City of Regina Project – Park Maintenance and Citizen Satisfaction

EN553.600 Final Project Report

Iris (Yiyao) Zhang, Shiyao Lei, Yang Ye, Youqi Wu

May 7, 2023

Contents

1 Executive Overview

The city of Regina has been facing a significant challenge in managing the large number of park maintenance requests. With around 300 parks in the city, it has become impossible for the city to manually review all of the requests. Facing this issue, the City of Regina presented their problem to Dr.Bertran De Lis, who is the Director of Analytics at the Johns Hopkins University Center for Government Excellence. In our project, both the City of Regina and Dr.Bertran De Lis are our clients. Our clients have approached us to find a process that can help them deal with the massive amount of park visitor comments, understand them, and prioritize them. The final goal for our clients is to increase overall citizen satisfaction with the service and reduce the number of citizen complaints.

Upon analyzing the data provided by our clients, we found that visitors made 5000 comments/complaints from January 2021 to October 2022 without knowing the name of the corresponding park. This underscores the urgent need for an effective solution. To address these concerns, we have identified five critical issues and have recommended a solution and an implementation plan for each one. By implementing these solutions, our client will be able to track urgent maintenance issues without manual effort and increase overall citizen satisfaction.

Our solutions will not only help our client to manage the park maintenance requests effectively, but will also enable them to provide a better experience to the visitors. By prioritizing the maintenance issues based on the visitor comments, our clients can ensure that the visitors have a better experience in the parks. This, in turn, will lead to an increase in overall citizen satisfaction with the service.

We are confident that our recommendations will be effective in addressing the concerns of our clients and improving the overall park maintenance process in the city of Regina.

2 Critical Issues

2.1 Data Processing Challenge

One major challenge faced in managing park services is the processing of large datasets. A possible solution is to explore the use of distributed computing frameworks to handle large dataset sizes, in combination with automated data cleaning and aggregation pipelines using machine learning techniques. Further optimization of the data processing workflow can be achieved through the investigation of data compression techniques that reduce storage costs and speed up processing times.

2.2 Missing Location Data

Another issue arises with unstructured location data, as it can be difficult to accurately associate visitor comments with park locations. A potential solution is to develop natural language processing (NLP) algorithms that can extract location information from visitor comments, as well as implementing a standardized labeling system to improve the accuracy of association with park locations. Additionally, the use of location data visualization tools can aid in identifying patterns and relationships in the unstructured data.

2.3 Issue Prioritization

Effective issue prioritization is essential for resource allocation and effective issue resolution. Machine learning models can be applied to categorize and prioritize issues based on complaint frequency and severity, and a data-driven decision-making framework can be established to allocate resources accordingly. Data visualization tools can also assist in tracking issue resolution progress and communicating results to stakeholders.

2.4 Urgent Issue Identification and Maintenance Constraints

Urgent issue identification and maintenance constraints can be addressed through the establishment of automated monitoring systems that detect urgent issues in real-time, and the utilization of predictive maintenance techniques to prevent issues before they occur. It's also essential to establish efficient communication channels between maintenance teams to ensure prompt and effective resolution.

2.5 Citizen Satisfaction and Complaint Reduction

Finally, to improve citizen satisfaction and reduce complaints, sentiment analysis can be conducted on citizen feedback to identify areas of improvement for park services. Targeted communication and outreach campaigns can then be implemented to address common citizen concerns, and data analytics can measure the impact of implemented strategies and continuously improve satisfaction levels. By addressing these challenges, park services can provide more effective and efficient services to their visitors.

3 Analysis

3.1 Data Description and Visualization

Our original data contains 5043 rows of park visitor comments and 14 columns with information that categorizes and explains the comments from 2021-01-01 to 2022-10-21. Despite the data having columns of Street Number and Street Name, the comments are not directly associated with the park location they are leaving a comment for, as some of the Street Number and Street Name information could be the park visitors' home address instead of the park address.

3.2 Location Mapping

The location data contains 306 parks and their address. By using Google geocoding api of park addresses, we get the latitude and longitude data columns. Google geocoding api converts address into longitude and latitude.

Building upon the previous geocoding API mapping, we have calculated the latitude and longitude coordinates for each park on the list. Ultimately, we have identified the specific park

location associated with each comment by assigning the nearest park to the corresponding comment coordinates.

3.3 Word Cloud

We use Python build-in package to create word cloud with raw comment data and keywords-extracted comment data. The word cloud (figure 3) build with keywords-extracted comment data gives us a better understanding of key issues in park maintenance.
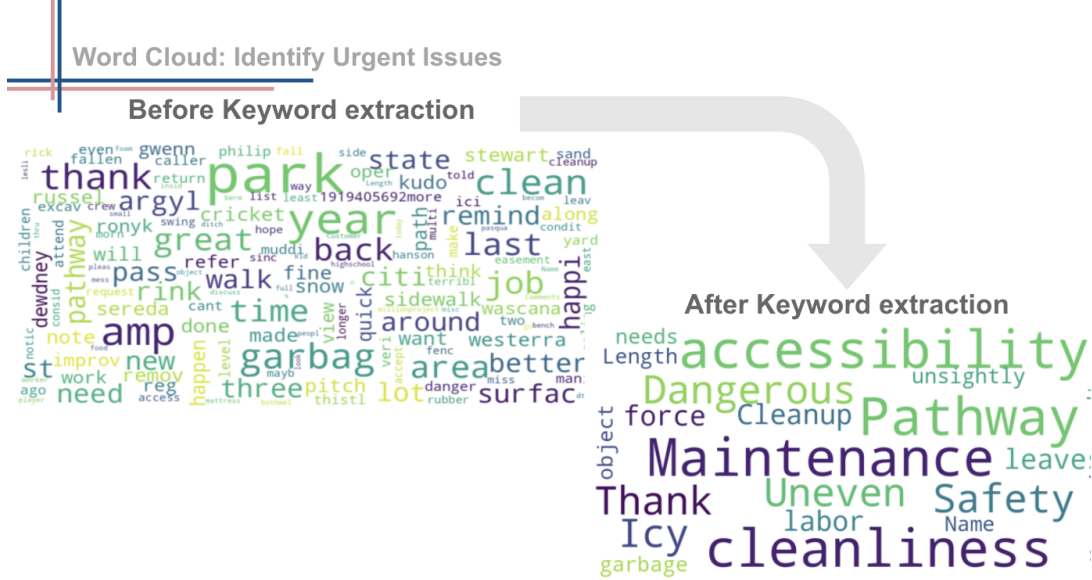
Figure 1: Sub Request Types

3.4 Sub Request Types

Revised analysis part: In our analysis, we examined the sub request types in the original dataset and counted the number of appearances for each type (as shown in Table 1). However, upon further investigation into the comments and sub request types, we discovered that the sub request type did not align well with the actual customer comments (as depicted in Figure 2). This limited our understanding of the underlying issues and hindered effective analysis.

To address this issue of incompatible sub request types, we utilized the ChatGPT API to extract key issues from customer comments. This approach yielded more accurate and detailed results compared to the original sub request type column in the dataset.

Table 1: Service Request Type and Count

ParksMnt	1371	ParksTrePr	172	CemetMntnc	75	BranchPick	6	PestsOther	2
TurfMntnc	558	Ease&Blvd	152	ParksDrai	53	BrokenBran	5	TreePla	1
IrrigMtnc	526	PrkGrbEmpt	144	DeadAnPU	49	TreeInsp	4	LightRep	1
PathMntnc	396	LitPubProp	132	FoxtalWeed	26	PesNoxWd	4	DeadAnimal	1
VehcleAcce	248	OutdrRink	103	TreePrunin	11	TreeObst	2	HortInq	1
WeedsCity	231	PlayEquMnt	94	TreeRemova	8	RodContrl	2	BirdInquir	1
No Subtype	202	WeedsDande	91	AdoptGree	7	TensCoMnt	2	MonInquiry	1

4

Sub Request Type	Customer Comments	Keyword
PathMntnc	**Lights** at outdoor rink **do not work**. Please check and repair.	Lights
OutdrRink	I was awakened this fine morning to that **most joyous sound of the back up "beep, beep, beep"**, flashing yellow beacons and truck lights shining into my bedroom window as the city worked on the skating rink at Mable Brown Park AT 6:45AM!!! Do I have to remind the city that there is a noise bylaw from 10pm to 7am (and 7am is a full hour too early in my opinion). Good Morning to you and Thanks guys.	Noise
OutdrRink	McNab Park Rink **lights** have not worked properly since the power went out in the storm 2 days ago - also asking to have the **heat** turned up some inside the shack	Broken lights, no heat
PathMntnc	RPS called - there was a **vehicle accident** that took down the post at the entrance of St Thomas Park on Tyvan Ave - Police File #RA21011380	Car accident

Figure 2: Sub Request Types

3.5 Topic Modeling

In our analysis, we utilized the BERT topic modeling technique in Python to identify the common complaint issues shared by customer comments. BERT topic modeling is a built-in package that helps us identify the most common topics within the dataset.

By applying this technique, we were able to cluster the key issues into issue category level 3. Subsequently, we manually categorized these issue category level 3 topics into level 2 and level 1 categories.

This approach allowed us to gain insights into the most prevalent and recurring issues mentioned by customers. By clustering the key issues and categorizing them into higher-level categories, we were able to better understand the broader themes and trends present in the dataset.

4 Possible Solutions

As previously mentioned, our goal is to provide our clients with several potential solutions for gathering and processing visitors' complaints for parks in the city of Regina, and to optimize these solutions. However, we have encountered some difficulties during the progress of our project.

4.1 Comment Web Page Design for Visitors Using ChatGPT.

Solution 1 involves designing a web page where visitors can input comments and receive immediate replies via ChatGPT. While this solution has the potential to be very effective, we have encountered a challenge in that some comments are inputted by the park client manager after receiving visitors' phone calls. This means that visitors do not have access to the web comment system, and many prefer to give their complaints through phone calls. This reduces the efficiency of this solution, but we believe that as the web comment system becomes more widely used, it will be a more effective method for gathering and processing complaints.

4.2 Streamline Data Processing

Solution 2 involves streamlining the data processing and continuously outputting a prioritized issue list. However, we are currently facing a challenge in terms of communication with the client. Specifically, we are uncertain about which tools are available to them, and which input formats are available to us. Additionally, we are unsure of our client's long-term preferences, which could impact the success of this solution. For example, if the format of the data changes from "complaints" to "customer reviews," our tool may not be able to process the data. We are actively working with the client to resolve these issues and ensure that this solution is effective.

4.3 Optimize Maintenance Schedule by Urgency and Severeness

Solution 3 involves optimizing the maintenance schedule based on the urgency of each comment. One challenge we are facing is that we are uncertain about how each issue is ranked. Currently, we are ranking the issues based on the number of times they are mentioned, but we understand that this approach may not be realistic. We are exploring other methods for ranking issues and ensuring that the maintenance schedule is optimized effectively.

4.4 Cluster Parks into Smaller Zones

Solution 4 involves clustering parks into smaller zones by their location and making maintenance issues more specific. This solution has the potential to be very efficient, but currently, we do not have detailed information about the maintenance schedule, such as the number of park maintenance centers in the city of Regina, their locations, or their responsible areas. We are actively working to gather this information and ensure that this solution can be implemented effectively.

5 Recommended Solutions

5.1 Solution for Data Processing/Missing Location Challenge:

One of the major challenges in this project is to structure the raw comment data. Our recommended solution is to use Python to process the raw Excel comment data. The output will be a structured Excel comment data with its corresponding park name. This will make the data more manageable and easier to analyze. The major advantages of our solution is its simplicity and preciseness - it is easy to implement with Google Geocoding API. All that is required is to input the raw data and run the iPython file.

5.2 Solution for Issue Prioritization

Another critical issue in this project is how to rank the urgency of park maintenance. Our recommended solution is to use a combination of natural language processing and data aggregation techniques. We use ChatGPT to summarize the comments and extract key issues. Then, we aggregate the issues by type and count the number of times they were mentioned by zone and

season. Based on this information, we create a prioritized list of maintenance issues. The final output will be an Excel or Google Sheet with issues that were mentioned over 10 times.

The major advantage of our solution is its simplicity. All that is required is to input the raw data and run the iPython file. Our approach is also highly customizable, allowing for different ranking standards to be applied to the data based on specific client needs.
Below are our solution excel outputs:

- Prioritized issue excel list based on Zone

- Prioritized issue excel list based on Season

- Comment level excel list

5.3 Solution for Urgent Issue Identification and Maintenance Constraints

Another critical pain point for our client is the lack of understanding of urgent issue distribution and their maintenance capacity. Our recommended solution is to create a Tableau interactive map that will help our client keep track of urgent issues. This map will show the number of maintenance issues by zone, season, issue category level 1, and issue category level 2.

By using the interactive map, our client will be able to quickly identify areas with a high concentration of urgent issues and allocate their maintenance resources accordingly. The major advantage of our solution is the ease of use and the ability to visualize complex data in a simple and user-friendly way.
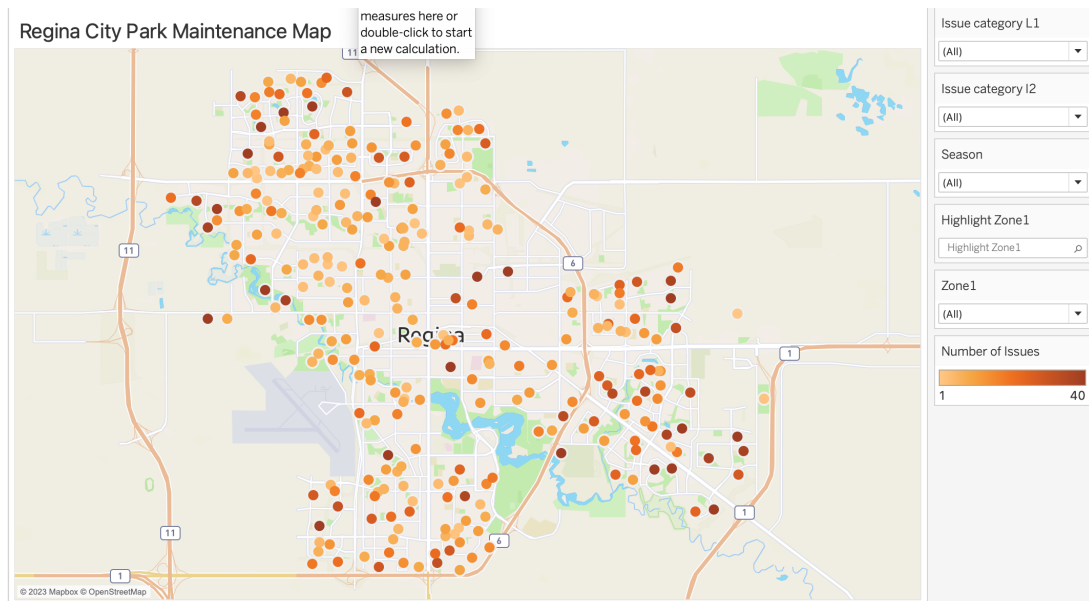

Figure 3: Sub Request Types

5.4 Solution for Citizen Satisfaction and Complaint Reduction

The city of Regina aims to enhance citizen satisfaction and minimize the number of citizen complaints. Our recommended solution is to implement a well-designed ChatBot using ChatGPT

to reply to visitor comments immediately. The ChatBot will use natural language processing techniques to identify key issues in visitor comments and provide citizens with a corresponding maintenance schedule. By addressing citizen concerns and providing them with up-to-date information, this solution could improve customer satisfaction and reduce the number of citizen complaints.

The major advantage of our solution is the ability to provide immediate and personalized responses to citizen comments. The ChatBot can handle a large volume of comments efficiently and provide citizens with accurate and helpful information, which could help to build trust and goodwill among the community.
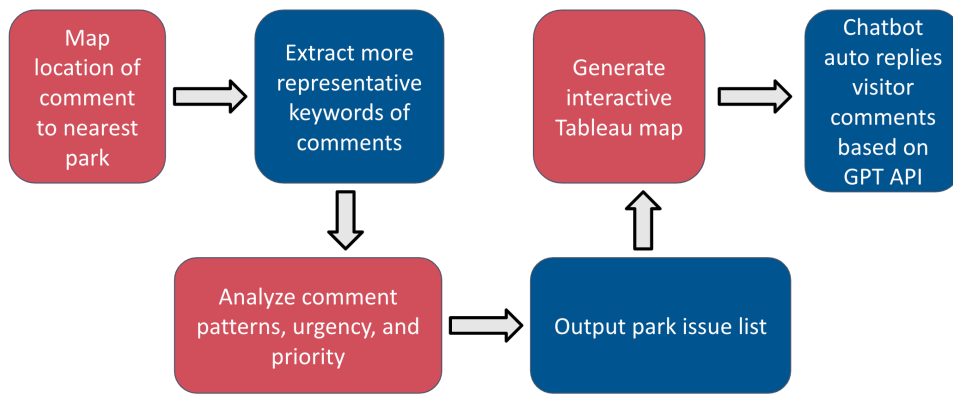
6 Implementation Plan

Figure 4: Implementation Plan

Our implementation plan is a one-time solution that deals with raw data in Excel format. The following steps outline our proposed approach:

- Data processing: We will use Python to process the raw Excel comment data and extract key issues using ChatGPT for issue summarization. We will then aggregate the issues by type and count the number of times they were mentioned by zone and season. This will result in a structured Excel comment data with its closest park name and a prioritized list of urgent issues.

- ChatBot implementation: We will design and implement a ChatBot using ChatGPT to reply to visitor comments with corresponding maintenance schedules. The output will be a comment level excel data with ChatBot replies

- Tableau implementation: We will use Tableau to create an interactive maintenance map that shows the number of maintenance issues by zone, season, issue category level 1, and issue category level 2. Our client can input the comment level excel data and get the maintenance map

The major advantage of our implementation plan is the ability to provide our client with a comprehensive and integrated solution that addresses their pain points related to park maintenance.