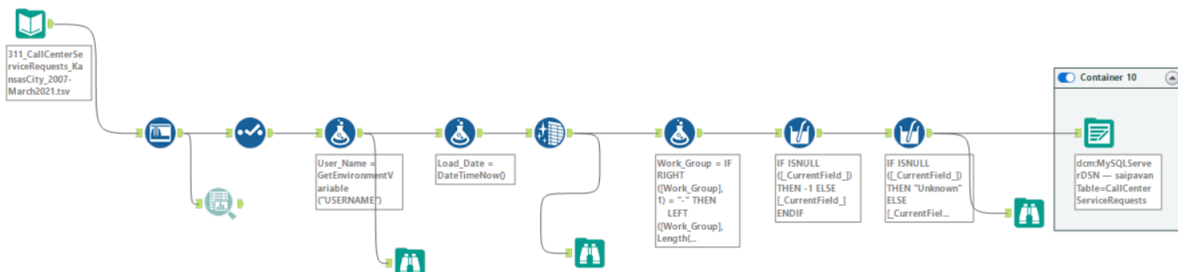# Data Cleaning & Preparation
Call Center Service Requests | Kansas City (2007–Mar 2021)

Below is a concise yet comprehensive outline of the 11-step ETL cleaning process implemented in Alteryx, demonstrating technical precision and business relevance.

1. **Data Ingestion & Field Sizing**

   o **Tool**: Input Data + Auto Field

   o **Action**: Loaded raw TSV data and auto-detected optimal data types and field sizes for all columns.

   o **Impact**: Ensured performant, schema-accurate processing from the outset.

2. **Field Selection & Renaming**

   o **Tool**: Select

   o **Action**: Removed irrelevant fields and prepared a streamlined schema for downstream analysis.
   (No renames were applied beyond pruning unused columns.)

3. **Audit Trail Enrichment**

   o **Tool**: Formula (User_Name)

   o **Action**: Created a User_Name field via GetEnvironmentVariable("USERNAME").
   (Captures the operator executing the workflow.)

4. **Load Timestamp Generation**

   o **Tool**: Formula (Load_Date)

   o **Action**: Added Load_Date field using DateTimeNow().
   (Stamps each record with execution timestamp for traceability.)

5. **Dataset Previews**

   o **Tool**: Browse (×3)

- o **Action**: Interactively inspected the data at three critical junctures—post-ingestion, post-user audit, and post-timestamp—to validate transformations.

6. **Categorical Cleanup**

   - o **Tool**: Formula

   - o **Actions & Expressions**:

   - o **Work_Group**: IF RIGHT([Work_Group],1)="-" THEN LEFT([Work_Group],LENGTH([Work_Group])-1) ELSE [Work_Group] ENDIF
     (Removes trailing hyphens.)

   - o **Status**: IF [Status]="Resol" THEN "Resolved" ELSE [Status] ENDIF
     (Standardizes abbreviated statuses.)

   - o **Department**: IF Contains([Department],"Parks") AND NOT Contains([Department],"Parks & Rec") THEN "Parks & Recreation" ELSE [Department] ENDIF
     (Unifies park-related department names.)
     (Result: Harmonized categorical values to support accurate segmentation.)

7. **Null Handling – Numeric Fields**

   - o **Tool**: Multi-Field Formula

   - o **Expression**: IF IsNull([_CurrentField_]) THEN -1 ELSE [_CurrentField_] ENDIF
     (Replaces all missing numeric values with -1 to preserve row counts in aggregations.)

8. **Null Handling – Text Fields**

   - o **Tool**: Multi-Field Formula

   - o **Expression**: IF IsNull([_CurrentField_]) THEN "Unknown" ELSE [_CurrentField_] ENDIF
     (Fills all empty text fields with "Unknown" to avoid gaps in categorical analyses.)

9. **Final Validation**

   - o **Tool**: Browse

   - o **Action**: Conducted a last interactive review of the fully transformed dataset.
     (Confirmed correct application of all cleaning rules.)

10. **Export to Downstream Systems**

    - o **Action**: Published the cleaned dataset to MySQL, creating a reliable source for Tableau dashboard data feeds.

11. **Business Impact**

  o **Outcome**: Delivered a fully audit-ready dataset with consistent schema, enabling executive-level dashboards to accurately track monthly trends and resolution rates.

This structured cleaning framework underscores my ability to implement end-to-end ETL processes that align with both technical rigor and business reporting needs.