[Document subtitle]

**ASSignment 1: data visualisatiaon**

**An analysis of the world of nations gross domestic profit through Microsoft Excel and Data Visualisation with Tableau in accordance with the GDPR and Data Protection Act regulations**

Contents

[0](#_Toc144655319)

[Contents 1](#_Toc144655320)

[Assignment Part 1- Policies and Procedures in Accordance with GDPR and Data Protection 1](#_Toc144655321)

[Page 2 4](#_Toc144655322)

# Contents

# Task 1- Policies and Procedures in Accordance with GDPR and Data Protection

**Task 1**: To research the policies and procedures when handling data under law, privacy, and protection of handling client data within the UK and the general regulations to follow when working as a data analyst handling client, customer, and company data.

**Answer:** For this task students were asked to conduct research on policies and procedures in accordance with handling and working with data. I conducted research, and this brought me to the [www.gov.uk/data-protection](http://www.gov.uk/data-protection) regulations. I learned about the Data Protection Act 2018 which offers regulations and guidelines to how employees handle personal information in organisations, businesses, or governments.

Upon researching these guidelines, I learned that the information for everyone handling personal data includes using their data lawfully and transparently. The data must only be used for specified, explicit purposes and used in a way that it is only necessary for the given professional organization. Furthermore, the data must be removed when it is no longer necessary as it introduced a right for individuals to request for their details to be removed.

Reflected throughout my professional experience working as a recruitment administrator and database systems assistant an example here then might be if a client was asking for their database profile on our recruitment software to be closed including all their personal details and contact details removed. Under certain circumstances, a company can refuse to delete data however if it is that is being used if it meets within the guidelines of being necessary and limited to use if the user repeatedly requests to the delete the data they have shared if it an unfounded reason or excessive.

Systems used include proper authorising processing and security software to ensure that there is no loss or destruction. Furthermore, there is a strong emphasis on the legal protection of race, ethnic background, religious information, sex life or orientation and biometrics, where failing to comply could result in a criminal offence and prosecution.

Further researching this, an individual under the Data Protection 2018 has the right to ask the handler in question how their data is being used and represented as well as access to their personal data, with access to restriction and confidentiality protection.

They can also ask for the data to be erased or restrict the processing of the data if the need be data. Furthermore, approval from a business or company needs to be gained so we can they can request the data to be open source so it can be used for different purposes. The data that we use as analysts are used for decision making processes, and predicting of behaviours across for example, customer data.

For the specific given task, the wealth of nations data being used here has been approved to be used in an open-source format across projects, conducting a filtering and analysis of the GDP gained for the given countries in the source. The data that I worked with across this project also met the guidelines given that the identities of those who provided the data, and their sensitive information was handled with care. The data is not used for anything more than providing an Excel analysis of the GDP across countries in terms of wealth to produce a data visualisation report for a GITHUB project using Tableau, therefore is complies with the policies and procedures highlighted in the Data Protection Act. The data is being used by multiple students training on the bootcamp I took part in to create this data visualisation and should this data be changed and removed from public access, we would no longer be doing work with the data.

On a closing note, I also searched what procedural practices would be best as an analyst, part of the procedures includes making sure the data we are using is presented and collated in a format that reflects good data quality. General presentation of data in procedures should include removing duplicates and errors, and visualisations produced to apply to both a technical and non-technical audience, with accessibility options for audiences who may have visual impairments and other disabilities, making it accessible. This also means using a data processor that is safe to handle the data of the business and client and ensuring that consent is taken from the subject in handling their data. This concludes my research on how employees in the workplace in accordance with law must handle personal, organizational, and business data in accordance with the Data Protection Act 2018, further elaborating on the etiquette and the procedural usage of ensuring the data is presented and handled with care by a data analyst.

## Task 2- Excel Analysis of Wealth of Nations Data

**Task 2:** For this task, I was asked to take the Wealth of Nations data that contained the gross domestic product ranked with 228 countries in 2019, create a visualisation for it, then filter and sort it to the Top 20 countries with the highest GDP, produce another visualisation there. I was then asked to create three macros, Print, Save, Copy, and then using the copy macro to copy the sheet there that I had with the filtered table and the visualisation to a word document and call it Excel Gross Domestic Report 1, finally I password protected the worksheet, in accordance with the data security and protection previously studied from the first section of the project and closed the file, as it was ready to be transported to Tableau for data visualisation aspect of the task.

To ensure that the data was of the highest quality I first formatted the Gross Domestic Product column to show the correct currency, changing it to British pounds to reflect the correct format for the currency as it was previously in dollars, after checking the data was in the right format types, I then turned the range into a table. The first task was to create a table to sort and filter the information here. I created a table and filtered the table to show only the information of the GDP PP in British Pounds for 2019, with each country Rank during that year and country name.

The procedure followed off as to create a visual to represent the data for 2019, of which I initially struggled with given the fact that 228 countries from 2019 would be quite difficult to present on a chart that can only show so much data. This resulted in experimentation with different visual types, I found as the task required me to only show Rank, GDP, and country, that the combo type chart did not work because the year was also represented on that visual. I then tried a pie chart, but there were too many slices to make the visualisation clear. Thus, I concluded for a bar chart would work the best there with a massive legend, a filter would work but then it would only show the GDP per that specific country and if a different rank was selected with a different country, at the time I didn’t have the knowledge of how to link the relationship between the two so the filter would display the rank and the GDP for each country. I then gave the chart a title name and coloured it with appropriate visuals using a multi coloured colour for the bars there to represent each difference in colour for the legend for each country in 2019. There was also a view issue, with the chart being too massive to view fully, so I zoomed out and tweaked the fonts so that it could be seen when the zoom out was to its max, even then the visual could not be seen without side scrolling, but it came up with the best result.

Following through, I proceeded to filter the table to show the Top 20 ranking countries with GDP, I did this by filtering the GDP column in the table to show the Top 20, which resulted in the Top 20 ranked countries and ranks there. From there I was asked to produce a chart there on the same sheet to represent the GDP amounts along with country and rank there. This ended with a bar graph there using a dark blue tone and an easy to visualise format to represent this. I then went further and experimented with another visualisation that would work with the Top 20 countries and found a heat map worked very well there, demonstrating creativity and different ways to represent data.

The final stage of the project involved the creation of three macros. I created three shapes there and formatted them to match the colour scheme of the rest of the sheet with the table and the graphs, I then labelled them “Save”, “Print” and “Copy”. I had previously not used macros, so it was a good learning experience. I assigned a macro for each of the following shapes there, recorded the macro of me copying, saving, and printing the file and then assigned them to each shape there. This resulted in me being able to click each of the buttons there and perform the following actions. I then used the copy macro to copy and select the entire sheet with the GDP Top 20 table and the chart there and then pasted it onto a word document, I was asked to name the file there Excel Gross Domestic Report 1 and saved the file and closed it. I also made alterations to the header inside the Excel File there, placing in the given project names there. With this part of the task finished, the data filtered and ready to create a Tableau visualisation dashboard.

## Task 3- Tableau Dashboard

Link to Dashboard: <https://public.tableau.com/views/TableauGDPTop20Countries-2019/Dashboard2?:language=en-GB&publish=yes&:display_count=n&:origin=viz_share_link>

**Task 3:** To create a visualisation of the wealth of nations data from the Excel analysis conducted upon it. The visualisation must use visuals that work with those who are colour blind and must be clear, concise, and easy to read displaying 4 visualisations with the data from the Top 20 countries for 2019 sorted table chart.

**Answer:** I conducted research when designing my visuals to meet the requirements for a colour-blind client. After much research I found that the tones of black, white, and blue, red, or orange were ideal. Certain colours of brown also worked. I further researched the properties of the Tableau software colour palettes and found that indeed there was a colour palette that was catered for colour blind users so that they could see the information and visualisations formatted clearly. I opted for a minimalistic dashboard that was less so dynamic but conveyed important areas of information with direct analysis instead as I found my dataset was limited, so rather than a generalised view there with visuals affecting each other through filters, I chose to represent certain discrete pieces of information involving facts and figures from the data. I first created a bar chart using the colour scheme there that showcased the Top 20 countries and the Gross Domestic Product Per Capita that they were outputting. I used white for the background and red for the titles across the visualisation, for if the report was printed in black and white the client viewing it could still see the differences in the colour shades as red would appear darker than the light grey and white overtones of the rest of the canvas there. I made sure the labels were clear and the axis was labelled appropriately so that they knew which information to view and sorted the bar chart into ascending order. The second visualisation was a heat map consisting of the Top 5 countries with the highest GDP output, but instead of displaying their capita I opt to choose to display their percentage over the top 5 instead, with Monaco taking the highest of £32 percent there, again opting for the same colour-blind palette consisting of reds, blues, oranges, and browns.

For this I used the percentage and rank data labels and then also the filters there to display the Top 5 for GDP per capita and their percentage of the total 100 percent that made up this percentile. The third visual I deemed appropriate was a map but again decided to drill down into the data for a closer look at the bottom 5 lowest economies of the data there. I once again used a filter, locked the map using the same visual palette and went with a dark map there that was grey with not too many colours to cause confusion, rather labelling the 5 countries with the lowest GDP across our dataset there, allowing the user to also hover over the map there and receive information about that country. The final visualisation was in fact a table, with the calculated average of the GDP across the Top 20 countries. When conducting an analysis, this figure would be useful, as the average GDP output would allow the world to know what average figure is being outputted by 20 countries across the study from their total, giving them an idea of what the world’s economy looks like.

This concludes my final part of the project.

## Reflective Account

This task was beneficial in helping me learn about the GDPR and Data Protection Act in accordance with how the government and the UK law requires data to be handled regarding organisations and companies. Furthermore, I learned about the policies and procedures of making sure that the data quality is fit for presentation in the manner of data quality, removal of errors and accessibility for all as seen by the third task where we were asked to create a visualisation for a colour-blind client.

During the second task some learning points would include applying and increasing my Excel skills of filtering and producing visualisation as well as being able to apply my understanding of Macros throughout the Just IT skills Data Technician Bootcamp and how to automate processes in Excel used in the “Save”, “Print”, and “Copy” section of the task there.

Admittedly I struggled with the initial first task of fitting 228 countries on a legend into a visualisation, after some experimentation however I managed to create a pivot chart and adjust the sizing so that most of the chart could be viewed. I also then went further and created a heat map to experiment with the different ways that unfiltered data could be used when there were too many countries on the legend, in the future I would hope that I could find an effective manner of representing large sources of information on charts with a limited visual capacity. After sorting and filtering for the Top 20 highest ranking countries of GDP per capita, I produced a visualisation there thereby completing the word part of the task as well, learning how to transfer information from Excel spreadsheets and represent them into different formats for printed use.

The final stage of the task put my Tableau skills to use. I was challenged to research the correct colour scheme for the client so that the dashboard would look appealing to them. After research I found that red, orange, blue or grey and white colour schemes worked the best since these schemes could easily be identified so opted to use the colour-blind scheme in the Edit Colours section of Tableau. Although for the most part the visualisations were simply me applying the skills, I learnt on the bootcamp, I struggled with the composition of the full dashboard and to fit all the visualisations upon it so that it could be viewed in the aspect ratio. After some research I found that sometimes changing the aspect radio and the view from Entire View to Fit to Height or Fit to Width worked better to have each visualisation fit and after testing both the floating options and the tiled options for the visualisation placements, I eventually found a dashboard view that worked. Another challenge that I came across was due to having such a limited dataset of Top 20 ranking countries with their GDP information, I was unsure how to represent a visual that uniquely identified a different piece of information there without copying the same information being represented in a different visual.

This is the reason why the filters are all locked on the visual and I opted for more discrete information, splitting my dataset for closer analysis. I chose to analyse the Top 5 of the countries there and their percentages that made up the Top 5, to demonstrate which of the countries were producing the highest percentage of GDPs and the ranks. The bar chart meanwhile produced a general overview with clear data labels of each GDP and country name whilst the map focused on the 5 lowest earners there in the table, which again used the filter function there through the cards that I placed into the filter section there in the options with “Top” there. The final visualisation was a table, with the aggregation of the average GDP across the Top 20 countries there formatted in clear blue that matched and contrasted with red and grey overtones of the visualisation. In the future, I should like to improve the dynamic and interactivity concepts of the dashboard there and find creative methods to analyse and demonstrate visualisations with a limited dataset.

To conclude, I have learned under what policies and procedures an analyst must work under when cleaning and formatting data as well as producing visualisations for clients, the challenges faced when deducing what kind of questions to ask from data and how I can improve upon that aspect and how to create visualisations that are accessible to all, including that with colour blindness in an easy to understand format.