311 Service Request Data Analysis for Kansas City, MO

**Executive Summary**

This report offers a comprehensive analysis of the 311 service request data from Kansas City, MO, spanning from the inception of the system until March 2021. The dataset contains around 1.56 million records, each representing a distinct service request. The purpose of this analysis is to gain insights into the data's structure, quality, and to identify trends and patterns that can inform operational improvements.

**Dataset Overview**

The dataset comprises 30 columns and 1,563,215 rows, capturing a wide range of information about each service request, including identifiers, sources, departments, request types, dates, statuses, and geographic details.

**Data Profiling Summary:**

Number of Variables: 23

Number of Observations: 1,563,215

Missing Cells: 2,595,087 (7.2%)

Duplicate Rows: 0

Total Memory Size: 274.3 MiB

Average Record Size: 184.0 B

Variable Types:

Numeric: 6

Categorical: 7

Text: 6

DateTime: 3

Data Quality and Cleaning

Key Metrics:

CASE ID: Each service request is uniquely identified by a CASE ID, ensuring no duplication and allowing precise tracking.

SOURCE: The data includes 21 distinct sources of service requests, indicating a diverse range of entry points for citizens' concerns. Major sources include phone (1,204,236 requests), web (211,721 requests), and email (80,585 requests).

STATUS: The status of service requests is categorized into six distinct values, providing clear visibility into the current state and progress of requests.

**Data Cleansing Steps:**

Unpivoting Categories: The categories (CATEGORY1, CATEGORY2, CATEGORY3) were unpivoted into a unified category and category description field to simplify and enhance the accuracy of analysis.

Normalization: All text fields were standardized to a consistent case format to eliminate ambiguities caused by case sensitivity.

Status Categorization: Status values were refined from abbreviations (e.g., canc, dup, assig) to full descriptors (e.g., Canceled, Duplicate, Assigned) for improved readability and analytical utility.

Dimensional Modeling

The dataset was structured into a dimensional model using ER Studio, facilitating efficient querying and analysis. The model includes several dimension tables and a central fact table:

**Dimension Tables:**

Dim Source: Details about the origin of each service request.

Dim Status: Standardized statuses of service requests.

Dim Parcel: Geographic parcel information.

Dim Date: Comprehensive date information, with an added SEASON\_NAME column.

Dim Location: Geographic details including neighborhood, county, and police district.

Dim Category: Consolidated category information from the original dataset.

Fact Table:

Fact Case: Contains detailed service request information linked to all dimension tables.

Advanced Analysis

**Talend Transformations:**

1.Kansas\_city\_staging table

A screenshot of a computer

Description automatically generated

2.Dim\_Category:  
A screenshot of a computer

Description automatically generated  
3.Dim\_Source:  
A screenshot of a computer

Description automatically generated

4.Dim\_Parcel:  
A screenshot of a computer

Description automatically generated

5.Dim\_Location:  
A screenshot of a computer

Description automatically generated

6.Dim\_Date:  
A screenshot of a computer

Description automatically generated

7.Dim\_Status:  
A screenshot of a computer

Description automatically generated

8.Fact\_Case:  
A screenshot of a computer

Description automatically generated

Seasonal Trends:

A new calculated column named SEASON\_NAME was added to the Dim Date table to categorize service requests by season based on the month of creation. This categorization helps in identifying seasonal patterns in service requests, allowing the city to anticipate and better manage fluctuations in demand.

**Source Optimization:**

Analyzing the distribution of service request sources reveals that phone calls are the predominant mode of communication, followed by web and email submissions. This insight can guide resource allocation to ensure that the most commonly used channels are well-supported and optimized for efficiency.

**Geographic Analysis:**

The geographic data, including street addresses, ZIP codes, neighborhoods, and police districts, allows for a granular analysis of service request hotspots. Identifying these areas enables targeted interventions and resource deployment, ensuring that high-demand areas receive appropriate attention.

**Insights and Recommendations**

Seasonal Trends: The analysis of seasonal trends can inform resource planning and allocation. For example, if a higher volume of requests is observed during certain seasons, the city can proactively allocate additional resources during these periods to maintain service levels.

Source Optimization: Understanding the predominant sources of service requests can help the city enhance its service delivery channels. For instance, improving the efficiency of phone and web-based request handling could significantly reduce response times and improve citizen satisfaction.

Geographic Targeting: By identifying geographic hotspots for service requests, the city can prioritize interventions and allocate resources more effectively. This targeted approach ensures that areas with higher service demand are adequately supported, improving overall service quality.

**Conclusion**

This analysis of the 311 service request data provides a detailed understanding of the dataset's structure, quality, and key insights. By leveraging data profiling and dimensional modeling, Kansas City can enhance its operational efficiency, improve resource allocation, and ultimately provide better services to its residents. As the city transitions to a new record management system, continued monitoring and integration of the new data will ensure seamless service delivery and ongoing improvements in citizen satisfaction.