# Opportunity Identified

## Proposal for software

From the analysis of data quality issues and the review of database management systems the author identified that analysing SQL database records for nulls can help identify

* completeness issues in data
* data that might produce unexpected results from queries issued that do not account for SQL’s three value-logic.

Performing this type of analysis on a database with many tables and numerous records can be tedious and time consuming, it also requires SQL programming skills. Hence, the proposal to create software that will autonomously analyse and report on null values in a SQL database.

## Target Database to be Analysed

There are serval SQL databases on the market but considering the limited time available for the project the scope of the software is narrowed to only analyse a single brand of SQL database. After a comparison between SQL Server, Oracle, PostgreSQL and mySQL it was decided to aim the initial development at mySQL on the bases that SQL Server and Oracle are not open source and mySQL offers superior performance vs PostgreSQL for applications with high volume of read transactions, as will be the case for the proposed software (Digitalocean.Com 2017).

## Development Environment

The following factors contributed to the selection of the software developing environment:

* The preference to deploy as a stand-alone application.
* Requirement for fast processing of multiple transactions.
* Restricting the number of technologies involved in the deployment in order to enable a faster uptake in learning and reduce the development time required.

Based on the criteria above and considering the skill set of the available resources, the only two software languages under consideration is Java or PHP/CSS/HTML, of the two options Java provides the closest match to the requirements of the project.

# myNullSQL Software.

## Description of myNullSQL

myNullSQL software will enable a user to identify and analyse nulls in mySQL databases. The software will analyse a mySQL database and produce reports which will allow the user to gain an understanding of the extent and distribution of nulls in the relations, attributes and tuples of the database.

## Functional Requirements

1. The software shall allow the user to configure connection parameters for mySQL databases hosted locally.
2. The software shall allow the user to test connection parameters and display the result of the test, indicating if the specified database could be found and if the login credentials has sufficient access to allow the software to perform the analysis
3. The software shall allow the user to initialise analysis of the database structure after successful test of connection parameters.
4. While performing initial analysis of the database structure the software will display a progress bar indicating progress
5. After initial analysis of the structure of the database, the software will list the names of relations that exist in the database as well as a count of attributes and tuples in each relation. Only non-system relations will be listed. See Figure 6-1 as an example.

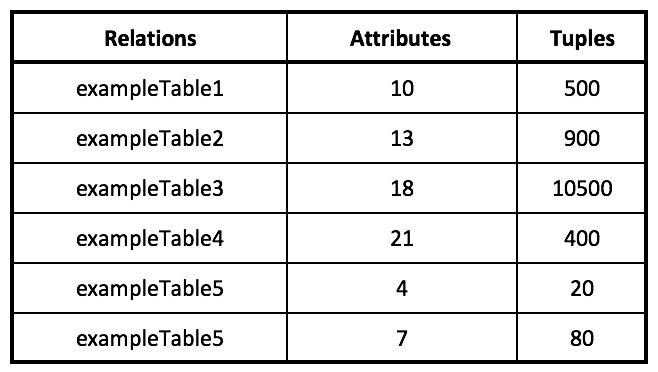


Table 2‑1 List of relations found in a mySQL database

1. The user will be able to select all, or any number of, relations listed after the initial analysis of the database in which to identify nulls and or blank values.
2. While analysing the selected relations for nulls and or blank values the software will display a progress window indicating:
   * The count of relations selected for analysis
   * The count of relations already analysed (in the current session)
   * The name of the relation currently being analysed
   * The count of relations still to be analysed
3. After the identification and analysis of nulls and or blank values in the selected relations the software will list the names of the selected relations a, count of attributes and tuples in each relation and a value for every relation indicating which percentage of all the fields in the relation contains nulls. See Figure 6-2 as an example.

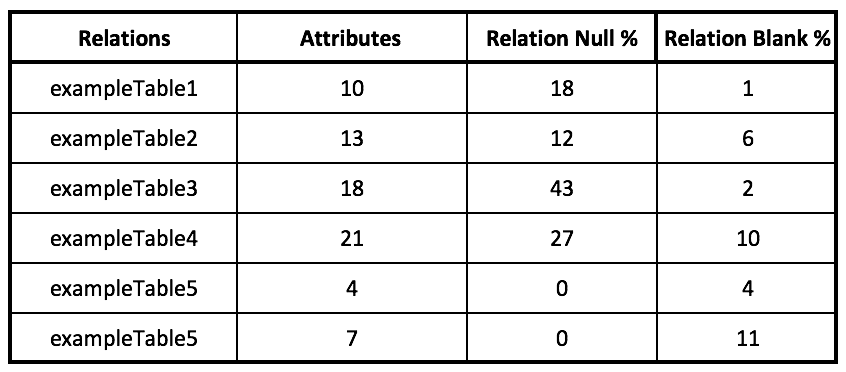


Table 2‑2 List Of Relations With Percentage Null And Blank

1. The system will allow the user to produce a graph of the selected relations showing the percentage of nulls in each relation. See Figure 6.1 as an example.

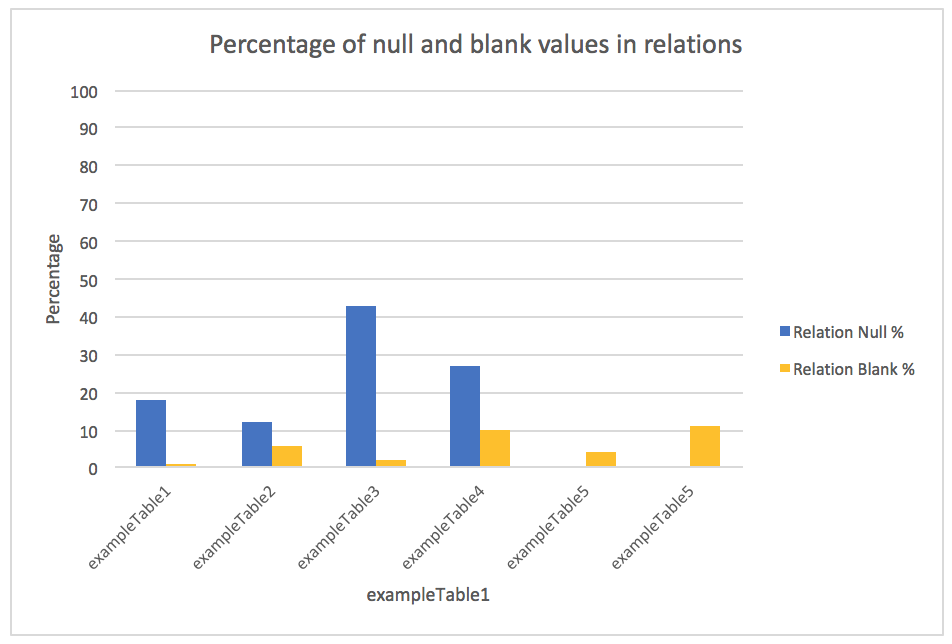


Figure 2‑1Graph Displaying Percentage Of Nulls And Blank Values In Relations

1. The user will be able to select a specific relation from the list of analysed relations to view an analysis of nulls by attribute for the selected relation. See Table 6-3 as an example.

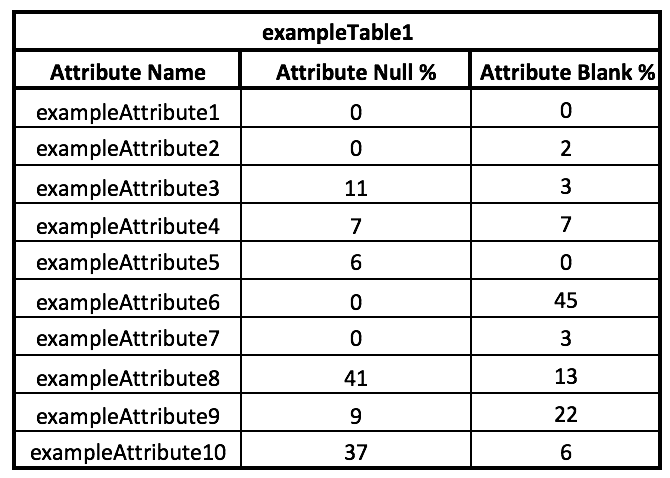


Table 2‑3 Analysis Of Nulls And Blank Values Per Attribute

1. The system will allow the user to produce a graph of the analysis of nulls by attribute for the selected relation. See Figure 6-2 as an example.

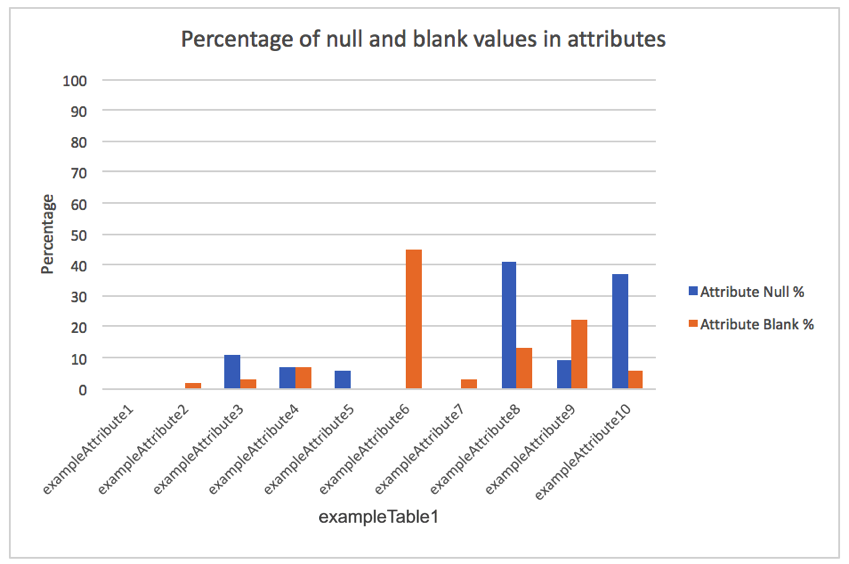


Figure 2‑2 Percentage Of Nulls And Blank Values In Attributes

1. The user will be able to select a specific relation from the list of analysed relations to view a percentage of tuples by number of attributes with nulls or blank values. See Table 6.4 and Table 6.5 as examples.

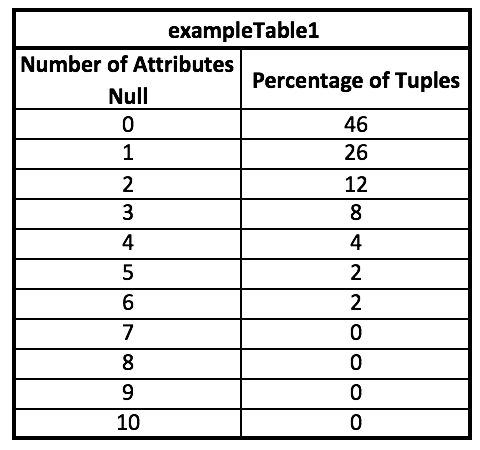


Table 2‑4 Percentage Of Tuples By Number Of Attributes With Nulls Value

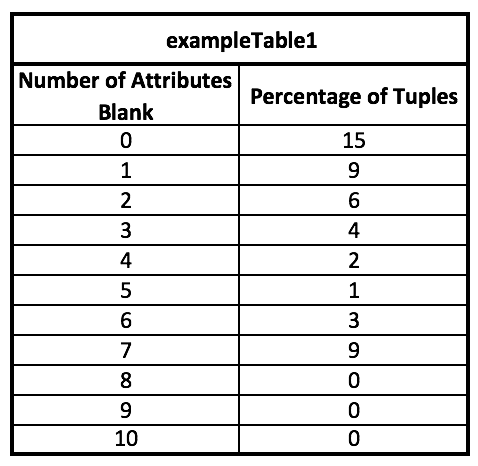


Table 2‑5 Percentage Of Tuples By Number Of Attributes With Blank Values

1. The system will allow the user to view data in a relation. The data view function will allow the user to specify which attributes to include in the view as well as specify conditions applied to values in each attribute selected (i.e. apply a filter) to restrict the tuples viewed to only those that the user want to investigate.
2. The system will allow the user to specify the maximum number of tuples to display in the data view (regardless of any conditions placed on data in attributes).
3. The system will allow the user the export data in the data view to files in CSV format.
4. The system will allow the user to determine the top 5 values for data in an attribute that appear most frequently in the tuples of a relation.
5. The system will allow the user to determine the median and average values of numeric attributes.
6. The system will allow the user to specify a new value for an attribute to replace the current value (including nulls) for tuples that meet conditions specified by the user (this function may only be available in future releases).