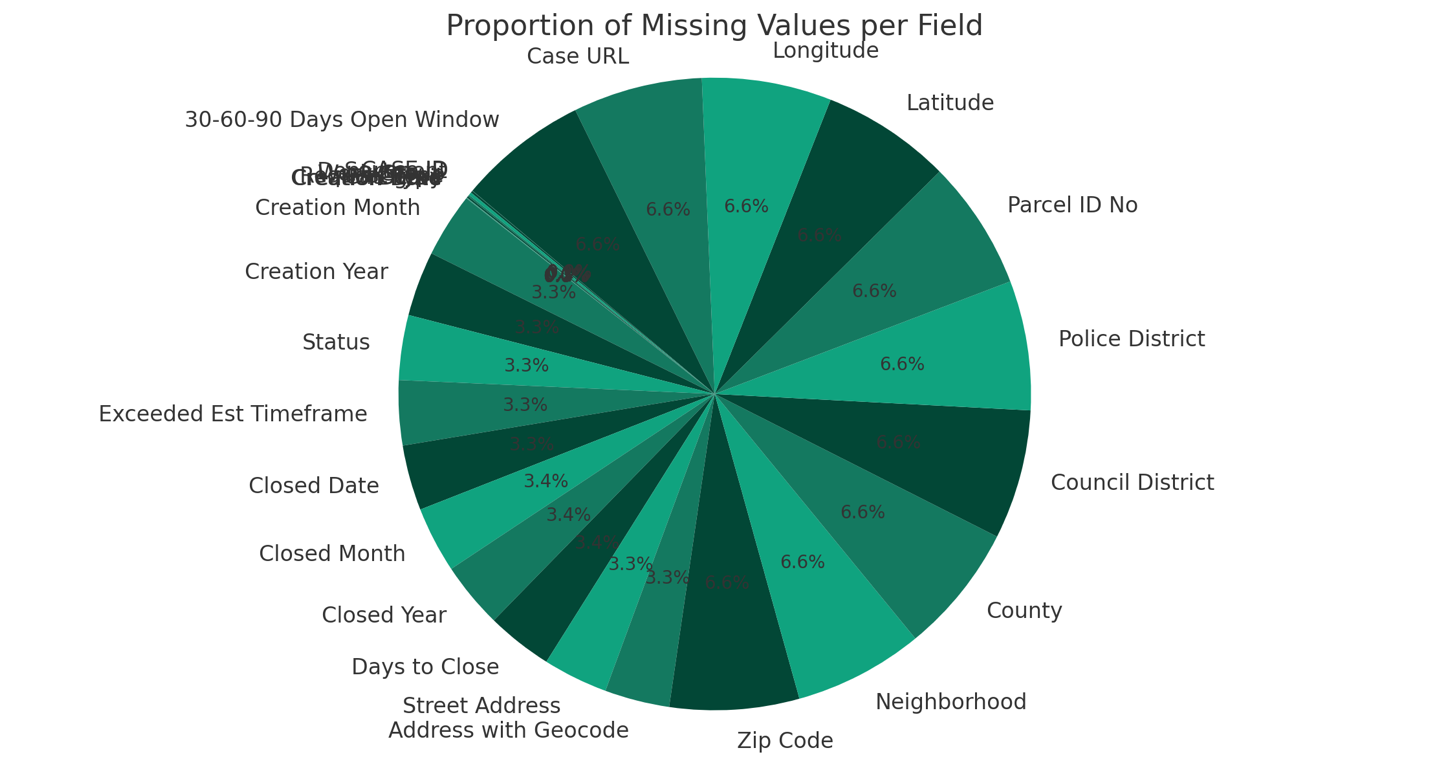
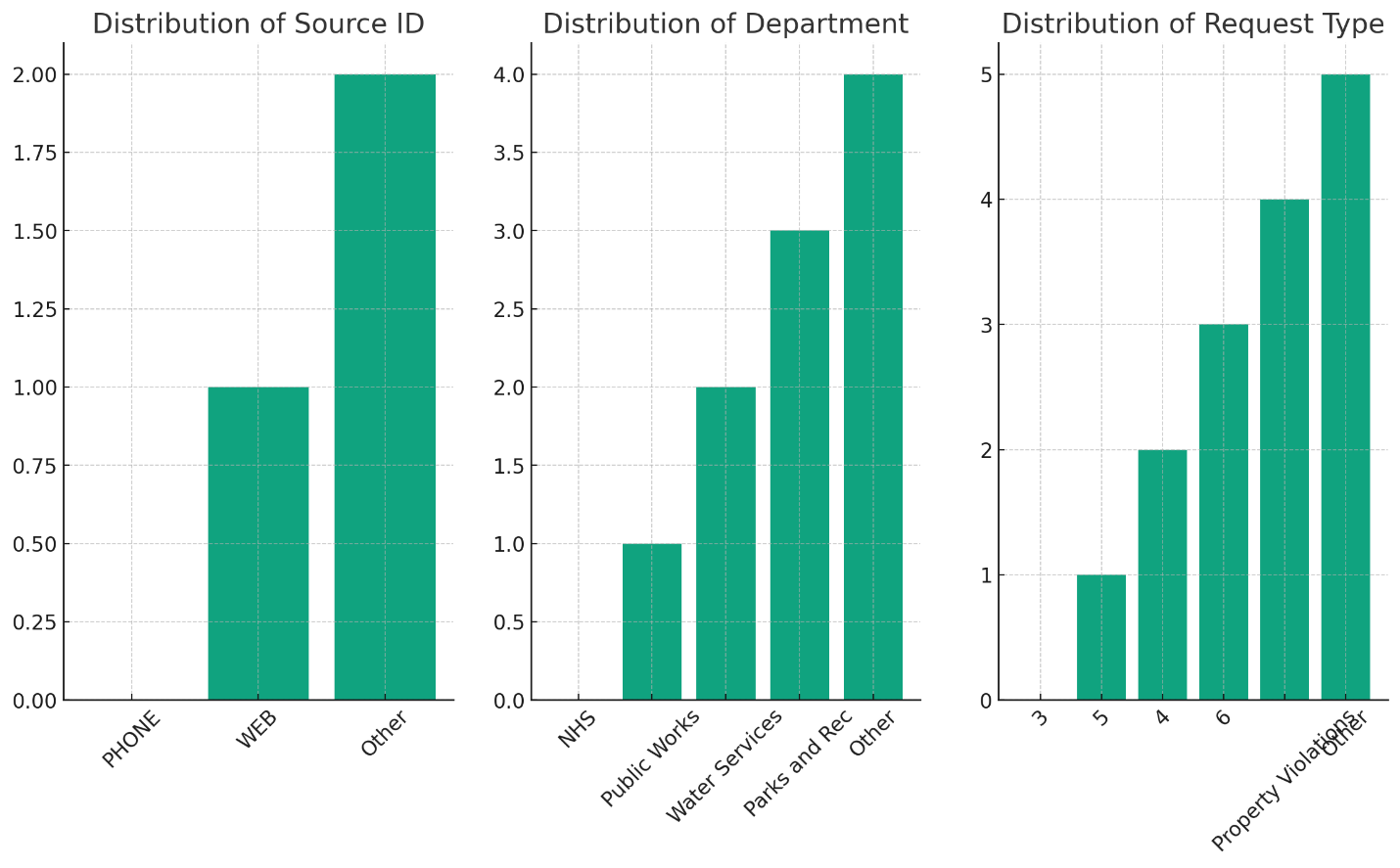
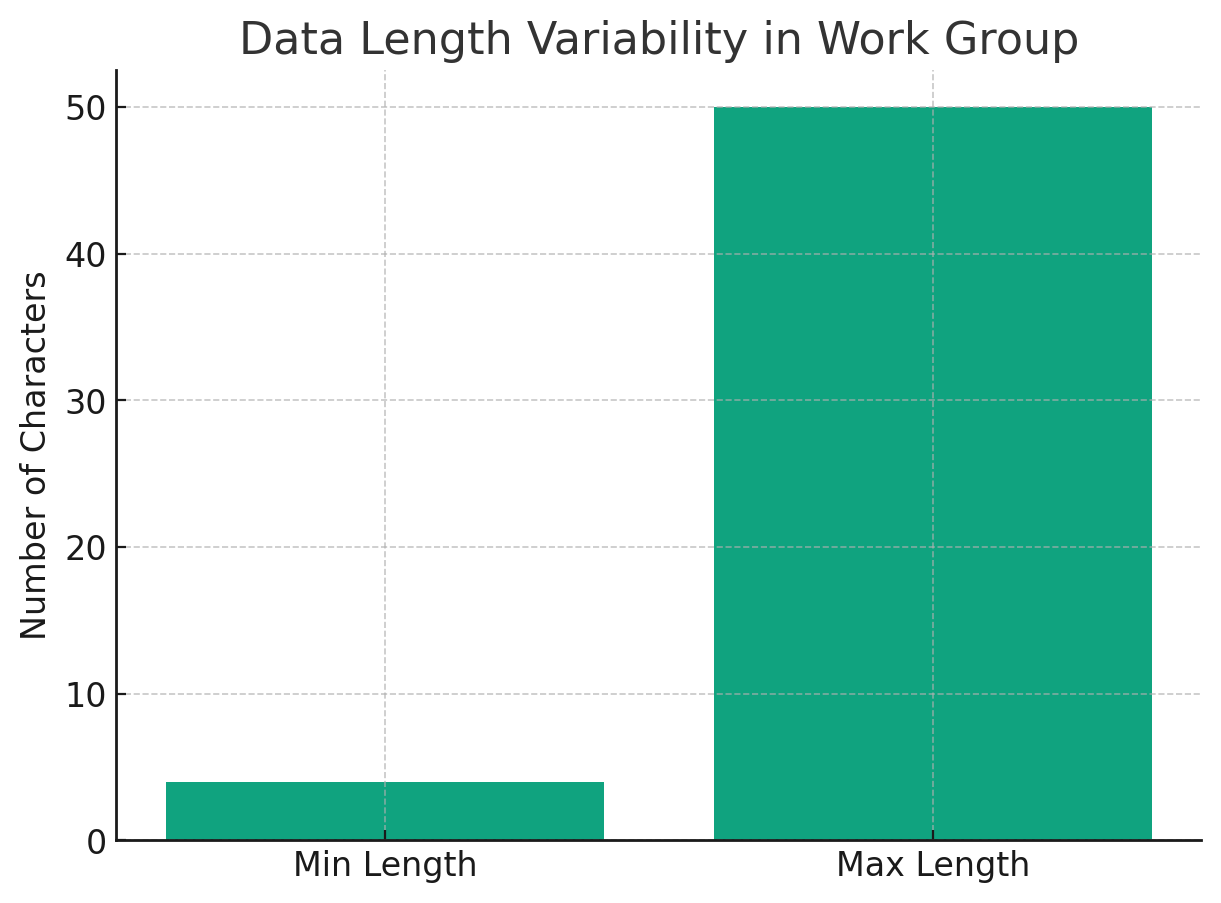
## **Report for the 311 Call Center Service Requests: 2007 - March 2021**

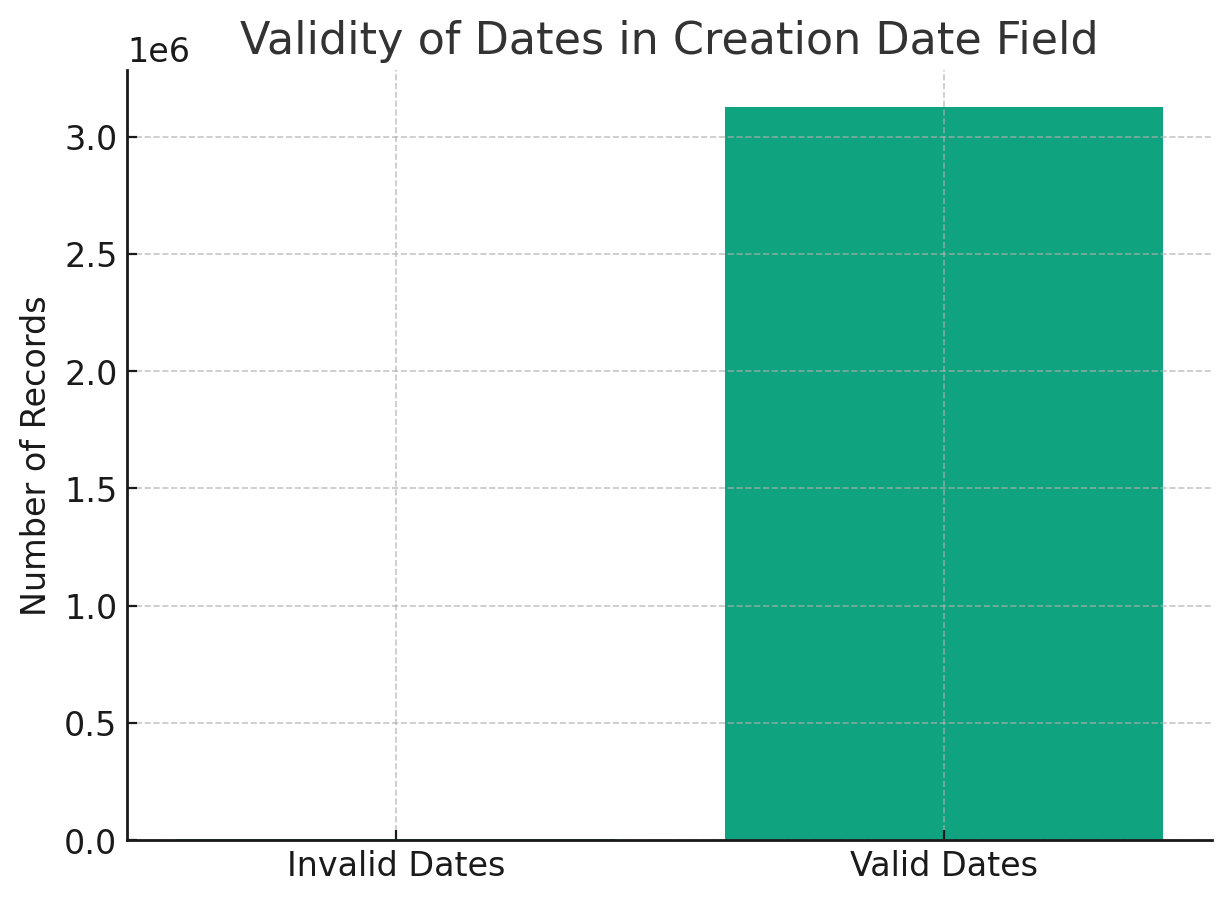
**General Overview -**

The dataset summary indicates a collection of over 3 million records with several fields related to case management, such as case IDs, source information, departmental categorizations, work groups, request types, and more. There are notable issues with data quality, including a significant number of missing values across multiple fields, inconsistencies in data formatting (especially in date-related fields where strings are used instead of date formats), mixed types within single fields (e.g., numerical values mixed with text in categorical fields), and the use of placeholder or out-of-range values. The 'Creation Date' and 'Creation Time' fields show particularly concerning anomalies with non-date/time data present. Additionally, there are fields with nearly all values missing, such as 'Zip Code', 'Neighborhood', and '30-60-90 Days Open Window', which suggests a substantial amount of data might be unrecoverable or was never captured. This summary highlights the need for significant data cleaning and validation to ensure the dataset's usability for analysis or operational purposes.









**Some of the insights about the dataset provided with the help of Alteryx are-**

1. CASE ID: There are 10 missing (null) values and an inconsistency in the formatting. The longest value looks like a coordinate pair rather than a typical ID, which might indicate incorrect data entry.

2. Source ID: There are 6,096 missing values. Also, some values that should be categorical like "PHONE" and "WEB" are mixed with numerical values like "64130", which may indicate data entered in the wrong format or column.

3. Department: There are 51,298 missing values and an unusually long value for a department name, which could suggest a concatenation error or incorrect data entry.

4. Work Group: There are 72,149 missing values. The length of the work group names varies significantly, with some as short as 4 characters and some as long as 50 characters.

5. Request Type: There are 40,434 missing values and a wide range of lengths, suggesting inconsistent categorization or description.

6. Category: There are 40,457 missing values, and the length statistics suggest possible inconsistency in category naming.

7. Type : There are 5,213 missing values and a wide range of lengths (1 to 48), indicating potential inconsistency in data entry.

8. Detail: There are 5,213 missing values and a wide range of lengths (2 to 48), which may point to inconsistent detail descriptions.

9. Creation Date: The data type is shown as string, which should ideally be a date format. There are 5,213 missing values, and some entries appear to be coordinates rather than dates.

10. Creation Time: It appears to be a mix of URL links and time values, which is inconsistent. There are 5,213 missing values.

11. Creation Month/Year: There's many missing values (over 50%), and some entries are out of the expected range (e.g., a month value of 90).

12. Status: Over 50% missing values and a limited range of unique values indicate a potential data quality issue.

13. Exceeded Est Timeframe: There are 1,568,426 missing values and only two unique entries, suggesting limited variability in this field.

14. Closed Date/Month/Year: There are over 50% missing values in each of these fields, and the formats are inconsistent.

15. Days to Close: There's a high number of missing values and a negative value, which might not make sense contextually.

16. Street Address: There are 1,568,427 missing values and inconsistencies in formatting (e.g., "ST" vs "St").

17. Address with Geocode: Similar issues to the street address with a high number of missing values.

18. Zip Code: Nearly all values are missing, with only 15 valid entries.

19. Neighborhood: Almost all values are missing.

20. County: Almost all values are missing, with only two entries for "Jackson".

21. Council District, Police District, Parcel ID No, Latitude, Longitude: These fields suffer from a very high number of missing values.

22. Case URL: Almost all values are missing, and it's unclear why the URLs have been shortened.

23. 30-60-90 Days Open Window: All values are missing.

For the columns with a significant number of missing values, it's critical to understand why the data is missing and whether it can be recovered or should be imputed. The presence of special characters, mixed data types, and inconsistent formatting all point to a need for data cleansing and standardization. It's also important to establish proper data validation rules to prevent these issues in future data entries.

**Some Data on Cleaning:**

In Alteryx, data cleaning and preparation tasks can be handled efficiently using a variety of tools and functions.

1. CASE ID:

- Use the Data Cleansing tool to remove null values or replace them with a placeholder if needed.

- Apply the RegEx tool to extract or validate the format of the CASE ID.

2. Source ID:

- Utilize the Filter tool to separate numeric values from text values.

- Use the Formula tool to correct or reformat values as needed.

3. Department:

- Handle missing values with the Imputation tool or Data Cleansing tool.

- Use the Select tool to ensure the data type is set correctly.

4. Work Group:

- Manage missing values with the Data Cleansing tool.

- Use the Select tool to trim the string lengths if necessary.

5. Request Type:

- Address missing values with the Data Cleansing tool.

- Standardize categories using the Formula tool or Find Replace tool.

6. Category:

- Clean missing values with the Data Cleansing tool.

- Standardize categories using the Formula tool.

7. Type:

- Use the Data Cleansing tool for missing values.

- Standardize entries with the Formula tool.

8. Detail:

- Address missing values using the Data Cleansing tool.

- Use the Select tool to adjust string lengths and ensure consistency.

9. Creation Date:

- Correct the data type using the DateTime tool to convert strings to date format.

- Handle missing or incorrect entries with the Formula tool or Data Cleansing tool.

10. Creation Time:

- Use the RegEx tool to extract time if mixed with URLs.

- Convert string times to the time data type with the DateTime tool.

11. Creation Month/Year:

- Use the Data Cleansing tool to address missing values.

- Validate and correct out-of-range values with the Formula tool.

12. Status:

- Use the Data Cleansing tool for missing values.

- Standardize status entries with the Formula tool.

13. Exceeded Est Timeframe:

- Address the high volume of missing values using the Data Cleansing tool.

- Use the Select tool to ensure the data type is logical (e.g., boolean).

14. Closed Date/Month/Year:

- For inconsistent formats, use the DateTime tool to parse and standardize dates.

- Impute or cleanse missing data with the Data Cleansing or Imputation tool.

15. Days to Close:

- Use the Data Cleansing tool to address missing values.

- Correct any negative values with the Formula tool, as they may be data entry errors.

16. Street Address:

- Apply the Data Cleansing tool to handle missing values.

- Use the Select tool to standardize text cases and formats.

17. Address with Geocode:

- Similar to Street Address, use the Data Cleansing tool for missing values.

- The Select tool can also help in standardizing the format.

18. Zip Code:

- With almost all values missing, consider if the field is necessary or if it can be recovered from another source.

- Use the Formula tool to format zip codes correctly.

19. Neighborhood:

- Use the Imputation tool to fill in missing values, if applicable.

- Standardize the entries using the select tool.

20. County:

- Decide if the field is necessary, given almost all values are missing.

- If yes, use the Imputation tool or an external data source to fill in the missing data.

21. Council District, Police District, Parcel ID No, Latitude, Longitude:

- Address the missing values with the Data Cleansing tool.

- Ensure correct data types with the Select tool.

22. Case URL:

- For the missing values, consider if they can be reconstructed or if they should be left blank.

- Use the Formula tool to ensure the URLs are correctly formatted.

23. 30-60-90 Days Open Window:

- As all values are missing, decide if this field can be populated with correct data or should be removed.

In Alteryx, we can also use the Multi-Field Formula tool to apply a formula to multiple fields at once, which is useful and time saving.

SQL Query Results-

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a phone number

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a data

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated