Designing Advanced Data Architecture for Business Intelligence  
**Individual Project- Finding Inconsistencies in the :**[***311\_CallCenterServiceRequests\_KansasCity\_2007-March2021.tsv.zip***](https://northeastern-my.sharepoint.com/:u:/g/personal/naveenhks_northeastern_edu/EYRwin79shxJgh_u6Kd7yTQBDyKfg0Ekjgrdb0cs0e62gg?e=PNc49f) **dataset**

**By Shamin Chokshi**

Inconsistencies in the Dataset which need to be cleaned are

* Field- CREATION DATE, CLOSED DATE are by default getting selected as string whereas they should be converted to date datatype
* Field- CASE ID and PAECEL ID NO should be converted to Number datatype as it is being considered as a string
* Field LATITUDE and LONGITUDE need to be converted to number datatype with decimal values as it would be asier for visualization purposes
* Field CREATION TIME needs to be converted to time format
* EXCEEDED EST TIMEFRAME can be converted to Boolean (Not required but good to have) so it occupies less space
* Days to close field has a few negative values
* After applying Auto feild in alteryx

|  |  |
| --- | --- |
| Name | Type |
| CASE ID | NUMBER |
| SOURCE | V\_String |
| DEPARTMENT | V\_String |
| WORK GROUP | V\_String |
| TYPE | V\_String |
| DETAIL | V\_String |
| CREATION DATE | DATE |
| CREATION TIME | TIME |
| STATUS | V\_String |
| EXCEEDED EST TIMEFRAME | V\_String (COULD BE BOOLEAN) |
| CLOSED DATE | DATE |
| DAYS TO CLOSE | NUMBER (DECIMAL) |
| STREET ADDRESS | V\_String |
| ZIP CODE | V\_String |
| NEIGHBORHOOD | V\_String |
| POLICE DISTRICT | V\_String |
| PARCEL ID NO | NUMBER |
| LATITUDE | NUMBER(DECIMAL) |
| LONGITUDE | NUMBER(DECIMAL) |
| CATEGORY1 | V\_String |
| CATEGORY2 | V\_String |
| CATEGORY3 | V\_String |

* The data seems pretty much clean but there are a few null values in the CATEGORY2 and CATEGORY3 fields  
  That seems because Category1 is the category of an issue that is being addressed at the customer support  
  Category 2 is the Subcategory of that particular issue  
  Category 3 is a further classification of category 2   
  It seems that if Category 1 cant be further classified, category 2 and 3 will be null and if category 2 cannot be further classified, category 3 will be null.  
  A screenshot of a computer

  Description automatically generatedA screenshot of a computer

  Description automatically generated
* In the Detail field there are many values that are “\*SELECT ONE”- That needs cleaning
* The Field “Type” has a few Numeric values that need to be cleaned  
  A screenshot of a computer

  Description automatically generated
* Neighborhood, County and Police district have a few null values

## Alteryx Workflow Screenshot

A screenshot of a computer

Description automatically generated

* Downloaded the TSV file
* Imported it on alteryx
* Applied data cleansing and auto field for appropriate datatypes and browsed data for inconsistencies
* Did data output to SQL Server Database by “damg7370shamin” user using SQL server authentication

## Screenshot of Alteryx Success

* A screenshot of a computer

  Description automatically generated

## Screenshot of Data populated in SQL Server

## A screenshot of a computer Description automatically generated

Screenshot of row count (Matches the dataset)  
A screenshot of a computer

Description automatically generated