30-39%	All requirements attempted but the overall level of understanding and performance is poor.
0-29%	There are requirements missing or completed to a very inadequate standard which indicates a very poor or non-existent level of understanding.

The report should be succinct and so **must not contain more than 8 visualisations**, although you may use the technique of faceting (i.e. a number of subplots in a single figure). 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.