A picture containing text

Description automatically generated

**Assessment**

**Data Analysis.**

# Module 1:

### Outline:

During the first three weeks of your Bootcamp, you will be exploring Data Analyst roles, what they involve, and typical duties involved in data analysis. For this module, you will be required to define the role, tools and common workflows of a data analyst through multiple choice questions, as well as a project.

## Week 1 Task (1.1, 1.3, 1.4, 1.5)

### Project Outline:

Research Data Analyst jobs online. This should include junior as well as senior positions in the UK. Note down **five** job roles that have caught your eye and discuss what stood out for you. Complete the table below and submit this workbook onto EQUAL. An example has been provided for you.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Job Title** | **Salary** | **Level** | **Technical skills** | **Tools** | **Interpersonal skills** | **Location** | **Website link** | **Your discussion of the job and what has stood out for you** |
| Data Analyst Apprentice | Unspecified | Junior or senior | Administering databases, data validation | Microsoft product, Excel, etc. | Good communicator | London | <https://www.totaljobs.com/job/data-analyst/manchester-airport-job98214482> | Great company to work for, achievable with some practice or finishing Bootcamp. Might need to look at learning specified tools/methods soon. |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Job Title** | **Salary** | **Level** | **Technical skills** | **Tools** | **Interpersonal skills** | **Location** | **Website link** | **Your discussion of the job and what has stood out for you** |
| Junior Data Analyst | 23K-27K | Junior | Microsoft Office and Data Analysis | SQL,Python,Power BI,JavaScript, HTML and CSS. | Teamwork, Problem-Solving, Proactive Approach, Continuous Improvement. | Manchester | [Junior Data Analyst - Manchester - Reed.co.uk](https://www.reed.co.uk/jobs/junior-data-analyst-manchester/51715271?source=searchResults&filter=/jobs/jobs-in-didsbury&utm_campaign=google_jobs_apply&utm_source=google_jobs_apply&utm_medium=organic) | There is Flexitime/Hybrid working – 2 days in Manchester office, social events, and bonus scheme. |
| People Data Analyst | 32K | Senior | Database Knowledge, Research Skills, and Data Analysis | PowerBi | Communication, Empathy, Problem-Solving. | Preston | [People Data Analyst - Preston PR2 - Indeed.com](https://uk.indeed.com/viewjob?jk=18f66cd51df15c51&tk=1hgoojh9qkj9d802&from=serp&vjs=3) | There is a people-centric culture and inclusivity for both customers and employees. |
| Data Analyst | Unspecified | Not Applicable | Data Management and Data Manipulation | Microsoft office, SQL, VBA, and Power BI | Active-Listening, Presentation, Communication, Attention to Detail, Numeracy and Time Management. | Manchester | [Collinson Grant hiring Data Analyst in Manchester, England, United Kingdom | LinkedIn](https://uk.linkedin.com/jobs/view/data-analyst-at-collinson-grant-3775705371?utm_campaign=google_jobs_apply&utm_source=google_jobs_apply&utm_medium=organic) | It has a Results-Driven Environment which focus on delivering value and long-term client relationships. |
| Data Analyst | 22K-52k | Senior | Database Knowledge, Data visualisation, Data mining (Basic Level), and Pattern Recognition. | SQL | Communication, Attention to Detail, Problem-Solving and work under pressure. | Manchester | [Data Analyst Job in Greater Manchester, Career, Permanent Jobs in Altus Group (onlydatajobs.com)](https://onlydatajobs.com/job/300787/Data-Analyst-Greater-Manchester-Manchester?utm_campaign=google_jobs_apply&utm_source=google_jobs_apply&utm_medium=organic) | There is career progression with 150,000 hours of learning materials, Hybrid working, and a good bonus packages. |
| Data Analyst | 25-35K | Junior or Senior | Data Management and Scripting experience (PowerShell) | SQL | Attention to Detail, Communication and Problem-Solving. | Salford | [Data Analyst Salford - Reed.co.uk](https://www.reed.co.uk/jobs/data-analyst/51698930?source=searchResults&filter=/jobs/jobs-in-wythenshawe&utm_campaign=google_jobs_apply&utm_source=google_jobs_apply&utm_medium=organic) | AJ Bell is known to be one of the best 100 companies to work for six consecutive years, there are many employee benefits, and equal employment opportunities. |

## Week 2 Task (1.2)

### Project Outline:

Create a diagram to discuss the CRISP DM data mining lifecycle. Explain what happens at each stage of the life cycle from an e-commerce business perspective. Submit Word documents as evidence.

<https://www.ibm.com/docs/en/spss-modeler/18.2.0?topic=dm-crisp-help-overview>

As a minimum, you should discuss the following for each stage:

* Typical business questions based on business objectives (better product recommendations, better customer profiling, achieve higher average order value, etc.)
* What tools or techniques would be needed to achieve these
* Which departments would need to work together to achieve these goals.

Add the diagram to the following page or upload screenshots. Submit this workbook onto EQUAL.

|  |
| --- |
|  |

## Week 3 Task (4.4)

### Project Outline:

Understanding the differences in statistical thinking and methodologies is really important to be able to deal with key business issues.

The head of marketing has asked you to create a data analysis plan to forecast the next Black Friday sales online. Your task is to choose the right model and decide on the steps to take, and what kind of data you will need to feed your forecast. Refer to Week 2’s data mining life cycle and use the CRISP DM model to discuss the individual steps, data and stakeholders involved. Use the CRISP DM model to outline the stages, then discuss what will happen at each stage, as well as time estimates for delivering each stage. Use this template to outline your stages.

Complete the following table and submit this workbook onto EQUAL.

|  |  |  |
| --- | --- | --- |
| **Stage** | **Deliverable** | **Timescale** |
| Business understanding | What are the historical sales trends as black Friday approaches? | One week |
| Data understanding | May need Historical sales data for black Friday sales, the website traffic data or even the marketing campaign data. | Two weeks |
| Data preparation | There might be missing or an incomplete data set, there may be data inconsistencies across these sources or maybe the need of data cleaning | Three weeks |
| Modelling | To have regression models in place for quantitative analysis, this will help to predict the sales figures. | Four weeks |
| Evaluation | It provides the accurate sales insight for black Friday | Two weeks |
| Deployment | The model can then be merged into online sales platform to help promote more sales. | One week |

# Module 2 Practical Statistics

## Week 1 Task (2.1)

### Project Outline:

For this task, you are required to identify the kinds of structured and unstructured data a fashion e-commerce business might use. Please list examples of structured and unstructured data that might need to be collected about customer behaviour, e.g. social media comments or time on-page (minimum **five** sources of data).

Complete the following table and submit this workbook onto EQUAL.

|  |  |  |
| --- | --- | --- |
| **Data sources** | **Structured or unstructured** | **Use case** |
| Spreadsheets .csv, .xlsx. | Structured | Normally use internally to track performance and send out weekly reports to show sales performance. |
| Customer Purchase History | Structured | To analyse the buying pattern and recommending personalised products and make the targeted marketing to meet customer needs. |
| Customer Reviews | Unstructured | To analyse the product feedback, to identify the strengths and weaknesses from this and to make the improvements needed. |
| Image and Video Data | Unstructured | For this you use Al for the visual search, to understand the customer preferences and to enhance the product display for customers. |
| Time on Page | Structured | For this you see the user engagement and evaluate it. This helps to identify any popular products to try and optimise the website content. |
| Social Media Comments | Unstructured | Where you would sentiment analysis, try and understand the feedback and based on this to improve the brand reputation. |

## Week 2 Task (2.2, 2.3)

### Project Outline:

The UK GDPR sets out **seven** key principles:

* Lawfulness, fairness and transparency
* Purpose limitation
* Data minimisation
* Accuracy
* Storage limitation
* Integrity and confidentiality (security)
* Accountability.

Your task is to research and discuss how the UK GDPR principles would apply to a fashion e-commerce business. You should write an example for each of the principles, e.g. Accuracy could be ‘an e-commerce business should have the most up-to-date information about a customer's address so that the item will be posted to the correct address.’ Complete your work in the box, or use a separate Word document as evidence. Upload your work to EQUAL.

|  |
| --- |
| 1. Lawfulness, Fairness and Transparency  * An example would be the e-commerce business must clearly communicate the data collection practices to the customers during the initial registration process. This may include informing the users about the types of personal data that is being collected, the purpose of the collection and the need to get their explicit consent before processing the data.  1. Purpose Limitation:  * The business should only collect the customers’ data for legitimate purposes only. This may be to manage accounts and provide customer support. The data is not used without explicit consent.  1. Data Minimisation:  * It is important for the business to only collect the minimum amount of personal data for its purpose. An example would be if a customer is going through the checkout process, the information that should be requested is shipping and payment, and to avoid any unnecessary details that are not needed during the transaction.  1. Accuracy:  * The business should ensure that the customer data, such as the shipping address, is correct and up to date on the system. So, that there are no issues due to outdated addresses.  1. Storage Limitation:  * The business should have policies in place to state how long customers’ data will be stored. For example, they may need to retain an order history for a certain period. This may be due to refunds or returns. But any data that doesn’t require to be stored should be deleted.  1. Integrity and Confidentiality (Security)  * There must be security measures in place to protect any customer data from being accessed by unauthorised users. This may include the need to encrypt sensitive information, for example card details.  1. Accountability:  * The business should ensure that there is a Data Protection Officer (DPO) to ensure GDPR compliance. There should be regular audits of any data processing activities and also there should be staff training put in place. |

## Week 3 Task (3.5)

### Project Outline:

Your manager requires you to list personally identifiable data and how it would be protected, e.g. spreadsheet, pdf file, etc. Create a report on how data types and associated personally identifiable data, as well as business secrets, would need to be safely stored, archived, and safely distributed across the business; e.g. personally identifiable data only being transmitted over VPN or using TLS/SSL certificates over email and web. Create your list in the box provided or on a separate Word document. Submit your work to EQUAL.

|  |
| --- |
| 1. Introduction:  * To outline any protocols and methods to store, archive and distribute personally identifiable data (PID) any business secrets in the organization.  1. Personally Identifiable Data (PID)   Types of PID – Names, Social Security numbers, addresses, contact info and financial information.   * Storage: It is important to have encrypted databases and to have strict access controls so that only some authorised users can view, edit, or delete the PID. * Archive: There is the importance of regular backups to ensure data integrity and availability. You must ensure that you use password-protected and have encrypted archives for any long-term storage. * Distribution: The PID should only be transmitted over a secure VPN and the emails should be encrypted with the TLS or SSL certificates.  1. Business Secrets:  * Storage: Important to restrict access so only some authorised users have access to business secrets, and the importance of having secure servers to host any business secrets with multi-factor authentication. * Archive: It is important to have version control in place to track any changes and to ensure the integrity of the business secrets. There is the need to maintain backups in diverse locations in the case of data loss. * Distribution: Secure file sharing platforms are needed to share business secrets; these are encrypted channels which are important to ensure that any data is kept confidential. It is important that employees or third parties sign the NDA before gaining any access to personal information. |
|  |

# Module 3 Data Wrangling

## Week 1 Task (3.1, 3.2)

### Project Outline:

Your manager is currently looking for a new database system as the current one is very slow, and wants to see what alternative databases are available, how they would benefit a business and what would be classed as a good use case. Research into new databases and give examples of their pros and cons and what activity they would be most useful in. The table has been started for you. Complete the table and submit this workbook onto EQUAL.

|  |  |  |  |
| --- | --- | --- | --- |
| **Type of database** | **What activity would this be used for?** | **Pros** | **Cons** |
| Graph databases | When trying to analyse a network of people for a social media application. | It is efficient to handle any complex relationships and the network structures. | It may not be suitable for simple, tabular data structures. |
| Relational databases | It can be used for applications that are complex queries or transactions | It is a mature technology and has a strong consistency. | It may have a learning curve and may also be less flexible for any hierarchical data. |
| In-memory databases | For real-life analytics | There is a fast data retrievals where it supports various data structures. | There is limitation to the RAM size, which means that data may not persist on the disk. |
| Column-family stores | Used for large-scale data | It is highly scalable | There is complexity in managing clusters which may mean that it cant be used for small-scale applications. |
| Cloud databases | It helps to deploy and manage database in the cloud environment. | There is high availability and scalability. | There is limited control over the infrastructure which could mean a vendor lock in. |

## Week 2 Task (3.4)

### Project Outline:

Your task is to select appropriate data types for the following e-commerce database tables (list the corresponding values like in the example):

**Orders table**

Order\_ID

Order\_date- DATE (example)

Order\_value

Customer\_ID

**Customers table**

Customer\_ID

First\_Name

Address

Phone

Email

Age

Submit a spreadsheet that reflects the work you have been doing. You should submit all work to EQUAL.

|  |
| --- |
|  |

## Week 3 Task (3.3)

### Project Outline:

Using this e-commerce database (<https://www.programiz.com/sql/online-compiler/>), complete the following tasks.

1. Create a query which asks for all customers/records named ‘John’ from the ‘Customers’ table.
2. Create a query to return all the columns from this table for only customers in the UK who are 22 years old.

Structure your query like so:

SELECT \_\_\_\_\_

FROM \_\_\_\_\_

WHERE \_\_\_\_\_ AND \_\_\_\_\_

1. Generate a report which displays all customers ordered in ascending order by age.
2. Create an inner join which returns Customer’s First Name and Amount Spent On An Order. This should be joined at Order ID and Customer ID.
3. Add an entry to the Customers database with Boris Johnson as a customer. You should insert an ID with his first name, surname, age and country.
4. Delete the row you have added to the database with Boris Johnson's details.

Your final report should include screenshots of each query input and result, and a brief evaluation of how you structured your SQL code. You should submit this workbook to EQUAL.

|  |
| --- |
|  |
|  |

# Module 4 Data Visualisation and Data Modelling

## Week 1 Task (4.1, 4.2, 4.3)

### Project Outline:

Working with the **E-commerce** **Customers-cleanse.csv**, which can be found in the ‘Learn More’ tab, identify data errors in the data set and cleanse it from any errors. Make sure to look for empty fields and replace them with appropriate values (e.g. averages replacing missing values).

Submit a cleansed spreadsheet onto EQUAL.

|  |
| --- |
| D, |

## Week 2 Task (4.5)

### Project Outline:

Your manager has asked you to create a dashboard to track performance. Using a free version of Tableau, PowerBi, Excel or any other free visualisation software, create a dashboard from the following data set, which can also be found in the ‘Learn More’ tab:

<https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls>

This dashboard should identify:

* The best performing states by sales and profit
* Least profitable categories by profit
* Identify customers with most orders
* Most popular dates for sales.

Your dashboard should include:

* Bar charts
* Scatter plots
* Pie charts.

Optional:

Include a map of states which are least profitable.

Your dashboard can be more than one page to fit the views.

Links to software for Data Visualisation:

<https://powerbi.microsoft.com/en-us/getting-started-with-power-bi/>

<https://public.tableau.com/en-us/s/download>

<https://www.microsoft.com/en-gb/microsoft-365/try>

<https://datastudio.google.com/>

Submit your dashboard as a PDF file. You should submit all work onto EQUAL.

|  |
| --- |
|  |

# Congratulations!

You have now completed your Data Analyst assessment workbook.

Please ensure that you have uploaded the full booklet to EQUAL for your final assessment.

You should also keep a copy of your workbook for your own records as evidence of your work. This may be useful in the future when you are applying for jobs and experience in your chosen field.