

Document your data cleansing issues. What were the anomalies identified and what you did to correct them?

The original table **Uploadfile.CurrentSemesterCourseOfferings** contains an aggregate of various data that are broken down and divided between the relational tables that are created for the project.

<u>Identified Anomalies</u>	<u>Corrections</u>
The column Course (hr, crd) contains 3 aggregated data that can be broken down and divided between the Department table and Course table.	Using a similar format of SUBSTRING([Course (hr, crd)], 0, CHARINDEX(' ', 0, [Course (hr, crd)])) to extract the department name, hours and credits fields.
None of the original table column table had no NULL values . Instead there were place holder for them in the form of blank spaces, dashes(-) and commas (,) , making data filter have very specific condition for each columns.	Updated the empty fields by replacing it all with 'TBA' using the UPDATE and SET query on the conditions that matches the patterns associated with missing data .
The original table's Instructor column contains the concatenated first and last name separated by a comma (,) of that instructor which needs to be used by the Instructor table.	Used RIGHT([Instructor], LEN([Instructor]) - CHARINDEX(' ', [Instructor])) to extract the First Name and SUBSTRING([Instructor], 0, CHARINDEX(' ', [Instructor])) to extract the Last Name
The original table's Location column contains the concatenated Building Name and Room Number which is to be used by the BuildingLocation table and RoomLocation table.	Using the SUBSTRING function to extract the building name followed by the room number.
The Day column contains a denormalized format of multiple days, blank fields and duplicates.	Using normalization to summarize the days and empty fields into an atomic table.
The Room Number for the class is not of INT value since it contains a mix of INT and CHAR.	In order to JOIN the ClassDetail table and use the section number to filter out the result, the section number had to be changed to type VARCHAR
The extracted data of Credits from the column Course (hr, crd) is of type INT and JOINS with the other table with credit as filter condition .	Either change the type to VARCHAR or use the CAST function in order to compare it to the substring of the original table.