ETL Project Proposal

We have decided to look at 311 requests made in Chicago. We have a variety of datasets based on request type including Graffiti Removal, Alley Lights Out, Street Lights - All Out, and Street Lights – One Out, amongst others. We will clean the data and combine the appropriate fields in each dataset based on ZIP code or Ward number.

We are going to clean each dataset so that each contains the following;

* Creation Date
* Service Request
* Type of Service
* ZIP Code
* Ward
* Police District
* Community Area

Graffiti Removal will also contain;

* “Where is the Graffiti Located?” Column
* Rename Service Request to Graffiti Request

Street Lights -All Out

* No Additional Columns
* Rename Service Request to All Lights Request

Street Lights – One Out

* No Additional Columns
* Rename Service Request to One Light Request

Alley Lights Out

* No Additional Columns
* Rename Service Request to Alley Light Request

There were many additional columns in each table that were dropped such as latitude and longitude, historical ward numbers, census tracts, etc.

We chose to use a SQL database for this because the data was already in a table format. We also thought it would be the easiest way to query the relationship of these types of 311 calls based on the different location parameters.