**Creation a database (using Microsoft SQL Server) from MS Excel files**

This piece of work is to create a database and associated queries for a MS Excel file. The intention is to load the Excel files into MS SQL Server and use SSMS to query the database. The database is to be updated daily.

**Background**

At present I have 1 Excel file, which contains 8 worksheets. Analysing the data in the worksheets involves copying and pasting from one worksheet to another and then using formulas between the worksheets do the analysis. This is laborious, time consuming and is prone to data corruption due to mistakes.

**Description of the Excel file (See attached)**

A sample of the current MS Excel file is attached. This contains a sample of the data to be uploaded into the database as well as the data to be analysed.

* **Sheet 1 (Results Database):** contains historical data and serves as the database (historical data (the Results Database) and the other current data.
* **Sheet 2 (Current File):** contains daily data that requires analysis
* **Sheet 3 (Full Match):** analyses data copied from the Current File worksheet (uses formulas to carry out the analysis)
* **Sheet 4 (First Half):** analyses data copied from the Current File worksheet (uses formulas to carry out the analysis)
* **Sheet 5 (Second Half):** analyses data from the Current File worksheet (uses formulas to carry out the analysis)
* **Sheet 6 (Full Match Column Description):** Details the output required for the proposed query required to analyse the **Full Match** worksheet
* **Sheet 7 (First Half Column Description):** Details the output required for the proposed query required to analyse the **First Half** worksheet
* **Sheet 8 (Second Half Column Description):** Details the output required for the proposed query required to analyse the **Second Half** worksheet



**Steps for the current process**

1. I start by copying data in Columns A-BT from the **Current File** (Sheet 2) worksheet
2. I paste it in the corresponding columns of the **Full Match** (Sheet 3)**, First Half** (Sheet 4) **and Second Half** (Sheet 5) worksheets. These are the sheets that will query the database (**Results Database** Worksheet)
3. I drag the cells drag down the cells (in the three worksheets) that contain the formulas that will do the analysis and return the output. This can take up to more than one hour depending on the amount of data in the **Current file** (Sheet 2)**.**
4. The following day when a new file is received, the data in the **Current File** (Sheet 2) worksheet is copied into the **Results Database** (Sheet 1) Worksheet (under the corresponding columns). The new data is then copied and pasted in the **Current File** worksheet and steps1-3 above are repeated.

**Requirements**

What is required is as follows:

* **Requirement 1:** Create utility (or similar) to update database
* **Requirement 2:** SQL scripts to allow the user to query/analyse the database

**Requirement 1a – (Full Match Worksheet)**

Create SQL scripts to query the database and generate the output in the **Full Match** worksheet as described in the **Full Match Column Description** worksheet.

The returned query should display the records under the same column headings as those in **Full Match** worksheet and should be Order by **league**

***\*\*\*Last x no of rows to be summed (as stated in the Full Match Description sheet) is variable i.e. changes and can be anything from 3-20. The query should allow for this to be changed as at when required.***

**Requirement 1b – (First Half Worksheet)**

Create SQL scripts to query the database and generate the output in the **First Half** worksheet as described in the **First Half Column Description** worksheet.

The returned query should display the records under the same column headings as those in **First Half** worksheet and should be Order by **league**

***\*\*\*Last x no of rows to be summed (as stated in the First Half Column Description sheet) is variable i.e. changes and can be anything from 3-20. The query should allow for this to be changed as at when required.***

**Requirement 1c – (Second Half Worksheet)**

Create SQL scripts to query the database and generate the output in the **Second Half** worksheet as described in the **Second Half Column Description** worksheet.

The returned query should display the records under the same column headings as those in **Second Half** worksheet and should be Order by **league**

***\*\*\*Last x no of rows to be summed (as stated in the Second Half Column Description sheet) is variable i.e. changes and can be anything from 3-20. The query should allow for this to be changed as at when required.***