

Student Name

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Email Address

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Course

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Group

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Module tutor \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Communication Protocol:***

*module staff will reply to student questions within a reasonable time but this will*

*normally be within office hours only. Students are advised to check this Handbook and also to see if there are any*

*online announcements or FAQ answers that deal with thei*

*r enquiry before contacting staff.*

**Advanced Databases**

**Assessment**

**and Module**

**Handbook**

2022/23

Students

**Level**

**6**

**, Semester A**

)

Credits

(20

Contents

Contents

[1 What this Module is about 2](#_Toc111532710)

[1.3 Module Learning Activities 3](#_Toc111532711)

[1.4 Graduate Attributes Developed and Assessed 4](#_Toc111532712)

[2 Schedule of Work 5](#_Toc111532713)

[At a glance schedule 5](#_Toc111532714)

[3 Key Resources to Support Learning (see reading list on VLE) 7](#_Toc111532715)

[4.1 Introduction to the Assignment using the Police case study 8](#_Toc111532716)

[1. The Police Reporting Crime System (PRCS) (oracle database) 8](#_Toc111532717)

[2. Police System - Wales (see Appendix A also) 8](#_Toc111532718)

[3. A speadsheet of crimes in Leeds (see Appendix A also) 8](#_Toc111532719)

[Data Mart (DM) – Design and ETL considerations 9](#_Toc111532720)

[Assignment 1 – tasks -1-3 9](#_Toc111532721)

[Assignment 2 – Tasks 1-2 11](#_Toc111532722)

[Data Mart (DM) – OLAP, dashboard and DW approach 11](#_Toc111532723)

[Appendix A 13](#_Toc111532724)

[The police case study – data sets 13](#_Toc111532725)

[1) The Police Reported Crime System (PRCS) 13](#_Toc111532726)

[2) The Police System – Wales (PS-Wales) 14](#_Toc111532727)

[4.4 Reassessment 22](#_Toc111532728)

[4.5 Feedback 22](#_Toc111532729)

[5 Understanding Your Assessment Responsibilities 23](#_Toc111532730)

# 1 What this Module is about

**1.1 Welcome to Advanced Databases!**

On behalf of the module team, I would like to welcome you to this module, which we hope you will find both challenging and rewarding.

The module is based around the use of data for business decision making and business intelligence solutions and data analytics dashboards and applications. We will follow a database mart/ warehouse development lifecycle by firstly looking at what we want to investigate in the data mart along with the types of reports (requirements). With this information we can identify the data required and use it to design an appropriate data model (design), part of the design will involve checking (and dealing with) the data quality. The implementation will involve the creation of a data mart data model and writing SQL based ETL scripts to ‘extract, transform and load’ the data from the original data sources into the data warehouse model. At this point we can produce some reports and consider the best visualisation tools for this.

Along the way topics such as data management, data integration, data quality, the data dictionary, data maintenance, and data ethics will be discussed. The module has a good mixture of database expertise and business understanding – you will write SQL and you will discuss business strategies! As your module tutors, we aim to provide you with a coherent set of learning opportunities, which will enable you to develop your skills and knowledge in databases.

We hope this will be a valuable learning experience for you.

Jackie

Module Aims

* This module aims to build understanding and practical capabilities of database technologies for effective management and utilisation of an organisations data resource.
* The principles, techniques and concepts of data integration, extraction and data quality when implementing a database system are addressing theoretically and practically.
* Your knowledge of technical and practical development and the end-purpose of using a database system for decision support making will be developed.

**1.2 Module Learning Outcomes:**

On completion of this module the student should be able to:

1. Demonstrate a critical understanding of the theories underpinning a range of database design and management issues, including methods, technologies and emerging trends
2. Critically evaluate the role of development tools in the design, development and management of database systems
3. Assess the organisational context, roles and tools for effective utilisation of an enterprise’s data resource.

## 1.3 Module Learning Activities

Keynote lectures will be used to develop knowledge and understanding and to provide a discussion of students’ own research finding.

Tutorial sessions will be lab-based and will be mainly problem or enquiry based, allowing students to analyse, evaluate and discuss these technologies using case study scenarios, and to use and appraise applications and technologies. These activities will be both individual and group-based.

Students will discuss and present ideas to their peers, enabling tutor and peer feedback. Independent research and application will be expected as students will need to read around the subject in order to gain a wider understanding of the theory application of the technologies covered.

## 1.4 Graduate Attributes Developed and Assessed

* The Enterprise graduate attribute is both developed and assessed via problem based activity.
* The digital literacy graduate attribute is both assessed and developed via the investigation and selection of appropriate tools for the activities.
* The global outlook graduate attribute is developed in that students are required to consider the case study activities for an international market.

# 2 Schedule of Work

You will need to revise the material from L4 and L5 and remind yourself of database design and modelling. An introduction to database systems by CJ Date Chapters 1-4 is good – although others listed in the reading list are preferred by some students. There is also material on the VLE under Unit 1.

General background and appreciation to this module can be found in the book Big Data: A Revolution That Will Transform How We Live, Work and Think[Viktor MayerSchonberger](http://www.amazon.co.uk/s/ref=ntt_athr_dp_sr_1?_encoding=UTF8&field-author=Viktor%20Mayer-Schonberger&search-alias=books-uk&sort=relevancerank)  and [Kenneth Cukier](http://www.amazon.co.uk/Kenneth-Cukier/e/B00C47ZFSY/ref=ntt_athr_dp_pel_2) . Copies are in the library.

### At a glance schedule

|  |  |
| --- | --- |
| 1  26/9/22 | Decision Support systems, Analytics and the Data Warehouse. |
| 2  03/10/22 | OLAP and OLTP – Decision support  (Understand the Business and Data) |
|  | Formative upload 1 – upload SQL queries (see assessment guide by 10/10) – contributes to portfolio |
| 3  11/10/22 | Data Marts and Star schema  (Understand the Business and Data) |
|  | Formative upload 2 – upload draft star schema (see assessment guide by 17/10) – contributes to portfolio |
| 4  17/10/22 | ELT - Data Integration, Data Quality, Data Prep |
| 5  24/10/22 | Data Transformation and ETL (ETL)  (Data Prep and data model) |
| 6  31/10/22 | Slowing Changing Dimensions |
| 7  07/11/22 | Assignment 1 support |
|  | Assignment 1 – tasks 1-3 due in by 14/11/2022 |
| 8  14/11/22 | Using tools, tableau for data prep and data analysis  (data prep, data analysis)  Extracting data from the Data Mart for analysis – in Access/Excel and using Oracle Apex (data analysis)  Extracting data from the Data Mart for analysis – for dashboards Data Maintenance |
| 9  21/11/22 | PL/SQL – functions and Procedures |
| 10  28/11/22 | PL/SQL – Packages, applications, process and application |
| 11  05/12/22 | Data Management reflection  Assignment support |
| 12  12/12/22 | Assignment support by appointment or email |
|  | Assignment 2 – tasks 4-5 due in by 13/1/2023 |

# 3 Key Resources to Support Learning (see reading list on VLE)

Please see online reading list.

**4 ASSIGNMENT REQUIREMENTS - Design and considerations of a Data Mart -**

## 4.1 Introduction to the Assignment using the Police case study

The Police Force has a number of stations throughout England, each station is located in a regional area such as ‘Yorkshire’, or ‘Lancashire’. When a crime is reported it is immediately assigned to a ‘station’, this is based on the area where the reported crime occurred.

Reported Crimes may be new or may be existing crimes which are ‘open’ or ‘closed’. A reported crime tends to last between 1 month and a year, although some will last years. The crime will be reviewed yearly after the reported date unless it has a ‘closed’ date. All crimes belong to a specific station and have a Lead Police Officer. A Lead Police Officer could be managing more than one crime at a time. Some crimes may be escalated to a higher level that the ‘station’ in which case their status is marked as ‘escalated’.

You have been given a number of data sources.

### 1. The Police Reporting Crime System (PRCS) (oracle database)

A database management system used by the Police in England to record crimes (see Appendix A). The script to create these tables and load test data in on MyBeckett **PRCS.sql and PRCS\_insert\_test\_data.sql**

and

### 2. Police System - Wales (see Appendix A also)

This is a database management system for Wales only. It records similar information, but is slightly different. The script **PS\_Wales.sql** sets up the tables and some data for this system.

and

### 3. A speadsheet of crimes in Leeds (see Appendix A also)

A spreadsheet of summary data of crimes in Leeds.

Note: The data you have been given is very limited, this helps to understand it and to be able to verify your results. However, depending on the reports implemented you may end up with very little data. This is fine. However, if you would like you may add extra data.

Your role is as an analyst/ developer on a Data Mart (DM) project to support the design, analysis and collection of information relating to this Police case study.

### Data Mart (DM) – Design and ETL considerations

**The Police have a number of KPI’s (Key Performance Indicators) they would like to consider. These include:**

#### KPI 1: Reduce crime

This KPI is concerned with reducing crime. Examples of the types of reports they would like are:

* Number of crimes per year
* Number of crimes per crime type per year

#### KPI 2: Close crimes

This KPI is concerned with ensuring crimes are solved and complete.

* Number of closed crimes per year

#### KPI 3: Identify areas with crime hotspots

This KPI is concerned with understanding crime patterns. Examples of the types of reports they would like are:

* Number of crimes per station

Choose one of the KPIs above to focus your assignment around. Address all the tasks with respect to this KPI.

## Assignment 1 – tasks -1-3

**Task 1: Data Mart (DM) star schema design for your chosen KPI**

Identify 3-5 reports\* that your star schema will support.

* Document the star schema (SS) design model to support these reports – use QSEE.
* Use the data dictionary template from tutorials to document the data model the project.
* Select one of the reports\* you have suggested. Illustrate the expected data in the star schema to support the report - use Excel (or oracle or similar) to do this and add a few rows.

**[20 marks]**

**Task 2: Star Schema set up (DM environment) [5 marks]**

Use QSEE to forward engineer the database for the star schema (SS) database you have designed. Create and run a script to create the data mart tables (edited as appropriate).

Include QSEE generated script(s) as part of your upload along with screen shots as evidence of the code running successfully and documentation of any changes you have made or problems you encountered.

*See marking scheme for more direction.*

#### Task 3: Extract, Transform and Load (ETL) script to populate the Star Schema (DM) with data

The ETL is a script that puts the data into the DM tables. It does this by extracting the data from the original sources, transforming the data as required and then loading them into the DM tables.

Write an ETL script to:

* populate one - two dimension tables,
* the time\_dimension table and
* 5-10 rows of the fact table with measure(s) – this will depend on your own project.

To do this, identify one of the reports to support (ideally the one you have already planned the expected data for task 1). Your script should deal with 2-3 data quality issues, 1 transformation and include at least 1 measure/calculation for the FACT table. **[25 marks]**

Perform and provide evidence that you have successfully completed these tasks (via screen shots which show your student id or evidence of successfully run scripts, this work should be done in your own University apex account). Documenting any changes you have made or issues you have encountered.

**Assignment 1 upload: Please upload a word report addressing tasks 1-3, include any code as an appendix to this document.**

**Upload to MyBeckett by Monday 14th November 23:00.**

### Assignment 2 – Tasks 1-2

### Data Mart (DM) – OLAP, dashboard and DW approach

#### Task 1: Data Analysis/OLAP/Mining Investigation

#### Undertake some data analysis on the data from one of the data sources you have been given. In this report screen shot some key visuals and also include some written interpretation of the visual (show and tell). Either:

1. Upload the case study spreadsheet into MS Excel (or tableau). Create a pivot table and produce some interesting (and appropriate) reports using the charts and visualisations functionality. *Use literature of data analysis, business intelligence and OLAP to support and drive this task*. You may include (or discuss) external data as

well. **[20 marks]**

OR

1. Using Apex create a dashboard for The Police system. *Use literature on data analytics and dashboards to inform the design*. The dashboard should support your chosen KPI. You can use the data from the Data Mart tables or the source data tables.

**[20 marks]**

#### Task 2: Take your design and code further by using PL/SQL techniques

1. Identify, code, test a function, procedure and package to support the ETL task.

**[20 marks]**

S*ee marking scheme for more direction.*

**[30 marks]**

S*ee marking scheme for more direction.*

**Assignment 2 upload: Please upload a word report addressing tasks 1-2, include any code as an appendix to this document.**

**Upload to MyBeckett by Friday 13th January 2023 23:00.**

#### Evidence of portfolio uploads

There are 2 key formative uploads:

1. SQL practice – upload by 10/10/22

Upload your code and evidence of the code running successfully in oracle apex for these queries:

* Number of crimes per year per month
* Number of crimes per crime type per year
* Number of crimes per station per year

Aim to use useful titles to ensure the report is meaningful

1. Star Schema model – upload by 17/10/22

* Upload a list of 3 reports your star schema model supports and the star schema model to support these reports (created in QSEE and as a screen shot).
* This can be used for your assignment. You may complete more of task 1 for this if you like.

These will be reviewed, and formative feedback given.

## Appendix A

# The police case study – data sets

# The Police Reported Crime System (PRCS)

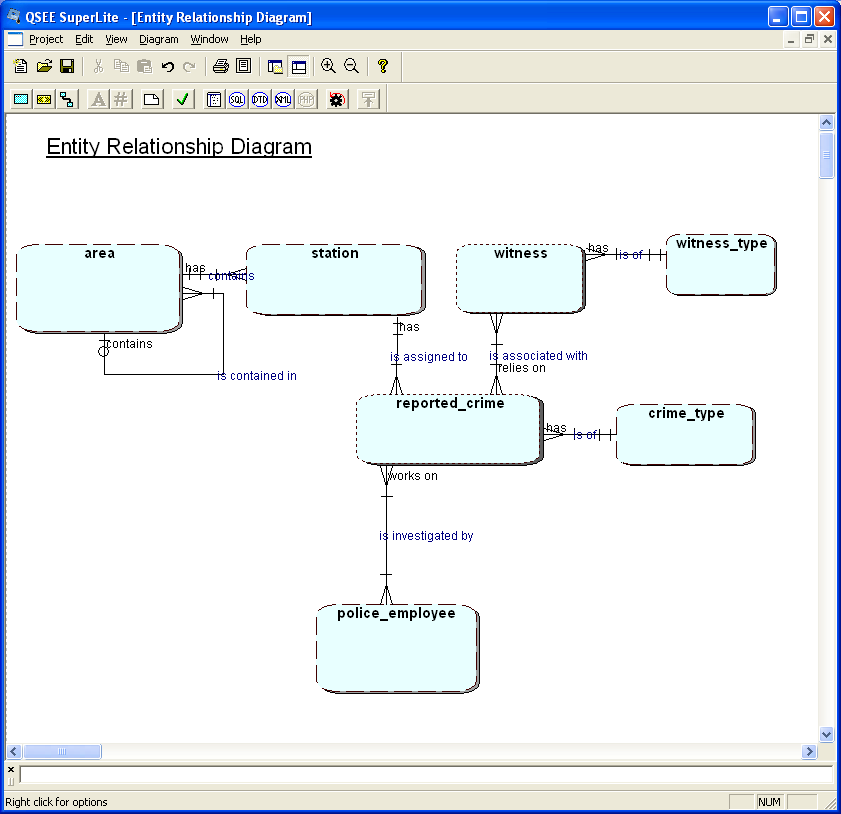
The Police Force has a number of stations throughout England, each station is located in a regional area such as ‘Yorkshire’, or ‘Lancashire’. When a crime is reported it is immediately assigned to a ‘station’, this is based on the area where the reported crime occurred.

Reported Crimes may be new or may be existing crimes which are ‘open’ or ‘closed’. A reported crime tends to last between 1 month and a year, although some will last years. The crime will be reviewed yearly after the reported date unless it has a ‘closed’ date. All crimes belong to a specific station and have a Lead Police Officer. A Lead Police Officer could be managing more than one crime at a time. Some crimes may be escalated to a higher level that the ‘station’ in which case their status is marked as ‘escalated’.

Attributes have not been documented. They can be inferred from the columns defined at the logical design stage. The script to create these tables and load test data in on MyBeckett **PRCS.sql and PRCS\_insert\_test\_data.sql**

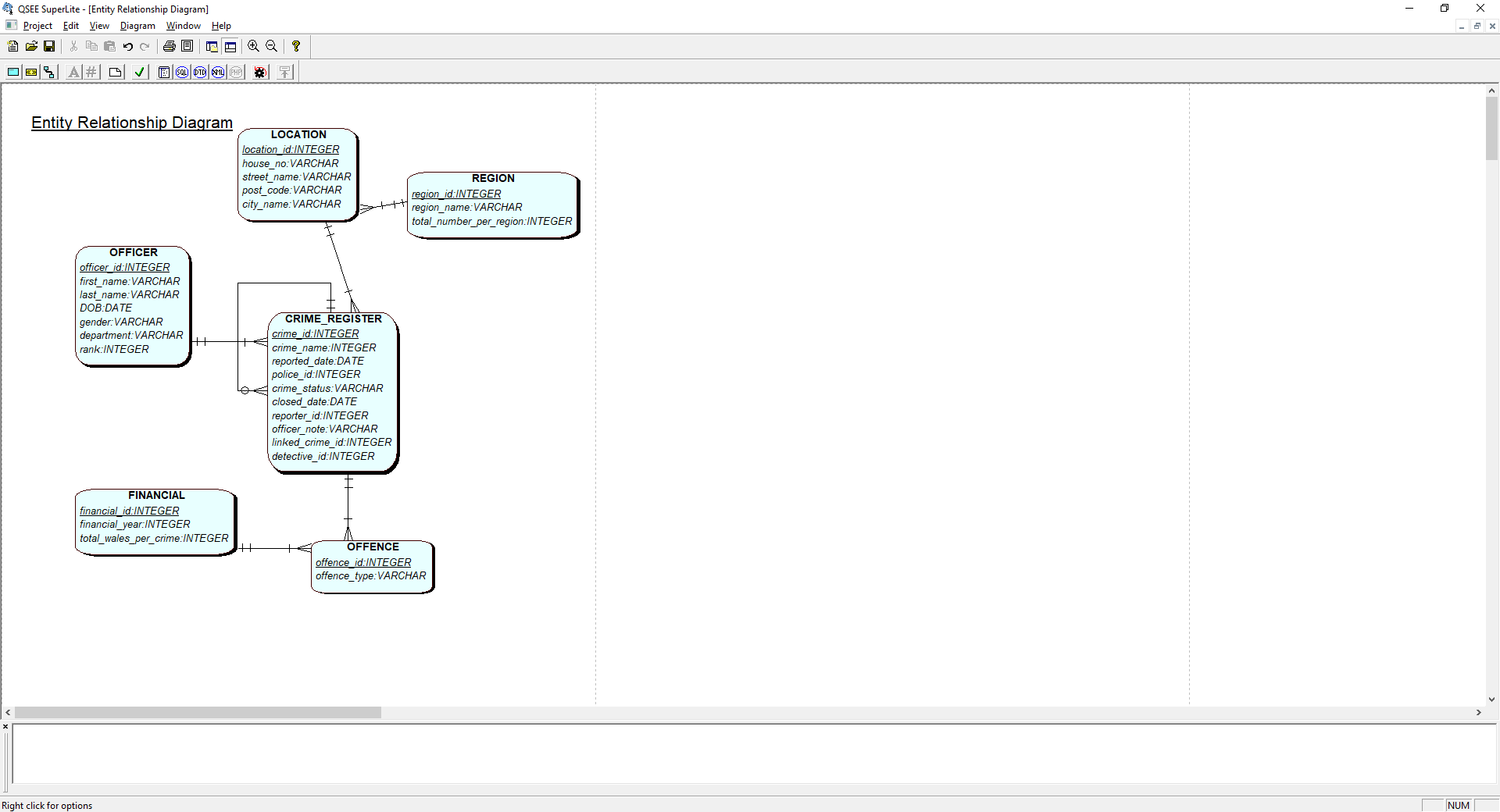
The conceptual model for PRCS (England only)

Note that for the logical and physical model M:N relationships will usually include a link table.



# The Police System – Wales (PS-Wales)

In Wales, there is a slightly different database system. See the model below:



The script to set up and run the Police System Wales is **PS\_wales.sql**

The online version of the PS-Wales (for info only)

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

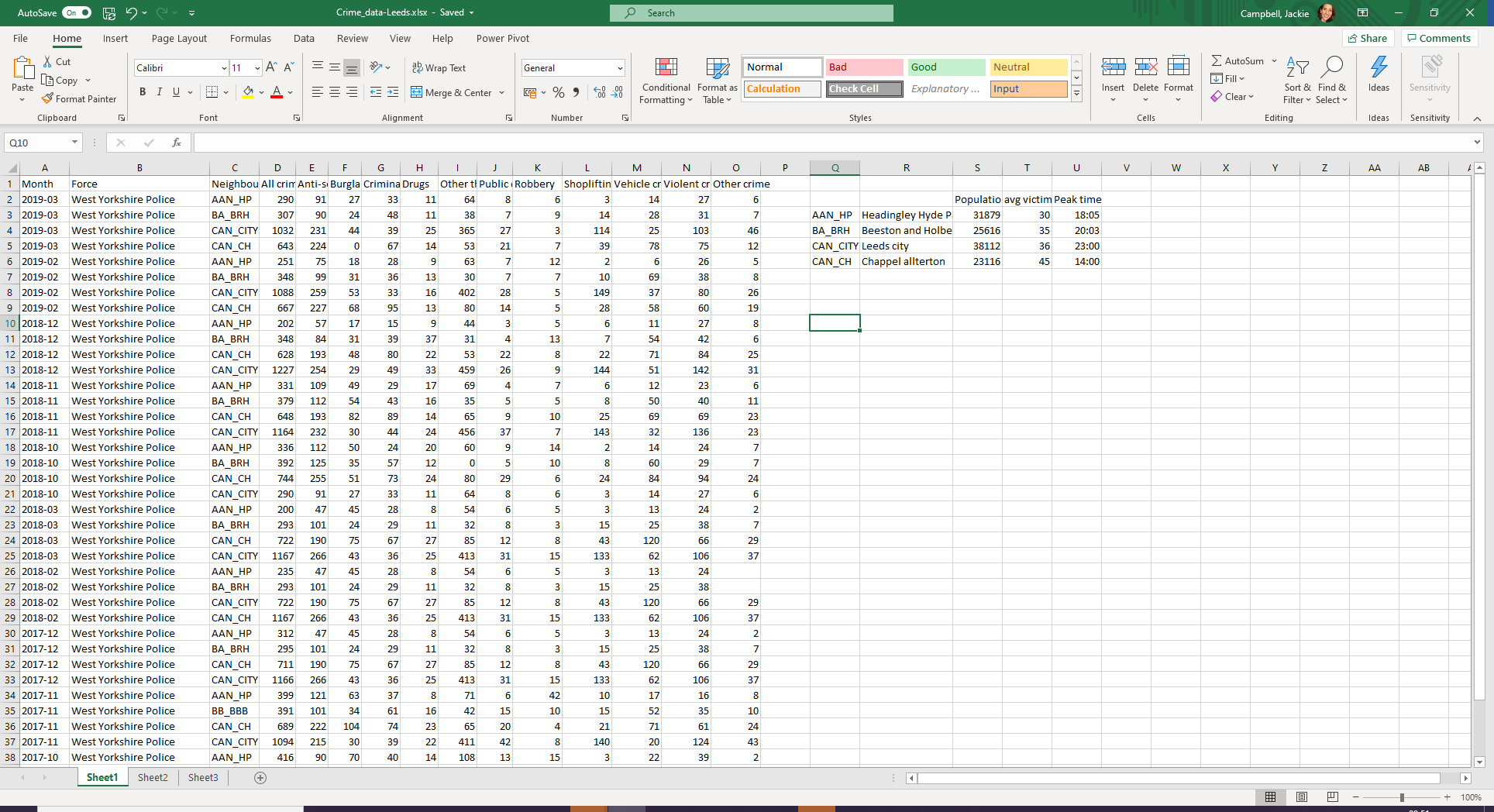
Graphical user interface, text, application, email

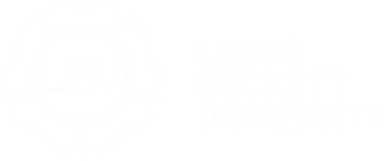
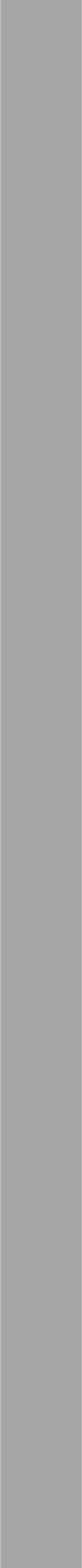
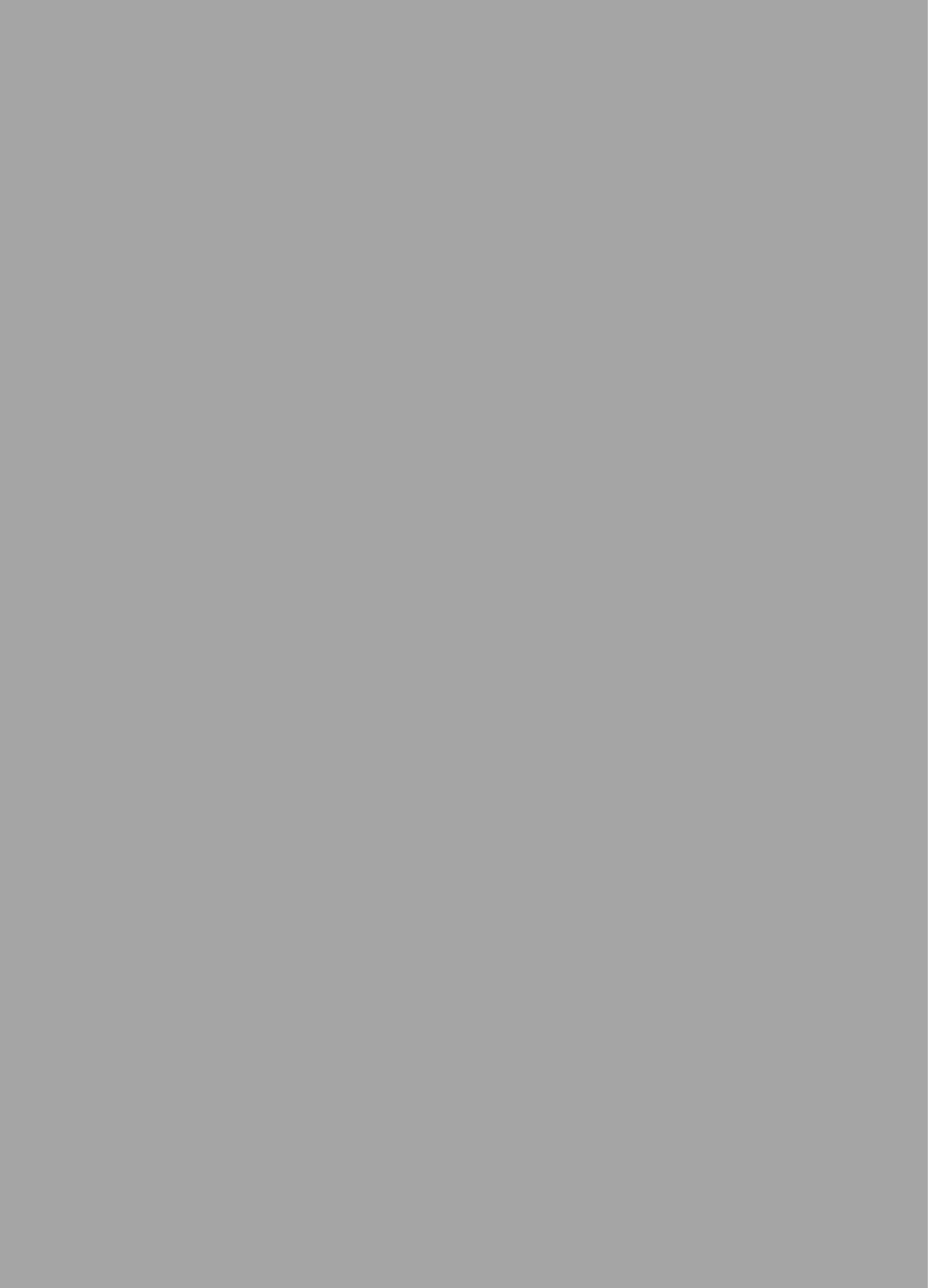
Description automatically generatedGraphical user interface, application

Description automatically generated

1. Spreadsheet of Crime\_data\_leeds

You have been given a spreadsheet of data, which is summerised data. You are not expected to use this the ETL or data mart tasks. You can use if for the dashboard task if you wish.





#### ASSIGNMENT MARKING SCHEME mark out of 100%

*Student no: Group: Date:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Criteria*** | ***Level: 70%+ 1st*** | ***60 – 69% 2:1*** | ***50 – 59% 2:2*** | ***40 – 49% 3rd*** | ***<39% Fail*** |
| **Task 1:**  **DM Star schema design**  **20 marks** | 3-5 reports of DM type listed and complement the chosen KPI.  Star schema fully specified and meets requirements specified design decisions discussed and  justified via the data dictionary(s). Data from both sources.  Advanced concepts included such as SCD  All attributes correctly documented.  Supports functionality and reports as documented.  Design considerations (Granularity design decisions and reflected in star schema design).  The data is illustrated for each dimension and the fact. With a least a few rows in each, M:N.The data is correct based on given case study data.  Shows full understanding.  Clear evidence of application of literature to the assignment tasks | Star schema fully specified and meets requirements, design decisions discussed and justified.  Mostly as for a first, some areas ambiguously defined, missing or generally simple, however excellent understanding evident.  Data from both sources.  Literature included ‘matched’ rather than applied. | Star schema mostly specified and meets most requirements, some design decisions discussed and justified via a data dictionary.  The reports and examples are very similar to those given in class or in assignment specification.  Generally well done with no major flaws.  Literature included, described rather than used to inform the tasks. | Star schema mostly specified and meets most requirements via a DD.  There are major flaws in the design, considerations or data illustrated.  Some literature. | Star schema doesn’t support queries. Little understanding evidenced.  Very little data, or incorrect data.  Little  understanding evidenced.. |
| **TASK 2:**  **SS- FACT table and 2 other tables as a minimum (5 marks)** | Excellent database produced. Evidence that CASE tool has been used. All aspects of the CASE database design issues addressed & dealt with. | Good database. Evidence that CASE tool has been used. Most aspects of the CASE database design issues addressed & dealt with. | Database produced. Using CASE tool .Some aspects of the database design addressed & dealt with. | Db produced. Little evidence that CASE tool have been used. Few aspects of the database design addressed & dealt with. | Little database, or evidence of CASE tool Database design not addressed and dealt with. |
| **Task 3: Design of the ETL and populate star schema**  **(25 marks )** | *All tables fully populated from both*  *“data-bases”. Data Extract, Quality, transformation, calculation of measures fully evidenced using mostly SQL. Competent use of appropriate, advanced SQL (eg sequences etc). Evidence of SQL running successfully.*  *Code is significantly developed from that given in tutorials. Uses standards and includes comments.*  Evidence of SQL running successfully.  *Excellent understanding evidenced* | *All tables populated from both “data-bases” as specified. Some*  *Data Extract,*  *Quality,*  *Transformation,*  *calculation of measures evidenced (mostly using SQL). Evidence of SQL running successfully.*  *Code is partially developed from that given in tutorials.*  Evidence of SQL running successfully.  *Good understanding evidenced.* | *All tables populated from both “databases” as specified. Some ETL stages addressed. Maybe errors in code or population. Evidence of SQL running successfully, may be very similar to that given in class.*  *Understanding evidenced.* | *Some tables populated.*  *Some evidence of SQL running successfully.*  Maybe lack of evidence of the SQL successfully running, or output that omits information.  *Some understanding evidenced* | *Little population of db*  *Little evidence of the SQL running successfully.*  *Little*  *understanding evidenced*. |

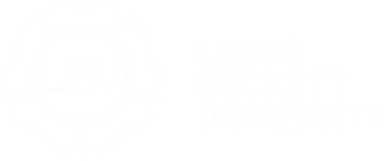
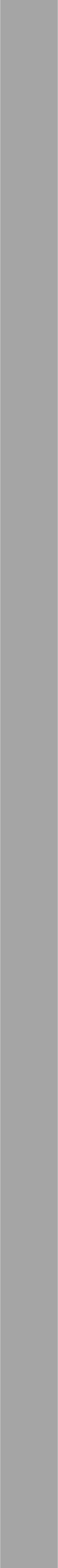
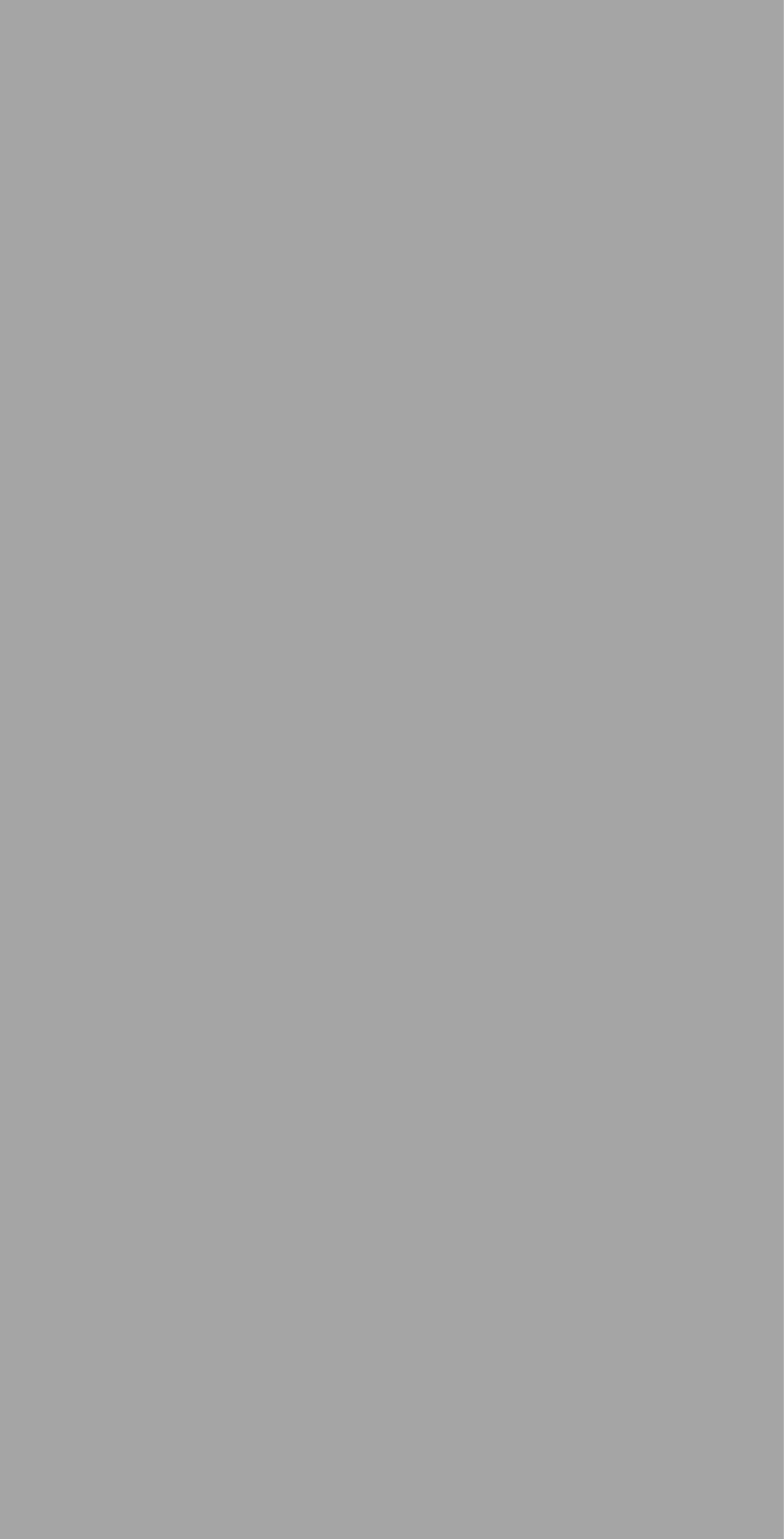
Feedback:

#### ASSIGNMENT 2 MARKING SCHEME

*Student no: Group: Date:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Criteria*** | ***Level: 70%+ 1st*** | ***60 – 69% 2:1*** | ***50 – 59% 2:2*** | ***40 – 49% 3rd*** | ***<39% Fail*** |
| **Task 1:**  **a) OLAP** | OLAP: An excellent pivot table produced with very relevant and suitable reports – | OLAP: A very good pivot table produced with very relevant and | OLAP: A pivot table produced with some reports – | OLAP: A pivot table produced with report – visualisations, | OLAP: little or not useful pivot table produced. |
| **OR**  **Apex dashboard**  **(20 marks)** | As before and excellent application of tools, research, understanding. | using appropriate visualisations, labels, titles etc.  An apex dashboard created to meet KPIs, excellent HCI, data consideration, advanced code.  Evidence of research being appropriately applied to drive the areas and types of investigation. Bibl’y  Excellent understanding. | suitable reports – using correct visualisations, labels, titles etc.  An apex dashboard created to meet KPIs, some HCI, data consideration, correct code.  Evidence of research some application. Bibliography.  Good understanding. | visualisations, labels, titles may not be useful.  An apex dashboard created to meet basic KPIs, some HCI, data consideration, basic SQL.  Some evidence of research – described rather than applied.Bibl’y.  Reasonable understanding *evidenced* | labels, titles may not be useful.  An apex dashboard created to meet very basic KPI(s), very basic SQL or code generated by Apex.  Little research or bibliography.  Some understanding *evidenced* |
| **Task 2:**  **PL/SQL**  **(30 marks) considerations** | Excellent design of package to support the ETL process (or part of).  One function and one procedure coded as part of the package. Code is moved on from that given in class examples and well tested. | Excellent design of package to support the ETL process (or part of).  One function and one procedure coded as part of the package. Code is moved on from that given in class examples and well tested. Not as advanced as for a first. | Package designed and includes procedure and function. All appropriate. Mostly coded, maybe some issues with fully testing. | A procedure and function designed coded and working.  Maybe package not entirely understood or used for testing. Code as given in class. | Little or no understanding of code ideas submitted.   * Design * Code * Test * Presentation of work |

Feedback:



**DETAILS OF THE REASSESSMENT**

### 4.4 Reassessment

Reassessment is to ‘re-do’ either or both components as required. A summary sheet listing all changes is required, this can be easily tracked using the Word, ‘tracking’ facility.

Reassessment date: : 7th April 2023

### 4.5 Feedback

Formative feedback will be given in tutorials. General feedback will be given approximately a week after the hand-in date and individual feedback upto 3 weeks after the hand-in date.

Formative feedback is feedback on “what you have done already”, this gives you the opportunity to apply the feedback to your final assignment. Students generally find this kind of feedback very useful and gain better marks as a result.

**Feedback schedule**

|  |  |  |  |
| --- | --- | --- | --- |
| **Assignment** | |  | **Feedback** |
| Assignment 1 | **14/11 2022 by 23:00** |  | Formative feedback available as indicated in schedule  General feedback provided via the VLE by week 11.  Individual feedback by end of week 11 |
|  |
| Assignment 2 | 13/1/2023 by 23:00 |  | Formative feedback available as indicated in schedule  General feedback provided within 3 weeks.  Individual feedback within 3 weeks. |

You are always welcome to make an appointment to discuss your feedback.

24

# 5 Understanding Your Assessment Responsibilities

Please refer to Course Handbook as appropriate

**Mitigation and Extenuating Circumstances**

If you are experiencing problems which are adversely affecting your ability to study (called 'extenuating circumstances'), then you can apply for mitigation. You can find full details of how to apply for mitigation at:

[http://www.leedsbeckett.ac.uk/studenthub/mitigation.htm](http://www.leedsmet.ac.uk/studenthub/mitigation.htm)

**Late Submission**

Without any form of extenuating circumstances, standard penalties apply for late submission of assessed work. These range from 5% to 100% of the possible total mark, depending on the number of days late. Full details (section C1.5.7) of the penalties for late submission of course work are available at:

[http://www.leedsbeckett.ac.uk/about/files/C1\_Assessment\_-\_General\_Provisions.pdf](http://www.leedsmet.ac.uk/about/files/C1_Assessment_-_General_Provisions.pdf)

**Academic Misconduct**

Academic misconduct occurs when you yourself have not done the work that you submit. It may include cheating, plagiarism and other forms of unfair practice. What is and what is not permitted is clearly explained in *The Little Book of Cheating, Plagiarism and Unfair Practice*, available at: [http://www.leedsbeckett.ac.uk/studenthub/plagiarism.htm](http://www.leedsmet.ac.uk/studenthub/plagiarism.htm)

The serious consequences of plagiarism and other types of unfair practice are detailed in section C9 of the Academic Regulations at: [http://www.leedsbeckett.ac.uk/about/academic-regulations.htm](http://www.leedsmet.ac.uk/about/academic-regulations.htm)

25