

MTH785P

PROGRAMMING FOR BUSINESS ANALYTICS

Input: csv files to be found in QM+

Output:

1. An Access database **MTH785_NameSurname_Question1.accdb** containing the work in Part I (where **Name** and **Surname** are to be replaced by your Name and Surname)
2. An Excel file named **MTH785_NameSurname_Question2.xlsx** containing the work in Part II.
3. An Excel file **MTH785_NameSurname_Question3.xlsm** containing the work in Part III.
4. A file **MTH785_NameSurname_Summary.pdf** where you explain your work.
 - a. For the Access/SQL part this should include screenshots of your database in Access and the SQL instructions you have used (Access sometimes changes the formatting of SQL instructions, I want to see the SQL you have written without these modifications).
 - b. For the Excel part you should include screenshots of the graphs and provide some explanation of any relevant business information they convey. Also indicate what version of Excel you have used (Office 365, Office 2016, Office for Mac...).
 - c. You do not need to write about the VBA part in the pdf file.

General observations

You should comply with standards that we have used in the course. For example, you should not use default names for anything and make sure that your variable names, file names, field names etc are informative and made of full English words. You should not have redundant old functionality. Your code should be properly spaced, commented, and indented. Any output should be sensibly formatted and made easy to understand.

If you use functionality or syntax that has not been explained in the course, you should document this with an explanation in the PDF file submitted

Submission deadline

The submission deadline is Friday 7th January 2022 at end of business day (5pm GMT)

QMUL Late submission penalty According to QMUL Academic Regulations (page 53 point 3.64.i): *“For every period of 24 hours, or part thereof, that an assignment is overdue there shall be a deduction of five per cent of the total marks available (i.e. five marks for an assessment marked out of 100). After seven calendar days (168 hours or more late) the mark shall be reduced to zero, and recorded as OFL (zero, fail, late)”*.

Plagiarism policy

Any suspicion of plagiarism in an assessment worth more than 30% of the total module needs to be reported to the QMUL Assessment Offences panel. Penalties can be very harsh (e.g., all modules failed and to be re-sat the next year with all grades capped at 50/100) so **be very careful.**

Part I Access/SQL

[35 marks]

This exercise uses Employee data used by the HR department in a hypothetical company. This data consists of a main table and auxiliary tables containing names for categories. Note that this data already has a column that should be primary key.

You can find these in QM+. The main table will be named with your student ID number. The auxiliary tables are: Education.csv, EnvironmentSatisfaction.csv, JobInvolvement.csv, JobSatisfaction.csv, PerformanceRating.csv, RelationshipSatisfaction.csv and WorkLifeBalance.csv. You should give these tables meaningful informative names.

Use them to create a database called HRDatabase with sensible table names and relationships.

Create SQL queries to achieve the following:

1. [2 marks] The Employee number and distance from home ordered by distance from home in descending order.
2. [3 marks] The average starting salary for an employee by gender and job level.
3. [5 marks] The minimum age of an employee with job level 4.
4. [5 marks] Using job level as parameter output the average salary and minimum and maximum job satisfaction for that job level.
5. [5 marks] The employee numbers, current salary, and descriptor for their job satisfaction for the five employees with the highest current salary (you should use the SQL instruction `TOP 5` which selects the first 5 records in a `SELECT` statement).
6. [5 marks] The count of employees for each performance rating (using descriptor).
7. [5 marks] Using the employee number as a parameter returns the Department and descriptor for Work Life balance.
8. [5 marks] One SQL instruction that you invent, and which responds a sensible business question you might have about the dataset.

Name your SQL queries by starting with the number of the question as in “01 Employee Number And Distance”, “02 Average Starting Salary”, ... This ensures that the queries appear in Access in the same order as above.

Make sure the display looks professional: informative labels, sensible format, ...

Part II Excel

[30 marks]

Your manager would like some visual analysis of the HR data in Part 1. For this you will use Excel. Connect Excel to the database in the previous part using Power Query (if your version of Excel does not have this then simply import the initial .csv file).

You are expected to produce an Excel worksheet with one tab (worksheet) for each table used, and one tab for each of the questions below. The spreadsheet should be of a professional standard (meaningful tab names, titles, headers, no hanging deprecated data, ...)

Your manager wants to see:

1. [6 marks] Total headcount by Educational field.
You should use the **UNIQUE** and **COUNTIF** functions.
Visualise the results using a horizontal bar chart.
2. [6 marks] Percentage male and female employees per department.
Visualise the results using a 100% Stacked bar.

3. [6 marks] Attrition by age band.
Visualise the results using a Line chart.
4. [6 marks] Starting Salary by gender and marital status.
Visualise the results using a sensible chart of your choice.
5. [6 marks] Average current salary by Age band.
Visualise the results using a funnel chart.

She defines the age bands as follows: 18-30, 30-40, 40-50, 50-60, 60+. The age band includes the lower age and excludes the upper age.

Each chart sheet should contain the table of data constructed to populate the chart and the chart itself. Whenever possible your tables should have a visible check to ensure you are not missing data. If any of the functionality above is not available in your version of Excel (e.g., Excel 2016 does not have Funnel charts) then choose a sensible alternative and document this in your pdf report.

More Details

1. You are expected to have several tabs (worksheets).
2. One tab should contain the main table in your Part I database formatted as an Excel table. Power Query just means that you should use the functionality in the Data tab to import this data and set a connection (so that it can be later refreshed).
3. You should then have a tab per question.
4. You are expected to have to use functions like COUNTIF, COUNTIFS, AVERAGEIFS. You can use pivot tables but if you do, only do so for at most a couple of questions.
5. In cases when this is possible you should add a simple check that the numbers are OK to help preventing any mistake. For example, in question 1 you could add the headcount per educational field and compare it with the total headcount to check there is no missing or misspelt department.
6. Staff with attrition field equals Yes are staff that will be shortly leaving the company.
7. You are not expected to conform with any specific Excel format but do make sure that your spreadsheet follows general professional spreadsheet guidelines we discussed in the lessons. Make sure to use informative labels for your data.

Part III VBA

[35 marks]

Excel build

1. [3 marks] Build a spreadsheet with 3 buttons as below.
2. [3 marks] A named range containing the names of the tables in your Part 1 database.
3. [3 marks] Named cells to contain a directory, a file name, and a table name.
4. [3 marks] The table name should be made to belong to the list in point 2 above.
5. [3 marks] Additionally, in the same column as the title in a row below the last row in List of Valid. Tables, name a cell that will contain the tables we will be querying.

Database Dump

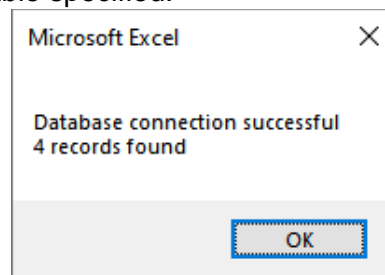
Test Database Connection	Load Data from Database	Clear Results
---------------------------------	--------------------------------	----------------------

Database directory C:\Documents\Progr Database file Assessment1.accdb Table JobInvolvement	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th style="text-align: left; padding: 2px;">List Of Valid Tables</th> </tr> <tr><td style="padding: 2px;">Education</td></tr> <tr><td style="padding: 2px;">Employee</td></tr> <tr><td style="padding: 2px;">Environment Satisfaction</td></tr> <tr><td style="padding: 2px;">JobInvolvement</td></tr> <tr><td style="padding: 2px;">JobSatisfaction</td></tr> <tr><td style="padding: 2px;">PerformqanceRating</td></tr> <tr><td style="padding: 2px;">RelationshipSatisfaction</td></tr> <tr><td style="padding: 2px;">WorkLifeBalance</td></tr> </table>	List Of Valid Tables	Education	Employee	Environment Satisfaction	JobInvolvement	JobSatisfaction	PerformqanceRating	RelationshipSatisfaction	WorkLifeBalance
List Of Valid Tables										
Education										
Employee										
Environment Satisfaction										
JobInvolvement										
JobSatisfaction										
PerformqanceRating										
RelationshipSatisfaction										
WorkLifeBalance										

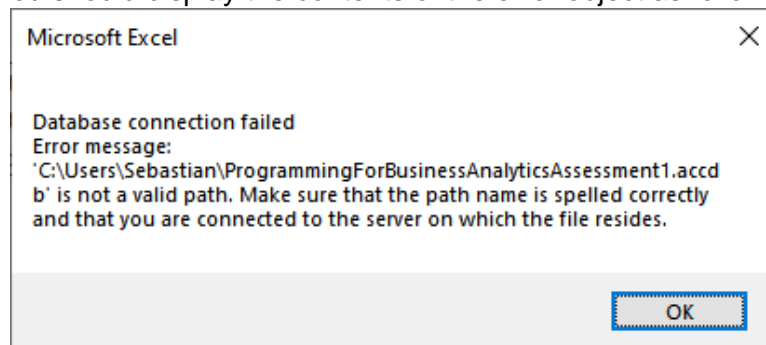
Education
Employee
Environment Satisfaction
JobInvolvement
JobSatisfaction
PerformqanceRating
RelationshipSatisfaction
WorkLifeBalance

VBA Build

- [3 marks] Write a subroutine connected with the button Test Database Connection that creates a new ADO connection to the file specified in the field and displays the number of records in the table specified.



- [3 marks] Add error trapping to manage the case where the database connection or query fails. You should display the contents of the error object as follows:



- [3 marks] Create a subroutine that is called upon hitting the "Load Data from Database" button. It should load the specified table from the database onto a Recordset object.
- [2 marks] Paste this Recordset object to the sheet on the range below the cell specified in item 5. of the Excel Build section.

Database Dump

Test Database Connection	Load Data from Database	Clear Results
--------------------------	-------------------------	---------------

Database directory	C:\Users\sebas\Drop
Database file	Assessment1.accdb
Table	Employee

List of Valid Tables	
Education	
Employee	
Environment Satisfaction	
JobInvolvement	
JobSatisfaction	
PerformanceRating	
RelationshipSatisfaction	
WorkLifeBalance	

41 Yes	Sales	1 2 Life Sciences	1 2 Female	2 Sales Executive	4 Sin
49 No	Research & Development	8 1 Life Sciences	2 3 Male	2 Research Scientist	2 Ma
37 Vac	Research & Development	2 2 Other	4 4 Male	1 Laboratory Technician	3 Sin

- [3 marks] Add a subroutine that clears the results and associate it with the button Clear Results. Call the subroutine that clears the results before you execute the Load Data From Database subroutine to avoid ending mixing results of different calls.
- [3 marks] Add Code to read the column headers from the Recordset object and display this above the data dump.
- [3 marks] Add code to add borders to the data and code in the Clear Results subroutine that removes these borders.

As in any project you will probably need to research and learn new functionality. You could consider:

- Recordset object properties: .Fields.Count, .Fields(i).Name.
- Range properties: CurrentRegion, .BordersLineStyle, .Clear
- VBA variables sometimes used: vbCrLf, xlNone, Excel.XlBorderWeight.xlThin.

Observations:

- You should make sure your spreadsheet and VBA look professional. In particular: do not have hanging unused cells, names, sheets, or VBA code, ... Do not use default names for anything.
- You should not use ActiveX controls.
- It should be sufficient to write one subroutine for question 1, one for question 4 and one for question 5. Do not write overcomplicated code.