Project Proposal

311 Service Requests Data Analysis and Visualization

DATA 601- Working with Data and Visualization

Team members

Anitha Joseph (30282217)
Megha Radhakrishnan Sanitha (30272516)
Jincy Thomas (30271046)

INTRODUCTION

311 is a non-emergency government service that helps citizens to report issues and access city information (311 Calgary, no date). It offers a wide range of services, including reporting problems like public property damages, abandoned vehicles, and noise complaints, as well as requests for services like debris removal, street cleaning, tree trimming and animal threats. 311 also provides general information on events organised by the city, building permits and licenses, public health information etc. It can be accessed by phone, online, or through mobile apps and often includes multilingual support and services for vulnerable populations like the senior residents. The 311 system collects all requests, which are then routed to the responsible city department for resolution. It operates 24/7 and allows users to track the progress of their requests, ensuring transparency and accountability. Thus the 311 services make the lives of the residents easier and saves their time by eliminating the need to contact multiple departments for resolving their concerns. The user-friendly interface of 311 also makes the communication with the local government effortless, convenient and accessible for everyone.

Analysis of 311 data over a long period of time helps us in determining the most frequent service requests which can be addressed more efficiently by allocating adequate resources. This can also be used to identify the efficiencies of different departments and the gaps where improvement is required. Seasonal patterns and long term trends can also be identified which can be used to predict the future demands and optimize the public services. By tracking types of requests, response times, and areas with frequent complaints, resources can be allocated more effectively, and operational inefficiencies can be identified and resolved. The residents will get a clear picture regarding the handling of their concerns which promotes transparency. Proactive planning is made possible, allowing for better resource distribution and improved service delivery. Ultimately, systemic issues can be identified, and solutions can be developed, leading to improved city planning and enhanced community satisfaction.

GUIDING QUESTIONS

The major guiding questions for the analysis of 311 service requests and the the ways in which the insights from answering these questions can be utilized are as follows:

1. Geographic Analysis

- Which community or location has the largest number of service requests?
- Are there any specific needs for certain areas?

High need areas can be prioritized and resources can be allocated where they are most required.

2. Seasonal Trends

- During which seasons do service requests occur most often?
- How do service requests change over seasons? Are there any identifiable patterns?

Seasonal patterns are identified thus helping to focus resources effectively.

3. Request Sources

- What is the primary source of service requests: phone calls or online submissions (web)?

Interfaces can be upgraded based on user preferences.

4. Types of Service Requests

- What is the service requested most frequently?

Helps to identify and allocate adequate resources for the most needed service.

5. Response Efficiency

- Which agency handles the most and least number of service requests?
- What is the average response rate and time for resolving for service requests?
- Who are the most efficient agents in terms of response and resolution times?
- How does the response efficiency vary across different years?

Track the department efficiencies and identify areas where improvement is required.

6. Trends Over Time

- How has the volume and type of service requests changed over the past five years?
- Are there noticeable trends in requests that could be used for future planning?

Future demands can be anticipated, allowing for better preparedness.

DATASET

Data description

The data for this analysis is sourced from The City of Calgary's open data portal, specifically the "311 Service Requests - Services and Amenities" dataset (The City of Calgary, 2025). It includes public service requests submitted via 311 from 2020 to the present, consisting of 2,585,662 rows and 15 columns, structured in a tabular format. Each row represents an individual service request.

Format and Structure

The data consists of several datasets depending on the city or organization's open data portal. Each dataset typically contains the following columns:

Column Name	Description	Data Type
service_request_id	The unique identifier for an individual request.	Text
requested_date	The date the request was submitted.	Floating Timestamp
updated_date	The most recent date the request was updated.	Floating Timestamp
closed_date	The date the request was closed.	Floating Timestamp
status_description	The current status of the request (e.g. open, closed).	Text
source	The channel used to submit the request.	Text
service_name	The type of service requested.	Text
agency_responsible	The department responsible for this request.	Text
address	The location of the service request (if applicable).	Text
comm_code	The community code associated with the service request location.	Text
comm_name	The community name associated with the service request location.	Text
location_type	The type of location information provided for this service request.	Text
longitude	The longitude of the service request.	Number
latitude	The latitude of the service request.	Number
point	The spatial coordinates based on latitude and longitude.	Point

The data is sourced from The City of Calgary's open data portal, 311 Service Requests - Services and Amenities.The data is publicly available and used with permission as per the open data policy. This data is provided by the City of Calgary at https://data.calgary.ca/Services-and-Amenities/311-Service-Requests/iahh-g8bj/about_data and all usage complies with their data usage and attribution requirements listed in the license URL https://data.calgary.ca/d/Open-Data-Terms/u45n-7awa**.**

TASKS

Data wrangling and visualization tasks will be performed to ensure that raw data is transformed into a clean format and then visualized effectively to develop valuable insights. Data wrangling (McKee.A, 2024) involves handling null values by imputing mean or median values in missing values of columns longitude or latitude (Nichani, P., 2020). Duplicate records are identified and removed, standardizing data types to ensure consistency, such as converting Floating Timestamp to datetime for the date columns (pandas pydata, no date). Data transformations like encoding categorical variables (e.g., encoding status_description "open" => 0 and "closed" => 1) and normalizing numerical data for easier analysis. Feature engineering to derive new columns, such as calculating response time (e.g., the difference between closed_date and requested_date will be the response time), and additionally, grouping and mapping subdivisions into broader categories, such as combining detailed service_name into high-level service groups for providing clearer insights (Turing, no date). When unusual values are found in numerical columns, these are outliers, which can be treated by capping or removal to prevent skewed results (Chirag Goyal, 2021).

Once the data is preprocessed, the next step is visualizing the data to identify patterns and insights. Distribution analysis using histograms and box plots for outlier detection. Exploring the relationships between variables can be done using scatter plots and line graphs (e.g., to show the trend of requests over time). Bar and pie charts for summarizing categorical columns (Govind, B. and Shivthare, 2024). Advanced visualizations like geographical columns, like longitude and latitude, and interactive dashboards will take the help of tools like Power BI or Tableau.

Task distribution for each members: *Anitha*: Focus on data cleaning, like handling null values, removing duplicates, standardizing data types, and treating outliers. *Jincy*: Handles data transformation and feature engineering, encoding, normalizing, deriving new features, and grouping categories. *Megha*: Performs data visualization using plots and charts and creates dashboards to identify valuable insights.

REFERENCES

1. 311 Calgary. (no date).*311 Calgary*. [online] Available at: https://www.calgary.ca/311.html
2. The City of Calgary (2025) *311 Service Requests*. [Online] Available at: https://data.calgary.ca/Services-and-Amenities/311-Service-Requests/iahh-g8bj/about_data (Accessed 18 Jan. 2025)
3. McKee. A. (2024). *A Beginner's Guide to Data Cleaning in Python*. [online] Datacamp.com. Available at: https://www.datacamp.com/tutorial/guide-to-data-cleaning-in-python (Accessed 17 Dec. 2024)
4. Nichani, P. (2020). *Appropriate Ways to Treat Missing Values*. [online] Analytics Vidhya. Available at: https://medium.com/analytics-vidhya/appropriate-ways-to-treat-missing-values-f82f00edd9be (Accessed 2 May 2020)
5. Pandas pydata. (no date). *pandas.to_datetime — pandas 1.3.4 documentation*. [online] Available at: https://pandas.pydata.org/docs/reference/api/pandas.to_datetime.html.
6. Turing. (no date). *Feature Engineering for Machine Learning with Python*. [online] Available at: https://www.turing.com/kb/feature-engineering-in-ml-with-python.
7. Chirag Goyal. (2021). *How to Detect and Remove Outliers | Outlier Detection and Removal*. [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2021/05/feature-engineering-how-to-detect-and-remove-outliers-with-python-code/. (Accessed 20 Dec. 2024)
8. Govind, B. and Shivthare. (2024). I*mpact of Data Visualization in Data Analysis to Improve the Efficiency of Machine Learning Models*. pp. 107–112. Available at: https://www.researchgate.net/publication/378841602_Impact_Of_Data_Visualization_In_Data_Analysis_To_Improve_The_Efficiency_Of_Machine_Learning_Models (Accessed March 2024)