Improving San José 311 Efficiency: Reducing Unlabeled Requests and Proactively Identifying Resident Needs

Author: Angela Zhao, Data Scientist

City of San José, Information Technology Department (ITD)

May 28, 2024

Acknowledgements

Abstract/Executive Summary

San José 311 (SJ311)[1], a platform for non-emergency City services launched in 2017 helps City residents access City services easily through the SJ311 website, the SJ311 mobile app, or through a phone call. City residents can report various issues into pre-defined categories or into the "Other Issues" category if the issue did not fit into a pre-defined category.

The goal of this research was to understand what kinds of requests residents report in the "Other Issues" category to identify new service types, redesign existing service types, and streamline internal processes. By reducing the number of unlabeled requests in the City's 311 backend system, City staff can focus on fixing issues instead of manually labeling issue requests.

The project successfully reduced the 2018-2023 requests with no category (also known as unlabeled requests and originally placed in the "Other Issues" category) from 397,223 requests, or 30% of all requests from 2018-2023, to 74,026 requests, or 6% of all requests from 2018-2023.

By adding subcategories, the SJ311 contact center team can track requests better, prioritize enhancements and future changes to the platform based on resident needs, and facilitate faster routing. This improves City services and enhances the experience for City residents.

Introduction

San José implemented its Smart City Vision to advance safety, sustainability, economic opportunity, and overall quality of life for the community. To help City residents more easily access City services, San José 311 (SJ311), a platform for non-emergency City services, was launched in 2017.

What is San José 311?

SJ311 is a consolidated platform for all non-emergency City services for San José residents. Created in 2017 to help all residents easily reach additional services, SJ311 has experienced an increase in demand as well as additional feedback in increasing its ease of use.

Three main ways through which San José residents can access 311:

1. Use the SJ311 website to log the request
2. Mobile SJ311 application Currently, the SJ311 platform has 18 different services. Of these 18 requests:
3. Call the SJ311 Customer Contact Center which can either connect them to
   a. A virtual agent to
      i. Log a request or
      ii. Forward the call to correct services
   b. A service representative who will log, categorize the request and forward the call if needed

---

[1] https://311.sanjoseca.gov/

Currently, the SJ311 platform has 18 different services. Of these 18 requests:

- Three are forwarded to portals/services that are outside of the SJ311 platform:
    - Pay Utility Bills
    - Rent Registry Tenant Portal
    - Affordable Housing
- Two are lookup services:
    - Street Sweeping
    - My Collection Schedule
- The remaining 13 are logged in the SJ311 backend system:
    - Vehicle Concerns
    - Graffiti
    - Illegal Dumping
    - Pothole
    - Streetlight Outage
    - Illegal Fireworks
    - Community WiFi
    - Eviction Prevention
    - Street Sweeping
    - Junk Pickup
    - Container Issues
    - Missed Collection
    - Services for New Homes
    - Other Issues

See Table 2 for a complete list of services and when they were added to the SJ311 platform.

Challenge

The "Other Issues" is a catch-all service type that allows residents to log any issue that does not fit in any existing category. While issues in the predefined service type categories are easily routed to the appropriate teams to be resolved, issues in the "Other Issues" category require human intervention to manually categorize and reroute the issue. This is an overhead that burdens SJ311 Customer Contact Center staff.

In 2020 there were six request types offered on the SJ311 platform: Abandoned Vehicles, Graffiti, Illegal Dumping, Streetlight Outage, Pothole, and Other Issues. An increase in various service types occurred that did not fit into the five predefined categories (Figure 1). In particular, the "Other Issues" category was the largest category with over 175,109 requests, or 32.6% of 537,636 requests from 2018-2020 (see Figure 2).

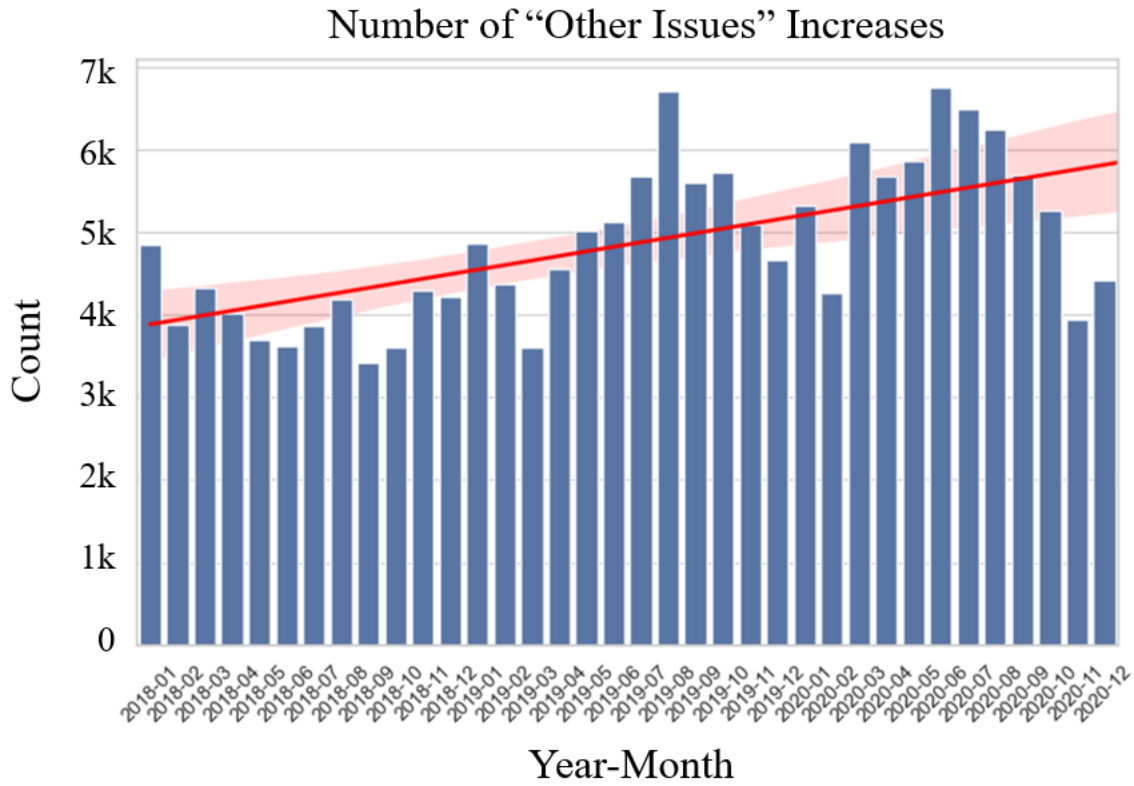## Number of "Other Issues" Increases

Figure 1: Count of "Other Issues" (unlabeled) requests received by SJ311
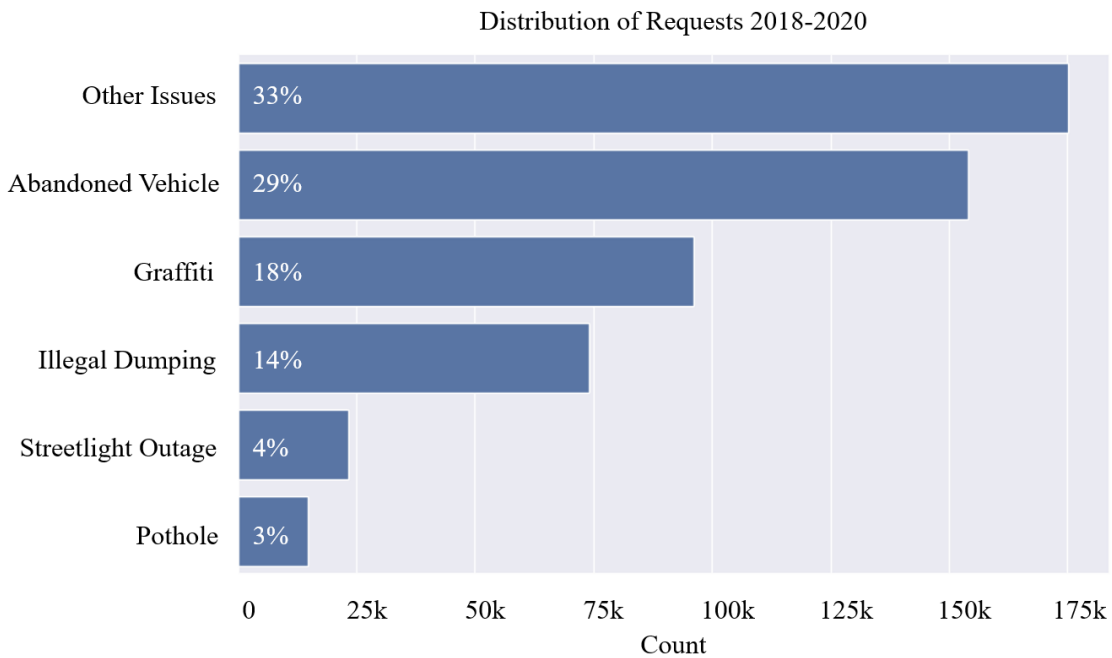
Distribution of Requests 2018-2020

Figure 2: Percentage of requests considered "Other Issues" (unlabeled) received by SJ311 from 2018-2020

Annually, over 50,000 "Other" requests had to be manually routed by 15 staff. If the number classified as "Other" is reduced, requests can be routed faster while freeing staff time for essential services.

This led the City to explore Natural Language Processing to read through the Descriptions logged in the "Other Issues" service requests to understand what residents were largely reporting on that could be used in:

1. Identifying new service types for the SJ311 platform
2. Redesigning of existing service types to meet resident needs
3. Streamlining internal processes to better handle "Other Issues" requests

Analysis

In 2021, the City engaged with a team of Data Scientists at Dell Technologies to help interpret the requests being logged in the "Other Issues" category. The Dell team used topic modeling to unearth new categories and themes that, when large enough, could be separated into its own category. In 2023, City's in-house team took on the work to grow, build, and sustain the work done by the Dell team.

Categorizing requests in Other Issues

The Dell team used Latent Dirichlet Allocation (LDA) and the in-house team used BERTopic, two widely used approaches for topic modeling. The topics unearthed by the two models formed the new categories.

Topic modelling is an unsupervised machine learning method that categorizes a set of documents into a set of underlying topics found in the documents. The two methods chosen – BERTopic and LDA are both widely used.

LDA produces more highly-interpretable topics at the expense of worse performance on documents with poorly defined topics. BERTopic performs well on nuanced and complex datasets, although topics can be less interpretable and less well-defined.

How Language Models work

LDA assumes that all documents are a random mixture of words selected from the mixture of topics and works backwards from this assumption. Each document is assigned a probability score for every topic and is sorted into the topic with the highest probability. The number of topics and the topic label must be chosen by a human.

When performing LDA, data preprocessing steps include cleaning, lemmatizing[2], and adding word-embeddings. The number of topics is a hyperparameter chosen by the researchers, and LDA outputs the topics as well as the probability for every document to be categorized in each

---

[2] Lemmatizing: the process of breaking down a word to its root form (ie running and ran are lemmatized to run). Source: https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html

topic. A topic is a probability distribution over a group of words found in the documents and each document is treated as a bag of words.

BERTopic clusters documents into groups of topics based on properties of the documents (i.e., word choice). This is a five-step process.

1. BERTopic takes advantage of the BERT model's ability to capture context to turn every document into a high-level embedding, such that similar documents have embeddings that are close to each other.
2. BERTopic reduces embedding size to make analysis more manageable.
3. BERTopic clusters the document embeddings so that clustered documents constitute single topic.
4. BERTopic chooses the number of clusters.
5. BERTopic identifies the main keywords inside each cluster to label the topic. BERTopic requires no human intervention for topic labeling.

While comparing topic labelling from different models is difficult and ill-defined as to "what constitutes a good topic labelling," the City compared the categories found by both teams. Similar and related topics crop up in both topic labels. This leaves us reasonably confident that the categories produced by both teams are on-target – they've hit the main key points in the uncategorized Other Issues – and that the City can trust these categories going forward.

Dell Research

The Dell Data Science Team utilized the free-text description attached to the Other Issues service-requests to reveal new service themes that weren't captured by the City's standard service categories. The Dell team used LDA for topic modeling and 247,905 "Other Issues" requests from April 2017 to June 2021.

Preprocessing included cleaning, lemmatization, and using word embeddings. First, each open text description is tokenized into words of two or more alphabetical characters long. All stop words were removed (including thank, please, and need). Each token was lemmatized, which meant it was reduced to its root form (i.e., "cars" becomes "car"). They were converted to lower case, and tagged the part of speech. Open text descriptions that were greater than 50 words in length (1% of descriptions), empty (0.2% of descriptions), or one word long (.4% of descriptions) were discarded.

Next, all the descriptions were converted into a matrix of token counts using a CountVectorizer[3] object. As before, only words with two or more alphabetical characters were considered. Stop words were excluded and all characters were converted to lowercase. All terms that occurred in less than 0.5% of documents or in more than 95% of documents were also excluded. All words were considered in both unigrams and bigrams (single words or phrases of two words each).

The corpus was run through the model 100 times to be categorized into 10 different topics. Each description is assigned a probability score for each topic, and the topic with the highest

---

[3] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

probability is assigned to the description. If there was no obvious topic with the highest probability score, then the description was removed from the analysis (totaling about 3% of descriptions).

The ten topics found were, in order of most frequent to least frequent:

1. Transportation hazard
2. Parking violation
3. Environmental hazard caused by humans
4. Garbage cart pick-up and replacement
5. Phone number inquiry
6. Billing inquiry
7. Abandoned vehicles and large item pick-up
8. Business related (+ covid-19)
9. Streets dispatch and sidewalk repaired
10. Junk pick-up and bill payment

Please see Figure 5 in the Appendix for the two most demonstrative examples from each topic.

In-house Research Part 1

San José continued the topic modeling with an in-house project using BERTopic as the model of choice, instead of LDA. Since BERTopic uses word embeddings and considers the context of the words, the data needed minimal text preprocessing. All requests of one word or less or written in a non-English language were removed.

To pilot the concept, San José chose to use a subset of requests to train a BERTopic model, which can be resource intensive. San José filtered the requests to directly from users interacting with the SJ311 website or mobile app from July 2021 to June 2022 for a total of 10,524 requests. All requests of one word or less (2% of requests) or written in a non-English language (0.2% of requests) were removed. This left 10,297 requests to be analyzed.

BERTopic was run multiple times. After the first pass, topics that were not salient were recategorized in one of the other categories. An ensemble strategy was used to assign descriptions as outliers – if a majority of the four strategies assigned the document as an outlier, then the document was an outlier. However, if the four strategies were split, then topic representation was based off the probabilities method. If there was no consensus between any two out of the four methods, the request would still be considered an outlier.

Since BERTopic generated 50 topics, these topics were consolidated into 13 topics by manually grouping topics by thematic content. Of the 10,297 requests, around 95% were categorized into one of fifty subtopics. Outliers were reduced from 3,556 to 479.

These topics were, in order of most frequent to least frequent:

1. Homelessness and Encampment Concerns
2. Road and Infrastructure issues

3. Illegal Parking
4. Garbage, Dumping, and Cleanup
5. Abandoned Vehicle and Extended Parking
6. Water Infrastructure Issues
7. Plant Hazard
8. Non-Descriptive Addresses
9. Loud Noises
10. People Concerns
11. Sidewalk and Curb Concerns
12. Animal Concerns
13. SJ311 Platform Issues

Results from this analysis showed the top concerns reported in Other Issues were regarding homelessness, road infrastructure, and illegal parking. Other common issues relate to encampments, plants/trees hazards, loud noises, animal concerns, people/resident issues, and the SJ311 platform.

In-house Research Part 2

After the success of the pilot model, San José built another BERTopic model. The goal was to understand how adding new categories impacted Other Issues. For the training data, 247,806 requests in the Other Issues category from 2018-2021 were pulled. All requests that were one word or less long (24% requests), written in a non-English language (3% of requests), or were repeated (2% of requests) were removed. 183,097 requests were used as training data.

BERTopic was run once. Topic labels with cosine similarity greater than 0.7 were merged together for a total of 79 topics.

While the previous pilot BERTopic model reduced outliers using an ensemble method, the new model did not reduce outliers using the same method due to the significant reduction of quality in topic labels. Topics suffered from overgeneralization – too many unrelated requests were grouped into the same topic.

The City then proceeded to use keywords to create more specific labels to better categorize requests. Then the City consolidated the topics into six categories and 15 sub-categories. These categories and sub-categories are intentionally aligned with the changes that were suggested by the first BERTopic model so that changes in topics over time could be more easily analyzed.

After creating the BERTopic model, the model was also used to categorize requests in the Other Issues category from 2022-2023. The requests were similarly preprocessed as the training data, so 132,391 out of 149,417 requests were labeled.

Please note that the subcategories only apply for the Other Issues category. These categories were:

1. Other issues
2. Garbage and Recycling

3. Junk Pickup
4. Water
5. Container Issues
6. Abandoned Vehicle

The subcategories of Other Issues were:

1. Building Permit
2. Business License
3. City Hall Info
4. Code Inquiry
5. E-Bill Express Technical Issue
6. External
7. Homeless Concerns
8. Hauler Complaint
9. Inquiry
10. Non-Emergency Police
11. Other
12. Parking Concerns
13. Sewer/Sanitation/Wastewater
14. Streets Dispatch
15. Utility And Billing

Comparison of findings

Both methods resulted in similar topic labeling – transportation and infrastructure issues as well as hazards caused by people were identified as main issues. While the labeling of the topics may be similar, sometimes the topics identified by the Dell team were split into multiple by the in-house team or vice versa. For example, what the Dell team considered "environmental hazard caused by humans" could also be labeled by the in-house team as "Homeless and Encampment Concerns" or "Illegal Dumping".

Improvements made to San José 311

Taking feedback from the two analyses, the City introduced Pay Utility Billing, Sewer Issues, and Street Sweeping as new service types. Sewer Issues is to be deployed; the other two are live. Residents logged issues around abandoned vehicles in the Other Issues event though it has its own category. This information combined with low customer satisfaction rating for abandoned vehicle service type and replaced Abandoned Vehicle with Vehicle Concerns. Based on the topics identified in the analyses, the City also modified the contact center's backend subcategories. These are used for tracking and prioritizing new service request types on the platform. When a subcategory becomes large enough, the subcategory is elevated into its own service request category.

New Categories

The new categories as identified by BERTopic modeling capture resident needs and can reduce the volume of requests labeled Other Issues. Figure 3 captures the change in the size of the Other Issues category in its original form and after requests were re-categorized using BERTopic.
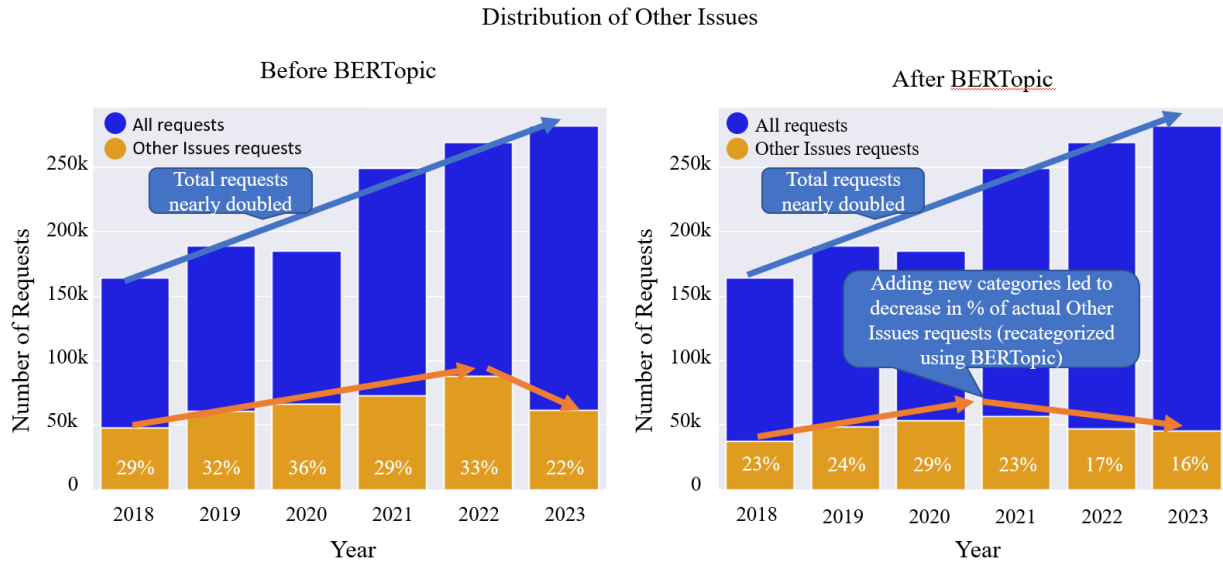
Distribution of Other Issues

Figure 3: Percentage of requests categorized as "Other Issues" before and after using BERTopic to recategorize topics.

The Other Issues category is still significantly high after adjusting based on the BERTopic findings. This steady size of the Other Issues category can be explained by several factors.

First, City's outreach efforts in 2022 and 2023 seem to have had an impact on the number of Other Issues category. Overall, Other Issues has decreased from 32.5% to 21.9% for all requests from 2019 to 2023.

Year	2018	2019	2020	2021	2022	2023
Abandoned Vehicle	33.9%	32.7%	20.3%	15.6%	11.8%	10.5%
Community Wi-Fi	N/A	N/A	N/A	N/A	0.04%	0.05%
Container Issues	2.2%	2.0%	2.1%	8.5%	8.5%	7.6%
Eviction Prevention	N/A	N/A	N/A	N/A	0.01%	0.01%
Garbage And Recycling	1.1%	1.0%	1.2%	1.4%	5.4%	9.5%
Graffiti	17.2%	15.5%	20.8%	9.6%	10.1%	9.9%
Illegal Dumping	13.9%	13.0%	14.5%	11.1%	8.5%	9.6%
Illegal Fireworks	N/A	N/A	N/A	N/A	0.6%	0.5%
Junk Pickup	1.1%	1.1%	1.4%	24.4%	30.1%	27.0%

Missed Collection	N/A	N/A	N/A	1.9%	1.7%	1.5%
Other Issues	23.0%	25.7%	29.0%	22.9%	17.4%	16.1%
Pothole	1.6%	2.7%	3.8%	0.6%	0.5%	0.9%
Start New Services	N/A	N/A	N/A	0.2%	0.2%	0.1%
Streetlight Outage	4.1%	4.0%	4.8%	2.7%	1.1%	2.3%

Table 1: Table of Service Distributions using category proportions after using BERTopic model. Some categories were too small to detect using BERTopic, which is why Illegal Fireworks is N/A and Garbage and Recycling is not for 2018.

Second, the addition of new request types draw more attention to those request types themselves, which is also why Other Issues hasn't changed as much as expected. This can be seen most clearly in Table 1 where the labeled data from 2018-2023 using the BERTopic model in the Junk Pickup and Container Issues categories. Junk Pickup averaged 1.2% from 2018-2020, and when it was added as a category in 2021, it jumped to 24.3% of all requests. Container Issues similarly averaged 2.1% from 2018-2020 and jumped to 8.5% of all requests when it was added as a category in 2021.

Finally, many of the issues that are in Other Issues can be more clearly fit into other categories. Table 1 demonstrates the breakdown of categories when using BERTopic to label requests in Other Issues. Adding new categories draw more residents who want to use SJ311 to resolve issues surrounding new categories, but also increases the likelihood that these requests will be incorrectly submitted under the Other Categories label. This is due to confusion surrounding request categories, and labeling a request as Other Issues is a safe alternative. The outreach for new categories is successful when seen in the increase of requests that are correctly labeled Other Issues. When using the labels given by the BERTopic model from 2018-2023, the Other Issues category is 22.4% of all requests when the requests are sorted again. However, without the model, the Other Issues category is nearly 50% larger at 30.1% of all requests.

Subcategory Labels
Additional labels were added in the backend for the Other Issues for SJ311 Customer Contact Center team to use when categorizing "Other Issues" requests that come through phone and chat. These are for categories that are not large enough to warrant their own categories and cannot be seen by residents. Many of these subtopics were pulled from the categories created by the Dell team's topic modeling.

These new subcategories were implemented in the backend system for better tracking and forwarding to other departments. Tracking these sub-categories helps track resident needs without overwhelming residents with many categories. Additionally, this helps SJ311 identify future categories if a subcategory becomes popular enough.

These subcategories are:

1. Building permit

2. Business Tax
3. City Hall info
4. Code Inquiry
5. County info
6. External
7. Homeless concerns
8. Non-emergency police
9. Other
10. Parking permit
11. SewerSanitation/WasteWater
12. Streets Dispatch
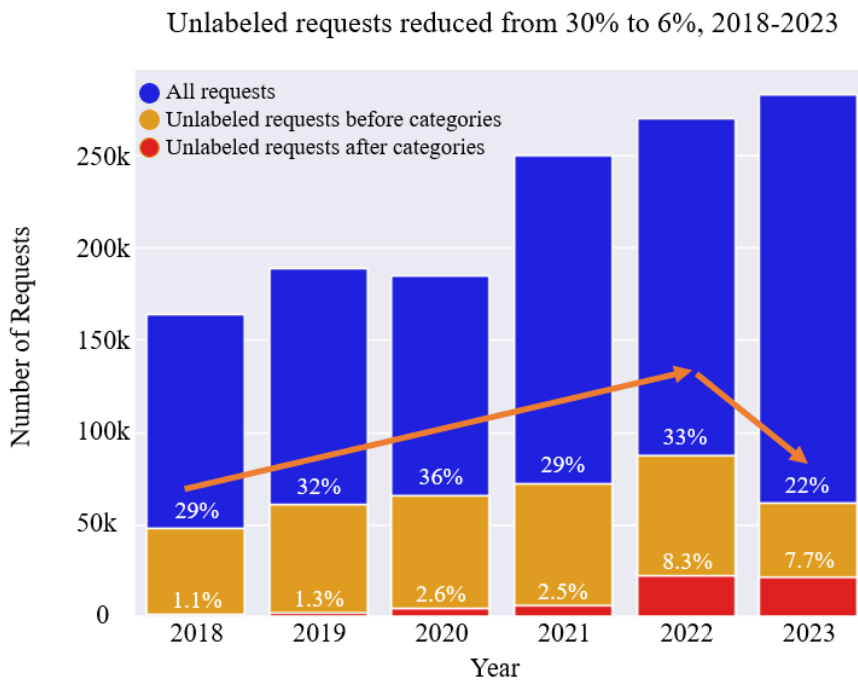13. Validation
14. Vehicle complaint

Figure 4: Percentage of requests that were unlabeled before and after using BERTopic and subtopic categories in the "Other Issues" category.

Of the five categories, Water proved to be the most useful in reducing the number of requests re-routed to the Other Issues category. Recall that two of the topics the Dell Team unearthed included billing inquiries, phone inquiries, and bill payments. Many of these were related to water services, such as water bills. Since the introduction of the Water category in 2023, over 7,000 requests have been logged as such.

Additionally, the use of the Abandoned Vehicles category was one of the sources that contributed to the push of creating a Vehicle Concerns category. Requests that were vehicle-related, such as graffiti on a vehicle, were frequently miscategorized into Other Issues as a

13

Vehicle Complaint. Since such requests would be better categorized as Vehicle Concern and so Abandoned Vehicles was renamed into Vehicle Concerns.

With these labels, the City can reduce Other Issues subcategory into 30.3% of the Other Issues category. Ultimately, this means that only 6% of all requests from 2018-2023 are ultimately uncategorized with labels – these are requests that are in the Other Issues category and have no subcategory label. This proportion used to be 30.1% of all requests that were uncategorized.

Conclusion

By analyzing the Other Issues category, common themes have been identified and used to create additional service request categories. These themes have helped identify backend labeling of Other Issues services for better tracking for City staff and for future improvements of SJ311 platform.

From 2021 to 2024, the SJ311 platform added six service categories and 14 subcategory labels. This has successfully reduced the percentage of requests with no label from 30% to 6% of all requests from 2018-2023. By adding subcategories, the SJ311 Customer Contact Center team experienced a decrease in reporting times and identify new categories to add to improve services for SJ311 residents.

This is intended to be an iterative improvement process for the City to improve its efficiency and respond times to SJ311. Next steps include taking the lessons learned from adding subcategories to the next customer relationship management system, such as ensuring that the uncategorized options within the "Other Issues" category is small by creating more subcategories early in the build.

Appendix

Year service was added	Service
2018	Abandoned Vehicle – renamed to Vehicle Issues
	Streetlight Outage
	Pothole
	Graffiti
	Illegal Dumping
	Other Issues
2021	Container Issues
	Start New Services
	Garbage and Recycling
	Junk pickup
	Missed Collection
2022	Illegal Fireworks
	Eviction Prevention
	Community WiFi
2023	Water
2024	Encampment Concerns*
	Street Sweeping Service
	Vehicle Concerns – previously Abandoned Vehicle

Table 2: The year services were added. Asterisk (*) denotes services that were added after the analysis.

Year Added	Subcategory
2018	City Hall Info
2021	Building Permit
	Business Tax
	Code Inquiry
	County Info
	External
	Homeless Concerns
	Non-Emergency Police
	Other
	Parking Permit

2022	Streets Dispatch
	Vehicle Complaint
2023	Sewersanitation/Wastewater
	Validation

Table 3: List of subcategories that were added for "Other Issues

Topic	Description
1: Transportation hazard	Cars continuously speed through our street as a short cut to get to flint ave and the schools on Flint ave. Please help us in putting a stop sign at Flintdale and Flintmont right in front of [address]. This shall help slow people down. Also helps in making it safe for people turning onto flintdale from flintmont. Busy times 745 am to 900am then again 200pm to 330pm
2: Parking violation	There is an abandoned car parked in front of my house and it hasn't moved for almost a week. The car is a Honda Civic, grey colored, paint damage and the license plate is [redacted].
3: Environmental hazard caused by humans	This corner lot has become not only an eyesore to our neighborhood but also a safety hazard. These RVs are occupied with people who are always causing nuisance, stray animals and an unsafe area that kids can get hurt at. The rv that is burned is the second one on this property and has been sitting there for months creating a safety hazard to the kids and people in the neighborhood. Please have the owner clean up this lot.
4: Garbage cart pick-up & replacement	Missed garbage pick up, broken recycle cart
5: Phone number inquiry	Caller wanted the number for GreenTeam.
6: Billing inquiry	garbage bill and water bill inquiry
7: Abandoned vehicles & large item pick-up	Customer checking on abandoned vehicle report.
8: business related (+covid-19)	Caller wanted to be connected to the Business Tax License Dept
9: streets Dispatch & sidewalk repaired	Sidewalk needs repaired
10: junk pick-up & bill payment	Caller wanted to schedule a free junk pick-up.

Table 4: The most demonstrative example from each topic as identified by the Dell Team. Identifying details have been redacted.

Topic labeling, in house findings in detail

San José continued the topic modeling with an in-house project using BERTopic as the model of choice, instead of LDA. Since BERTopic uses word embeddings and considers the context of the words, the data needed minimal text preprocessing.

Models were evaluated by topic saliency (how interpretable the topic was) and the number of outliers generated. The goal was to generate a set of non-overlapping interpretable topics. Therefore, the main parameter altered was the minimum sample size in a cluster that would be

labeled as topic. Minimum samples of 25-30 were ideal in creating clusters. Smaller minimum sample size led to more specific topics, but high redundancy and larger sample sizes led to less topic specificity.

Data preprocessing included removing all single word or empty reports, removing all reports that consist only of locations, and translating no-English reports using Google Translate. Uniform Manifold Approximation and Projection (UMAP), a dimension reduction technique, was used for dimension reduction, Hierarchical Density-Based Spatial Clustering of Application with Noise (HDBSCAN) for clustering, and Class- Based Term Frequency Inverse Document Frequency (C-TF-IDF), which is a measure of importance of a word to a document was used for key word extraction for cluster labeling.

BERTopic was run multiple times. After the first pass, topics that were not salient were recategorized in one of the other salient categories. BERTopic also categorizes descriptions as an outlier using C-TF-IDF, probabilities, distributions, and embeddings. An ensemble strategy was used to assign descriptions as outliers – if a majority of the four strategies assigned the document as an outlier, then the document was an outlier. However, if the four strategies were split, then topic representation was based off the probabilities method. If there was no consensus between any two out of the four methods, the request would still be considered an outlier.

Since BERTopic generated 50 topics, these topics consolidated into 13 topics by manually grouping topics by thematic content. Of the 10,297 requests, around 95% were categorized into one of 50 subtopics. Outliers were reduced from 3,556 to 479.

These topics are, in order of most frequent to least frequent:

1. Homelessness and Encampment Concerns
2. Road and Infrastructure issues
3. Illegal Parking
4. Garbage, Dumping, and Cleanup
5. Abandoned Vehicle and Extended Parking
6. Water Infrastructure Issues
7. Plant Hazard
8. Non-Descriptive Addresses
9. Loud Noises
10. People Concerns
11. Sidewalk and Curb Concerns
12. Animal Concerns
13. SJ311 Platform Issues

Results from this analysis showed the top concerns reported in Other Issues are regarding homelessness, road infrastructure, and illegal parking. Other common issues relate to homelessness/encampments, plants/trees hazards, loud noises, animal concerns, people/resident issues, and the SJ311 platform.

Pre- and Post-Pandemic topic changes

The team then analyzed the changes in requests made to SJ311 before the pandemic (2018-2019) and after the pandemic (2022-2023).

For this analysis, BERTopic labeling was used to label requests in the Other Issues category. Due to the random nature of BERTopic and the splitting of the data into pre- and post-pandemic categories, the labels are not the same as the labels produced in the previous use of BERTopic.

The dataset was split into a 70:15:15 train-validation-test. The V-measure score (a harmonic mean between homogeneity and completeness) and accuracy were the evaluation metrics for the goodness of cluster labeling. The V-measure score is the harmonic mean between the homogeneity score and the completeness score. Which is: $v = (1 + \beta)$ homogeneity × completeness (β × homogeneity + completeness). Beta (β) decides how homogeneity and completeness are weighted – it defaults to "1" to weight homogeneity and completeness equally. Given a set of data from n classes (the genuine clusters), and the clustering result of the same data set (m clusters), the homogeneity measures the "purity" of a cluster (how many of the data points within a class belong to that specific cluster). The completeness measures how many data points that belong to a class are included in a specific cluster. The highest score of V-measure is "1", which means perfect match, and the lowest is "0".

After the final clusters are generated and read into each cluster, the City can identify which clusters are the labeled classes within the validation set, the City can then verify the performance by calculating the accuracy of each labeled classes. The best model reached a V-score of 0.65.

Main findings include:

- Overall usage of the SJ311 mobile app has increased, while the number of "Other Issues" requests submitted through the SJ311 website has decreased
- Topics with the biggest increases post-pandemic include: water leakage, noise, and traffic/street light
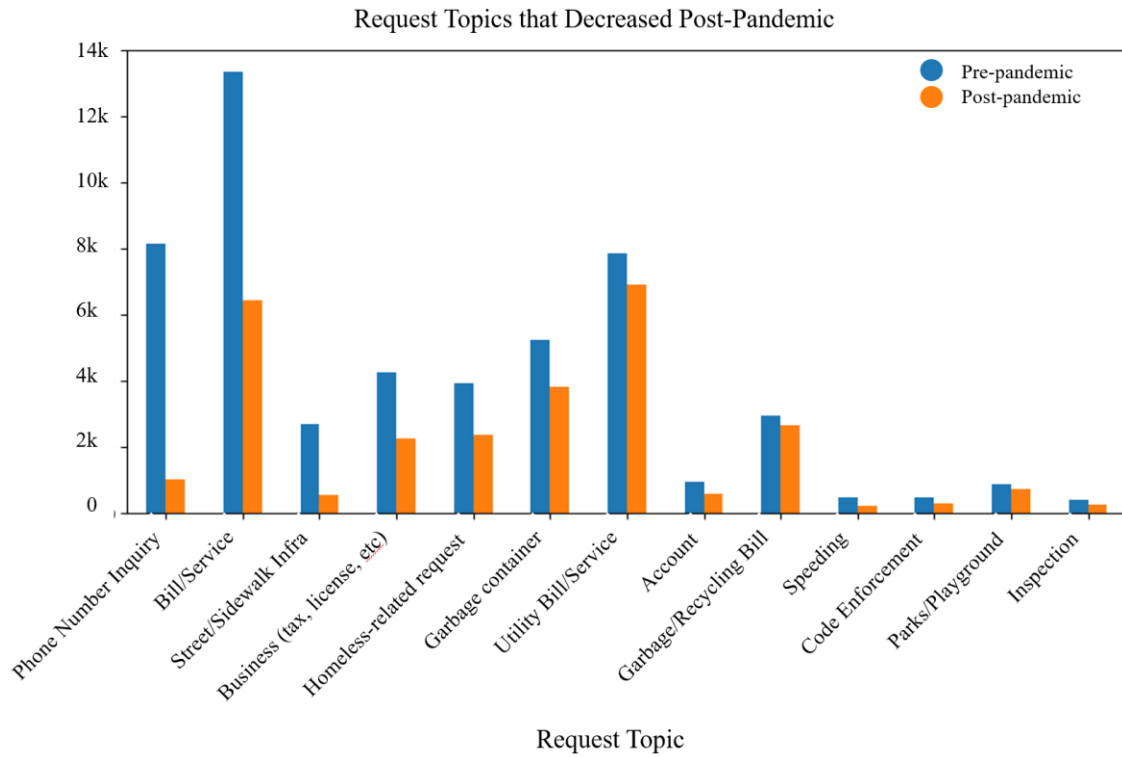- All district request counts have decreased since the pandemic

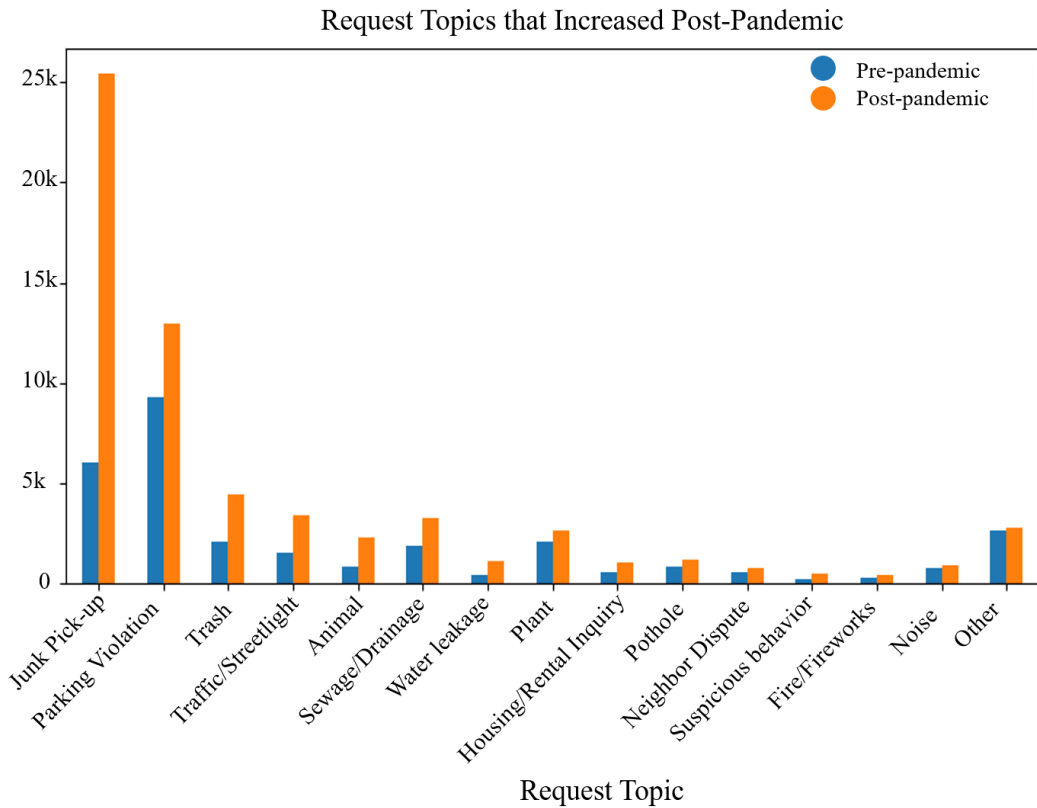*Figure 5: Requests that decreased after the pandemic.*

Figure 6: Types of requests that increased after the pandemic.