Name: Anushka Paradkar

NU ID: 002202598

DADA BI

**311 Call Center Service Request**

**Individual Project 1**

**Alteryx Workflow Explanation:**

1. Initially, the Input Data tool was employed to import the 311 Call Center Service Request dataset in the form of a TSV file into Alteryx.
2. Subsequently, the Browse tool was utilized to inspect the data, ensuring that all rows and columns were accurately loaded. The dataset comprises 1,563,215 records.
3. Additionally, the Auto Field tool was connected to the Input tool. This tool scans through all input records, optimizing field types to the smallest size suitable for the data within each column.
4. The DateTime Parsing tool was then applied to convert the creation date and closed date into a standardized format, enhancing their suitability for visualization purposes.
5. To enhance data manipulation, the Record ID and Select tools were linked. The Select tool facilitates the inclusion, exclusion, and reordering of columns for streamlined data flow within the workflow.
6. Data profiling and field summary tools were employed to scrutinize the data's characteristics and structure. These tools provide insights into the data's composition, simplifying the process of drawing inferences and making informed decisions.

**Problems with the dataset:**

1. **Missing Values:** Missing values can lead to biased analysis, incomplete insights, and hinder model performance.

* Category 2: This field has a great number of missing values 64.1%. This could lead to biased analysis.
* Category 3: Similarly, this field too contributes to 89.3% of missing values which will highly impact the insight creation for the data set.
* Police District: While the missing value percentage is slightly lower (2.3%), consider investigating these missing values as they might be relevant depending on the analysis goals.

1. **Data Granularity:** Many fields have too many unique values/or small value counts, suggesting that the data might be overly granular.

* TYPE: 295 unique values
* LONGITUDE: 126,281 unique values.
* DAYS TO CLOSE: 2,749 unique values.
* STATUS: Only 6 unique values, might be beneficial to further categorize these.
* SOURCE: Only 22 unique values.
* EXCEEDED EST TIMEFRAME: Only 3 unique values, potentially recode as binary (Yes/No).
* CREATION TIME: 1,440 unique values.
* WORK GROUP: 146 unique values.
* PARCEL ID NO: 175,953 unique values.
* DEPARTMENT: 27 unique values.

1. **Inconsistent Formatting**: Inconsistent formatting observed in the dataset leading to future analysis problems during visualizations.

* Creation Date: Different date formats observed in this column.
* Closed Date: Different date formats observed in this column as well.

1. **Outliers:** Outliers are data points which deviate highly from the rest of the values present in the dataset. They are basically extreme values which create a problem for analysis purposes.

* STREET ADDRESS: The presence of a single address with a high value count ("9300 NE Shoal Creek Valley Dr CLUBHOUSE") might be an outlier.
* Category 3: Extreme values also observed in this column which can be termed as outliers.
* DAYS TO CLOSE: This field has a minimum value of 3.0 and a maximum of 1154.0, suggesting there may be a few cases with significantly longer values, which leads to the presence of outliers.

**Ways to clean the data:**

1. To deal with missing values, we can implement imputation strategies depending on the nature of missing values. For e.g.: Replacing the missing values with mean or median of that specific column if they belong to same categories.
2. To deal with data granularity, we should combine similar values. Identifying and merging the categories with the same information will lower the granularity.
3. We can also aggregate the data into one larger overview for e.g.: instead of writing the entire address, we can combine and mention the city which will improve the data quality without mentioning unnecessary and larger details.
4. For dealing with inefficiencies like inconsistent formatting, we can standardize labeling techniques to overcome this. Develop a consistent labeling convention throughout the dataset.

Alteryx Screenshots:

A screenshot of a computer

Description automatically generated

Note: Getting that error because, I tried to run again after successful connection.

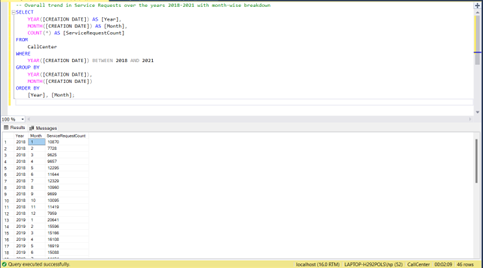
A computer screen with a white screen

Description automatically generated

A screenshot of a computer

Description automatically generated

1. **Service Requests Over Time:**



**2. Volume of service requests received from different sources:**

A screenshot of a computer

Description automatically generated

**3. Volume of service requests received by Department**

A computer screen shot of a computer

Description automatically generated

**4.Top 10 Performance Metrics (Response Time) per CATEGORY and Type of Request:**

A computer screen shot of a computer

Description automatically generated

**5.Geographical Visualization**

A computer screen shot of a computer

Description automatically generated

**6.Departmental Workload Comparison:**

A computer screen shot of a computer

Description automatically generated

**7. Response Time Analysis:**

A computer screen with a white screen

Description automatically generated

**(8) Service Request Status Composition:**

A computer screen with a white screen

Description automatically generated

**(9) Time to Closure Analysis:**

A computer screen with a white screen

Description automatically generated

**(10) Workload Efficiency:**

A computer screen with a computer screen

Description automatically generated

**Power BI Visualizations:**

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

**Tableau Visualizations:**

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screen shot of a graph

Description automatically generated